Application of detector precision characteristics and histogram packing for compression of biological fluorescence micrographs

Tytus Bernas[1,2], Roman Starosolski[3], J. Paul Robinson[4], Bartlomiej Rajwa[4]

[1] Department of Physiology and Medical Physics, RCSI, 123 St Stephens Green, Dublin 2, Ireland
[2] Department of Biochemistry, Biophysics and Biotechnology, Jagiellonian Univesity, Gronostajowa 7, 30-387 Krakow, Poland
[3] Institute of Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
[4] Purdue University Cytometry Laboratories, 1203 West State St., West Lafayette, IN, USA

Author for correspondence: Roman Starosolski, Institute of Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland, e-mail: rstarosolski@polsl.pl, phone: +48 322372151.

Abstract

Modern applications of biological microscopy such as high-content screening (HCS), 4D imaging, and multispectral imaging may involve collection of thousands of images in every experiment making efficient image-compression techniques necessary. Reversible compression algorithms, when used with biological micrographs, provide only a moderate compression ratio, while irreversible techniques obtain better ratios at the cost of removing some information from images and introducing artifacts. We construct a model of noise, which is a function of signal in the imaging system. In the next step insignificant intensity levels are discarded using intensity binning. The resultant images, characterized by sparse intensity histograms, are coded reversibly. We evaluate compression efficiency of combined reversible coding and intensity depth-reduction using single-channel 12-bit light micrographs of several subcellular structures. We apply local and global measures of intensity distribution to estimate maximum distortions introduced by the proposed algorithm. We demonstrate that the algorithm provides efficient compression and does not introduce significant changes to biological micrographs. The algorithm preserves information content of these images and therefore offers better fidelity than standard irreversible compression method JPEG2000.

Keywords: microscope imaging, intensity resolution, histogram packing, image compression, image coding

1. BACKGROUND

Digital imaging based on light microscopy has become an established technique in basic and applied biological sciences. Modern applications such as high-content screening (HCS), 4D imaging, and multispectral imaging may involve collection of thousands of images in one experiment. On the other hand, storage space and network transmission bandwidth which may accommodate these data is often limited. Therefore efficient image-compression techniques may be necessary to mitigate this problem. Several compression routines developed for digital photography and film, may be used for this purpose. Reversible compression algorithms (Deflate, LZW, RLE, Huffman encoding, etc.) neither introduce distortion to images [**1**] nor remove any information from images and therefore preserve the data integrity, as defined by 21 CFR part 11 [**2**]. However, reversible ("lossless") techniques when used with biological micrographs usually provide only a moderate compression ratio, which does not exceed 3:1. This is caused by the large dynamic range (4096 – 65336 intensity levels per 'color' component) of biological micrographs, and the fact that in a typical case the full extent of the dynamic range is used and all the intensity levels are populated. Furthermore, no typical intensity distribution may be constructed for these images.

More efficient compression can be obtained with irreversible ("lossy") techniques, which use vector quantization [**3**, **4**], discrete cosine transform [**5**], wavelet transform [**6**, **7**] or fractal coding [**3**, **8**]. However, these algorithms by definition remove some information from images and introduce artifacts. In the field of digital photography these distortions are considered acceptable as long as the essential perceptual image quality is not affected [**9**, **10**, **11**, **12**]. In other words, all intensity differences between pixels are regarded as significant if they are detectable by human observers [**9**]. This approach is established (sometimes mandatory) in medical applications of brightfield microscopy (including pathology, cytology and hematology) owing to massive amounts of digitized data. For instance, the whole-slide imaging (WSI), where the entire micrograph area is digitized, may generate daily terabytes of uncompressed data.

However, procedure of verification of compression integrity by a panel of experts is time-consuming and may be specific for different applications (biomedical assays). Thus, the procedure difficult to standardize and to apply systematically across a range of specimens and imaging regimes. This problem hinders use of automated data analysis/interpretation methods in the context of medical imaging. Therefore, one might adopt a different approach and focus on the information content of the input image data in order to establish acceptable compression limits and criteria. Owing to the presence of noise in the microscope images, not all intensity differences can be considered significant from a statistical standpoint. Thus, the number of meaningful intensity levels may be lower than the nominal dynamic range (corresponding to 12- or 16-bit precision) provided by the AD converters of the cameras. Therefore, to establish an irreversible yet information-preserving compression mechanism that does not use perceptual quality as a criterion, one should first construct a model of noise that is a function of signal in an imaging system. Then, intensity binning can be applied to discard insignificant intensity levels. The processed images are characterized by sparse intensity histograms; therefore reversible coding and intensity depth-reduction algorithms should be employed to provide efficient compression.

Herein we report the design and implementation of a compression pre-processing scheme which takes advantage of the noisy nature of microscope images by preserving only statistically significant levels of intensity. We provide results of local and global measures of intensity distribution to demonstrate that the alterations introduced by this algorithm to biological micrographs range from none to minor. We also demonstrate that the algorithm offers better fidelity than JPEG2000 (a standard "lossy" compression routine) at the same compression ratio.

2. MATERIALS AND METHODS

2.1. Cells and fluorescence labeling

FluoCells prepared slide #2 (Molecular Probes) was used in all experiments. The slide contained fixed bovine pulmonary artery endothelial cells in which microtubules were labeled using mouse anti-bovine α-tubulin monoclonal antibodies in conjunction with BODIPY FL goat anti-mouse IgG antibody; the cell nuclei were labeled with DAPI.

2.2. Microscope imaging

Images of the endothelial cells were collected using a Nikon E1000 wide-field fluorescence microscope equipped with a Nikon 40x Fluor oil-immersion objective lens (NA 1.3) and a 100-W Hg arc lamp. The BODIPY FL fluorescence was registered using a 475- to 495-nm excitation filter (band-pass), a 505-nm long-pass dichroic mirror and a 525- to 565-nm emission filter (band-pass). A monochrome CCD camera (Retiga 4000R, Qimaging, Burnaby, Canada) was used for image acquisition. A 16x neutral density (ND) filter was used to attenuate the flux of excitation light. The microscope aperture diaphragm was fully open, whereas the field diaphragm was adjusted to match the field of view of the objective. Image collection was carried out at room temperature. The camera was cooled to 25°C below ambient.

A time series of 128 images of stationary (fixed) cells were collected using full frame (no binning) at 5-s intervals. The series was registered at three gain settings and for 0.25-s or 0.75-s acquisition times. Image acquisition was controlled using ImagePro Plus v 5.1 (Media Cybernetics, Silver Spring, MD, USA).

2.3. Calculation of noise levels and background signal

Levels of background (dark) and fluorescence signals together with their respective variances (corresponding to total noise) were calculated as described elsewhere [**13**]. Briefly, for every image pixel the fluorescence intensity changes in time were modeled with three components: a systematic trend (related to photobleaching), a periodic component (associated with fluctuation of the excitation light source), and an irregular component (representing noise). Following [14] we studied our system using a simple univariate version of the unobserved components (UC) model:

$$y_t = T_t + S_t + e_t \tag{1}$$

where t denotes the value of the associated pixel intensity at the t^{th} time point, y is the observed value, T is the trend (or low-frequency component), S is the periodic (or "seasonal" component), and e is the irregular component. All the calculations were executed on a pixel-by-pixel basis utilizing the CAPTAIN modeling toolkit operating within the environment of Matlab [**14**]. First, the stochastic trend component was estimated using the integrated random walk (IRW) model:

$$\begin{aligned} I_t &= T_t + e_t \\ T_t &= 2T_{t-1} - T_{t-2} + \eta_t \end{aligned} \tag{2}$$

where I_t is registered fluorescence intensity (at the t^{th} time point), T_t is the smoothed intensity at the t^{th} time point, T_{t-1} and T_{t-2} are values of T_t at two previous time points, e_t is measurement noise (zero mean, variance σ^2_e), and η_t is the system disturbance (zero mean, variance σ^2_η) .

The trend was subtracted from the observed intensity (I_t). The de-trended data were used to isolate periodic components of intensity changes with the dynamic harmonic regression (DHR). Subsequently, the IRW and optimal order DHR were used jointly to fit the trend and periodic component to the initial fluorescence intensity data (I_t). One should note that the noise variance ratio ($\sigma^2_\eta / \sigma^2_e$) was

optimized in this step as well to minimize residual variance globally. The sum of the trend and periodic components represented the true instantaneous fluorescence intensity (signal, S^i_t) at every time point. Hence, the instrumental noise (for a pixel at a given time point) and its variance (for a signal level) were estimated as:

$$N^i_t = \left| S^i_t - I^i_t \right|$$
$$V_{S=F} = \frac{\sum_{i,t} \delta_{SF} (N^i_t)^2}{\sum_{i,t} \delta_{SF}} \quad (3)$$

where N is the noise, S is the signal, I is the fluorescence intensity registered at the i^{th} and t^{th} points of the image time.

Estimates of the other two components of a time series (periodic component and trend) can be further used to characterize the stability of the light source and the photobleaching rate of the fluorochromes used in the experiment. However, they were utilized here only to provide an estimate of total signal level (fluorescence and background) and thus to calculate the corresponding level of total noise.

In order to estimate the background signal, uniform dark image regions were identified for each time series. These regions (represented using binary masks) were comprised of pixels characterized by fluorescence intensity and local fluorescence heterogeneity that were smaller than 10% of the respective maxima. The heterogeneity was measured using the algorithm described in [**15**]. Average intensity (I_b) calculated in dim and homogenous regions was taken as the background (i.e., the pixel value of an image registered in the absence of fluorescence). The noise variance (V, Equation 3) was plotted against the signal corrected for background (S_c=S-I_b). A quadratic function was fitted to these data in order to characterize signal-noise dependency:

$$V = A + PS_c + MS_c^2, \quad (4)$$

where M, P, and A are estimators of the signal variance associated with the multiplicative, Poisson (photonic), and additive noise components. The standard deviation of I_b ($\sqrt{B}$) was calculated to estimate the background noise.

2.4. Calculation of significant intensity levels and histogram binning (HB)

Owing to the presence of noise in the images, not all intensity differences can be considered significant. Thus, the number of meaningful intensity levels is lower than the nominal dynamic range provided by the cameras (12 bits, 4096 levels). Hence, the significant levels were calculated iteratively using the following algorithm:

1. Input I_b, A, P, M
2. Set k=0
3. Do:
    a. Set k=k+1
    b. Set $I^k_{med}=I_b$
    c. Set $\sigma(I^k_{med})=\sqrt{(A+ I^k_{med} *P+ I^k_{med} * I^k_{med}*M)}$
    d. Set $I^k_{high}=I^k_{med}+1.96*\sigma(I^k_{med})$
    e. Set $I^{k+1}_{low} = I^k_{high}$
    f. Calculate I^{k+1}_{med} so that $I^{k+1}_{med} - I^{k+1}_{low}=1.96*\sigma(I^{k+1}_{med})$
4. Loop while $I^{k+1}_{med}<4095$
5. Terminate. Output scalar s=k and vector $I=[I^1_{med} I^2_{med} \ldots I^s_{med}]$.

The algorithm produces a set of I^k_{med} for which $I^k_{med}-I^{k-1}_{med}=1.96(\sigma(I^k_{med})+\sigma(I^{k-1}_{med}))$. The I^k_{med} values smaller than I_{max} represent intensity levels significantly different from one another with 95% probability (confidence) in the sense of Student's t test (hence the factor of 1.96). The choice of confidence interval was arbitrary. However, similar calculations can be performed for every confidence level. The set of values was used to segment the representative images by setting all pixel intensities (I_r) to the nearest significant level. This operation is hereinafter referred as histogram binning (HB).

2.5. Simple depth reduction and histogram packing

Reversible image-compression algorithms used on data characterized by sparse intensity histograms (such as those generated by HB) may exhibit poor performance [**16**]. To alleviate this problem a simple depth reduction (SDR) transform, first introduced in [**17**] was used. SDR maps pixel intensity levels in the following way:

$$I_{SDR} = I_0 s / I_{\max}, \tag{5}$$

where I_{SDR} denotes intensity after SDR transform, I_0 is the original intensity, s is the number of significant levels, and $I_{\max}$ is the intensity of the highest significant level.

To perform the inverse SDR transform one needs to know s and $I_{\max}$. The compression ratios were calculated taking into consideration this necessary overhead (4 bytes).

The performance of reversible image-compression algorithms may also be improved using histogram packing (HP) [**16**]. HP maps all the significant levels to the lowest part of the nominal intensity range using order-preserving one-to-one mapping. The transform is reversible provided that information that permits histogram expansion after decompression is encoded with compressed images. It is sufficient to encode significant intensity levels for histogram expansion, and there is no need to specify how many times a level was used. An array of histogram-encoding methods is available [**18**]. When the number of significant levels is low, a simple mapping table (MT) may be used. Briefly, if n significant intensity levels are found in the image the MT represents sorted intensity values as contiguous indexes:

$$\mathrm{MP}(I_n) = (I_0 \mapsto 0, I_1 \mapsto 1, \ldots, I_{n-1} \mapsto n-1) \tag{6}$$

The compression ratios reported in this work were calculated taking into consideration the size of MTs encoded according to the JPEG-LS standard. Therefore the MTs were encoded with $5+2s$ bytes, where s is the number of significant levels. This form of histogram packing used together with histogram binning (HB) is hereinafter referred as histogram binning/packing (HBP). Note, that instead of storing original histogram with each compressed image, one could store it only once for all images that share the same detector characteristic, or just store the detector parameters used to perform Histogram Binning. An implementation illustrating this possibility and allowing to experiment with the HB and HBP methods has been prepared (http://sun.aei.polsl.pl/~rstaros/hbp/index.html).

2.6. Image-compression algorithms

We estimated compression efficiency of two still-image coding methods: JPEG-LS [**19**] and JPEG2000 [**6**], developed by the JPEG committee. These algorithms are recent coding standards of ISO/IEC and ITU organizations, and are incorporated in the DICOM standard of NEMA [**20**]. JPEG-LS describes a low-complexity predictive compression algorithm with entropy coding using a modified Golomb-Rice [**21**, **22**] code. The technique is based on the LOCO-I algorithm [**23**]. We used the SPMG/UBC implementation (version 2.2, ftp://ftp.netbsd.org/pub/NetBSD/packages/distfiles-/jpeg_ls_v2.2.tar.gz). JPEG2000 is based on wavelet-transform image decomposition and arithmetic

coding [**24**]. This algorithm provides progressive transmission and region-of-interest coding [**25**]. We used JasPer implementation by Adams (version 1.700.0, http://www.ece.uvic.ca/~mdadams/jasper/). We also tested the performance of a universal data-compression algorithm Deflate (RFC 1951 [**26**]). The algorithm is based on the Ziv-Lempel dictionary coding LZ77 [**27**] and is incorporated in the DICOM standard. We used the gzip implementation (version 1.2.4, http://www.gzip.org/).

All the compression methods were used with default coding parameters. Compression ratio is defined as U/C, where C is the size, in bytes, of a compressed image (including header); U is the size of the original image, defined as $U = bn/8$, where b is the image bit depth (12 in our case), and n is number of pixels in the image. One should note that in practice space savings may be greater than implied by the compression ratio since uncompressed image file formats store image pixels on whole bytes, so each 12-bit pixel actually occupies 16 bits.

One should note that both JPEG2000 and JPEG-LS may be used for reversible and irreversible coding. While the reversible compression ratio depends on the image contents, it may be set in arbitrary manner when the irreversible mode is used. However, high compression ratio is obtained in that case at the expense of the fidelity of decompressed image (micrograph). Typical artifacts include blurring and generation of spurious image details (Fig. 1).

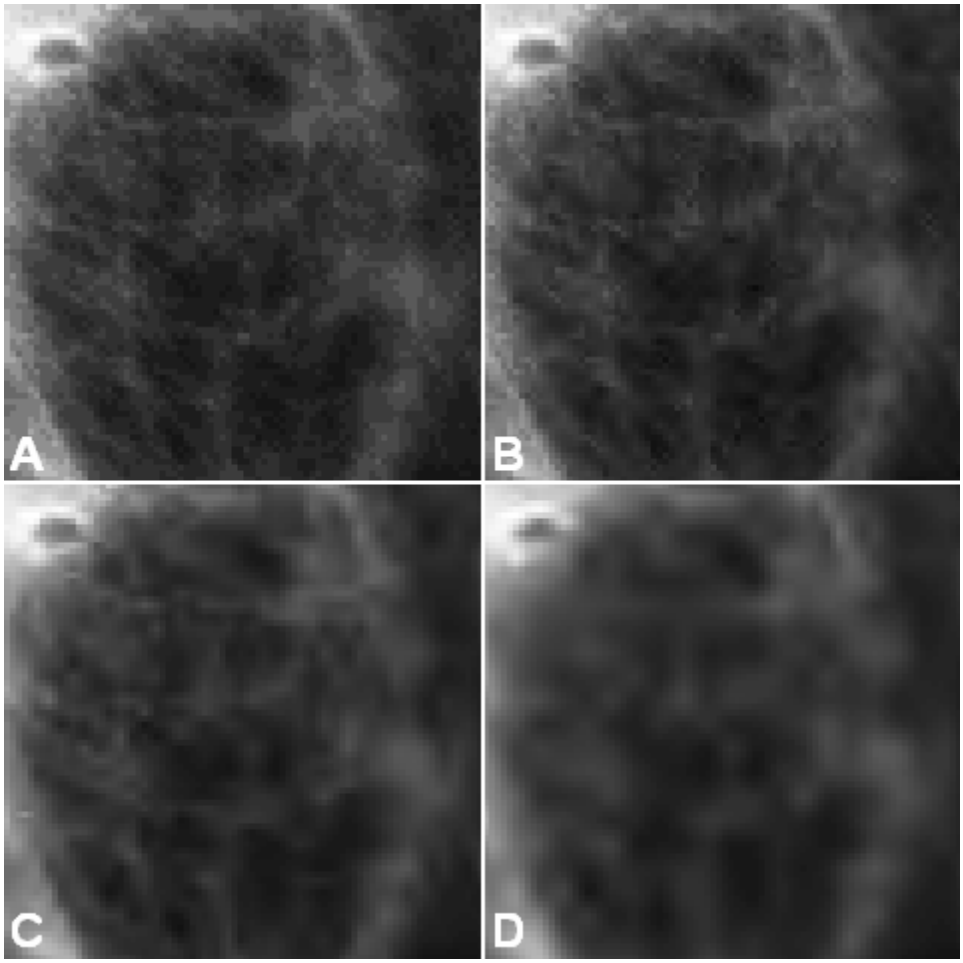

Fig. 1. JPEG2000 compression at different ratios. Enlarged fragment of micrograph: raw image (A), for which the reversible JPEG-LS ratio is 1.75, and images compressed with irreversible JPEG2000 at ratios 10 (B), 25 (C) and 50 (D).

2.7. Estimation of global changes in fluorescence intensity distributions

We used the earth mover's distance (EMD) algorithm in order to establish whether compression introduced changes in the total fluorescence intensity distributions [**28**]. We compared image histograms of raw images with their counterparts processed with SDR, HB, or irreversible JPEG2000. The minimal average intensity change (per pixel) needed to transform histograms of a compressed image into the respective histogram of an uncompressed (reference) image was computed for every such image pair.

2.8. Alterations of local fluorescence distributions

To verify whether reduction in the number of intensity levels altered spatial fluorescence intensity distributions, raw images were compared with their counterparts subjected to HB. The distributions were compared using texture parameters (features): Haralick features based on the grey-level co-occurrence matrix (GLCM), gradient-based features, run-length matrix parameters, and wavelet energy. Detailed descriptions of these parameters are provided in [**29**, **30**]. The GLCMs were calculated at distances from 1 to 9. The gradient-based features and run-length matrix parameters were calculated at 0 (horizontal), 45, 90 (vertical), and 135 degrees. The wavelet energy was calculated at first, second, and third decomposition levels. Calculations were performed for the areas where the fluorescence intensity was higher than background. The texture parameters of images subjected to histogram binning were divided by the respective values for their raw counterparts. The total number of 137 normalized texture parameters was subjected to step-wise linear discriminant analysis. The Mahalanobis distance was used to establish the parameters characterized by the highest discriminant power. The parameters were added to and removed from the analysis set using probability of F (0.05 for entry and 0.10 for removal). An identical set of texture parameters was calculated for images processed with SDR and compressed with irreversible JPEG2000.

3. RESULTS

3.1. CCD noise and background

Square roots of Poisson and additive noise coefficients (Equation 4) were analyzed as a function of gain for the monochrome (single-channel) images registered Retiga 4000R CCD camera at 12-bit (4096 levels) precision., as described in detail elsewhere [**13**]. Briefly, the amount of both types of noise depended linearly on the gain for the CDD, but was not affected by acquisition time (Table 1). On the other hand, the background signal (see Materials and Methods) increased linearly with acquisition time and gain [**13**]. There was no significant difference between the additive noise (√A) and background noise (√B). Hence both these parameters may be regarded as estimators of the dark noise. Consequently, dependence between signal and noise for the CCD may be accurately expressed using Equation 4 [**13**].

Table 1. Dependence of Poisson and additive noise on CCD gain for the set of 12-bit single-channel images used in experiments. The respective fit coefficients (√P, √A, eq.2) calculated for the variance (V, eq. 2) as the function of gain are given with their standard errors (for details see [**13**]).

	acq. time [s]	**slope (SL_p)**	**intercept (IN_p)**	**correlation (r^2)**
Poisson noise	0.250	0.0948 ± 0.007	-0.028 ± 0.005	0.98
	0.750	0.0986 ± 0.006	-0.018 ± 0.002	0.99
Additive noise	0.250	2.030 ± 0.150	-1.07 ± 0.10	0.98
	0.750	2.087 ± 0.371	-0.37 ± 0.22	0.99

3.2. Calculation of significant intensity levels and histogram binning

The CCD cameras registered images with 4096 nominal intensity levels (12-bit digitization). However, owing to the presence of noise not all differences in intensity between pixels may be considered significant (see Materials and Methods), and consequently practical dynamic resolution is lower than nominal. Hence, the number of significant intensity levels (with 0.95 probabilities) was calculated at several values of CCD settings (gain, acquisition time and, offset) and are presented in Table 2. These numbers of levels were used to perform SDR (Equation 5). Furthermore, the respective vectors of intensity values corresponding to significant levels were used to segment series of images registered at given values of gain, acquisition time, and offset (see Materials and Methods) by setting all pixel intensities (I_r) to the nearest level (histogram binning, HB). Result of these transformations is illustrated by a representative image in the Fig. 2. One may note that no gross image distortions were introduced by this operation.

3.3. Coding efficiency of sparse histogram images

Fluorescence microscopy images characterized by various levels of noise and background were coded reversibly using JPEG2000, JPEG-LS, and Deflate compressors (Table 3). Only moderate ratios of compression were obtained by digital photography dedicated algorithms such as JPEG2000 and JPEG-LS. Nonetheless, these ratios were higher than those provided by the Deflate algorithm (which is not optimized for image data). Reduction in the number of intensity levels achieved by HB operation resulted in a slight increase in compression ratio for JPEG-LS, and a large improvement for the Deflate algorithm. These results indicate that HB improved compressibility of microscopic images. However, HB caused deviation from typical image data characteristics necessary for optimal performance of standard image compression algorithms. As expected, application of HBP operation markedly increased compression ratios of JPEG2000 and JPEG-LS, but did not significantly affect the performance of Deflate. The increase in compression was similar for JPEG2000 and JPEG-LS. The resultant ratios were better by about 45% comparing with the Deflate algorithm. Even greater improvement in compression efficiency was obtained when SDR was used. However, the standard deviation of the mean compression ratio was greater in SDR than in HBP, indicating that only some specific images were compressed more efficiently with SDR than with HBP.

Table 2. Total number of significant (p=0.95) intensity levels of the CCD calculated with histogram binning (HB) for the detector and acquisition parameters used in experiments. The highest numbers of levels are indicated with asterisks, the lowest with crosses.

acq. time [s]	gain	Offset [AU]					
		0		150		400	
		HB	SDR	HB	SDR	HB	SDR
0.250	5	74	66	72	61	70	52
	10	36	32	36	32	34	26
	15	24	22	24	21	**23**$^{+}$	**19**$^{+}$
0.750	2	**190***	**165***	186	146	178	124
	5	74	65	72	60	69	55
	10	36	32	36	26	34	28

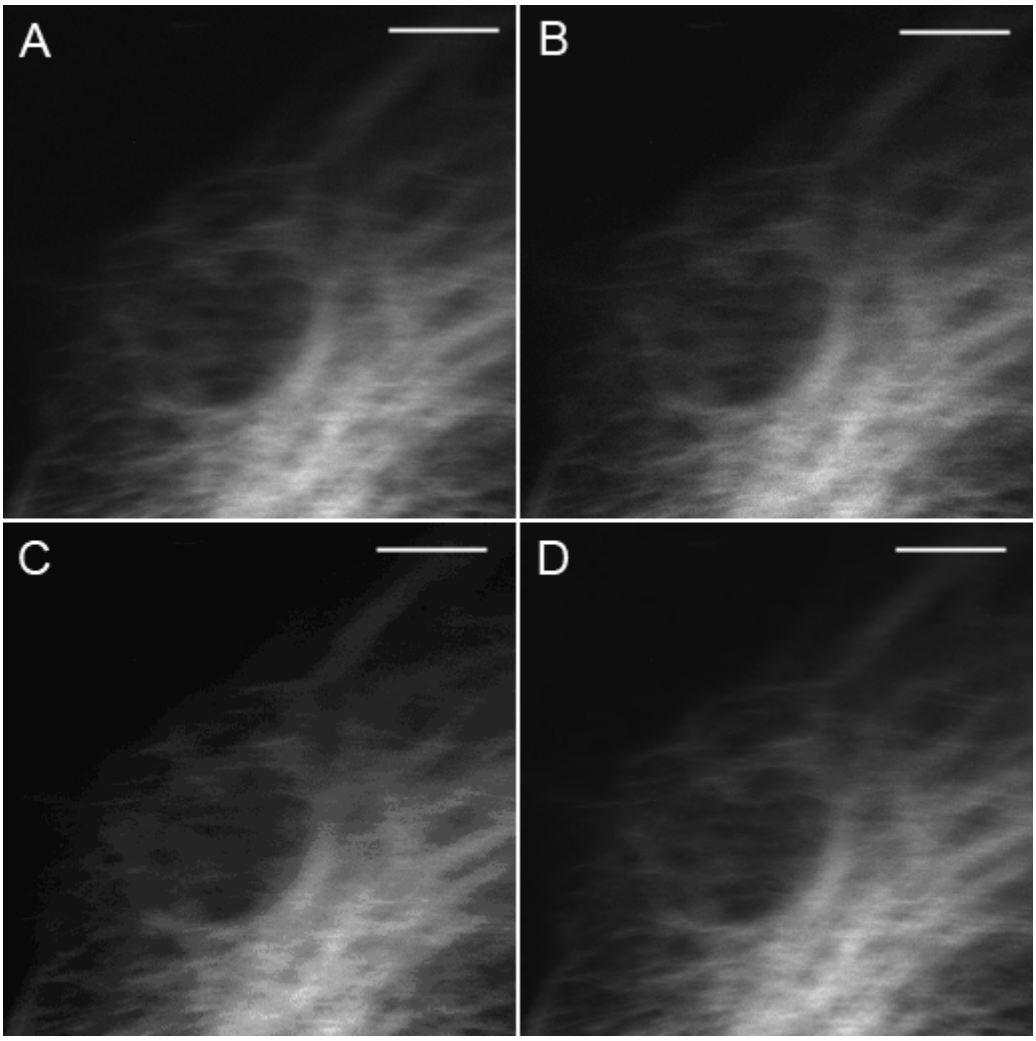

Fig. 2. Micrograph of fluorescently (BODIPY) immunostained tubulin in a fibroblast, registered at gain of 10, acquisition time of 0.75 s, and black level of 150 units. The image is shown using nominal (A) and reduced number of intensity levels, calculated with histogram binning, HB (B), and simple depth reduction, SDR (C). For comparison a raw image compressed using irreversible JPEG2000 (with the same ratio as obtained with HBP and reversible JPEG2000 coding) is shown (D). Image segmentation was performed using HB with p=0.95. Scale bar 10 μm.

Table 3. Compression efficiency of micrographs with full and reduced number of intensity levels. The images were reversibly coded with JPEG2000, JPEG-LS, and Deflate. Compression ratios are expressed as the average ± standard deviation.

Intensity depth reduction method	Coding algorithm		
	JPEG 2000	**JPEG LS**	**Deflate**
None	1.80 ± 0.21	1.79 ± 0.20	1.13 ± 0.08
Histogram binning (HB)	1.94 ± 0.21	2.87 ± 0.26	6.46 ± 0.52
Histogram binning + packing (HBP)	9.45 ± 0.47	9.55 ± 0.50	6.53 ± 0.51
Simple reduction (SDR)	11.44 ± 1.79	11.48 ± 1.53	8.29 ± 1.08

3.4. Conservation of image information in histogram binning and simple depth reduction

3.4.1. Alteration of global intensity distribution in compression

The intensity histograms of the processed images were compared quantitatively to the histograms of their raw counterparts. Average EMD values for HB-processed images with histogram binning (HB) and their raw counterparts (see Table 4) were small comparing to the average distance between nearest significant intensity levels (the nominal intensity range divided by the number of significant levels, see Table 2). This notion indicates that HB did not introduce gross changes to global intensity distributions. Greater alterations were introduced by simple depth reduction (SDR). On the other hand, compression of raw images with irreversible JPEG2000 (compression ratio was adjusted to match efficiency of HBP with reversible JPEG2000) produced much smaller histogram alterations than HB did.

3.4.2. Alteration of local intensity distribution in compression

Local intensity distribution in the images was characterized using Haralick texture parameters based on GLCM, wavelet energy, and gray-level runlength (see Materials and Methods). The parameters which provided strongest discrimination between raw and HB-processed images were identified using linear discriminant analysis (LDA). As a result of LDA the following parameters were studied in detail: wavelet energy (LH and HH bands), runlength (fraction and short run emphasis), and Haralick features (contrast, correlation, inverse difference moment, difference average).

3.4.3. Wavelet energy

Application of HB resulted in a modest increase of wavelet energy in five out of six decomposition bands (Fig. 3A-E). No change of energy was observed in the LH band at the third decomposition level (Fig. 3F). This band corresponded to details 8 to 16 pixels in size. A larger increase in the energy in the former five bands was detectable when SDR was used (Fig. 3A-E). One should note that this method produced some distortions detectable at the highest (third) decomposition level (Fig. 3EF). Furthermore, the dispersion of the relative energy values was larger with SDR than with HB, indicating that the effect of image content on the magnitude of the distortion was greater with SDR than with HB. Compression with irreversible JPEG2000 (with the same ratio as obtained with HBP and reversible JPEG2000) generated no significant change of the wavelet energy in the bands at the third decomposition level corresponding to 8-pixel (Fig. 3E) and 8-16 pixel (Fig. 3F) details. However, at the second decomposition level (the HH [4-pixel details] and LH bands [4-8 pixel details]) the energy change was similar in magnitude to that produced by HBP and SDR, but opposite in sign (Fig. 3CD). Furthermore, at the first decomposition level (the HH [2-pixel details] and LH bands [2-4 pixel details]) irreversible JPEG2000 generated larger distortions than HPB (Fig. 3AB). One should note that significant dispersion (manifested at the 2-pixel detail level, Fig. 3A) indicates that the distortion was affected by the image content and was not uniform within images (as describer further).

Table 4. Alterations of fluorescence intensity distributions (histograms) in irreversible compression, as measured using EMD. Data are expressed as the average ± standard deviation.

Method	EMD
Histogram binning (HB)	19.73 ± 8.81
Simple depth reduction (SDR)	59.62 ± 27.64
JPEG2000	1.14 ± 0.60

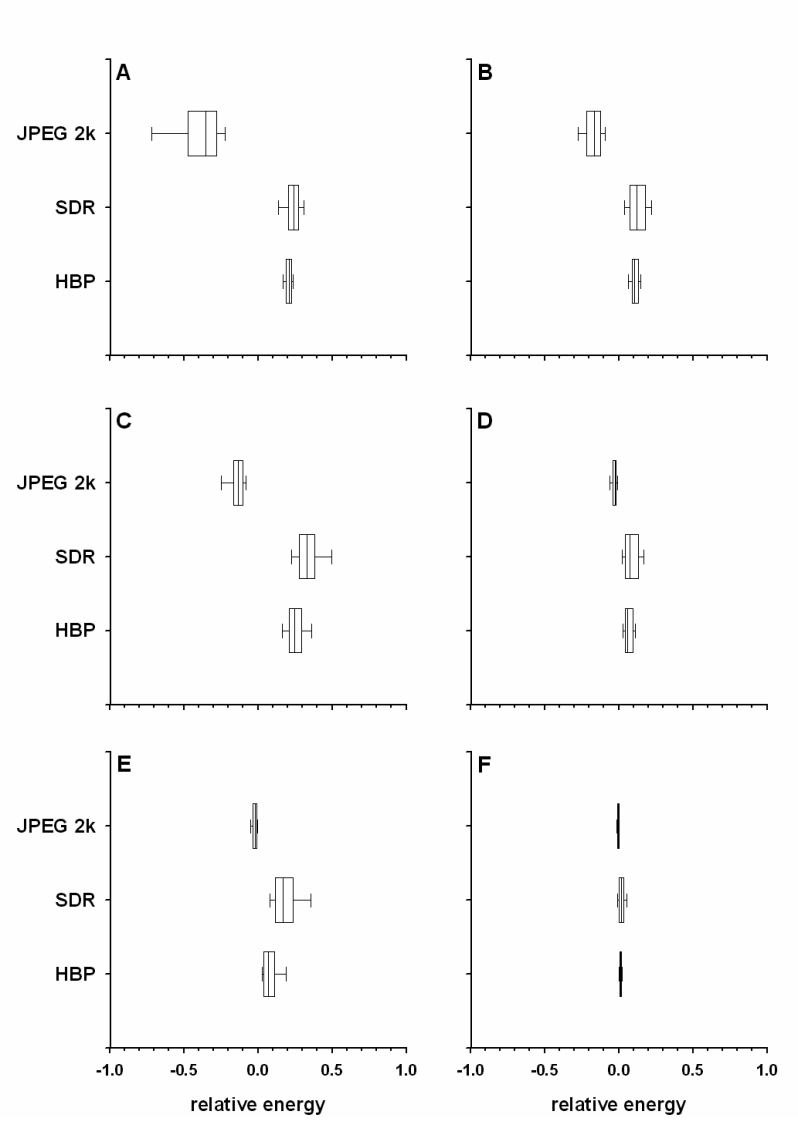

Fig. 3. Change of relative wavelet energy in compression with HBP, SDR, and irreversible JPEG 2000 (at the same compression ratio as obtained with HBP). The energy was calculated in HH (A, C, E) and LH (B,D,F) bands at first (A,B), second (C,D), and third (E,F) levels of decomposition. The data boxes represent medians with 75th percentiles, while error bars indicate corresponding 90th percentiles.

3.4.4. Haralick features

Reduction of the number of intensity levels and coding with HBP resulted in a moderate increase of GLCM contrast at the 1-pixel distance (Fig. 4A). HB did not produce changes of this parameter at 4-pixel (Fig. 4C) and 7-pixel (Fig. 4E) distances. When compression of raw images (with ratio identical to that obtained with HBP) was performed using irreversible JPEG2000, a decrease in the contrast was detected at the 1-pixel distance, whereas no changes were observed at larger distances. One should note that the decrease was similar in magnitude to the change produced by HBP at the same distance. Application of SDR resulted in an increase of the contrast detectable at all pixel distances (Fig. 4ACE). The increase was larger than that generated by HBP. Furthermore, dispersion of the correlation was greater with SDR than with HBP. One should note that the former algorithm altered GLCM correlation at all pixel distances (Fig. 4BDF). No such effect was observed when HBP or irreversible JPEG2000 was used.

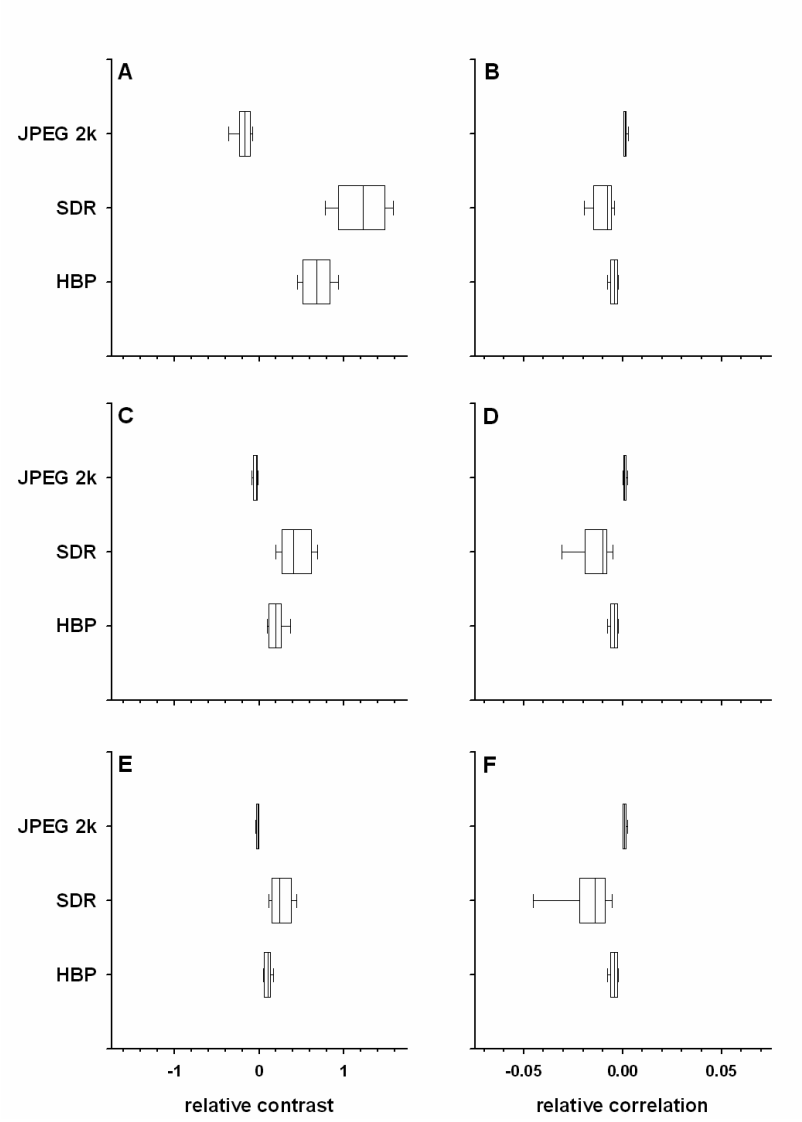

Fig. 4. Change of relative GLCM contrast (ACE) and correlation (BDF) in compression with HBP, SDR, and irreversible JPEG 2000 (at the same compression ratio as obtained with HBP). The GLCM parameters were calculated at 1-pixel (AB), 4-pixel (CD) and 7-pixel (EF) distances. The data boxes represent medians with 75th percentiles, while error bars indicate corresponding 90th percentiles.

Application of HBP produced a marked increase in inverse difference moment at distances of 1 pixel (Fig. 5A), 4 pixel (Fig. 5C) and 7 pixels (Fig. 5E). The magnitude of this increase was similar at all distances. Only small dispersion of GLCM inverse difference moment values was observed (at all distances) when HBP was used. Larger median increase and dispersion were detected when SDR was applied (Fig. 5ACE). Both these parameters were dependent on pixel distance when this algorithm was used. On the other hand, application of irreversible JPEG2000 resulted in increase of the inverse difference moment only at the 1-pixel distance. Furthermore, the magnitude of this change was smaller (by a factor of 10) than that generated by HBP. The opposite situation could be observed when texture alterations were studied using GLCM difference average. Application of HBP did not generate any significant changes of this parameter (Fig. 5BDF), whereas irreversible JPEG2000 produced a marked decrease at the 1-pixel distance and a smaller decrease at the 4-pixel distance. One should note that

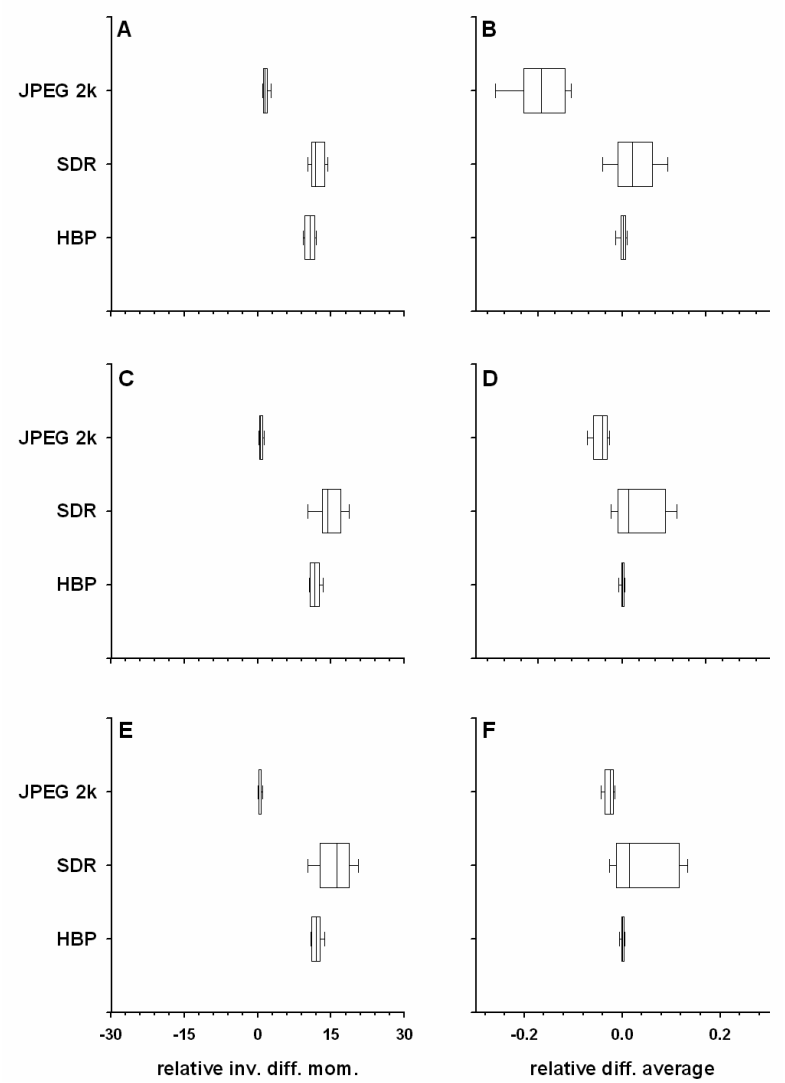

Fig. 5. Change of relative GLCM inverse difference moment (ACE) and difference average (BDF) in compression with HBP, SDR, and irreversible JPEG 2000 (at the same compression ratio as obtained with HBP). The GLCM parameters were calculated at 1-pixel (AB), 4-pixel (CD) and 7-pixel (EF) distances. The data boxes represent medians with 75th percentiles, while error bars indicate corresponding 90th percentiles.

large dispersion of this parameter was detectable at the 1-pixel distance. This indicates that the distortion was affected by the image content (as described further). As with HBP, use of SDR did not result in a significant median change of the correlation. However, dispersion of this parameter was large at all pixel distances.

3.4.5. Runlength parameters

Application of HB and SDR resulted in a decrease of runlength short-length emphasis (Fig. 6). The median decrease was similar in all directions (compare panels A,B,C, and D in Fig. 6) and the dispersion was small in all cases. Compression with reversible JPEG2000 did not alter median the value of this parameter. However, dispersion short-length emphasis indicated that changes in texture could be

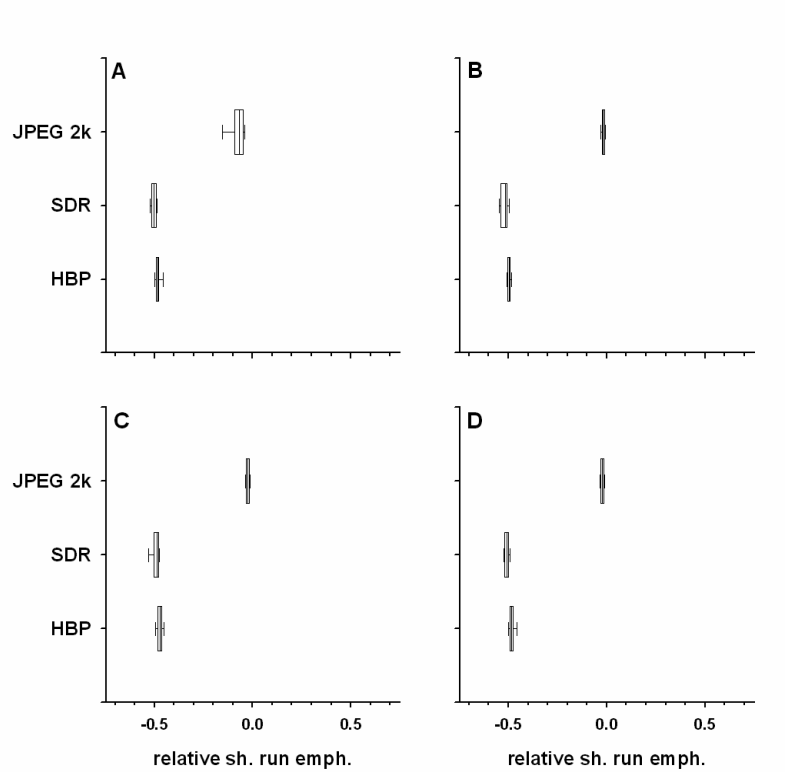

Fig. 6. Change of relative runlength short-run emphasis in images compressed with HBP, SDR, and irreversible JPEG 2000 (at the same compression ratio as obtained with HBP and reversible JPEG2000 coding). The parameter was calculated at directions corresponding to 0 degrees (A), 45 degrees (B), 90 degrees (C), and 135 degrees (D). The data boxes represent medians with 75th percentiles, while error bars indicate corresponding 90th percentiles

present in some of the images. Furthermore, the dispersion was detectable only at 0 degrees (Fig. 6A), and not at 45, 90, or 135 degrees (Fig. 6B-D). This notion indicates directional character of these alterations. Like the previous parameter the runlength fraction was decreased when images were compressed using HB and SDR (Fig. 7). Again the magnitude of the decrease was similar in all directions and the values exhibited small dispersion. No changes of this parameter were detected in the images compressed with irreversible JPEG2000.

3.4.6. Local vs. global image distortion

The presented measures of image distortion deliver one estimate which takes into account all pixels in a studied image. However, one may hypothesize that the distortions are not distributed uniformly in images. Closer investigation of images reveals no such effect in the images compressed with HBP (Fig. 8B) and SDR (Fig. 8C), where dim and bright image regions are altered in similar manner. However, in the images compressed with irreversible JPEG2000 pronounced smoothing is detectable in the dim but not in the bright image regions (compare Fig. 8A and 8D). Furthermore, texture artifacts (in the form of crosses) are introduced in the dim image regions when this algorithm is used. This notion indicates that compression with irreversible JPEG2000 resulted in both removal of original and introduction of new texture elements.

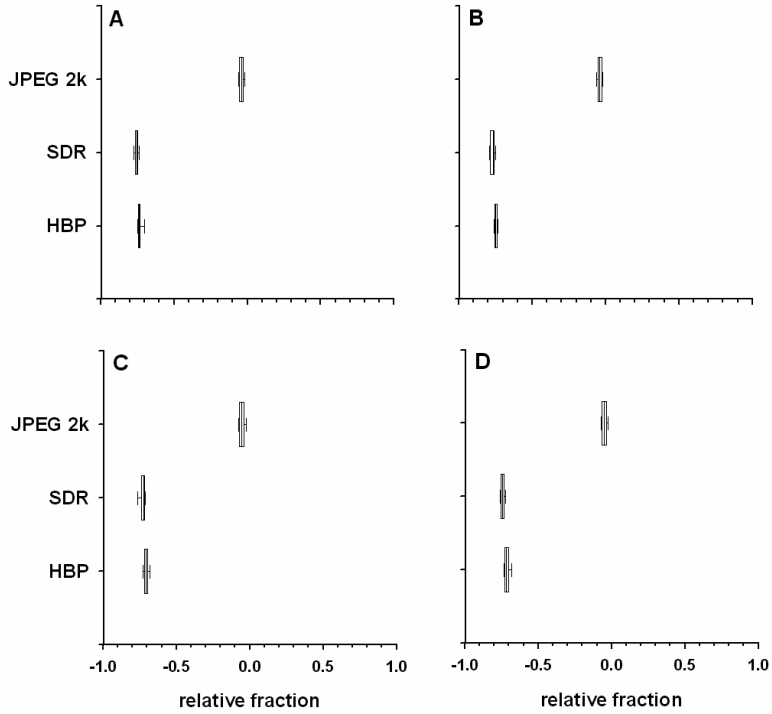

Fig. 7. Change of relative runlength fraction in images compressed with HBP, SDR, and irreversible JPEG 2000 (at the same compression ratio as obtained with HBP). The parameter was calculated at directions corresponding to 0 degrees (A), 45 degrees (B), 90 degrees (C), and 135 degrees (D). The data boxes represent medians with 75th percentiles, while error bars indicate corresponding 90th percentiles

4. DISCUSSION

Owing to the heterogeneity of imaged specimens and the presence of detector noise, typical biological micrographs are characterized by a large number of intensity levels. Using characteristics of noise as a function of signal one may reduce the number of levels with intensity-binning (HB) procedure. The resultant images have sparse histograms, i.e., a small number of intensity levels distributed throughout the nominal intensity range. One should note that image-compression algorithms are not optimized for this type of histograms. Indeed, the presence of sparse histograms has been reported to worsen compression ratios of images coded with reversible algorithms [**16**]. Therefore, to take advantage of a small number of intensity levels, histogram-packing techniques should be used with reversible coding.

This idea is in agreement with the observation that only a small increase in compression ratios was obtained when images with sparse histograms were coded using reversible algorithms optimized for image data (JPEG-LS and JPEG2000). On the other hand, compression efficiency was significantly improved when a universal coding algorithm (such as Deflate) was used. As expected, the highest compression ratios for biological fluorescence micrographs were obtained when histogram packing (employing HBP or SDR) was combined with reversible image coding (JPEG-LS and JPEG2000). The ratios were nearly an order of magnitude higher than those typically obtained for the reversible coding used with raw data. One should note that histogram packing did not improve compression efficiency for universal coding. Usually, higher ratios were obtained with SDR than with HBP. However,

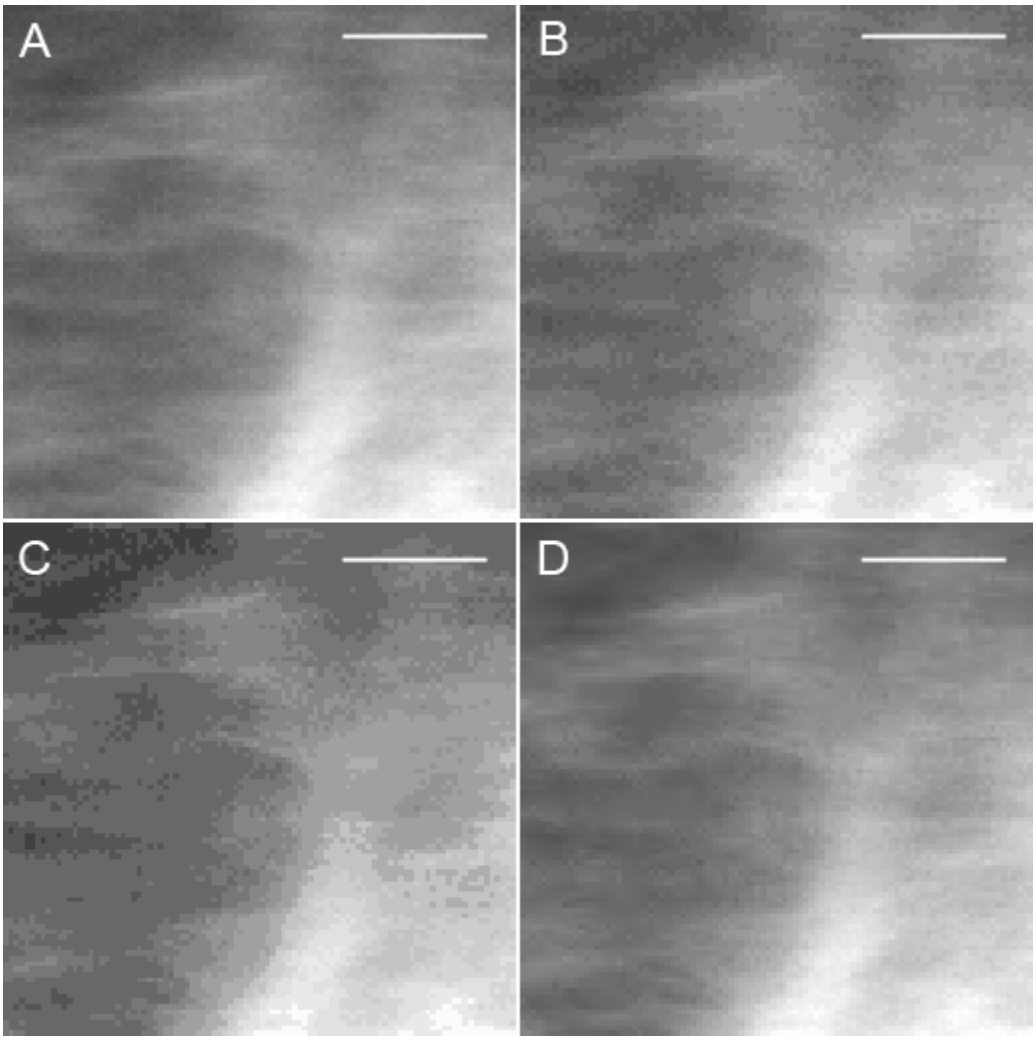

Fig. 8. Enlarged central region of micrograph of fluorescently immunostained fibroblast (presented in the Fig. 2). The image is shown with gamma of 1.3 and using nominal (A) and reduced number of intensity levels, calculated with histogram binning, HB (B), and simple depth reduction, SDR (C). For comparison a raw image compressed using irreversible JPEG2000 (with the same ratio as obtained with HBP and reversible JPEG2000 coding) is shown (D). Image segmentation was performed using HB with p=0.95. Scale bar 5 μm.

this notion was true for images characterized by high background (i.e., registered at high offset values), so it may not be regarded as a general rule.

Compression pre-processing with HBP does not markedly affect the global intensity distribution (image histogram) as estimated with EMD. The average change in pixel intensity was 0.48% of the nominal intensity range (4096 levels). Greater distortion was observed when SDR was used (1.46% of the intensity range). However, both these values were smaller than the average distance between nearest intensity levels in the images processed with HBP and SDR (2.57%). Histogram alterations generated by irreversible JPEG2000 were negligible when compared to SDR and HBP (0.03% respectively).

Processing images with HBP and SDR does significantly alter local intensity distribution measured with wavelet energy, regardless of decomposition band. On the other hand, distortions detectable in the decomposition bands corresponding to fine details were produced when irreversible

JPEG2000 was used. One should note that these distortions were dependent on image content and were not uniform within images. Irreversible JPEG2000 compression and HBP/SDR produced changes in Haralick contrast and correlation similar in magnitude but opposite in sign. Larger dispersion of these values was noted only when SDR was used, indicating some dependence on image content. Irreversible JPEG2000 generated no texture changes detectable with GLCM inverse difference moment, whereas such changes were observed for HBP and SDR. Conversely, JPEG2000 generated changes in GLCM correlation at small distances (corresponding to fine details), whereas HBP did not influence this value. Large dispersion of the correlation values suggests dependence of JPEG2000-generated distortion on image content. Similar large dispersion was noted for SDR, albeit with smaller median compared to the respective value for JPEG2000.

HBP and SDR produced large decreases in runlength fraction and short-run emphasis. However, this change was similar in all the images (as indicated by small dispersion) and independent of direction. No significant change of median values of these parameters was introduced by irreversible JPEG2000. A large dispersion fraction values indicated directional character of the changes.

It should be emphasized that the texture parameters were specifically picked to detect changes introduced to the images by reduction in the number of intensity levels. Therefore the presented results may be considered to be upper estimates of image distortions introduced by SDR and HBP. Therefore, this specific set of texture parameters might not detect all distortions introduced by irreversible JPEG2000. Nonetheless, performance of HPB was superior to irreversible JPEG2000 in the sense of many texture parameters. What is more, the effects produced by the former were similar within the image set and uniform within single images. Hence, possible distortions were introduced in a consistent and non-arbitrary manner. This is not surprising since reduction of the number of intensity levels was performed using the detector characteristics but not image-content information. JPEG2000 and other popular irreversible image-compression algorithms do not offer these advantages as they are optimized to preserve perceptual image quality.

Although objective metrics of image fidelity are used in this work, one may note also that the changes introduced to images by tested methods could be detected on visual inspection. These changes were clearly manifested on the large scale SDR method (Fig. 2 and 8). Similar effects of HBP or JPEG2000 (at ratio identical to HBP) could be noticed when pixels become distinguishable with naked eye (small scale, Fig. 1 and 8). Since these differences may not always be evident in print or pdf file (as creating these is an operation of image processing) we provide original (uncompressed bitmap) figures upon request.

Intensity binning presented in this work was executed so that the levels were different from one another with 95% probability (confidence). The choice of confidence interval was arbitrary. Similar calculations can be performed for every confidence level. Therefore, one may optimize the statistical significance of intensity differences to obtain a desired compression ratio. In other words limited storage space or network transmission bandwidth may be used optimally so as to accommodate maximum amount of scientific information. Since such optimization depends only on detector characteristics and is not affected by image content management of the storage and bandwidth resources may be facilitated by intensity binning. This advantage may be especially important in the storage systems consisting of separate volumes connected by a network. Furthermore, image data corresponding to various levels of statistical significance may be encoded sequentially providing progressive image transmission over a network. One should note that data may be transmitted in this scheme in the order of statistical significance (as opposed to ("visual importance"). This may be an important advantage in the fields of telemicroscopy [**31**, **32**] and telemedicine [**32**, **33**].

The proposed HBP algorithm can be used directly with JPEG2000 and JPEG-LS compressors/decompressors. In fact, direct support for histogram packing, designed initially for palette images, is part of the JPEG-LS baseline standard [**19**]. Moreover, the 2nd part of the JPEG2000 standard [**7**] describes two generic non-linear transformations that may be applied to decoded pixel intensity levels: the piece-wise linear function and the gamma-style function. Encoding packed histogram as a mapping table is a special case of the latter. Therefore these algorithms are capable of reconstructing an image and its original histogram employing their standard functionality (i.e., no additional step of histogram expansion is required). Furthermore, the JPEG-LS and JPEG2000 standards define only the decoding parts of the respective algorithms and the format of the compressed data stream. Provided that histograms are encoded in the required format, the whole process of applying HBP and core JPEG-LS/JPEG2000 image-coding routines is compliant with these standards.

The SDR and HP algorithms, applied in this study for monochrome images, might also be used with color data. It should be noted that color channels are usually transformed (Multi-Component Transform) to remove their mutual correlation. Therefore, if such operation was applied directly to multi-channel images processed with HB an increase of the number of levels would be expected. Consequently, the compression results could be worsened. This problem may be avoided by applying SDR/HP to individual channels of color image after subjecting them to HB and prior to Multi-Component Transform.

The presented algorithms were used with commercial package for modeling of camera noise. In the next stage open-source implementation of SDR and HP together with the noise analysis routine is envisaged.

5. CONCLUSIONS

The results presented show that HBP pre-processing provides significant enhancement in compression efficiency and does not introduce significant changes to biological micrographs. The algorithm has better fidelity than irreversible JPEG2000 (at the same compression rate) and can be easily implemented within the context of the current image-compression standards. This makes the proposed procedure an attractive enhancement to data-acquisition packages developed for biological imaging systems.

Acknowledgement

This work was supported by the Polish Ministry for Science and Higher Education (MNiSW) grant Nr N N301 463834. (TB).

References

[1] T. Bernas, J. P. Robinson, Z. Darzynkiewicz, W. Hyun, A. Orfao and P. Rabinovitch, Basics of digital microscopy, J. P. Robinson (Editor), Wiley-Liss, New York, 2005, pp. 1-14.

[2] US Federal Government, 21CFR11 Code of Fedaral Regulations Title 21 Chapter I Part 11—Electronic Records; Electronic Signatures, USA, 2004.

[3] C. C. Chen, On the selection of image compression algorithms, Proceedings of the 14th Int'l Conference on Pattern Recognition, Brisbane, Australia, 1998, pp. 1500-1504.

[4] P. C. Cosman, R. M. Gray and R. A. Olshen, Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy, Proceedings of the IEEE 82 (1994), pp. 6919-932.

[5] ISO/IEC, The basic JPEG standard, ISO/IEC International Standard 10918-1 and ITU-T recommendation T.81 part 1, 1994.

[6] ISO/IEC, Information Technology – JPEG2000 image coding system: Core coding system, ISO/IEC International Standard 15444-1 and ITU-T recommendation T.800, 2004.
[7] ISO/IEC, Information Technology – JPEG2000 image coding system: Extensions, ISO/IEC International Standard 15444-2 and ITU-T recommendation T.801, 2004.
[8] B. Wohlberg and G. de Jager, A review of the fractal image coding literature, IEEE Transactions on Image Processing 8 no. 12, 2002, pp. 1716-1729.
[9] F. Ebrahimi, M. Chamik and S. Winkler, JPEG vs. JPEG2000: An objective comparison of image encoding quality, Proc. SPIE Applications of Digital Image Processing 5558 (2004), pp. 300-308.
[10] S. Grgic, M. Mrak, M. Grgic and B. Zovko-Cihlar, Comparative study of JPEG and JPEG2000 image coders, Proceedings of the 17th Conference on Applied Electromagnetics and Communications (2003), pp. 109-112.
[11] J. Paz, M. Perez and I. Miranda, Diagnostic quality of high resolution JPEG 2000 compressed CT and MR brain images, IFMBE Proceedings 25/5 (2009) pp. 334-337
[12] F. Zanca, J. Jacobs, and H. Bosmans, Evaluation of clinical image processing algorithms used in digital mammography, Med. Phys. 36, (2009), pp. 765-776
[13] T. Bernas, E. K. Asem, J. P. Robinson and B. Rajwa, Estimation of imaging precision for microscope calibration and image compression., J. Microscopy 226 (Pt 2), (2007), pp. 163-74.
[14] P. Young, Environmental Modelling and Software, Data-based mechanistic modeling, 13 (1998), no. 2, pp. 105-122.
[15] A. Amer, E. Dubois and A. Mitiche, Reliable and fast structure-oriented video noise estimation, Proc. IEEE Int. Conf. On Image Proc. 1 (2002), pp. 840-843.
[16] A. J. Pinho, On the impact of histogram sparseness on some lossless image compression techniques, International Conf. on Image Proc., 2001, pp. 442-445.
[17] T. Bernas, B. Rajwa, E. K. Asem and J. P. Robinson, Compression of fluorescence microscopy images based on the signal-to-noise estimation, Micr. Res. Tech. 69 (2006), no. 1, pp. 1-9.
[18] R. Starosolski, Compressing images of sparse histograms, Proc. SPIE Medical Imaging 5959, 2005, pp. 209-217.
[19] ISO/IEC, Information Technology – Lossless and near-lossless compression of continuous-tone still images – baseline, ISO/IEC International Standard 14495-1 and ITU-T recommendation T.87, 1999.
[20] National Electrical Manufacturers Association (NEMA), Digital imaging and communications in medicine (DICOM) Part 5: Data Structures and Encoding, NEMA Standard PS 3.5-2009, Rosslyn, USA, 2009.
[21] S. W. Golomb, Run-length encodings, ieee trans. On information theory, IEEE Trans. Inf. Theor. IT-12 (1966), pp. 399-401.
[22] R. F. Rice, Some practical universal noiseless coding techniques – part iii, Jet Propulsion Laboratory JPL 91-3, 1979.
[23] M. J. Weinberger, G. Seroussi and G. Sapiro, A low complexity, context based, lossless image compression algorithm, Proc. IEEE Data Compression Conference, 1996, pp. 140-149.
[24] A. Moffat, R. M. Neal and I. H. Witten, Arithmetic coding revisited, ACM transactions on information systems, 16 (1998), no. 3, pp. 256–294.
[25] C. Christopoulos, A. Skodras and T. Ebrahimi, The JPEG2000 still image coding system an overview, IEEE trans. on consumer electronics 46 (2000), no. 4, pp. 1103-1127.
[26] P. Deutsch, Network working group, request for comments: 1951 – deflate compressed data format specification version 1.3, 1996.

[27] J. Ziv and A. lempel, A universal algorithm for sequential data compression, IEEE Trans. Inf. Theor. 32 (1977), no. 3, pp. 337-343.

[28] Y. Rumner, C. Tomassi and J. G. Leonidas, Earth mover's distance as a metric for image retrieval, Int. J. Comp. Vis. 40 (2000), pp. 99-121.

[29] A. Materka and M. Strzelecki, Texture analysis methods – a review, COST B11 report (1998), pp. 1-33.

[30] M. Tuceryan, A. K. Jain, Texture analysis, in Handbook of Pattern Recognition and Computer Vision (2nd Edition) by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), World Scientific Publishing Co., 1998, pp. 207-248.

[31] P. J. Hines, Telemicroscopy - a review, Australian Conference on Microscopy and Microanalysis ACMM19, Sydney, Australia, 2006.`

[32] H. K. Huang, Telemedicine and Teleradiology in PACS and Imaging Informatics: Basic Principles and Applications, Wiley-Blackwell; 2 edition (2010), pp. 451 - 468

[33] M. G. Strintzis, A review of compression methods for medical images in PACS, International Journal on Medical Informatics 52 (1998), pp. 159 - 165.