Low tensor train and low multilinear rank approximations of 3D tensors for compression and de-speckling of optical coherence tomography images

Ivica Kopriva[1,7], Fei Shi[2] , Mingyig Lai[3], Marija Štanfel[4], Haoyu Chen[5], and Xinijan Chen[2,6]

[1] Division of Electronics, Ruđer Bošković Institute, Zagreb, Croatia
[2] School of Electronics and Information Engineering, Soochow University, Suzhou, China
[3] Shenzen Eye Hospital, Shenzen, China
[4] Department of Ophthalmology, University Hospital Centre Zagreb, Zagreb, Croatia
[5] Joint Shantou International Eye Centre, Shantou University and the Chinese University of Hong Kong, Hong Kong, China
[6] State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China
[7] Author to whom any correspondence should be addressed

E-mail: ikopriva@irb.hr

Keywords: 3D optical coherence tomography, compression, de-speckling, tensor train rank, multilinear rank, Schatten p-norm, surrogate of Schatten-0 norm

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

Objective. Many methods for compression and/or de-speckling of 3D optical coherence tomography (OCT) images operate on a slice-by-slice basis and, consequently, ignore spatial relations between the B-scans. Thus, we develop compression ratio (CR)-constrained low tensor train (TT) - and low multilinear (ML) rank approximations of 3D tensors for compression and de-speckling of 3D OCT images. Due to inherent denoising mechanism of low-rank approximation, compressed image is often even of better quality than the raw image it is based on. *Approach*. We formulate CR-constrained low rank approximations of 3D tensor as parallel non-convex non-smooth optimization problems implemented by alternating direction method of multipliers of unfolded tensors. In contrast to patch- and sparsity-based OCT image compression methods, proposed approach does not require clean images for dictionary learning, enables CR as high as 60:1, and it is fast. In contrast to deep networks based OCT image compression, proposed approach is training free and does not require any supervised data pre-processing. *Main results*. Proposed methodology is evaluated on twenty four images of a retina acquired on Topcon 3D OCT-1000 scanner, and twenty images of a retina acquired on Big Vision BV1000 3D OCT scanner. For the first dataset, statistical significance analysis shows that for CR≤35, all low ML rank approximations and Schatten-0 (S_0) norm constrained low TT rank approximation can be useful for machine learning-based diagnostics by using segmented retina layers. Also for CR≤35, S_0-constrained ML rank

approximation and S_0-constrained low TT rank approximation can be useful for visual inspection-based diagnostics. For the second dataset, statistical significance analysis shows that for CR≤60 all low ML rank approximations as well as S_0 and $S_{1/2}$ low TT ranks approximations can be useful for machine learning-based diagnostics by using segmented retina layers. Also, for CR≤60, low ML rank approximations constrained with S_p, $p \in \{0, 1/2, 2/3\}$ and one surrogate of S_0 can be useful for visual inspection-based diagnostics. That is also true for low TT rank approximations constrained with S_p, $p \in \{0, 1/2, 2/3\}$ for CR≤20.
Significance. Studies conducted on datasets acquired by two different types of scanners confirmed capabilities of proposed framework that, for a wide range of CRs, yields de-speckled 3D OCT images suitable for clinical data archiving and remote consultation, for visual inspection-based diagnosis and for machine learning-based diagnosis by using segmented retina layers.

1. Introduction

Optical coherence tomography (OCT) is noninvasive imaging modality widely used in diagnostic ophthalmology (Fang *et al* 2015, Huang *et al* 1991). It is founded on the measurement technique of interferometry (Schmitt *et al* 1999, Huang *et al* 1991), that itself relies on spatial and temporal coherence of optical waves backscattered from a tissue. Thus, quantification of optical properties of the tissue enables discrimination of different pathological states of tissue (Cheong *et al* 1990, Wilson and Jacques 1990, Chen *et al* 2015). However, the same coherence gives rise to speckle, a fundamental property of signals acquired by narrowband detection systems (Schmitt *et al* 1999). Speckle reduces contrast and makes boundaries between constitutive tissues more difficult to resolve (Schmitt *et al* 1999, van Soest G *et al* 2012, Karamata *et al* 2005). That in, return, stands for a major obstacle in quantitative OCT image analysis. Since OCT itself can be classified as a special adaptation of electronic speckle pattern interferometry (Leith *et al* 1991), one can argue that complete suppression of speckle is not desirable (Schmitt *et al* 1999). That is what makes speckle suppression a challenging problem to deal with. It is, therefore, a subject of many research efforts. In particular, speckle suppression methods belong into two groups: physical compounding and digital filtering. The first group suppress speckle by incoherent summation of different realizations of the same OCT image (Szkulmowski *et al* 2012). The second group of methods is based on post-processing of observed OCT image (Ozcan *et al* 2007). To that group also belong matrix decomposition methods (Kopriva *et al* 2017, Baghaie *et al* 2015, Chen *et al* 2017) and matrix completion method (Cheng *et al* 2016). Significant performance advance in OCT image enhancement is brought by sparse representations in learned dictionary (Abassi *et al* 2018, Fang *et al* 2015, Fang *et al* 2017). These methods require high signal-to-noise ratio images for dictionary learning. Their performance is maximal when work with small patches (8x8 pixels) and small step size (1 pixel). That makes them computationally expensive (see Table 2). Recent advances in the increase of the performance of OCT image enhancement are brought by algorithms that are based on deep learning (Ma *et al* 2018, Shi *et al* 2019, Wang *et al* 2021, Zhou *et al* 2022).

To improve clinical analysis ophthalmologists need large field-of-view and high spatial resolution OCT images (Fang *et al* 2015). However, storage and transmission of high spatial and temporal resolution OCT images creates a major obstacle for clinical data archiving systems and remote consultation, diagnosis and monitoring of diseases, such as age-related macular degeneration detection (Rasti *et al* 2018, Guo *et al* 2020). Hence, development of efficient image compression strategies is of great importance. At disposal are standard (general purpose) image compression methods such as the joint photographic expert group (JPEG) (Wallace 1991), which adapts discrete cosine transform (DCT) (Ahmed *et al* 1974), JPEG 2000 that is based on discrete wavelet transform (DWT) (Skodras *et al* 2001), the 3D SPIHT image compression method (Said and Pearlman 1996), or singular value decomposition (SVD) based image compression methods (Laxmi *et al* 2019, Yeganegi *et al* 2018, Andrews and Patterson 1976). As pointed out in (Fang *et al* 2015), JPEG and JEPG200 are designed for 2D images and do not take into account neighborhood information (correlations) among the B-scans. Furthermore, DCT and DWT are built on fixed mathematical models (Rubinstein *et al* 2010), i.e. they lack adaptability to specific classes of images such as retinal OCT volumes. SVD-based image compression methods are also matrix-based and do not take into account tensorial nature of the 3D OCT images. Regarding sparsity- and patch-based image compression, it is known that compression ratio (CR) is proportional to the square of the step size and proportional to the sparsity of the code, see section 3.2. In order to achieve high CR patch size as well as sparsity of learned code have to be increased. In that case, however, quality of the reconstructed OCT image deteriorates significantly (see section 3.3). That is why sparsity- and patch-

based image compression are presented in (Fang *et al* 2015) for CR=10. For the sake of comparative performance analysis, we modified nonlocal weighted sparse representation (NWSR) algorithm (Abassi *et al* 2018) for testing with CR≤10. SVD-based image compression also works with 2D images with CR that is proportional to the patch (block) size, see section 3.2 for details. For large CR a larger patch size is required. To achieve CR=60, that is possible with the low tensor rank based methods proposed herein, or CR=80 as in deep network based method (Guo *et al* 2020), SVD-based compression has to be implemented on the whole image as a patch. In addition to speckle suppression, deep networks are also useful in OCT image compression (Guo *et al* 2020). However, deep learning approach in the training phase depends on the prior knowledge such as segmentation of the retina layers in data preprocessing stage, and high quality reference images for estimation of the multi-scale structural similarity image metric (Wang *et al* 2003).

Motivated by limitations of OCT image compression methods discussed above, we propose low tensor rank approach to 3D OCT image compression. It treats the 3D OCT image as a whole, and not on a scan-by-scan basis. It is also motivated by the observation that sparsity-based representations poses inherent denoising mechanism (Fang *et al* 2015). Since low-rank representations are actually sparsity-based representations in singular values domain, this observation is also relevant for SVD-based OCT image compression and for low tensor rank approach to 3D OCT image compression. In particular, we propose low tensor rank approximation of 3D OCT images through compression with the pre-specified CR. Our approach is based on tensor train (TT) model (Oseldets 2011) and low multilinear (ML) model (Tucker 1966, Lee and Cichocki 2018). Unlike deep learning OCT image compression method (Guo *et al* 2020), it does not require training phase. Unlike sparsity- and patch-based image compression methods (Fang *et al* 2015, Abassi *et al* 2018) it enables high CRs. Related optimization problems are based on alternating direction method of multipliers (Boyd *et al* 2010) for which convergence towards stationary points is proven theoretically. Since it is known that intelligent selection of singular values is of key importance for the quality of SVD-based image compression and reconstruction (Andrews and Patterson 1976), selection of singular values in optimization problems related to unfolded matrices in modes of 3D OCT tensor is carried out through the specific class of thresholding operators, a.k.a. proximal operators. Thereby some of them are data adaptable. Proposed low tensor rank 3D OCT image compression methods are also computationally efficient, see Table 2, due to scheme that avoids direct calculation of related tensor decompositions. For mathematical aspects/details of proposed algorithms, we refer interested readers to the section 1 in the Supplement. Statistical significance analysis conducted on images from two 3D OCT scanners shows that at least all low ML rank approximations for CR≤35 and Schatten-0 (S_0) norm constrained low TT rank approximation for CR≤35 yield de-speckled images that can be used for machine learning-based diagnostics by using segmented retina layers. Furthermore, at least $S_{2/3}$ norm constrained low ML rank approximation for CR≤35 yields de-speckled images that can be used for visual inspection-based diagnostics. For a more comprehensive presentation of the results, the readers are referred to Tables 1 and 3.

The rest of the paper is organized as follows. Section 2 presents low TT- and low ML rank approximations of 3D tensors with the focus on OCT imaging and its corresponding issues of compression and speckle suppression. Section 3 presents experimental results and comparisons with the state-of-the-art methods for image compression, including OCT image compression, and no-compressions speckle suppression methods. In particular, the methods are compared in terms of accuracy of segmentation of retina layers and expert-based image quality metrics. Discussion is given in section 4, and conclusion in section 5.

2. Methods

2.1 Methods overview

In this paper we propose low TT- and low ML rank approximations of 3D OCT image for compression and speckle suppression. Due to the character of the journal detailed mathematical derivations are presented in section 1 of the Supplemement. Here we present the methods on elementary level. We denote 3D OCT image tensor as $\underline{\mathbf{X}} \in \mathbb{R}_0^{I_1 \times I_2 \times I_3}$ where I_1 and I_2 stand respectively for number of rows and columns, and I_3 stands for number of B-scans. Multilinear tensor model is given with (Tucker 1966, Lee and Cichocki 2018):

$$\underline{\mathbf{X}} \cong \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \tag{1}$$

where $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is called core tensor and $\left\{\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R_n}\right\}_{n=1}^{3}$ are called factor matrices. $\underline{\mathbf{G}} \times_n \mathbf{A}^{(n)}$ denotes the mode-n

product between the tensor $\underline{\mathbf{G}}$ and the matrix $\mathbf{A}^{(n)}$. (R_1, R_2, R_3) is called the multilinear rank of Tucker decomposition (Lee and Cichocki 2018). CR-constrained low ML rank approximation of a tensor $\underline{\mathbf{X}}$ is addressed by the optimization problem:

$$\hat{\mathbf{X}}_{(n)} = \min_{\mathbf{W}_n} rank\left(\mathbf{W}_n\right) \quad \text{such that: } \mathbf{X}_{(n)} = \mathbf{W}_n + \mathbf{E}_n, \quad CR \geq CR^* \quad n = 1,2,3. \tag{2}$$

$\{\mathbf{E}_n\}_{n=1}^3$ denote the approximation errors and $\mathbf{X}_{(n)}$ denotes the mode-n matricization, a.k.a. mode-n unfolding, of tensor $\underline{\mathbf{X}}$, see definition 4 in the Supplement, and $\hat{\mathbf{X}}_{(n)}$ denotes estimate of $\mathbf{X}_{(n)}$. CR^* stands for predefined (target) value of compression ratio. The low ML rank approximation of 3D OCT image tensor $\underline{\mathbf{X}}$ is obtained as output of the Algorithm 2 in the Supplement as:

$$\hat{\underline{\mathbf{X}}} = \frac{1}{3}\sum_{n=1}^{3} fold_n\left(\hat{\mathbf{X}}_{(n)}\right) \tag{3}$$

where $fold_n$ denotes mapping from matrix to tensor operator. For ML model, CR is obtained as:

$$CR = \frac{I_1 I_2 I_3}{I_1 R_1 + I_2 R_2 + I_3 R_3} \tag{4}$$

Tensor train model is given with (Oseldets 2011, Lee and Cichocki 2018):

$$\underline{\mathbf{X}} \cong \mathbf{G}^{(1)} \bullet \underline{\mathbf{G}}^{(2)} \bullet \underline{\mathbf{G}}^{(3)} \tag{5}$$

where the third order core tensors are of the sizes $\{R_{n-1} \times I_n \times R_n\}_{n=1}^3$, and '$\bullet$' denotes contracted tensor product. It is assumed R_0=R_3=1. (R_1, R_2) is called the tensor train rank. CR-constrained low TT rank approximation of a tensor $\underline{\mathbf{X}}$ is addressed by the optimization problem:

$$\hat{\mathbf{X}}_{[n]} = \min_{\mathbf{W}_n} rank\left(\mathbf{W}_n\right) \quad \text{such that: } \mathbf{X}_{[n]} = \mathbf{W}_n + \mathbf{E}_n, \quad CR \geq CR^* \quad n = 1,2. \tag{6}$$

$\{\mathbf{E}_n\}_{n=1}^2$ denote the approximation errors, $\mathbf{X}_{[n]}$ stands for the mode-(1,2,3) matricization, a.k.a. mode-n canonical unfolding, of tensor $\underline{\mathbf{X}}$, see definition 6 in Supplement, and $\hat{\mathbf{X}}_{[n]}$ denotes estimate of $\mathbf{X}_{[n]}$. The low TT rank approximation of 3D OCT image tensor $\underline{\mathbf{X}}$ is obtained as output of the Algorithm 1 in the Supplement as:

$$\hat{\underline{\mathbf{X}}} = \frac{1}{2}\sum_{n=1}^{2} fold_n\left(\hat{\mathbf{X}}_{[n]}\right). \tag{7}$$

For TT model, CR is obtained as:

$$CR = \frac{I_1 I_2 I_3}{I_1 R_1 + R_1 I_2 R_2 + R_2 I_3} \tag{8}$$

Solutions of optimization problems (2) and (6) by Algorithms 2 and 1 are obtained without direct computation of corresponding tensor decompositions and that makes them fast. The key role in quality of OCT image $\hat{\mathbf{X}}$ reconstructed after compression play surrogates of the rank function in optimization problems (2) and (6), see sections 1.5 and 1.8 in the Supplement. For that purpose we use the S_0 quasi-norm as an exact measure of the rank, its surrogates in terms of S_p norms, $p \in \{1/2, 2/3, 1\}$ and the function:

$$f_\sigma\left(x\right) = 1 - e^{-\sigma|x|}, \ \sigma > 0. \tag{9}$$

Thresholding operators associated with these norms play a key role in smart selection of singular values in related low-rank optimization problems. We refer readers interested in more details to sections 1.2 and 1.4 in the Supplement.

The observation is made in (Fang *et al* 2015) that sparsity-based representation has inherent denoising property. Similar property is observed in sparsity-constrained blind image deconvolution (Kopriva 2005). It is clear from the above exposition that low tensor rank approach to approximation of the 3D OCT image tensor is based on sparse approximation of the vector of singular values of corresponding unfolded matrices. Thus, the low tensor rank approximation/compression method also has inherent denoising property. That was known long time ago for SVD-based compression of 2D images (Andrews and Patterson 1976). Unlike patch- and sparsity-based approach, proposed low tensor rank approach enables compression with the high CRs at the much smaller computational costs. Unlike SVD-based approach that works on a scan-by-scan basis, it enables compression of 3D OCT image tensor as a whole.

2.2 B-scan alignments method

The involuntary movement of the eye during OCT acquisition may cause the position of the retina to jump between B-scans. That compromises the B-scan wise spatial correlations. Therefore, we perform a pre-processing step called B-scan alignment (Shi *et al* 2015). The position of the retinal upper boundary, detected by a multi-resolution graph search method, is used to estimate the vertical displacement of the retina across B-scans. Specifically, a mean height value is calculated by averaging the height of the left most and the right most 20% points on the retinal upper boundary. The center part is excluded from the above calculation considering the fact that the concave-shaped fovea appears in the center of some B-scans. Then each B-scan is shifted up or down so that the mean height values of the peripheral upper boundary become the same for all B-scans. Experiments showed that this pre-processing step significantly improved performance of the proposed tensor approximation methods.

2.3 Evaluation metrics

We segmented retina layers from each enhanced 3D OCT image as well as from original images. We compared corresponding segmented layers with the ones obtained by manual segmentation. Segmentation error (SE) is calculated as:

$$SE = \frac{1}{I_2 |S| L} \sum_{i_2=1}^{I_2} \sum_{i_3 \in S} \sum_{l=1}^{L} \frac{\left| AS^l(i_2, i_3) - MS^l(i_2, i_3) \right|}{T(i_2, i_3)} \tag{10}$$

where $AS^l(i_2, i_3)$ and $MS^l(i_2, i_3)$ denote the vertical position of the *l*th segmenting surface obtained by the automatic method and the manual segmentation, respectively. $T(i_1, i_2)$ stands for the total thickness of the retina at position (i_2, i_3). Since manual segmentation of the retina layers is tedious, we used L=8 surfaces and a subset of B-scans with indices belonging to $S = \{i \times 6\}_{i=1}^{10}$ for the SE estimation (Chen *et al* 2013). Thus, the SE represents the mean absolute vertical difference between the automatic and manual segmentation relative to the total thickness of a retina. The most often used figure of merit in OCT image de-speckling is CNR (van Soest 2012). It is inversely proportional to the speckle fluctuation and defined as: $CNR = \mu_l(\underline{\mathbf{X}}) / \sigma_l(\underline{\mathbf{X}})$, where $\mu_l(\underline{\mathbf{X}})$ and $\sigma_l(\underline{\mathbf{X}})$ respectively correspond to the mean and standard deviation in some selected homogeneous part of the image $\underline{\mathbf{X}}$. Experimental results reported in section 3.3 were estimated in the region that corresponds with the topmost layer in the OCT image of a retina, indicated in Figure 5 by an arrow (Chen *et al* 2013). Furthermore, three experienced ophthalmologists assessed visual quality of original and de-speckled images. Mean and standard deviation of the expert image quality metric (EIQM) were estimated from the three grades of three B-scans of the three 3D OCT images de-speckled by each considered method as well as the original images. The EIQM ranges from 1 to 5, where 1 means the poorest quality and 5 the highest quality.

3. Experimental results

3.1 Data

Proposed algorithms were compared on two sets of 3D OCT images acquired by two scanners. The first set is comprised of twenty-four macular-centered 3D OCT images of normal eyes acquired with the Topcon 3D OCT-1000 scanner (Topcon, Tokyo, Japan). 3D OCT image tensor is comprised of I_3=64 B-scan images with the size of I_1=480 × I_2=512 pixels. The scanning range covered a retinal area of 6 mm × 6 mm. First dimension refers to width of each B-scan, so that not only the macular center but also the optical nerve head is included. These images were used previously for the study for optical intensity analysis, where they were segmented into retina layers (Chen *et al* 2013). They are publicly available online at: https://pan.baidu.com/s/1edkG7k8W3Wkjhq8vffCYng (password: frpa). The second set is comprised of twenty macular-centered 3D OCT images of normal eyes acquired with the non-commercial in-house BV1000 3D OCT scanner (BigVision, Suzhou, China): https://www.bigvisiontech.com/services/bv1000. 3D OCT image tensor is comprised of I_3=100 B-scan images with the size of I_1=1000 × I_2=512 pixels. The scanning range covered area of 12 mm × 9 mm. All images were analyzed with the software written in the MATLAB® (the MathWorks Inc., Natick, MA) script language on PC with Intel i7 CPU with the clock speed of 2.2 GHz and 16GB of RAM.

3.2 Other methods for comparison

We compared the CR-constrained 3D tensor approximation algorithms with the JPEG2000 standard for image compression and the 3D SPIHT image compression method (Said and Pearlman 1996). Since low-rank approximation is actually sparsity-based optimization in the singular values domain it is evident that observation made in (Fang *et al* 2015), sparsity-based representation has inherent denoising property, is relevant to the SVD-based image compression (Laxmi *et al* 2019, Yeganegi *et al* 2018, Andrews and Patterson 1976) as well as to low tensor rank based approaches proposed herein. Therefore, we also apply SVD-based approach to 3D OCT image compression. Due to the same reason, we adopt nonlocal weighted sparse representation (NWSR) approach to retinal image reconstruction (Abassi *et al* 2018) for compression of 3D OCT image. The concept is similar to the sparsity based OCT image compression in (Fang *et al* 2015). SVD and NWSR methods operate on the scan-by-scan basis. Thereby, the NWSR method is computationally very demanding, see Table 2. Patch- and sparsity-based OCT image compression (Abassi *et al* 2018, Fang *et al* 2015) represents each patch in an overcomplete learned dictionary. CR of this approach is approximately obtained as:

$$CR \approx \frac{s^2}{n_p \left\|\boldsymbol{\alpha}\right\|_0 + 1} \tag{11}$$

where s stands for the step size (patch shift), $\left\|\boldsymbol{\alpha}\right\|_0$ stands for sparsity of the code (number of non-zero entries) of the patch $\mathbf{x}$ (in vectorized form) in learned dictionary $\mathbf{D}$, i.e. $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, and n_p stands for number of patches. Let's $p \times p$ be the patch size in number of pixels. Then evidently $1 \le s \le p$. It follows from (11): $s^2 \ge CR \times (n_p \times \left\|\boldsymbol{\alpha}\right\|_0 + 1)$. It also follows that requirement on code sparsity is approximately given with $\left\|\boldsymbol{\alpha}\right\|_0 \approx \frac{s^2 - CR}{n_p \times CR}$. Hence, large CR implies large value of s, and that further implies increased value of p and small value of $\left\|\boldsymbol{\alpha}\right\|_0$. It was verified that patch size of 8×8 pixels in combination with the step size (patch shift) of s=1 yields the best performance of reconstructed image (Abassi *et al* 2018, Fang *et al* 2015). When p=8 is used in above calculations it turns out that maximal CR=10 is possible with s=8 and $\left\|\boldsymbol{\alpha}\right\|_0$ =1. For CR=7, we obtain s=7 and $\left\|\boldsymbol{\alpha}\right\|_0$ =1. For CR=4 we obtain s=7 and $\left\|\boldsymbol{\alpha}\right\|_0$ =2, and for CR=2 we obtain s=5 and $\left\|\boldsymbol{\alpha}\right\|_0$ =2. We adopted the NWSR algorithm with those parameters in 3D OCT image compression experiments reported in section 3.3. Regarding SVD-based approach to image compression, expression for CR is obtained as:

$$CR = \frac{I_1 I_2}{n_p k \left(p_x + p_y + 1\right)} \tag{12}$$

where $p_x \times p_y$ stands for patch size, and k stands for number of singular values (rank) to be retained. Hence, for given CR we obtain:

$$k = \left\lfloor \frac{I_1 I_2}{CR \times n_p \left(p_x + p_y + 1 \right)} \right\rfloor \tag{13}$$

where $\lfloor \bullet \rfloor$ denotes the floor operator. Again, for large CR the patch size has to be large (to reduce n_p). Thus, in the experiments reported in section 3.3 we assumed n_p=1, p_x=I_1, p_y=I_2.

We also compare CR-constrained 3D tensor approximation algorithms with no compression de-speckling methods: 2D bilateral filtering (BF), 2D median filtering (MF), ELRpSD (Kopriva *et al* 2016), and maximum *a posteriori* (MAP) estimation based on local statistical model for OCT image denoising (Li *et al* 2017). For BF, the parameters are chosen so that the mean contrast-to-noise-ratio (CNR) is comparable to the low TT rank tensor denoising algorithm with CR=10. For MAP, we used the parameters already tested in cited reference. The ELRpSD method was compared extensively with additive low-rank plus sparse matrix decomposition methods for de-speckling (Ma *et al* 2018). The ELRpSD algorithm yielded the best performance based on a combination of the CNR and signal-to-noise-ratio (SNR) measures. Thus, we used the ELRpSD as a representative for low-rank sparse additive matrix decomposition-based de-speckling methods.

3.3 Comparative experiments

3.3.1 First dataset

Figure 1 shows the SE vs. the CR for 3D OCT images approximated with the CR-constrained low TT- and low ML rank methods for S_p, $p \in \{0, 1/2, 2/3, 1\}$, norms, and a variant of ℓ_0 surrogate (9), $\frac{1}{\sigma} f_\sigma$, for $\sigma \in \{10/7, 10\}$. CR-constrained low ML rank methods yield smaller SE than corresponding CR-constrained low TT rank methods. That is consequence of greater approximation capability of ML model due to the fact that the core tensor allows interaction between a factor with any factor in other modes. We compare in Figure 2, for CR=10, SE of selected low ML rank method constrained with S_0, with the SE of methods listed in section 3.2. We comment that SE of JPEG2000 remains approximately the same across the CR interval 2 to 60, while SE of 3D SPIHT remains approximately the same across the CR interval 10 to 60, see Figure S16. SE of NWSR- and SVD method are high due to the reasons discussed in section 3.2. See Figure S16 for SE of NWSR- and SVD method for others CRs. In NWSR-based experiment, we used dictionary made available at: https://github.com/ashkan-abbasi66/NWSR. The reason was that OCT images acquired on Topcon 3D-OCT 1000 scanner are of comparable quality with those used in: (Abassi *et al* 2018, Fang *et al* 2015). In doing so we also tested how sensitive NWSR-like OCT image compression methods are on unavailability of perfectly matched dictionary. In that regard, we comment that results presented in Figure S16 were obtained on 67 out of 96 B-scans. That is because only those 67 scans were of high enough quality to enable the layer segmentation algorithm to segment retina layers. Figure 3 shows the EIQM vs. the CR for 3D OCT images approximated with the CR-constrained low TT- and low ML rank methods with the same regularizers as in the case of Figure 1. Figure S4, in the supplement, shows EIQM for JPEG2000, 3D SPIHT, NWSR and SVD as a function of CR. Results comparable with those on Figure 2 are for EIQM shown in Figure 4, while scores by individual experts are shown in Figures S9 (a) to (c). Figures S5 (a) and (b), and S6 (a) and (b) respectively show comparable results for CNR and SNR. Figures S7 and S8 (a), (c) and (e) show individual EIQM of the three ophthalmologists for low TT rank, low ML rank and JPEG2000, 3D SPIHT, NWSR and SVD algorithms. Results comparable with Figures 2 and 4 are for CNR shown in Figure S14 (a) and for SNR in Figure S15 (a). As can be seen in Figure 2, no compression MF method yielded lowest SE. Thus, we conducted two-sample *t*-test of all the methods relative to the MF method as a reference. We used MATLAB `ttest2` command with the null hypothesis that two independent random variables come from normal distributions with equal mean and unequal variance. We were interested to find out which results agree with this hypothesis within 95% confidence interval. Afterward, we repeated statistical significance analysis by taking original images as a reference for all metrics. Our results are summarized in Table 1. Important finding is that S_0 low MLR approximation method for CR≤35, yielded segmentation error statistically insignificantly different from those obtained from no compression MF method. The same is true when original images are used as a reference. S_0 low MLR approximation method for CR≤35 also yields statistically insignificant difference in terms of the EIQM. As can be seen in Figures S5 (a) and S6 (a), S_0 low MLR also yields, in respective order, statistically significantly higher CNR value for CR≤60, and statistically insignificantly different SNR value for CR≤10. All other methods, that compare fairly in terms of SE and/or EIQM, are inferior either in terms of CNR, see Figure S14 (a), or SNR, see Figure S15 (a). We present in Table 2 computation times averaged over twenty-four images for selected 3D OCT image enhancement methods. With the exception of 3D SPIHT algorithm, that was implemented in C language, all algorithms were implemented

in MATLAB. As can be seen, all low tensor rank-based methods are computationally competitive. As expected, the NWSR-based compression and reconstruction method exhibited high computational complexity. For the purpose of illustration, we show in Figure 5 one exemplary B-scan from one original 3D OCT image and corresponding 3D OCT images enhanced/compressed with various methods discussed above.

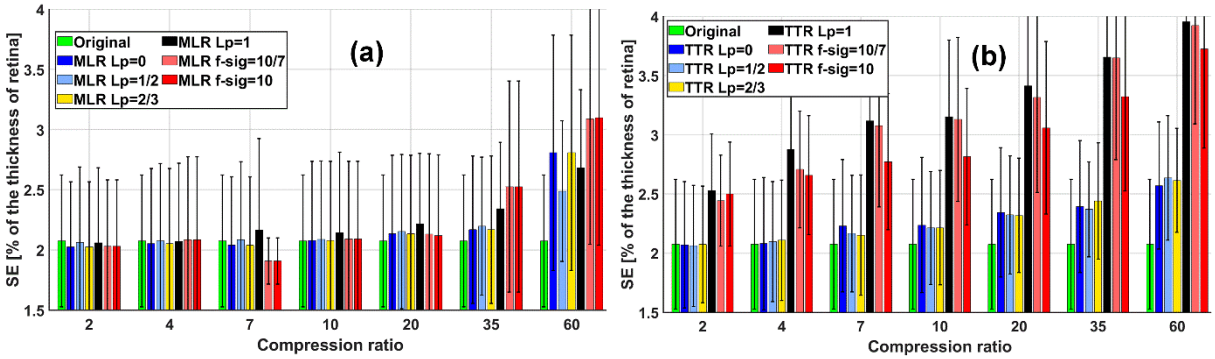

Figure 1. Retina layers segmentation error in terms of the thickness of retina (mean±standard deviation): (a) multilinear rank; (b) tensor train rank.

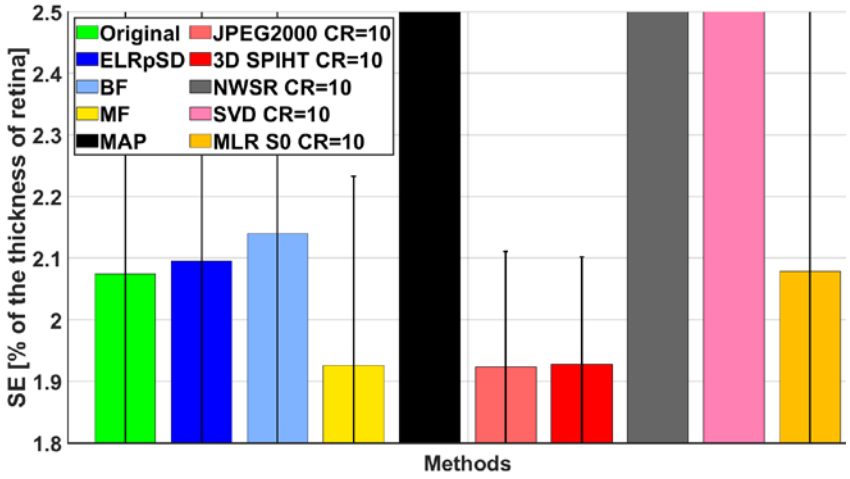

Figure 2. Best seen in color. SE (10) (mean ± standard deviation). SE of MAP is 4.5%, NWSR is 5.37%, and SVD is 4.25%.

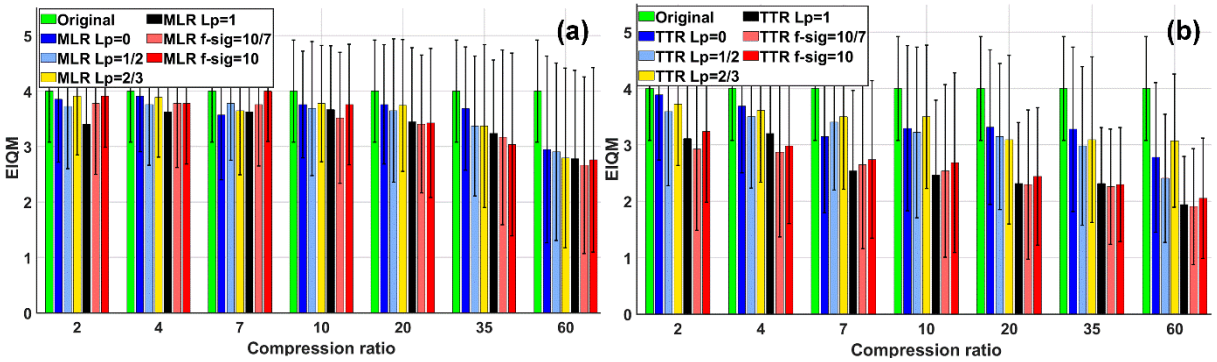

Figure 3. Expert images quality metric evaluated by three ophthalmologists (mean±standard deviation): (a) multilinear rank; (b) tensor train rank.

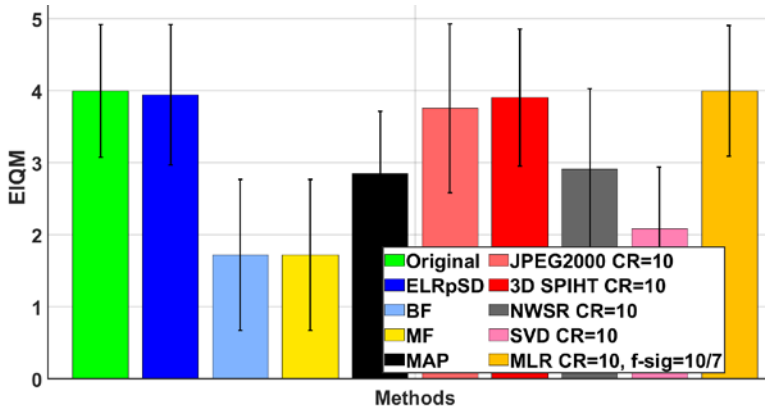

Figure 4. EIQM (mean ± standard deviation).

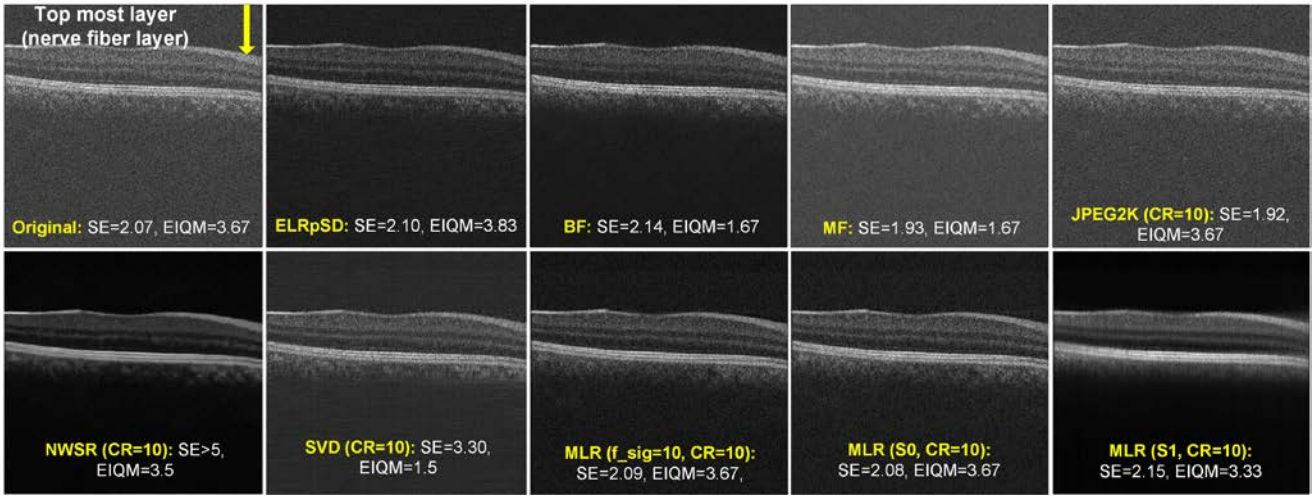

Figure 5. Best seen in color. 3D OCT image B-scan 27. SE is in percentage of the total thickness of the retina. For visual comparison, B-scans were mapped to [0, 1] interval with the MATLAB command `mat2gray`.

Table 1. Statistical significance analysis for various evaluation metrics for the first dataset. The most competitive results related to low tensor rank methods are highlighted in grey. Method not mentioned are significantly worse.

Metric	No statistically significant difference
SE, reference MF image	No compression: BF, ELRpSD; Compression: 3D SPIHT, JPEG200 for all CR; MLR: $S_{p\in\{0,1/2,2/3\}}$ for CR≤35, S_1 for CR≤10, two S_0 surrogates for CR≤20; TTR: S_0 for CR≤4, $S_{p\in\{1/2,2/3\}}$ for CR≤7.
SE, reference original image	No compression: MF, BF and ELRpSD; Compression: JPEG2000 and 3D SPIHT for CR≤60; MLR: All for CR≤35; TTR: S_0 for CR≤35, $S_{p\in\{1/2,2/3\}}$ for CR≤20.
EIQM, reference original image	No compression: ELRpSD; Compression: JPEG2000 for CR≤60; 3D SPIHT for 10≤CR≤60; MLR: S_0 and $S_{2/3}$ for CR≤35; $S_{p\in\{1/2,1\}}$ and two S_0 surrogates for CR≤20; TTR: S_0 for CR≤4, $S_{2/3}$ for CR≤10, $S_{1/2}$ for CR≤4.
CNR, reference original	JPGE2000.
SNR, reference original	MLR S_0 for CR≤10.

Table 2. Computation times (mean±standard deviation) for 3D OCT images from first dataset

Method	Time [s]	Method	Time [s]
S_0 TT (CR=10)	2.26±0.07	2D BF	400.32±2.35
S_0 ML (CR=10)	3.39±0.07	2D MF	0.426±0.014
S_0 surrogate σ=10 TT, (CR=10)	1.98±0.12	ELRpSD	38.19±0.23
S_0 surrogate σ=10 ML, (CR=10)	3.49±0.16	MAP	47.56±0.31
3D SPIHT (CR=10)	2.40±0.02	NWSR	1132.2±0.9
JPEG 2000 (CR=10)	4.66±0.23	SVD	1.46 ±0.05

3.3.2 Second dataset

Figure S10 shows the SE vs. the CR for 3D OCT images approximated with the CR-constrained low TT- and low ML rank methods for S_p, $p\in\{0, 1/2, 2/3, 1\}$, norms, and ℓ_0-surrogate (9), $\frac{1}{\sigma} f_\sigma$, for $\sigma \in \{10/7,10\}$. As can be seen, both means and standard deviations of SE estimated on twenty images are large in comparison with corresponding values estimated from the first image set, see Figure 1. The reason is that noise in images acquired by BVS 1000 OCT scanner is significantly stronger than noise in images acquired by Topcon OCT-1000 scanner. That is visible from the SNR values presented in Figure S6. Thus, the high noise level present at low CR values affects highly performance of the segmentation algorithm. For higher CR values, blur is the main cause of high SE. Hence, prior to executing statistical significance analysis, see Table 3, six OCT images with very large relative SE (above 100%) were removed from dataset. Other evaluation metrics were estimated on the whole dataset. Consequently, we show in Figure 6 SE estimated from the remaining fourteen 3D OCT images. We compare in Figure 7, for CR=10, SE of low ML rank method constrained S_0 surrogates with σ=10, with the SE of methods listed in section 3.2. We comment that SE of JPEG2000 remains approximately the same across the CR interval 2 to 60, while SE of 3D SPIHT remains approximately the same across the CR interval 10 to 60, see Figure S17. As seen in Figures 7 and S17, SE of NWSR method remains below 3% for CR≤10, despite step size greater than 1. That, in one part, is a consequence of a dictionary learned on 40 "clean" OCT images. In another part, it is a consequence of computationally demanding image reconstruction process that in average takes 3900 s per 3D OCT image of the size 1000×512×100. SE of SVD method is above 10%, and in most cases greater than 20%. That is a consequence of the two facts: (i) SVD operates on a scan-by-scan basis, and (ii) it does not rely on smart selection of singular values such as regularized low tensor rank methods proposed herein. Figure 8 shows the EIQM vs. the CR for 3D OCT images approximated with the CR-constrained low TT- and low ML rank methods with the same regularizers as in the case of Figure 6. Results analogous to Figure 7 are for EIQM shown in Figure 9, while scores by individual experts are shown in Figures S11 (a) to (c). Figure S12 shows EIQM for JPEG2000, 3D SPIHT, NWSR and SVD methods as a function of CR. Figures S13 and S8 (b), (d) and (f) show, respectively, individual EIQM of the three ophthalmologists for low TT rank, low ML rank and JPEG2000, 3D SPIHT, NWSR and SVD algorithms. Due to the same reason discussed in relation with the SE, NWSR method yielded OCT images with the EIQM comparable with those associated with the low tensor rank-based compression methods. Figures S5 (c) and (d), and S5 (c) and (d) respectively show analogous results for CNR and SNR. Results analogous to Figure 9 are for CNR shown in Figure S14 (b) and for SNR in Figure S15 (b). CR-constrained low ML rank methods yield smaller SE than corresponding CR-constrained low TT rank methods due to the reasons already elaborated in section 3.3.1. We comment that SE of JPEG2000 remains approximately the same across the CR interval 2 to 60, while SE of 3D SPIHT remains approximately the same across the CR interval 10 to 60, see Figure S17. As can be seen in Figure 7, no compression MF method yielded lowest SE. However, SE of low ML rank methods constrained with S_o norm, and its surrogate (9), $\frac{1}{\sigma} f_\sigma$, for $\sigma \in \{10/7,10\}$ yielded very similar SE. NWSR method yielded SE slightly worse SE, while SE of SVD-based compression is high. Statistical significance analysis, with the results presented in Table 3, reveals that relative to the original image as a reference all low ML rank methods for CR≤60 achieve SE that is not significantly different. Hence, they can be used for retina layers segmentation-based diagnostics. It is also visible in Table 3 that in terms of the EIQM, low ML rank methods constrained with S_p norm, $p\in\{1/2, 2/3, 1\}$ as well as surrogate of S_o (9), $\frac{1}{\sigma} f_\sigma$, for σ=10, yielded either significantly better or not significantly worse results

relative to those obtained from the original images for CR≤60. For S_o constrained low ML rank that is true for CR≤10. NWSR-based OCT image compression and reconstruction slightly inferior SE and competitive EIQM but at very high computational cost of approximately 3900 s per 3D OCT image. Low ML rank methods achieve much higher CNR than no compression methods or compression JPEG2000 and 3D SPIHT, see Figure S5 (b) and (c) and Figure S14 (b). It is also seen in Figures S15 (b) that low ML rank methods achieved SNR comparable to the SNR achieved by JPEG2000, 3D SPIHT and no compression methods.

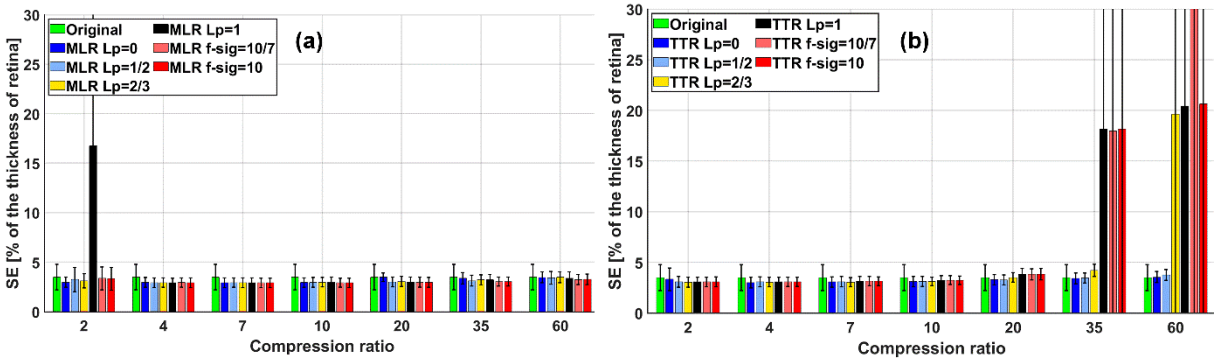

Figure 6. Retina layers segmentation error in terms of the thickness of retina (mean±standard deviation): (a) multilinear rank; (b) tensor train rank.

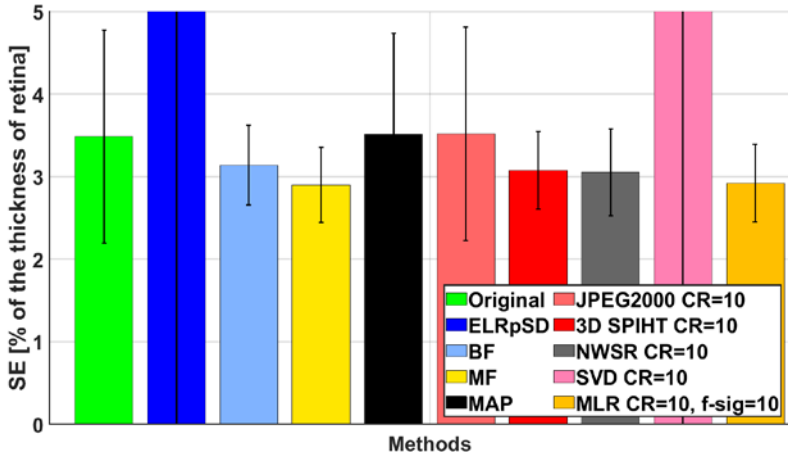

Figure 7. Best seen in color. SE (10) (mean ± standard deviation). SE of ELRpSD is 33.24%±75.17%. SE of SVD is 20.21%±50.57%.

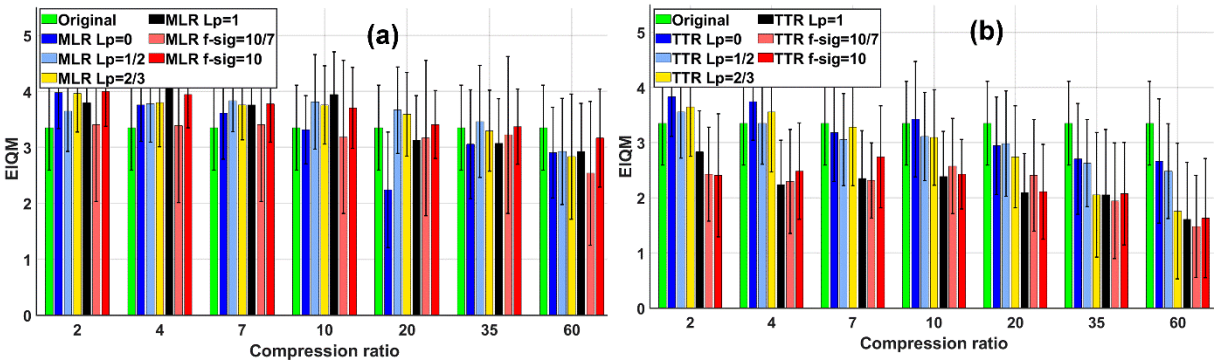

Figure 8. Expert images quality metric evaluated by three ophthalmologists (mean±standard deviation): (a) multilinear rank; (b) tensor train rank.

Thus, low ML rank methods constrained with S_p norm, $p \in \{1/2, 2/3, 1\}$ as well as surrogate of S_o (9), $\frac{1}{\sigma} f_\sigma$, for σ=10, can also be used for visual inspection based diagnostics. For the purpose of illustration, we show in Figure 10 one exemplary B-scan from one original 3D OCT image and corresponding 3D OCT images enhanced with various methods discussed herein.

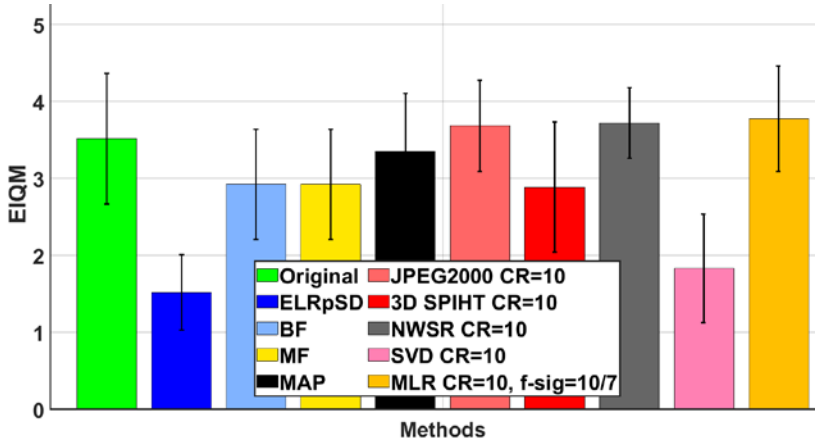

Figure 9. EIQM evaluated by three ophthalmologists (mean ± standard deviation).

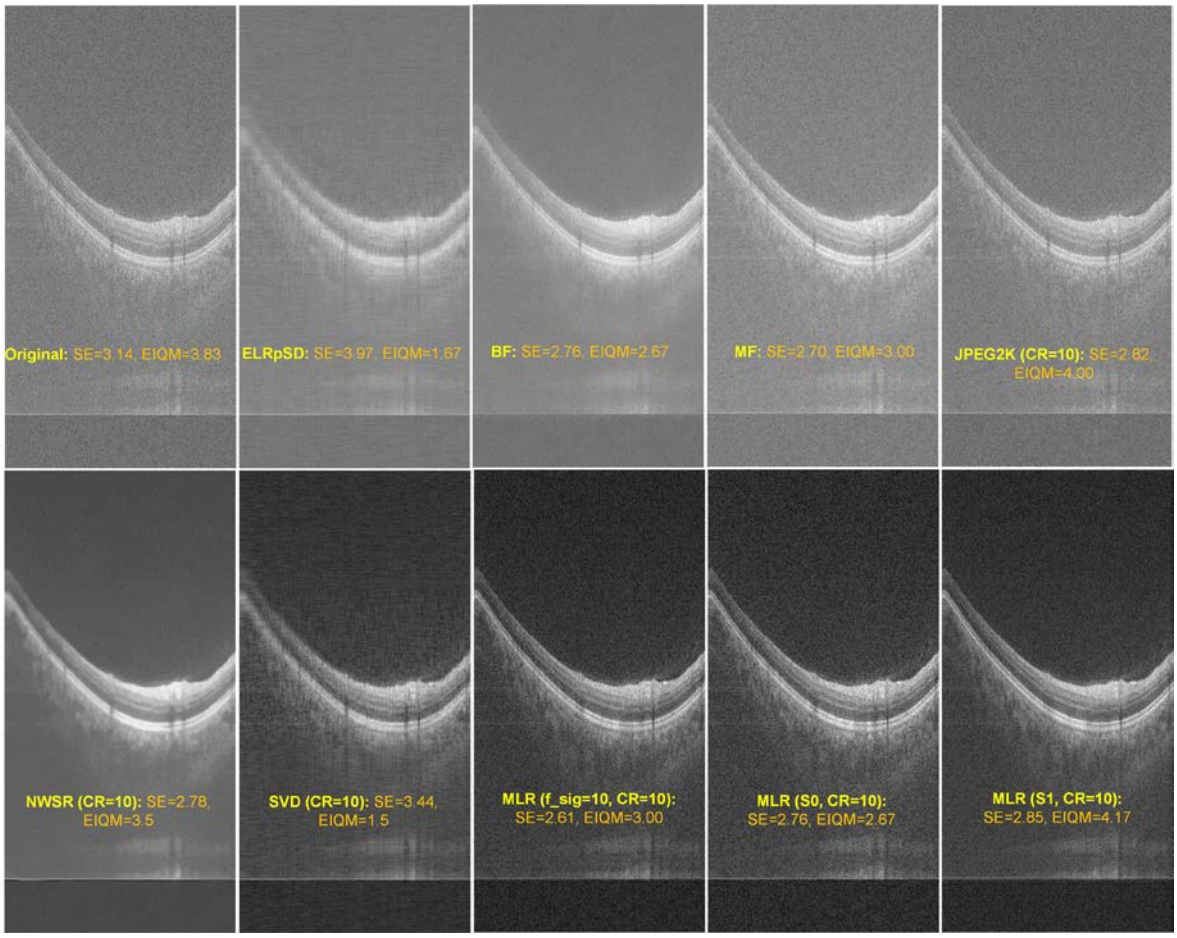

Figure 10. Best seen in color. 3D OCT image B-scan 28. SE is in percentage of the total thickness of the retina. For visual comparison, B-scans were mapped to [0, 1] interval with the MATLAB command `mat2gray`.

Table 3. Statistical significance analysis for various evaluation metrics for the second dataset. The most competitive results related to low tensor rank methods are highlighted in grey. Methods not mentioned are significantly worse.

Metric	No significant difference	Significantly better
SE, reference MF image	No compression: BF, MAP. Compression: 3D SPIHT for CR≤35, JPEG200 for CR≠4; NWSR for CR≤10; SVD for CR≤4; MLR: S_0 for CR≤10; $S_{p\in\{1/2,2/3,1\}}$ and two S_0 surrogates for CR≤35; TTR: $S_{p\in\{0,1/2,2/3,1\}}$ for CR≤20, two S_0 surrogates for CR≤10.	
SE, reference original	No compression: MF, BF and MAP. Compression: JPEG2000 for CR≠4 and 3D SPIHT for CR≤60; NWSR for CR≤10; SVD for CR≤10; All MLR for CR≤60; TTR: S_0 and $S_{1/2}$ for CR≤60; $S_{2/3}$ for CR≤35, S_1 and two S_0 surrogates for CR≤20.	
EIQM, reference original	No compression: MAP. Compression: JPG2000: CR=2, 6≤CR≤60; 3DSPIHT: 10≤CR≤35; NWSR for CR≤10; MLR: S_0 for 7≤CR≤10; $S_{1/2}$ for CR=2, 20≤CR≤60; $S_{2/3}$ for 20≤CR≤60; S_1 for 20≤CR≤60; S_0 surrogate with σ=10/7 for 2≤CR≤35; S_0 surrogate with σ=10 for 10≤CR≤60. TTR: S_0 for 4≤CR≤20; $S_{1/2}$ for 2≤CR≤35; $S_{2/3}$ for 2≤CR≤20.	JPG2000: CR=4; MLR: S_0 for 2≤CR≤4; $S_{1/2}$ for 4≤CR≤10; $S_{2/3}$ for 2≤CR≤10; S_1 for 4≤CR≤10; S_0 surrogate with σ=10 for 2≤CR≤7. TTR: S_0 for CR=2.
CNR, reference original	JPEG2000	All other methods
SNR, reference original	MLR: S_0 for CR≤10; SVD for CR≤4.	ELRpSD, S_0 TTR for CR≤4.

4. Discussion

Suppression of speckle artefacts in optical coherence tomography (OCT) is necessary for high quality quantitative assessment of ocular disorders associated with vision loss. It is also important for machine learning approach to diagnosis of retina diseases that is based on segmentation of retina layers. Furthermore, cross-platform software systems for clinical data archiving and remote consultation and diagnosis require storage and transmission of large number of high-resolution 3D OCT images. While many algorithms for compression and de-speckling of OCT images operate on a B-scan basis, they ignore spatial relations between B-scans. To preserve volumetric character of 3D OCT image, we developed algorithms for its low-rank approximations based on tensor train (TT) model and multilinear (ML) model. The TT- and ML tensor models are intimately related to the two unfolding schemes known respectively as mode-*n* canonical unfolding and mode-*n* unfolding. They enabled formulation of 3D OCT image compression and de-speckling problem as low-tensor-rank approximation problem. The algorithms were regularized by several Schatten-*p* norms, S_0 quasi-norm and its surrogates. Such smart selections of singular values of unfolded 3D OCT tensor matrices yield sparse vector of singular values, and that brings inherent denoising mechanism in low tensor rank-based 3D OCT image compression and, afterward, its tensor model-based reconstruction. Thereby, optimization algorithms estimate factors of related tensor models directly, i.e. without executing corresponding tensor decomposition algorithm, and that makes proposed algorithms fast. Moreover, corresponding optimization problems are additionally constrained by compression ratio (CR) that makes developed algorithms practically important for applications mentioned above. Unlike patch- and sparsity-based OCT image compression and reconstruction methods, proposed methods do not require clean OCT images for dictionary learning. Furthermore, since they operate on

whole 3D OCT image, they enable compression with high CR. Unlike deep learning based OCT image compression method, proposed methods do not require heavily supervised training phase.

Proposed algorithms were evaluated on 3D OCT images of a retina recorded by two different types of OCT scanners. Retina layers were segmented from de-speckled images for CR≤60, and corresponding segmentation errors were estimated. In case of first scanner, for 2≤CR≤35, all low ML rank approximations and S_0-constrained low TT rank approximations yielded segmentation errors that were statistically not different from errors corresponding to layers segmented from original images. In case of second scanner, for CR≤60, all low ML rank approximations and S_0 and $S_{1/2}$ low TT rank approximations yielded segmentation errors that were statistically not different from errors corresponding to layers segmented from original images. Thus, even at high CR segmented retina layers can be used for machine learning-based diagnosis.

Quality of visual perception was also estimated by three ophthalmologists from de-speckled images for 2≤CR≤60. In case of first scanner, for CR≤35, S_0-constrained low ML rank approximation and S_0-constrained low TT rank approximation yielded quality of perception that was either statistically better or not different from the one corresponding to original images. In case of second scanner, for CR≤60, low ML rank approximations constrained with S_p, $p\in\{0, 1/2, 2/3\}$ and one surrogate of S_0 yielded quality of perception that was statistically better or not worse than the one corresponding to original images. That is also true for low TT rank approximations constrained with S_p, $p\in\{0, 1/2, 2/3\}$ for CR≤20. Thus, even at high CR quality of de-speckled OCT images enables visual inspection based diagnostics.

Conducted analysis also shown that proposed low-tensor-rank 3D OCT image compression methods compared fairly or favourably with no-compression image enhancement methods, and JPEG2000, 3D SPIHT, NWSR and SVD image compression methods in terms of segmentation error of retina layers, quality of visual perception as well as in terms of SNR. They however, yielded OCT images with significantly higher CNR values, and in comparison with the NWSR exhibited more than two orders of magnitude lower computational complexity.

5. Conclusion

In this paper, we developed two algorithms for compression and de-speckling of 3D OCT images by using low tensor-train and low multilinear rank models constrained with Schatten-p norms, $p\in\{1/2, 2/3, 1\}$, S_0 quasi-norm and two surrogates of S_0 quasi-norm that yield unbiased estimate of the rank. Statistical significance of the results obtained from OCT images acquired on two different types of scanners confirmed that OCT images compressed and de-speckled by proposed algorithms yield segmented retina layers that can be used for machine learning-based diagnosis at the compression ratios as high as 35 or 60. The same images also have quality of visual perception that enables visual inspection-based diagnosis at the same compression ratios. Regarding the future work, we point out that core tensor of the multilinear model contains discriminative features that could be used for automated diagnosis (classification) of retina diseases. Furthermore, proposed compression ratio-constrained tensor train approximation can be extended to 4D tensors enabling feature extraction for classification of retina diseases.

Acknowledgements

This work has been supported by the Croatian Science Foundation Grant IP-2016-06-5235, the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS)**,** and the 8th Chinese-Croatian Inter-governmental S&T Cooperation Project.

Conflict of interest

The authors have no relevant conflicts of interest to disclose.

ORCIDiDs

Ivica Kopriva https://orcid.org//0000-0002-8610-8877

Xinjian Chen https://orcid.org/0000-0002-0871-293X

Fei Shi https://orcid.org/0000-0002-8878-6655

References

Abasi A, Monadjeni A, Fang L and Rabbani H 2018 Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation *J. Biomed. Opt.* 23 036011

Ahmed N, Natarajan T and Rao K K 1974 Discrete cosine transform *IEEE Trans. Comput.* 100 90-93

Andrews H C and Patterson C L 1976 Singular Value Decomposition and Digital Image Processing *IEEE Trans. Acoustics, Speech and Sig. Proc.* 24 26-53.

Baghaie A, Souza R M S and Yu Z 2015 Sparse Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction Retinal OCT Image," in *Proc. 2015 IEEE Int. Symp. on Biomed. Imaging* 226-30.

Boyd S, Parikh N, Chu E, Peleato B and Eckstein J Distributed optimization and statistical learning via the alternating direction method of multipliers 2010 *Found. Trends Mach. Learn.* 3 1-122

Chen X *et al* 2013 Quantitative analysis of retinal layers' optical intensities on 3D optical coherence tomography *Investigative Ophthalmology & Visual Science* 54 6846-51

Chen J *et al* 2016 Speckle Reduction in 3D Optical Coherence Tomography of Retina by A-Scan Reconstruction *IEEE Trans. Med. Imaging* 35 2270-79

Chen H *et al* 2019 Feature-oriented singular values shrinkage for optical coherence tomography images *Opt. Laser Eng.* 114 111-20

Chen H *et al* 2015 Quantitative analysis of retinal layers' optical intensities on 3D optical coherence tomography for central retinal artery occlusion *Scientific Reports* 9269

Cheong W F, Prahl S A and Welch A J 1990 A review of the optical properties of biological tissues *IEEE J. Quant. Elect.* 26 2166-85

Fang L, Li S, Kang X, Izatt J A and Farsiu S 2015 3-D Adaptive Sparsity Based Image Compression With Applications to Optical Coherence Tomography *IEEE Trans. Med. Imaging* 34 1306-20

Fang L, Li S, Cunefare D and Farsiu S 2017 Segmentation Based Sparse Reconstruction of Optical Coherence Tomography Images *IEEE Trans. Med. Imaging* 36 407-20

Guo P, Li D and Li X 2020 Deep OCT image compression with convolutional neural networks *Biomed. Opt. Express* 11 3543-54

Huang D, Swanson E A, Lin C P *et al* 1991 Optical coherence tomography *Science* 254 1178-81

Kopriva I 2005 Single Frame Multichannel Blind Deconvolution by Non-negative Matrix Factorization with Sparseness Constraint *Optics Letters* 30 3135-37

Kopriva I, Shi F and Chen X 2016 Enhanced low-rank + sparsity decomposition for speckle reduction in optical coherence tomography *J. Biomed. Opt.* 21 076008

Karamata B, Hassler K, Laubscher M and Lasser T 2005 Speckle statistics in optical coherence tomography *J. Opt. Soc. Am. A* 22 593-96

Laxmi P R M, Sasibhushana R G and Prabhakara R B 2019 Performance Analysis of Compression Techniques Using LM Algorithm and SVD for Medical Images in: *2019 6th Int. Conf. Sig. Proc. and Integrated Networks (SPIN)* 654-59

Lee N and Cichocki A 2018 Fundamental tensor operations for large-scale data analysis using tensor network formats *Multidimensional Systems Signal Processing* 29 921-60

Leith E *et al* 1991 Electronic holography and speckle methods for imaging through tissue using femtosecond gated pulses *Appl. Opt.* 30 4204-10

Li M, Idoughi R, Choudhury B and Heidrich W 2017 Statistical model for OCT image denoising *Biomed. Opt. Express* 8 3903-17

Ma Y, Chen X, Zhu W, Cheng X, Xiang D and Shi F 2018 Speckle noise reduction in optical coherence tomography images base on edge-sensitive cGAN *Biomed. Opt. Express* 9 5129-5146

Oseledets IV 2011 Tensor-Train Decomposition *SIAM J. Sci. Comp.* 33 2295-317

Ozcan A, Bilenca A, Desjardins A E, Bouma B E and Tearney G J 2007 Speckle reduction in optical coherence tomography using digital filtering *J. Opt. Soc. Am. A* 24 1901-10

Rasti R, Rabbani H, Mehridenhavi A and Hajizadeh F 2018 Macular ICT classification using a multi-scale convolutional neural network ensemble *IEEE Trans. Med. Imaging* 37 1024-34

Rubinstein R, Bruckstein A M and Elad M 2010 Dictionaries for sparse representation modelling *Proc. IEEE* 98 1045-57

Said A and Pearlman WA 1996 A new fast and efficient image codec based on set partitioning in hierarchical trees *IEEE Trans. Circuits Syst. Video Technol.* 6 243–50

Schmitt J M, Xiang S H and Yung K M 1999 Speckle in optical coherence tomography: an overview *J. Biomed Opt.* 4 95-105

Shi F *et al* 2019 DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images *Phys. Med. Biol.* 64 175010

Shi F *et al* 2015 Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images with Serous Pigment Epithelial Detachments *IEEE Trans. Med. Imaging* 34 441-52

Skodras A, Christopoulos C and Ebrahimi T 2001 The JPEG 2000 still image compression standard *IEEE Signal Proc. Mag.* 18 36-58

Szkulmowski M, Gorczynska I, Sylverstrzak M, Kowalczyk A and Wojtkowski M 2012 Efficient Reduction of Speckle Noise in Optical Coherence Tomography *Optics Express* 20 1337-59

van Soest *et al* 2012 Frequency domain multiplexing for speckle reduction in optical coherence tomography *J. Biomed. Opt.* 17 a076018

Tucker L R 1966 Some mathematical notes on three-mode factor analysis *Psychometrika* 31 279-311

Wallace G K 1991 The JPEG still picture compression standards *Commun. ACM* 34 30-44

Wang Z, Simoncelli E P and Bovik A C 2003 Multi-Scale Structural Similarity for Image Quality Assessment in *Proc. of the 37th IEEE Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 9-12, doi:10.1109/ACSSC.2003.1292216

Wang M *et al* 2021 Semi-Supervised Capsule cGAN for Speckle Noise Reduction in Retina OCT *IEEE Trans. Med. Imaging* 40 1168-1183

Wilson B C and Jacques S L 1990 Optical reflectance and transmittance of tissue-principles and applications *IEEE J. Quant Elec.* 26 2186-99

Yeganegi F, Hassanzade V and Ahadi S M 2018 Comparative Performance Evaluation of SVD-based Image Compression in: *26th Iranian Conf. Elec. Eng. (ICEE2018)* 464-69

Zhou Y *et al* 2022 Speckle Noise Reduction for OCT Image Based on Image Style Transfer and Conditional GAN *IEEE Trans. Med. Imaging* 26 139-50

Supplementary Material of "Low Tensor Train Rank and Low Multilinear Rank Approximations of 3D Tensors for Compression and De-speckling of Optical Coherence Tomography Images"

Ivica Kopriva, Fei Shi, Mingyig Lai, Marija Štanfel, Haoyu Chen and Xinjian Chen

1. Mathematical derivations of low tensor train- and low multilinear rank approximations of 3D tensors

1.1 Notations

N-order tensor is represented as $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with the elements $x_{i_1 i_2 \ldots i_N}$ where i_1=1,...,I_1, i_2=1,...,I_2, ..., i_N=1,...,I_N. Each index is called way or mode and $\{I_n\}_{n=1}^{N}$ are corresponding dimensions. $\mathbb{R}$ stands for the real manifold. That is standard notation adopted for use in multiway analysis (Kolda and Bader 2009). Matrices are denoted as $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ and vectors as $\mathbf{x} \in \mathbb{R}^{I_1}$, while italic font letter represents scalar, e.g. *x*. The *p*th power of the $\ell_{0<p\leq 1}$ norm is defined as $\ell_p^p(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum_{i=1}^{N} \|x_i\|^p$, and ℓ_0 quasi-norm as $\ell_0(\mathbf{x}) = \|\mathbf{x}\|_0 = \#\{x_i \neq 0, i = 1,..,N\}$. Here, # stands for the cardinality function. The $S_{0<p\leq 1}$ norms of a matrix **X** are defined as corresponding ℓ_p norms of a vector of singular values of **X**, i.e. $S_p^p(\mathbf{X}) = \|\sigma(\mathbf{X})\|_p^p$. $S_0(\mathbf{X}) = \|\sigma(\mathbf{X})\|_0$. Since minimization of $S_{0<p\leq 1}(\mathbf{X})$ and minimization of $S_{0<p\leq 1}^p(\mathbf{X})$ yield the same result, we refer to $S_p^p(\mathbf{X})$ minimization problems as $S_p(\mathbf{X})$ minimization problems.

1.2 Singular value function, proximity operators and low-rank approximation

Let *M*, *N* be two positive integers and $r = \min\{M, N\}$.

Definition 1 (Beck 2017). $\boldsymbol{\sigma}:\mathbb{R}^{M\times N}\to\mathbb{R}^{r}$ denotes the singular value function that assigns to each matrix $\mathbf{X}\in\mathbb{R}^{M\times N}$ the vector of singular values $\left(\sigma_1(\mathbf{X}),\sigma_2(\mathbf{X}),...,\sigma_r(\mathbf{X})\right)^{\mathrm{T}}$, where $\sigma_1(\mathbf{X})\geq\sigma_2(\mathbf{X})\geq...\geq\sigma_r(\mathbf{X})\geq 0$.

Definition 2 (Lewis and Sendov 2005). $f:\mathbb{R}^{r}\to(-\infty,\infty]$ is called absolutely symmetric function, if

$$f(x_1,x_2,...,x_N)=f\left(\left|x_{\pi(x_1)}\right|,\left|x_{\pi(x_2)}\right|,...,\left|x_{\pi(x_N)}\right|\right) \tag{1}$$

holds for any permutation π for [1, ..., *N*].

Definition 3 (Beck 2017). A proper function $g:\mathbb{R}^{M\times N}\to(-\infty,\infty]$ stands for a symmetric spectral function over $\mathbb{R}^{M\times N}$ if there exists a proper absolutely symmetric function $f:\mathbb{R}^{r}\to(-\infty,\infty]$ such that $g=f\circ\boldsymbol{\sigma}$. f is called the associated function of *g*.

Hence, the $S_{0\leq p\leq 1}$ norm is a symmetric spectral function over $\mathbb{R}^{M\times N}$ with the associated function equal to $\ell_{0\leq p\leq 1}$ norm. Likewise, $f_\sigma(x)$ in (11) and its variant $\frac{1}{\sigma}f_\sigma(x)$ are associated functions of the related symmetric spectral function. The proximity operator of closed proper lower semi-continuous function $f:\mathbb{R}\to\mathbb{R}\cup\infty$ is (Parikh and Boyd 2013):

$$prox_f^{\tau}(x)=\arg\min_u\left\{\frac{1}{2}\|u-x\|^2+\tau f(u)\right\}. \tag{2}$$

For a separable function with the vector argument, proximal mapping is computed element-wise. The following property is important to explain (un)biasedness of the proximity operator.

Property 1. If $f\equiv c$ for some $c\in\mathbb{R}$, then

$$prox_f^{\tau}(x)=\arg\min_u\left\{\tfrac{1}{2}\|u-x\|^2+\tau c\right\}=x. \tag{3}$$

Lemma 1. Proximity operator $prox_f\left(x\right)$ yields the unbiased estimate of the large values of x if $\lim_{|x|\to\infty} f\left(x\right)=c,\ c\in\mathbb{R}$.

Proof. Unbiased estimate of x by the proximity operator $prox_f\left(x\right)$ implies $\lim_{|x|\to\infty} prox_f\left(x\right)=x$. According to the property 1 that is achieved when $\lim_{|x|\to\infty} f\left(x\right)=c,\ c\in\mathbb{R}$. □

Remark 1. Lemma 1 implies that proximal operator of $\ell_{p>0}$ yields biased estimates of x. See also illustration in Fig. 1 for $p\in\{1/2, 2/3, 1\}$.

Low-rank approximation of a matrix $\mathbf{X}$ is obtained by minimizing the error between $\mathbf{X}$ and its model, whereat the appropriate constraints are imposed on singular value (Liu *et al* 2016):

$$\hat{\mathbf{X}}=F\left(\mathbf{X}\right)=\arg\min_{\mathbf{A}}\left\{\frac{1}{2}\left\|\mathbf{X}-\mathbf{A}\right\|_F^2+\tau f\left(\sigma\left(\mathbf{A}\right)\right)\right\}\ . \tag{4}$$

$f\left(\boldsymbol{\sigma}\left(\mathbf{A}\right)\right)$ stands for a function that is absolutely symmetric in its argument, and it is associated function to symmetric spectral function $g=f\circ\boldsymbol{\sigma}\left(\mathbf{A}\right)$. Existence of the low-rank approximation of matrix $\mathbf{X}$ is ensured by the Eckart-Young theorem (Eckart and Young 1936). It is based on the singular-value decomposition (SVD) of $\mathbf{X}=\mathbf{U}\,\mathrm{dg}\left(\boldsymbol{\sigma}\left(\mathbf{X}\right)\right)\mathbf{V}^{\mathrm{T}}$, where (Beck 2017):

$$\mathrm{dg}\left(\boldsymbol{\sigma}\left(\mathbf{X}\right)\right)_{i,j}=\begin{cases}\sigma_i\left(\mathbf{X}\right) & i=j\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

It is shown that $g=f\circ\boldsymbol{\sigma}$ constrained low-rank approximation of $\mathbf{X}$ is given with (Beck 2017):

$$\hat{\mathbf{X}} = \mathbf{U}\mathrm{dg}\left(T_f^{\tau}\left(\boldsymbol{\sigma}(\mathbf{X})\right)\right)\mathbf{V}^T \tag{6}$$

where $T_f^{\tau}\left(\boldsymbol{\sigma}(\mathbf{X})\right) = prox_f^{\tau}\left(\boldsymbol{\sigma}(\mathbf{X})\right)$ is the thresholding operator associated with f and applied entry-wise on $\boldsymbol{\sigma}(\mathbf{X})$. When $f \equiv \ell_{0\le p\le 1}$ analytical expression for $T_{\ell_p}^{\tau}(\circ)$ exists for $p\in\{0, 1/2, 2/3, 1\}$. The proximity operators $T_{\ell_0}^{\tau}(\circ)$ and $T_{\ell_1}^{\tau}(\circ)$ are, respectively, the hard-, (Blumensath and Davis 2008) and soft-thresholding operators (Donoho 1995). Analytical expressions also exist for $T_{\ell_{1/2}}^{\tau}(\circ)$) (Xu *et al* 2012), and $T_{\ell_{2/3}}^{\tau}(\circ)$ (Cao *et al* 2013). It is seen in Fig. 1, that the $\ell_{p>0}$ norms amplify large noise, i.e. $\lim_{|x|\to\infty} \ell_{p>0}(x) = \infty$. Thus, see remark 1, they also introduce bias in approximation of the large values. Therefore, in (11) we propose to use the unbiased surrogate for ℓ_0.

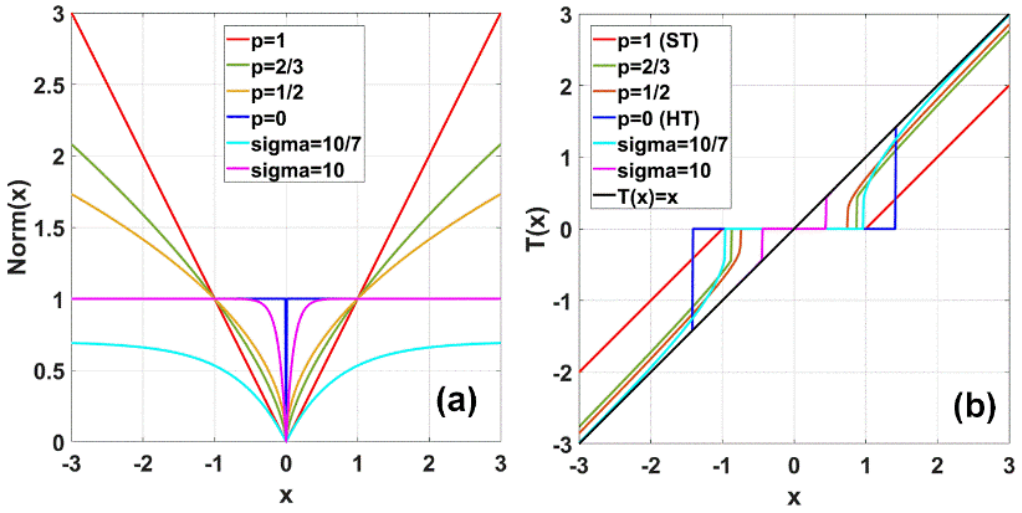

Figure S1. (a) ℓ_p, $p\in\{0, 1/2, 2/3, 1\}$, norms of x. Surrogate of the ℓ_0: $\frac{1}{\sigma}\left(1-e^{-\sigma|x|}\right)$, for $\sigma\in\{10/7, 10\}$. (b) Corresponding proximity operators for threshold value τ=1.

1.3 Tensors

Definition 4. The mode-*n* matricization, a.k.a. mode-*n* unfolding, reorders elements of the *N*th order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ for a fixed index $n \in \{1, 2, \ldots, N\}$ into a matrix:

$$\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N} \tag{7}$$

The mode-*n* matricization is associated with the Tucker model of the tensor (Tucker 1966).

Definition 5. The multilinear, a.k.a. Tucker, model decomposes a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ into a core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \ldots \times R_N}$ multiplied with factor matrices $\left\{\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R_n}\right\}_{n=1}^{N}$ on each mode (Tucker 1966):

$$\underline{\mathbf{X}} \cong \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)} \tag{8}$$

where $\underline{\mathbf{G}} \times_n \mathbf{A}^{(n)}$ denotes the mode-*n* product between the tensor $\underline{\mathbf{G}}$ and the matrix $\mathbf{A}^{(n)}$. (R_1, R_2, ..., R_N) is called the multilinear rank of Tucker decomposition (Lee and Cichocki 2018).

Definition 6. The mode-(1, 2, ..., *n*) matricization, a.k.a. mode-*n* canonical unfolding, reorders elements of the *N*th order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ for a fixed index $n \in \{1, 2, \ldots, N\text{-}1\}$ into a matrix (Oseldets 2011, Lee and Cichocki 2018):

$$\mathbf{X}_{[n]} \in \mathbb{R}^{I_1 I_2 \ldots I_n \times I_{n+1} I_{n+2} \ldots I_N} \tag{9}$$

The mode-(1, 2, ..., *n*) matricization is associated with the TT model (Oseldets 2011).

Definition 7. TT representation of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is given in terms of contracted tensor products as:

$$\underline{\mathbf{X}} \cong \mathbf{G}^{(1)} \bullet \underline{\mathbf{G}}^{(2)} \bullet \ldots \bullet \underline{\mathbf{G}}^{(N-1)} \bullet \mathbf{G}^{(N)} \tag{10}$$

where the third order core tensors are of the sizes $\left\{R_{n-1} \times I_n \times R_n\right\}_{n=1}^{N}$. It is assumed R_0=R_N=1. (R_1, R_2, ..., R_{N-1}) is called the tensor train rank.

1.4 Surrogate for ℓ_0 quasi-norm

We propose the following surrogate function for ℓ_0 quasi-norm (Chen and Gu 2014, Malek-Mohammadi *et al* 2016):

$$f_\sigma(x) = 1 - e^{-\sigma|x|}, \ \sigma > 0. \tag{11}$$

This surrogate satisfies: (*i*) $\lim_{|x| \to \infty} f_\sigma(x) = 1$, (*ii*) $\lim_{\sigma \to \infty} f_\sigma(x) = \|x\|_0$, and (*iii*) $f_\sigma(x) = 0 \Leftrightarrow x = 0$. Hence, it can substitute the ℓ_0 quasi-norm in sparsity-constrained problems. Due to (*i*) it follows from lemma 1 that $T_{f_\sigma}^{\tau}(x)$ yields unbiased estimate of large values of *x*. For $x \neq 0$ it also applies that $\lim_{\sigma \to 0} f_\sigma(x) \approx \sigma$ in which case $T_{f_\sigma}^{\tau}(x)$ yields again the unbiased estimate of large values of *x*. Thus, the $T_{f_\sigma}^{\tau}(x)$ behaves like hard thresholding operator with thresholding region controlled with σ. Due to the same reason, and unlike the case with the $\ell_{p>0}$ norms, occasionally large values of sparse noise are suppressed. In comparison with ℓ_0, f_σ amplifies less small errors around zero. We now focus attention to $\frac{1}{\sigma} f_\sigma(x)$ that is modification of (11). It is easy to show that $\lim_{\sigma \to 0} \frac{1}{\sigma} f_\sigma(x) = \ell_1(x)$. Hence, in the limit $\lim_{\sigma \to 0} T_{\frac{1}{\sigma} f_\sigma}^{\tau}(x)$ behaves like the soft-

thresholding operator. Also, $\lim_{|x|\to\infty} \frac{1}{\sigma} f_\sigma(x) = \frac{1}{\sigma}$. Thus, $T^{\tau}_{\frac{1}{\sigma} f_\sigma}(x) = \frac{1}{\sigma} T^{\tau}_{f_\sigma}(x)$ yields unbiased estimate of large values of *x* and behaves like hard thresholding operator with the thresholding region controlled with σ. Fig. 1 illustrates function $\frac{1}{\sigma} f_\sigma(x)$ for σ∈{10/7, 10} in comparison with ℓ_p norms, for p∈{0, 1/2, 2/3, 1}. Contrary to $T^{\tau}_{\ell_p}(x)$ for p∈{1/2, 2/3, 1}, one can observe unbiasedness of $T^{\tau}_{\frac{1}{\sigma} f_\sigma}(x)$. From the practical optimization point of view, it is important that an analytic expression for $T_{f_\sigma}(x,\tau)$ is obtained as the solution of the problem (2), (Malek-Mohammadi *et al* 2016). For the reason of completeness, we present herein the analytical formula:

$$T_{f_\sigma}(x,\tau) = \begin{cases} 0, & |x| \le \varphi(\tau) \\ 0, & L(x_1, x) \ge L(x) \\ \text{sign}(x)\left(\frac{1}{\sigma} W_0(z) + |x|\right) & \text{otherwise.} \end{cases} \tag{12}$$

where:

$$\begin{aligned} &\varphi(\tau) = \tfrac{1}{\sigma}\left(1 + \ln\left(\tau\sigma^2\right)\right) \\ &z = -\left(\tau\sigma^2\right)\exp\left(-\sigma|x|\right) \\ &x_1 = \tfrac{1}{\sigma} W_0(z) + |x|, L(x) = x^2/2 \\ &L(x_1, x) = (1/2)\left(x_1 - |x|\right)^2 + \tau f_\sigma\left(|x_1|\right). \end{aligned} \tag{13}$$

$W_0(z)$ in (12)/(13) refers to the principal branch of the Lambert *W* function (Corless *et al* 1996). *W*(*z*) is defined implicitly as $W(z)e^{W(z)} = z$. For real *z* it is double-valued on the interval [-1/e, 0). The branch satisfying -1≤*W*(*z*) is denoted by $W_0(z)$.

1.5 Optimization algorithm for CR-constrained TT rank approximation of a 3D tensor

CR-constrained low TT rank approximation of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is addressed by the optimization problem:

$$\min_{\mathbf{W}_n} rank\left(\mathbf{W}_n\right) \quad \text{such that: } \mathbf{X}_{[n]} = \mathbf{W}_n + \mathbf{E}_n, \quad CR \geq CR^* \quad n = 1, 2. \tag{14}$$

$\left\{\mathbf{E}_n\right\}_{n=1}^{2}$ denote the approximation errors, and CR^* stands for the pre-specified value of the compression ratio. For TT model, CR is obtained as:

$$CR = \frac{I_1 I_2 I_3}{I_1 R_1 + R_1 I_2 R_2 + R_2 I_3} \tag{15}$$

The *CR*-related inequality constraint in (14) ensures that *CR* value increases gradually and algorithm stops when the inequality constraint is satisfied. We now convert (14) into the objective function:

$$O_n\left(\mathbf{W}_n\right) = \frac{1}{2}\left\|\mathbf{X}_{[n]} - \mathbf{W}_n\right\|_F^2 + \tau_n g\left(\mathbf{W}_n\right) \quad n = 1, 2 \quad \text{such that}: CR \geq CR^*. \tag{16}$$

Here g stands for the symmetric spectral function and $\left\{\tau_n\right\}_{n=1}^{2}$ are regularization constants. Optimization problems $\left\{O_n\left(\mathbf{W}_n\right)\right\}_{n=1}^{2}$ in (16) can be run in parallel. By following (4) and (6) we convert (16) into:

$$\hat{\mathbf{X}}_{[n]} = \mathbf{U}_n \mathrm{dg}\left(\arg\min_{\mathbf{w}_n}\left\{\frac{1}{2}\left\|\mathbf{w}_n - \boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right)\right\|_F^2 + \tau_n f\left(\mathbf{w}_n\right)\right\}\right)\mathbf{V}_n^{\mathrm{T}} \quad n = 1, 2. \tag{17}$$

In (17) f stands for the absolutely symmetric associated function of spectral function g in (16). By introducing the splitting variables $\left\{\mathbf{q}_n\right\}_{n=1}^{2}$, we obtain minimization problem:

$$\arg\min_{\mathbf{q}_n,\mathbf{w}_n}\left\{\frac{1}{2}\left\|\mathbf{q}_n-\boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right)\right\|_F^2+\tau_n f\left(\mathbf{w}_n\right)\right\} \quad \text{such that: } \mathbf{q}_n=\mathbf{w}_n,\ n=1,2. \tag{18}$$

Augmented Lagrangian of (18) is:

$$\begin{aligned}\mathcal{L}_{\mu_n}\left(\mathbf{q}_n,\mathbf{w}_n,\boldsymbol{\lambda}_n\right)&=\frac{1}{2}\left\|\mathbf{q}_n-\boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right)\right\|^2+\tau_n f\left(\mathbf{w}_n\right)\\&\quad+\frac{\mu_n}{2}\left\|\mathbf{q}_n-\mathbf{w}_n\right\|^2+\left\langle\boldsymbol{\lambda}_n,\mathbf{q}_n-\mathbf{w}_n\right\rangle\quad n=1,2\end{aligned} \tag{19}$$

where $\left\{\mu_n>0\right\}_{n=1}^2$ are penalization constants and $\left\{\boldsymbol{\lambda}_n\right\}_{n=1}^2$ stand for the vectors of Lagrange multipliers.

Update rule for $\left\{\mathbf{q}_n^{k+1}\right\}_{n=1}^2$: Given $\left\{\mathbf{w}_n^k\right\}_{n=1}^2$, $\left\{\boldsymbol{\lambda}_n^k\right\}_{n=1}^2$ and $\left\{\mu_n^k\right\}_{n=1}^2$ we minimize (19) with respect to $\mathbf{q}_n$ yielding:

$$\min_{\mathbf{q}_n}\left\{\frac{1}{2}\left\|\mathbf{q}_n-\boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right)\right\|^2+\frac{\mu_n^k}{2}\left\|\mathbf{q}_n-\mathbf{w}_n^k\right\|^2+\left\langle\boldsymbol{\lambda}_n^k,\mathbf{q}_n-\mathbf{w}_n^k\right\rangle\right\}n=1,2.$$

From $\dfrac{\partial\mathcal{L}_{\mu_n^k}\left(\mathbf{q}_n,\mathbf{w}_n^k,\boldsymbol{\lambda}_n^k\right)}{\partial\mathbf{q}_n}=0$ we obtain:

$$\mathbf{q}_n^{k+1}=\frac{\boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right)+\mu_n^k\mathbf{w}_n^k-\boldsymbol{\lambda}_n^k}{1+\mu_n^k}\qquad n=1,2. \tag{20}$$

Update rule for $\left\{\mathbf{w}_n^{k+1}\right\}_{n=1}^2$: Given $\left\{\mathbf{q}_n^{k+1}\right\}_{n=1}^2$ and $\left\{\boldsymbol{\lambda}_n^k\right\}_{n=1}^2$ we minimize (19) with respect to $\mathbf{w}_n$ yielding:

$$\min_{\mathbf{w}_n}\left\{\tau_n f\left(\mathbf{w}_n\right)+\frac{\mu_n^k}{2}\left\|\mathbf{q}_n^{k+1}+\frac{\boldsymbol{\lambda}_n^k}{\mu_n^k}-\mathbf{w}_n\right\|^2\right\} \quad n=1,2. \tag{21}$$

Following (6) we obtain:

$$\mathbf{w}_n^{k+1}=T_f^{\tau_n/\mu_n^k}\left(\mathbf{q}_n^{k+1}+\frac{\boldsymbol{\lambda}_n^k}{\mu_n^k}\right) \quad n=1,2. \tag{22}$$

We estimate TT-rank as:

$$R_n^{k+1}=\left\|\mathbf{w}_n^{k+1}\right\|_0 \quad n=1,2 \tag{23}$$

and CR^{k+1} using (23) and (15).

Update of $\left\{\boldsymbol{\lambda}_n^{k+1}\right\}_{n=1}^2$: Given $\left\{\mathbf{q}_n^{k+1}\right\}_{n=1}^2$ and $\left\{\mathbf{w}_n^{k+1}\right\}_{n=1}^2$ we obtain:

$$\boldsymbol{\lambda}_n^{k+1}=\boldsymbol{\lambda}_n^k+\mu_n^k\left(\mathbf{q}_n^{k+1}-\mathbf{w}_n^{k+1}\right) \quad n=1,2. \tag{24}$$

Update of penalization constants:

$$\mu_n^{k+1}=\min\left(\rho_n\mu_n^k,\mu_n^{\max}\right) \quad n=1,2. \tag{25}$$

In our experiments we set $\{\rho_n = 1.3\}_{n=1}^{2}$, $\{\mu_n^0 = 1\}_{n=1}^{2}$ and $\{\mu_n^{\max} = 10^6\}_{n=1}^{2}$. When, after the iteration *k*, the algorithm stops we obtain low-rank approximations of mode-(1,2) matrices as:

$$\hat{\mathbf{X}}_{[n]} = \mathbf{W}_n = \mathbf{U}_n \mathrm{dg}\left(\mathbf{w}_n^k\right) \mathbf{V}_n^{\mathrm{T}} \quad n = 1, 2. \tag{26}$$

We obtain cores of the TT model (10) without executing TT decomposition. We obtain an estimate of $\mathbf{G}^{(1)}$:

$$\hat{\mathbf{G}}^{(1)} = \mathbf{U}_1 \mathrm{dg}\left(\mathbf{w}_1^k\right)^{1/2} \tag{27}$$

We obtain an estimate of $\mathbf{G}^{(3)}$:

$$\hat{\mathbf{G}}^{(3)} = \mathrm{dg}\left(\mathbf{w}_2^k\right)^{1/2} \mathbf{V}_2^{\mathrm{T}} \tag{28}$$

We obtain an estimate of $\underline{\mathbf{G}}^{(2)}$:

$$\hat{\underline{\mathbf{X}}} = \frac{1}{2} \sum_{n=1}^{2} fold_n\left(\hat{\mathbf{X}}_{[n]}\right) \tag{29}$$

$$\hat{\underline{\mathbf{G}}}^{(2)} = \hat{\underline{\mathbf{X}}} \times_1 \left(\hat{\mathbf{G}}^{(1)}\right)^{\dagger} \times_3 \left(\left(\hat{\mathbf{G}}^{(3)}\right)^{\mathrm{T}}\right)^{\dagger} \tag{30}$$

where "†" denotes the Moore-Penrose pseudoinverse. The algorithm for CR-constrained low TT-rank tensor approximation is summarized in the Algorithm 1.

Algorithm 1. CR-constrained low TT-rank 3D tensor approximation.

Inputs: The observed data tensor $\underline{\mathbf{X}}$, type of symmetric spectral function *g*, i.e. its associated function *f*. Regularization constants $\{\tau_n\}_{n=1}^{2}$, targeted value of CR^*. Set: $\{\rho_n = 1.3\}_{n=1}^{2}$, $\{\mu_n^0 = 1\}_{n=1}^{2}$, $\{\mu_n^{\max} = 10^6\}_{n=1}^{2}$, ε=10^{-6}, *itmax*=20.

1: **Initialize and precompute**:

Mode-*n* canonical unfolding of $\underline{\mathbf{X}}$ to obtain $\mathbf{X}_{[1]}$ and $\mathbf{X}_{[2]}$.

SVD $\left\{\mathbf{X}_{[n]}=\mathbf{U}_n \mathrm{dg}\left(\boldsymbol{\sigma}_n\right)\mathbf{V}_n^{\mathrm{T}}\right\}_{n=1}^{2}$. $\left\{\mathbf{w}_n^0 \leftarrow \boldsymbol{\sigma}_n / \left\|\boldsymbol{\sigma}_n\right\|_1\right\}_{n=1}^{2}$, $\left\{\boldsymbol{\lambda}_n^0=\mathbf{0}\right\}_{n=1}^{2}$, where $\mathbf{0}$ is the vector of all zeros. *stop*_1=0, *stop*_2=0, k=0.

2: **while** *stop*_1=0 $\vee$ *stop*_2=0 **do**: $n\in\{1, 2\}$

3: Compute $\mathbf{q}_n^{k+1}$ using (20).

4: Compute $\mathbf{w}_n^{k+1}$ using (22).

5: Compute CR^{k+1} using (15) and (23).

6: Compute $\boldsymbol{\lambda}_n^{k+1}$ using (24).

7: Compute μ_n^{k+1} using (25).

8: **If** $CR^{k+1}\geq CR^{*} \vee k=itmax$

9: *stop*_1=1, *stop*_2=1.

10: **End if**

11: **If** $\left\|\mathbf{q}_1^{k+1}-\mathbf{w}_1^{k+1}\right\|_{\infty} \leq \varepsilon$

12: *stop*_1=1.

13: **End i**f

14: **If** $\left\|\mathbf{q}_2^{k+1}-\mathbf{w}_2^{k+1}\right\|_{\infty} \leq \varepsilon$

15: *stop*_2=1.

16: **End if**

17: $k \leftarrow k+1$

18: **End while**

19: Use (26) to compute: $\left\{\hat{\mathbf{X}}_{[n]}=\mathbf{U}_n \mathrm{dg}\left(\left\|\boldsymbol{\sigma}_n\right\|_1 \times \mathbf{w}_n^k\right)\mathbf{V}_n^{\mathrm{T}}\right\}_{n=1}^{2}$.

20: Use (27) to (29) to compute estimates of the core tensors of TT representation.

Output: Low-TT rank approximation of 3D data tensor: $\hat{\underline{\mathbf{X}}} = \left[\!\left[\hat{\mathbf{G}}^{(1)}, \hat{\underline{\mathbf{G}}}^{(2)}, \hat{\mathbf{G}}^{(3)} \right]\!\right]$.

Scaling by $\left\{ \left\| \boldsymbol{\sigma}_n \right\|_1 \right\}_{n=1}^{2}$ in step 1 and re-scaling by the same quantity in step 19 is not obligatory. It makes tuning of the regularization constants easier due to $\left\{ 0 \le \tau_n \le 1 \right\}_{n=1}^{2}$.

1.6 Convergence analysis

We prove the global convergence of Algorithm 1. It is comprised of two parallel optimization problems. Hence, it suffices to prove convergence for one optimization problem only.

Theorem 1. Let for problem $n \in \{1, 2\}$, $D_n^k = \left\{ \mathbf{q}_n^k, \mathbf{w}_n^k, \boldsymbol{\lambda}_n^k \right\}$ be a sequence generated by Algorithm 1. For a sufficiently large μ_n^k Algorithm 1 converges globally.

Proof. We rewrite the problem (18) as:

$$\min_{\mathbf{q}_n, \mathbf{w}_n} h_1\left(\mathbf{q}_n\right) + h_2\left(\mathbf{w}_n\right) \quad \text{subject to: } \mathbf{A}\mathbf{q}_n = \mathbf{B}\mathbf{w}_n$$

where **A**=**I**, **B**=**I**, $h_1\left(\mathbf{q}_n\right) = \left(1/2\right)\left\| \mathbf{q}_n - \boldsymbol{\sigma}\left(\mathbf{X}_{[n]}\right) \right\|^2$ and $h_2\left(\mathbf{w}_n\right) = \tau_n f\left(\mathbf{w}_n\right)$. This problem is in canonical form for which convergence conditions A1 to A5 are established (Wang *et al* 2019). Associated function f of the singular value spectral function *g* is either ℓ_p norm, 0≤*p*≤1 or the ℓ_0 surrogate function (11). Both spectral functions are non-negative lower semi-continuous functions bounded from below. Thus, h_2 is non-negative lower semi-continuous and bounded function. Since h_1 is coercive and h_2 is coercive, $h_1 + h_2$ is also coercive. Therefore, assumptions A1 and A4 are satisfied. Since **A=I** and **B=I**, assumptions A2 and A3 are satisfied. Since $h_1\left(\mathbf{q}_n\right)$ is Lipschitz differentiable, assumption A5 is also satisfied. Thus, Algorithm 1 converges for sufficiently large μ_n^k. $\ell_{0 \le p \le 1}$ are semi-algebraic functions and satisfy the Kurdyka-Łojasiewicz (KL) inequality (Attouch *et al* 2013). $f_\sigma\left(x\right)$ in (11) as well as $\frac{1}{\sigma} f_\sigma\left(x\right)$ are real analytic

functions and, therefore, also satisfy the KL inequality (Łojasiewicz 1963). Thus, Algorithm 1 converges globally to the unique stationary point of (18), (Attouch *et al* 2013). □

1.7 Stopping criteria and computational complexity

The algorithm alternates either until convergence, or when $CR^{k+1} \geq CR^{*}$, or when maximum number of iterations, *itmax*, is reached. In all the experiments reported in section 3.3 we set *itmax*=20. We check the convergence by verifying $\left\|\mathbf{q}_1^{k+1} - \mathbf{w}_1^{k+1}\right\|_{\infty} \leq \varepsilon$ and $\left\|\mathbf{q}_2^{k+1} - \mathbf{w}_2^{k+1}\right\|_{\infty} \leq \varepsilon$. In all the experiments we set $\varepsilon=10^{-6}$. Regarding computational complexity, the major one is related to computing SVD of the unfolded matrices. For $\mathbf{X}_{[1]}$ it is $O\left(I_2 I_3 \left(I_1\right)^2\right)$ assuming $I_2 I_3 > I_1$, and that is true for 3D OCT images considered in this paper. For $\mathbf{X}_{[2]}$ complexity is $O\left(I_1 I_2 \left(I_3\right)^2\right)$ assuming $I_1 I_2 > I_3$, and that is also true for 3D OCT images considered in this paper. The steps 3, 4, and 6 have $O(I_1)$ complexity for n=1 and $O(I_3)$ complexity for n=2. Thus, assuming the algorithm convergence after k iterations the overall computational complexity amounts to $O\left(I_2 I_3 \left(I_1\right)^2\right) + O\left(I_1 I_2 \left(I_3\right)^2\right) + O(kI_1) + O(kI_3)$.

1.8 Optimization algorithm for CR-constrained low ML rank approximation of a 3D tensor

CR-constrained low ML rank approximation of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is addressed by the optimization problem:

$$\begin{aligned} \min_{\mathbf{W}_n} \; & rank\left(\mathbf{W}_n\right) \quad \text{such that: } \mathbf{X}_{(n)} = \mathbf{W}_n + \mathbf{E}_n, \\ & CR \geq CR^{*} \qquad n = 1, 2, 3. \end{aligned} \tag{31}$$

$\left\{\mathbf{E}_n\right\}_{n=1}^{3}$ denote the approximation errors. For ML model, CR is obtained as:

$$CR = \frac{I_1 I_2 I_3}{I_1 R_1 + I_2 R_2 + I_3 R_3} \tag{32}$$

ADMM-based optimization method for (31) is structurally equivalent to the one for low-TT rank problem (16) to (25) whereat the mode-(1, 2,...,*n*) matricization of $\underline{\mathbf{X}}$ is substituted with the mode-*n* matricization. When, after the iteration *k*, the algorithm stops we obtain low-rank approximations of mode-*n* matrices as:

$$\hat{\mathbf{X}}_{(n)} = \mathbf{W}_n = \mathbf{U}_n \mathrm{dg}\left(\mathbf{w}_n^k\right) \mathbf{V}_n^{\mathrm{T}} \quad n = 1, 2, 3. \tag{33}$$

We obtain low-ML rank approximation of 3D tensor $\underline{\mathbf{X}}$ as:

$$\hat{\underline{\mathbf{X}}} = \frac{1}{3} \sum_{n=1}^{3} fold_n \left(\hat{\mathbf{X}}_{(n)}\right) \tag{34}$$

We obtain cores of the ML model (8) without executing Tucker decomposition, such as Tucker alternating least squares algorithm (Kolda and Bader 2009). By using (34) and formulas for mode-*n* unfolded matrices, we obtain estimates of factor matrices as (Lee and Cichocki 2018):

$$\left\{\hat{\mathbf{A}}^{(n)} = \mathbf{U}_n \mathrm{dg}\left(\mathbf{w}_n^k\right)^{1/2}\right\}_{n=1}^{3} \tag{35}$$

and estimate of the core tensor as:

$$\hat{\underline{\mathbf{G}}} = \hat{\underline{\mathbf{X}}} \times_1 \left(\hat{\mathbf{A}}^{(1)}\right)^{\dagger} \times_2 \left(\hat{\mathbf{A}}^{(2)}\right)^{\dagger} \times_3 \left(\hat{\mathbf{A}}^{(3)}\right)^{\dagger} \tag{36}$$

The algorithm for CR-constrained low ML-rank tensor approximation is summarized in the Algorithm 2.

Algorithm 2. CR-constrained low ML-rank 3D tensor approximation.

Inputs: The observed data tensor $\underline{\mathbf{X}}$, type of symmetric spectral function *g*, i.e. its associated function *f*. Regularization constants $\{\tau_n\}_{n=1}^3$, targeted value of CR^*. Set: $\{\rho_n = 1.3\}_{n=1}^3$, $\{\mu_n^0 = 1\}_{n=1}^3$, $\{\mu_n^{\max} = 10^6\}_{n=1}^3$, ε=10^{-6}, *itmax*=20.

1: **Initialize and precompute**:

Mode-*n* unfolding of $\underline{\mathbf{X}}$ to obtain $\left\{\mathbf{X}_{(n)}\right\}_{n=1}^3$. SVD: $\left\{\mathbf{X}_{(n)} = \mathbf{U}_n \mathrm{dg}\left(\boldsymbol{\sigma}_n\right)\mathbf{V}_n^{\mathrm{T}}\right\}_{n=1}^3$. $\left\{\mathbf{w}_n^0 \leftarrow \boldsymbol{\sigma}_n / \left\|\boldsymbol{\sigma}_n\right\|_1\right\}_{n=1}^3$. $\left\{\boldsymbol{\lambda}_n^0 = \mathbf{0}\right\}_{n=1}^3$, where $\mathbf{0}$ is the vector of all zeros. $\{stop\_\mathrm{n}{=}0\}_{n=1}^3$, k=0.

2: **while** *stop*_1=0 $\vee$ *stop*_2=0 $\vee$ *stop*_3=0 **do**: $n \in \{1, 2, 3\}$

3: $$\mathbf{q}_n^{k+1} = \frac{\boldsymbol{\sigma}\left(\mathbf{X}_{(n)}\right) + \mu_n^k \mathbf{w}_n^k - \boldsymbol{\lambda}_n^k}{1 + \mu_n^k}.$$

4: $$\mathbf{w}_n^{k+1} = T_f^{\tau_n/\mu_n^k}\left(\mathbf{q}_n^{k+1} + \frac{\boldsymbol{\lambda}_n^k}{\mu_n^k}\right).$$

5: Compute CR^{k+1} using (32) with $R_n^{k+1} = \left\|\mathbf{w}_n^{k+1}\right\|_0$.

6: $\boldsymbol{\lambda}_n^{k+1} = \boldsymbol{\lambda}_n^k + \mu_n^k\left(\mathbf{q}_n^{k+1} - \mathbf{w}_n^{k+1}\right)$.

7: $\mu_n^{k+1} = \min\left(\rho_n \mu_n^k, \mu_n^{\max}\right)$.

8: **If** $CR^{k+1} \geq CR^* \vee$ *k*=*itmax*

9: *stop*_1=1, *stop*_2=1, *stop*_3=1.

10: **End if**

11: **If** $\left\|\mathbf{q}_1^{k+1} - \mathbf{w}_1^{k+1}\right\|_\infty$

12: *stop*_1=1.

13: **End if**

14: **If** $\left\|\mathbf{q}_2^{k+1} - \mathbf{w}_2^{k+1}\right\|_\infty \leq \varepsilon$

15: *stop*_2=1.

16: **End if**

17: **If** $\left\|\mathbf{q}_3^{k+1}-\mathbf{w}_3^{k+1}\right\|_\infty \le \varepsilon$

18: *stop*_3=1.

19: **End if**

20: $k \leftarrow k+1$

21: **End while**

22: Use (33) to compute: $\left\{\hat{\mathbf{X}}_{(n)} = \mathbf{U}_n \mathrm{dg}\left(\left\|\boldsymbol{\sigma}_n\right\|_1 \times \mathbf{w}_n^k\right)\mathbf{V}_n^{\mathrm{T}}\right\}_{n=1}^{3}$.

23: Use (34) to (36) to compute low rank estimates of the core tensors and factor matrices of the ML model (8).

Output: Low-ML rank approximation of a 3D data tensor: $\hat{\underline{\mathbf{X}}} = \left[\!\left[\hat{\underline{\mathbf{G}}}; \hat{\mathbf{A}}^{(1)}, \hat{\mathbf{A}}^{(2)}, \hat{\mathbf{A}}^{(3)}\right]\!\right]$

Regarding scaling and rescaling with $\left\{\left\|\boldsymbol{\sigma}_n\right\|_1\right\}_{n=1}^{3}$ in steps 1 and 22 the same comment applies as for the Algorithm 1. The Algorithm 2 is comprised of three parallel optimization problems. Thus, to prove convergence it is enough to prove convergence of one problem only.

Theorem 2. Let for problem $n \in \{1, 2, 3\}$, $D_n^k = \left\{\mathbf{q}_n^k, \mathbf{w}_n^k, \boldsymbol{\lambda}_n^k\right\}$ be a sequence generated by Algorithm 2. For a sufficiently large μ_n^k Algorithm 2 converges globally.

Proof. The proof is equivalent to the one for Theorem 1. Instead of problem (18), we now refer to the equivalent formulation of problem (31). □

Regarding the stopping criteria, the same comments apply as for the Algorithm 1 in section 2.5. By using the same reasoning as for the Algorithm 1 and assuming the algorithm converges after

k iterations, that the overall computational complexity amounts to $O\left(I_2 I_3 \left(I_1\right)^2\right) + O\left(I_1 I_3 \left(I_2\right)^2\right)$ $+ O\left(I_1 I_2 \left(I_3\right)^2\right)$ + $O(kI_1)$ + $O(kI_2)$ + $O(kI_3)$. For low-TT rank methods the approximate time is 2 seconds for Topcon OCT-1000 images, and 11 seconds for BV1000 images. For low-ML methods the approximate time it is 3.5 seconds and 16 seconds respectively, see Figure S2. Regularization constants had to be tuned in order that Algorithms 1 and 2 compress 3D OCT image with the pre-specified CR value. In that regard, they were tuned separately on twenty four 3D OCT images from Topcon OCT-1000 scanner and twenty 3D OCT images from BV1000 scanner for $CR^* \in \{2, 4, 7, 10, 20, 35, 60\}$. Regularization constants corresponding to other values of CR^*, are obtained by the shape-preserving cubic Hermite interpolation using MATLAB function pchip. Figure S3 shows mean values and standard deviations of the estimated CR^* vs. nominal CR^* after tuning.

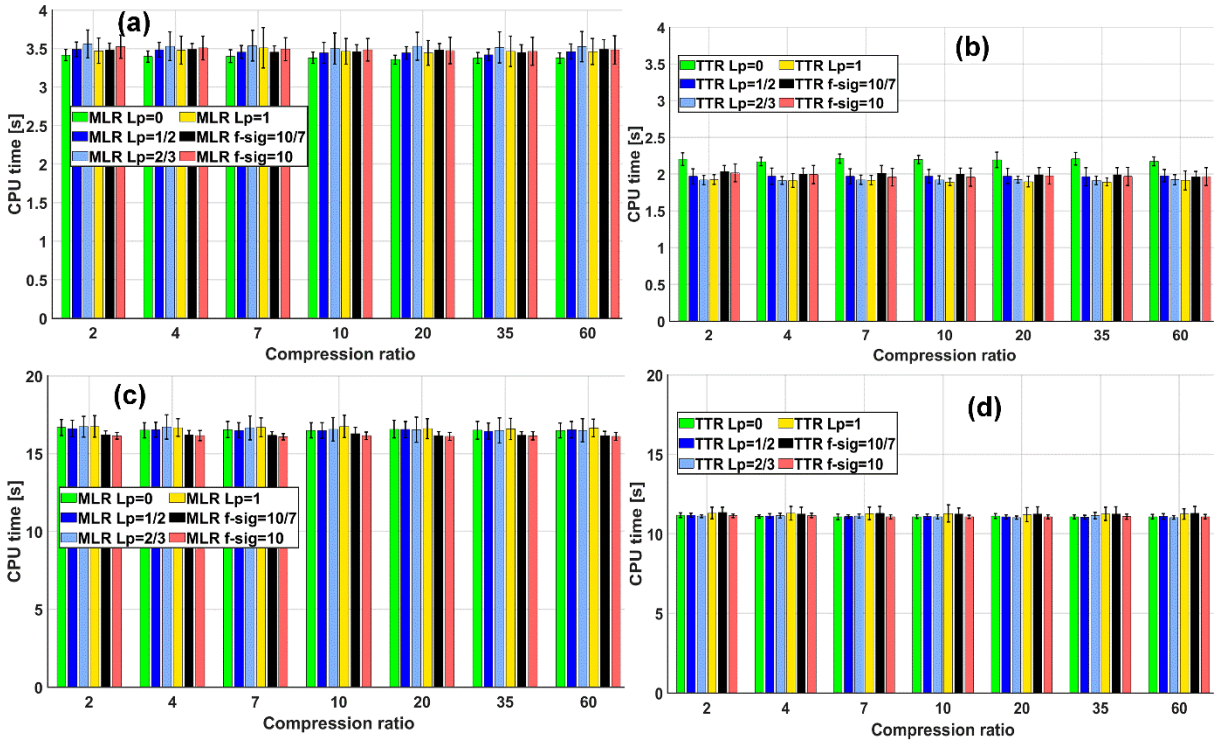

Figure S2. Computation times vs compression ratio (mean ± standard deviation). First data set (image size 480×512×64): (a) multilinear rank; (b) tensor train rank. Second data set (image size 1000×512×100): (c) multilinear rank; (d) tensor train rank.

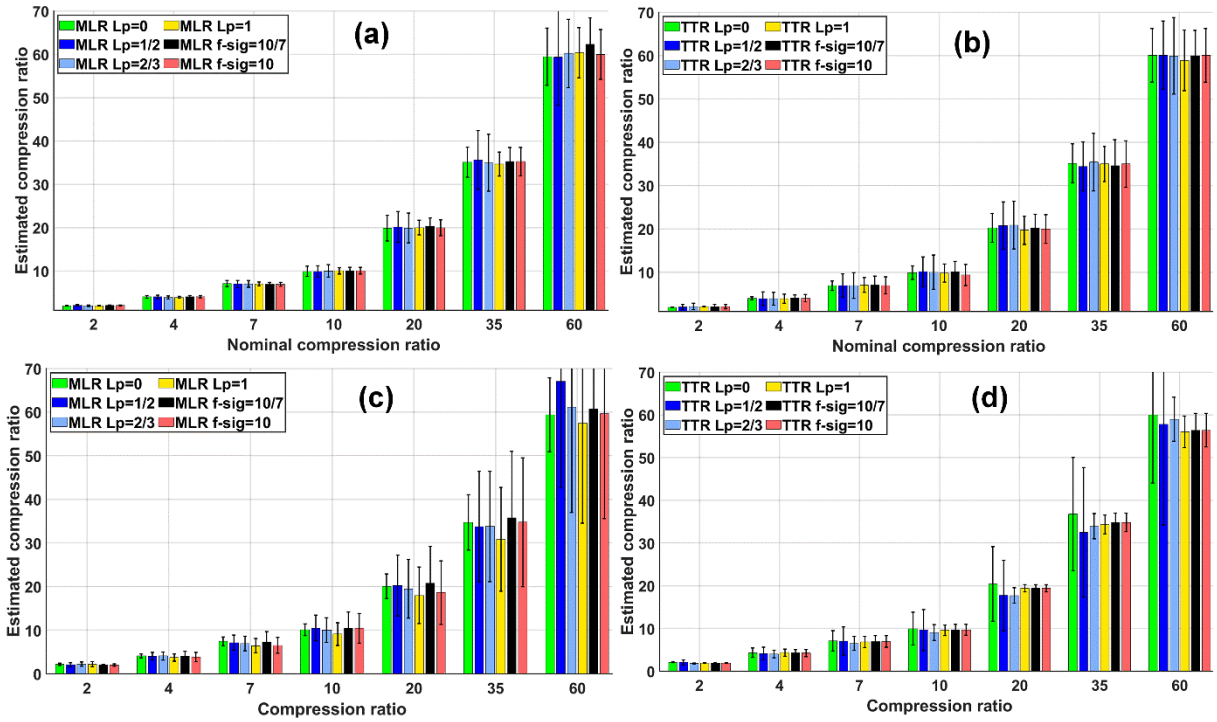

Figure S3. Estimate vs nominal compression ratio (mean ± standard deviation). First data set (image size 480×512×64): (a) multilinear rank; (b) tensor train rank. Second data set (image size 1000×512×100): (c) multilinear rank; (d) tensor train rank.

2. Results

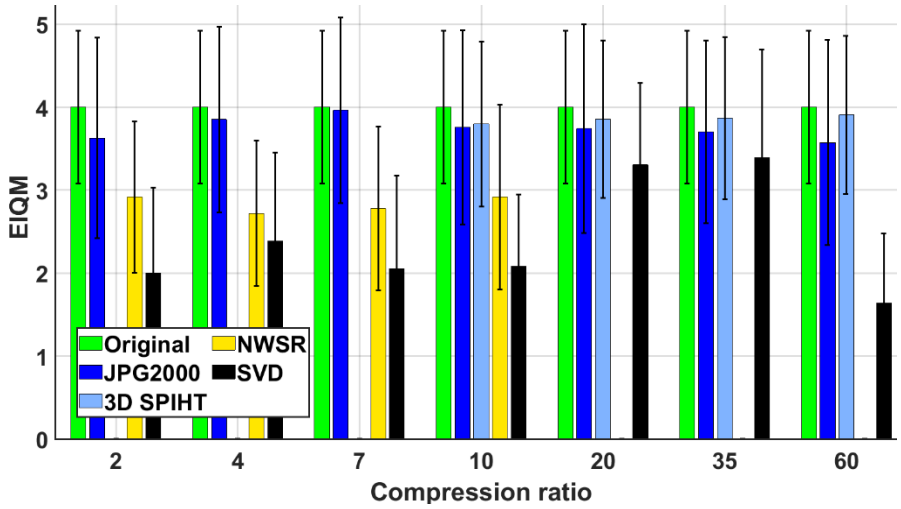

Figure S4. Best seen in color. EIQM (mean ± standard deviation) vs. selected values of CR for JPEG2000, 3D SPIHT, NWSR and SVD compressed images. Evaluations by three ophthalmologists. First data set (image size 480×512×64).

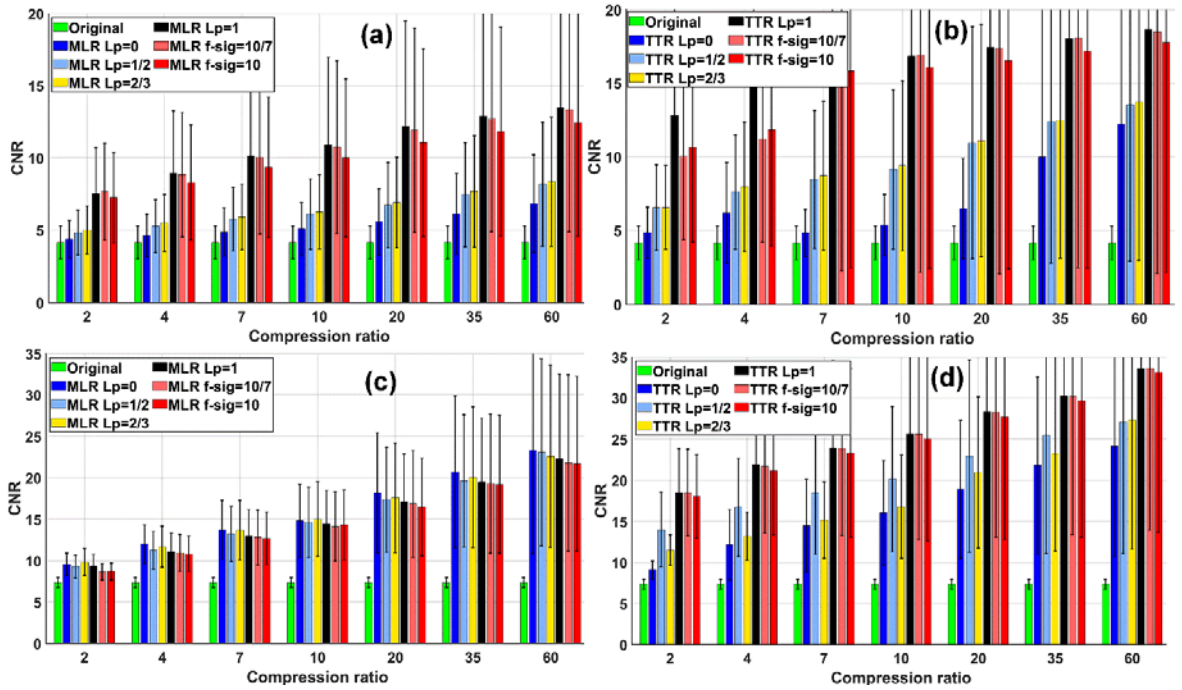

Figure S5. Best seen in color. CNR (mean ± standard deviation) vs. selected values of CR. First data set (image size 480×512×64): (a) multilinear rank; (b) tensor train rank. Second data set (image size 1000×512×100): (c) multilinear rank; (d) tensor train rank.

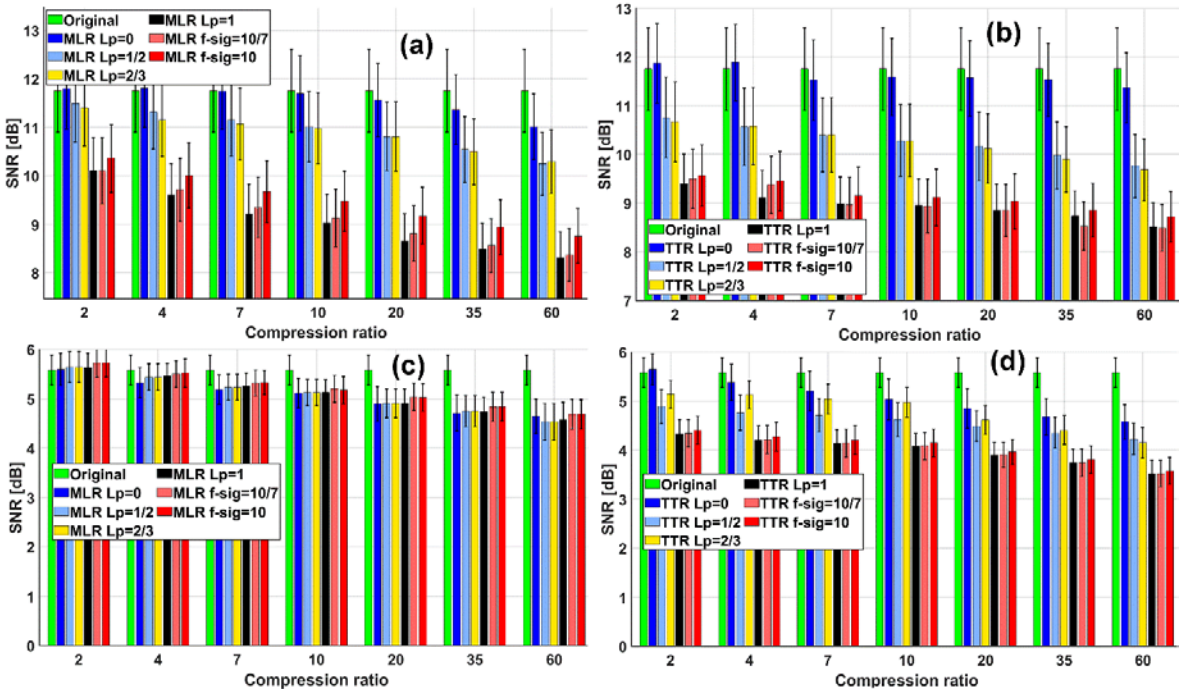

Figure S6. Best seen in color. SNR (mean ± standard deviation) vs. selected values of CR. First data set (image size 480×512×64): (a) multilinear rank; (b) tensor train rank. Second data set (image size 1000×512×100): (c) multilinear rank; (d) tensor train rank.

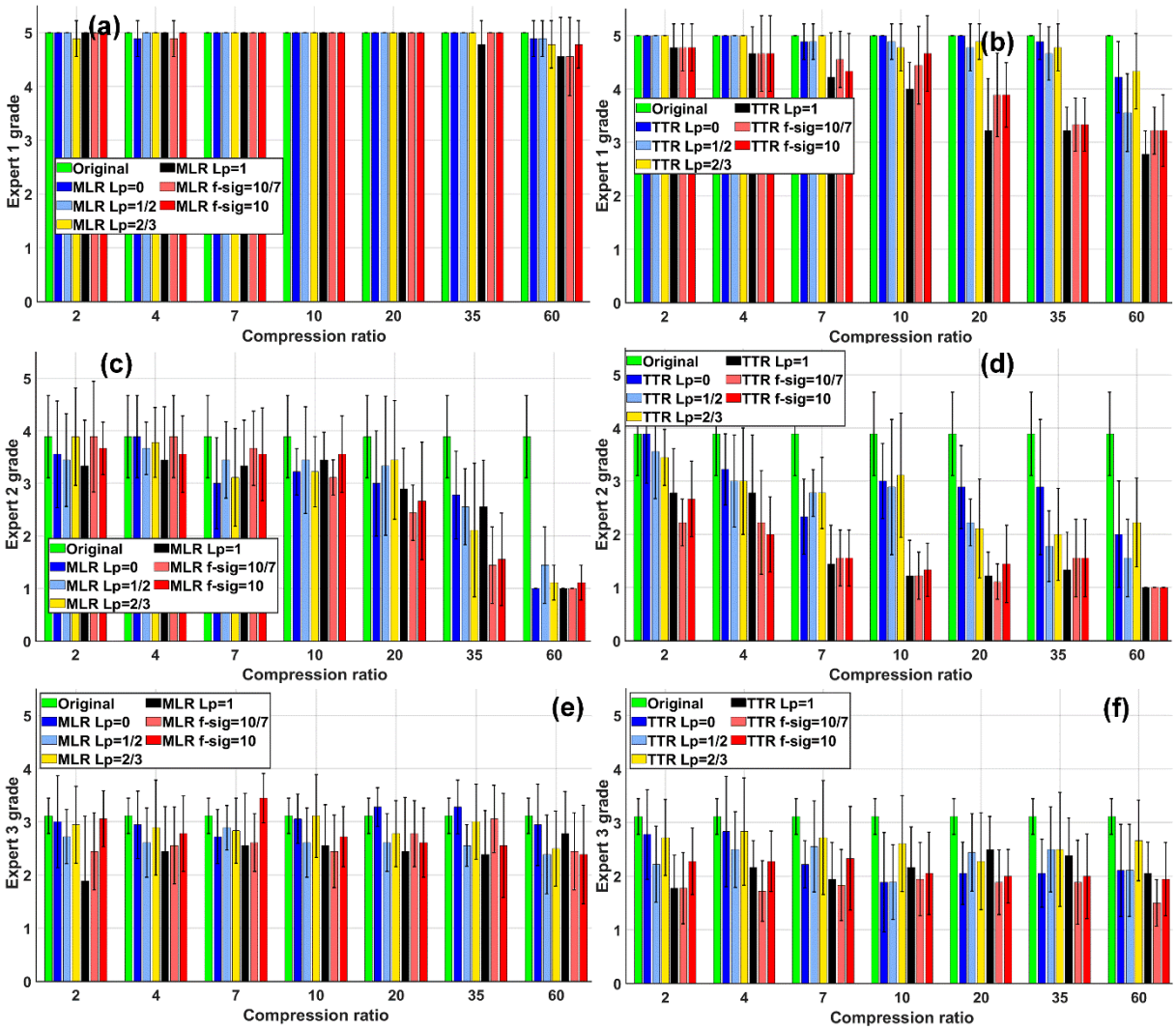

Figure S7. Best seen in color. EIQM (mean ± standard deviation) vs. selected values of CR. First data set (image size 480×512×64). Top: expert 1. Middle: expert 2. Bottom: expert 3. (a) expert 1 - multilinear rank; (b) expert 1- tensor train rank; (c) expert 2 - multilinear rank; (d) expert 2 - tensor train rank; (e) expert 3 - multilinear rank; (f) expert 3 - tensor train rank.

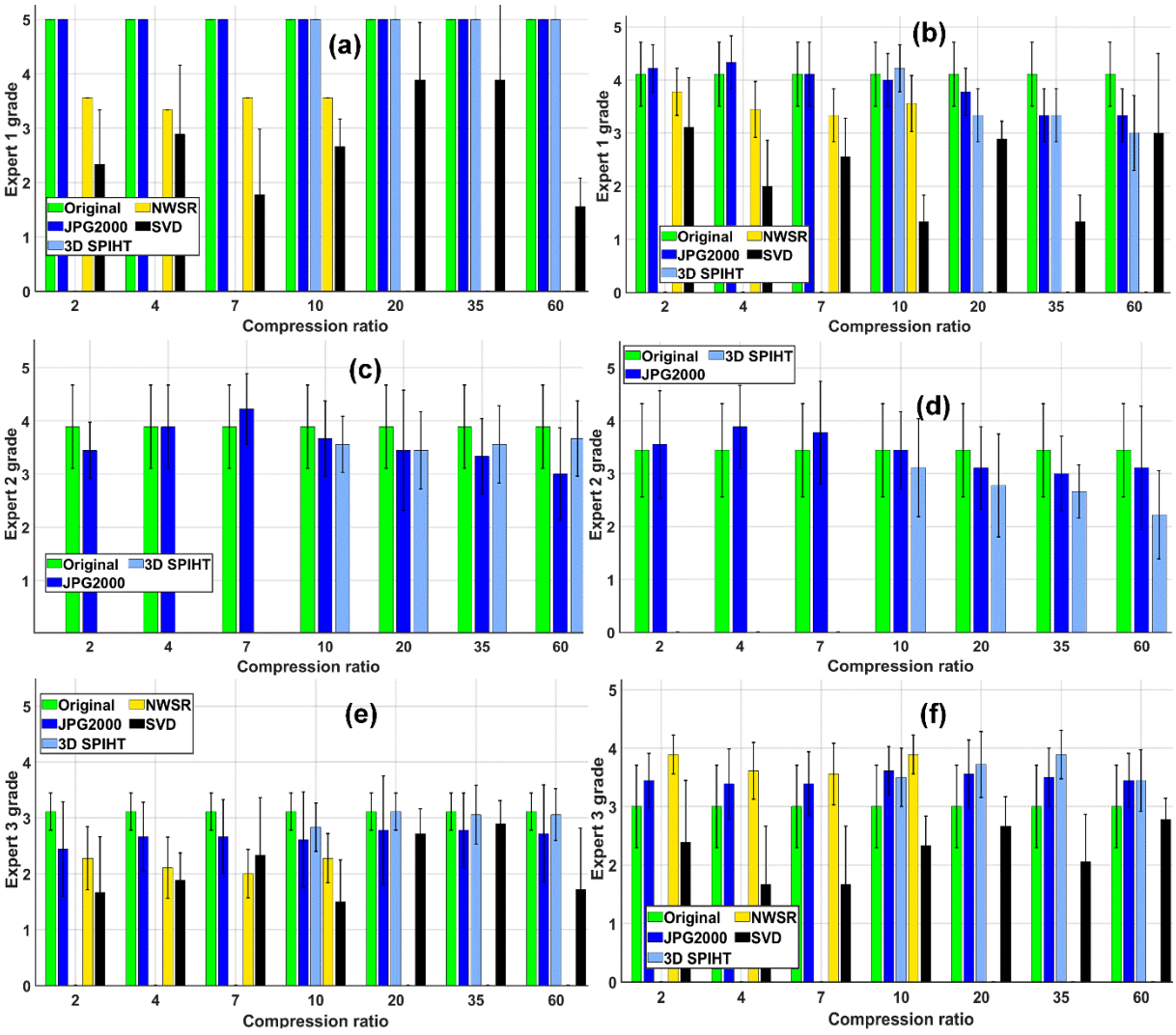

Figure S8. Best seen in color. EIQM (mean ± standard deviation) vs. selected values of CR for JPEG2000 and 3D SPIHT compressions. First data set (image size 480×512×64): (a) expert 1; (c) expert 2; (e) expert 3. Second data set (image size 1000×512×100): (b) expert 1; (d) expert 2; (f) expert 3.

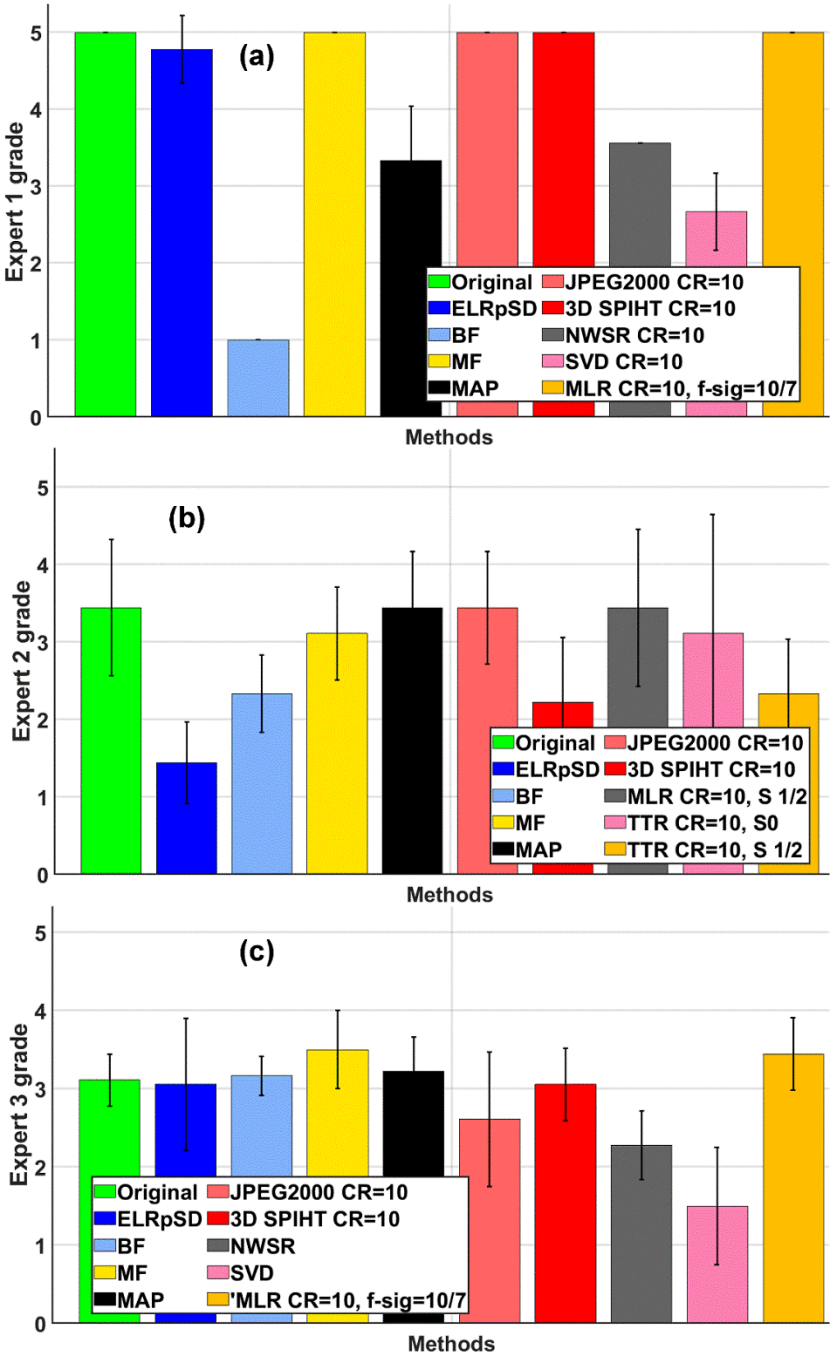

Figure S9. Best seen in color. EIQM (mean ± standard deviation) for original image, ELRpSD, BF, MF and MAP no compression methods, JPEG2000 and 3D SPIHT, and low multilinear rank approximations for CR=10. (a) expert 1. (b) expert 2. (c) expert 3. First data set (image size 480×512×64).

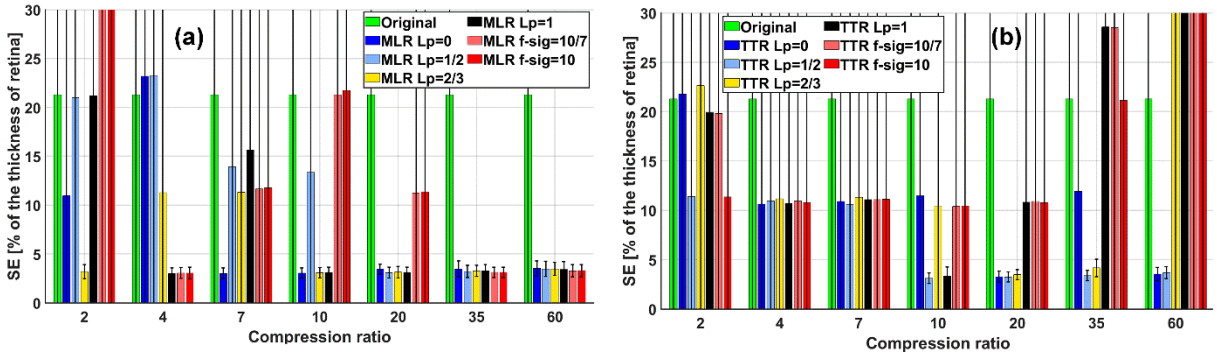

Figure S10. Retina layers segmentation error in terms of the thickness of retina (mean±standard deviation): (a) multilinear rank; (b) tensor train rank. Second data set (image size 1000×512×100).

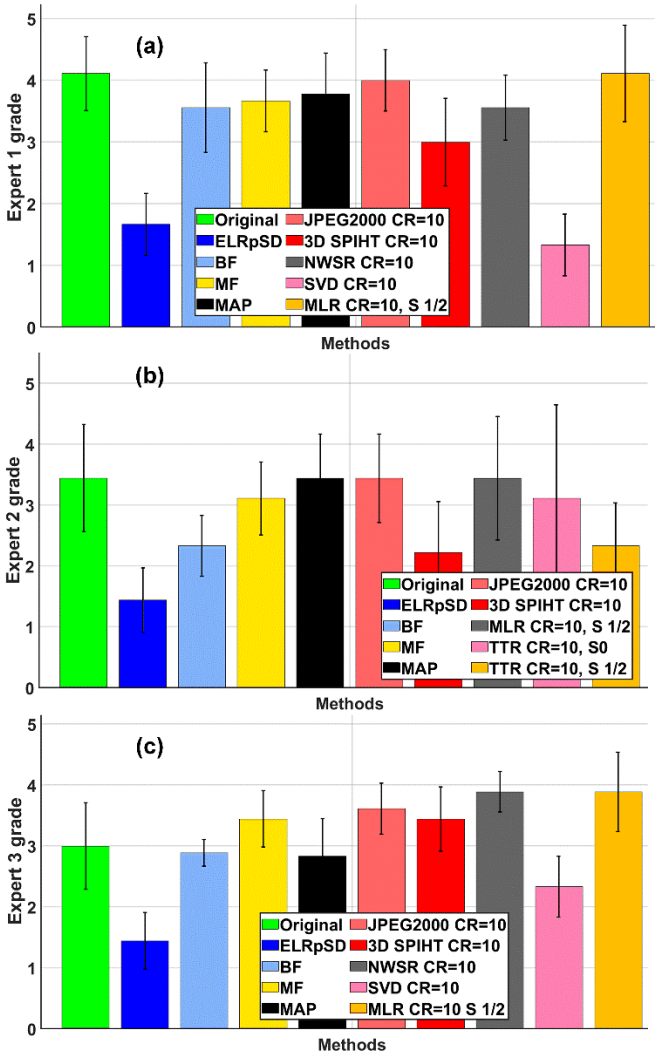

Figure S11. Best seen in color. EIQM (mean ± standard deviation) for original image, ELRpSD, BF, MF and MAP no compression methods, JPEG2000 and 3D SPIHT, and low multilinear rank approximations for CR=10. (a) expert 1. (b) expert 2. (c) expert 3. Second data set (image size 1000×512×100).

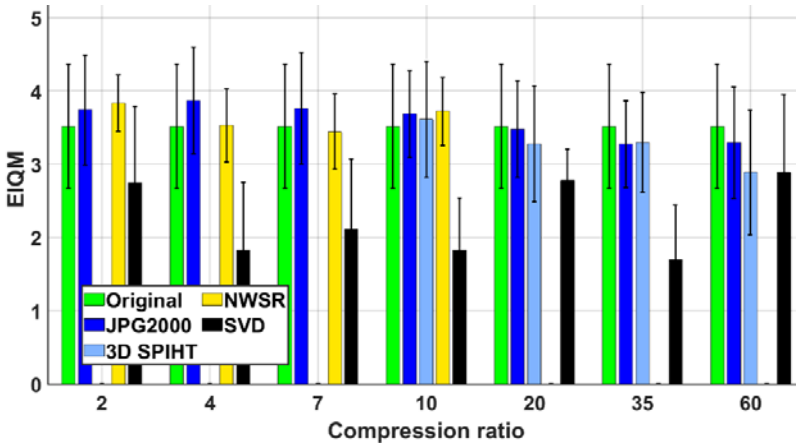

Figure S12. Best seen in color. EIQM (mean ± standard deviation) vs. selected values of CR for JPEG2000, 3D SPIHT, NWSR and SVD compressed images. Evaluations by three ophthalmologists. Second data set (image size 1000×512×100).

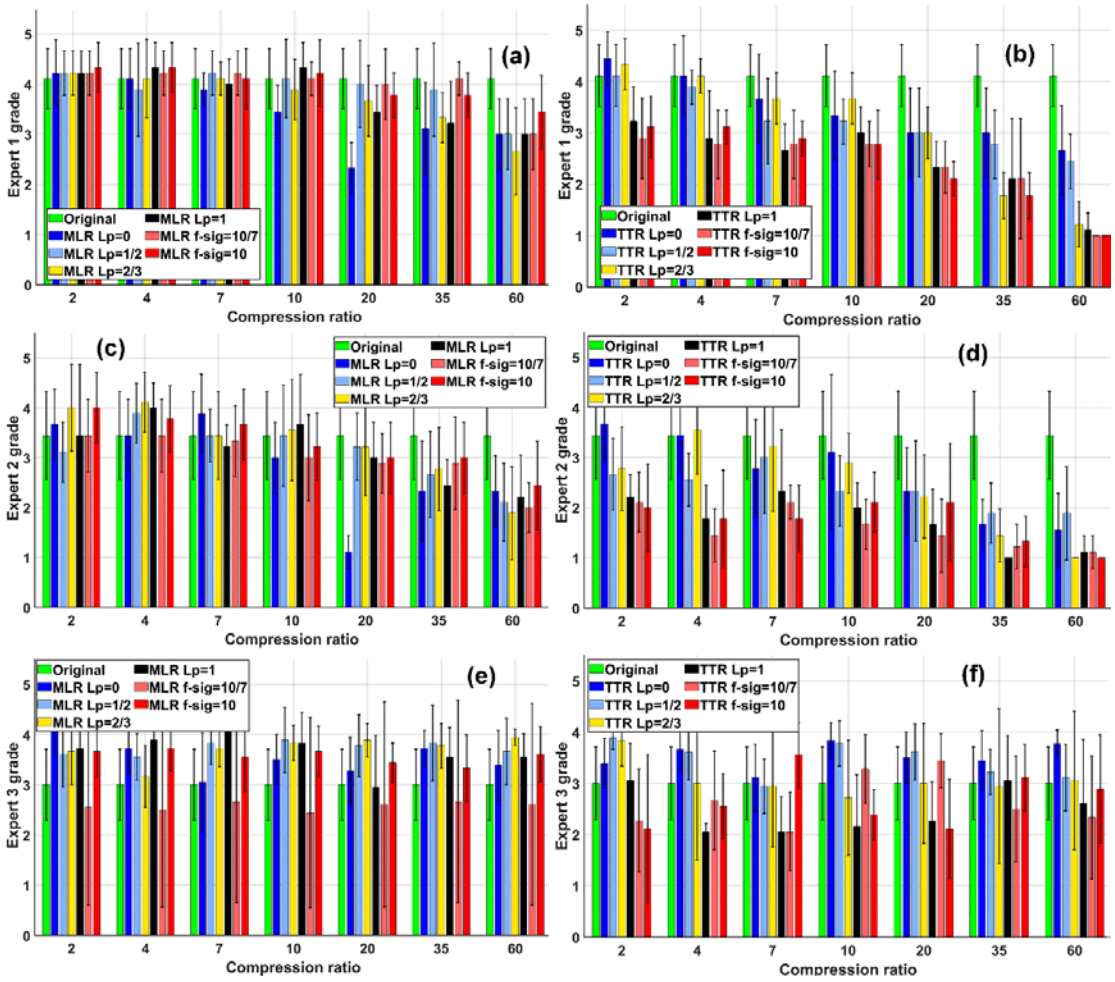

Figure S13. Best seen in color. EIQM (mean ± standard deviation) vs. selected values of CR. Top: expert 1. Bottom: expert 2. Second data set (image size 1000×512×100): (a) expert 1 - multilinear rank; (b) expert 1- tensor train rank; (c) expert 2 - multilinear rank; (d) expert 2 - tensor train rank.

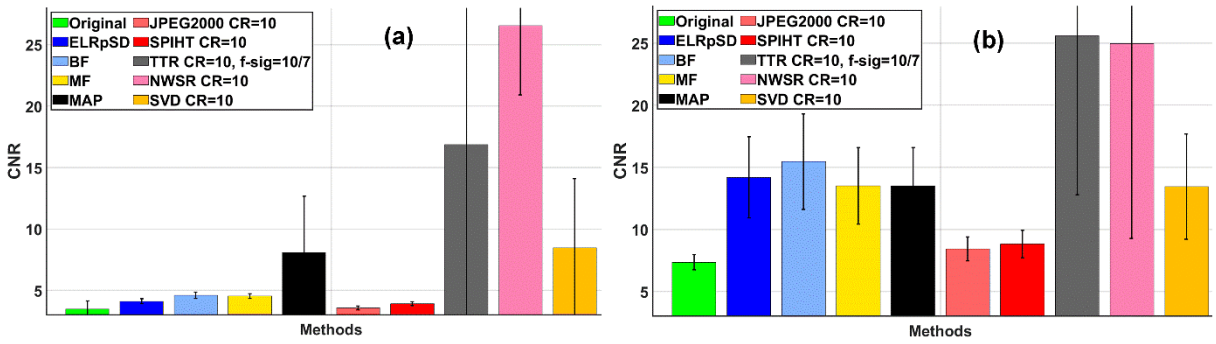

Figure S14. Best seen in color. CNR (mean ± standard deviation) for original image, ELRpSD, BF, MF and MAP no compression methods, JPEG2000, 3D SPIHT, NWSR, SVD and low tensor train rank approximations for CR=10. (a) First data set (image size 480×512×64). (b) Second data set (image size 1000×512×100).

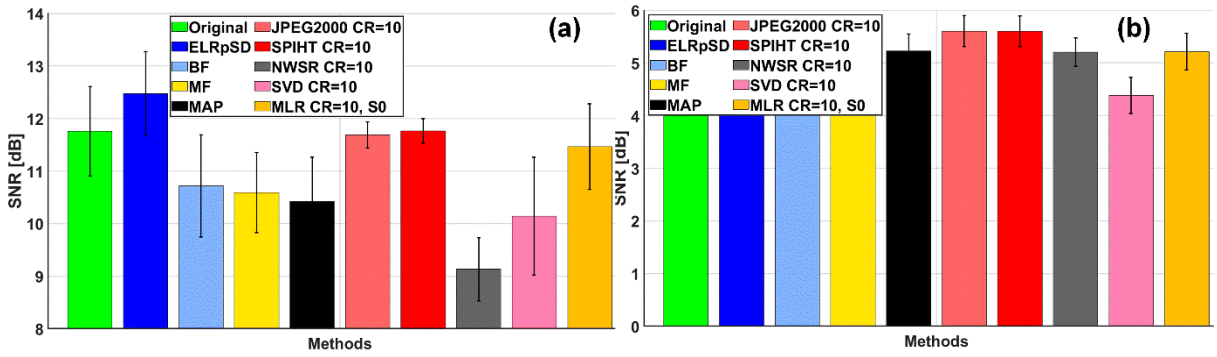

Figure S15. Best seen in color. SNR (mean ± standard deviation) for original image, ELRpSD, BF, MF and MAP no compression methods, JPEG2000, 3D SPIHT, NWSR, SVD and low multilinear rank approximations for CR=10. (a) First data set (image size 480×512×64). (b) Second data set (image size 1000×512×100).

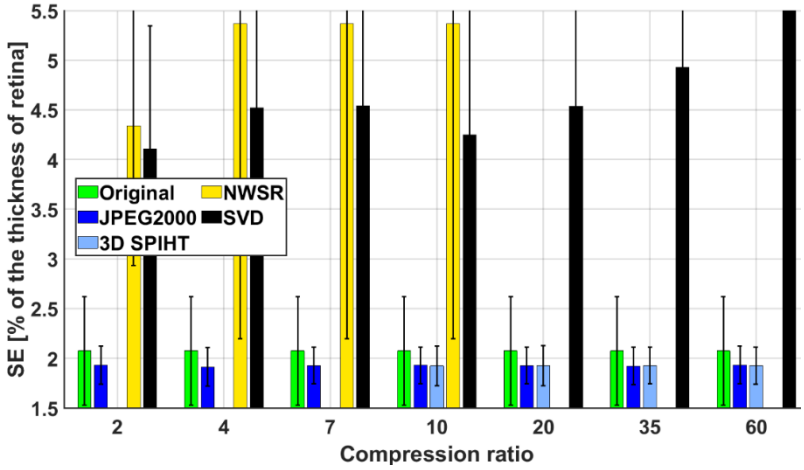

Figure S16. Best seen in color. SE (mean ± standard deviation) vs. selected values of CR for JPEG2000 compressed images and 3D SPIHT compressed images. First data set (image size 480×512×64).

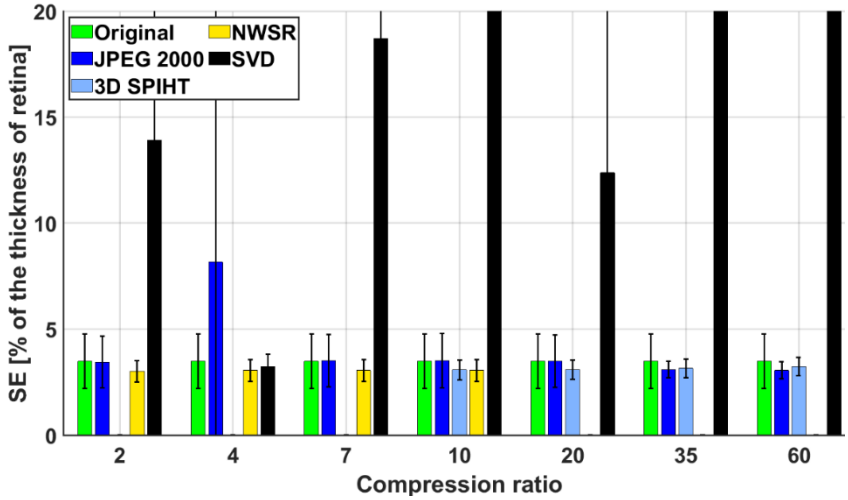

Figure S17. Best seen in color. SE (mean ± standard deviation) vs. selected values of CR for JPEG2000 compressed images and 3D SPIHT compressed images. Second data set (image size 1000×512×100).

References

Attouch H, Bolte J, and Svaiter BF 2013 Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods *Math. Program.* 37 91–129

Baghaie A, S'Souza RM and Yu Z 2015 Sparse and Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of Retinal OCT Image *IEEE Int. Symp. on Biomed. Imag.* 226-230

Beck A 2017 *First-Order Methods in Optimization* (Philadelphia, PA: SIAM)

Blumensath T and Davies ME 2008 Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.* 14 629–654

Boyd S, Parikh N, Chu E, Peleato B and Eckstein J Distributed optimization and statistical learning via the alternating direction method of multipliers 2010 *Found. Trends Mach. Learn.* 3 1-122

Candès EJ, Li X, Ma Y and Wright J 2011 Robust principal component analysis? *J. ACM*. 11 1-37

Cao W, Sun J and Xu Z 2013 Fast image deconvolution using closed-form thresholding formulas of $L_q\,(q=\frac{2}{3},\frac{1}{2})$ regularization *J. Vis. Commun. Image. Represent.* 24 31-41

Chandrasekaran V, Sanghavi S, Parillo PA and Willsky AS 2011 Rank-sparsity incoherence for matrix decomposition *SIAM J. Opt*. 21 572-596

Chen H, Fu S, Wang H, Lv H, Zhang C, Wang F and Li Y 2019 Feature-oriented singular values shrinkage for optical coherence tomography image *Opt. Laser. Eng.* 114 111-120

Chen L and Gu X 2014 The Convergence Guarantees of a Non-Convex Approach for Sparse Recovery *IEEE Trans. Sig. Proc.* 62 3754-3767

Chen X, Hou P, Jin C, Zhu W, Luo X, Shi F, Sonka M and Chen H 2013 Quantitative analysis of retinal layers' optical intensities on 3D optical coherence tomography *Investigative Ophthalmology & Visual Science* 54 6846-6851

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ and Knuth DE 1996 On the Lambert *W* function *Advances in Comput. Math.* 5 329-359

Donoho D 1995 De-noising by soft-thresholding *IEEE Trans. Inf. Theory* 41 613–627

Eckart C and Young G 1936 The approximation of one matrix by another of lower rank *Psychometrika* 1 211–218

Foucart S and Lai M 2009 Sparsest solution of underdetermined linear systems via ℓ_q-minimization for $0<q\leq 1$ *Appl. Comput. Harmon. Anal.* 26 395-407

Huang D, Swanson EA, Lin CP, et al. 1991 Optical coherence tomography *Science* 254 1178-1181

Kopriva I, Shi F and Chen X 2016 Enhanced low-rank + sparsity decomposition for speckle reduction in optical coherence tomography *J. Biomed. Opt.* 21 076008

Kolda TG and Bader BW 2009 Tensor decompositions and applications *SIAM Review* 51 455–500

Larsson V and Olsson C 2016 Convex low rank approximation *Int. J. Comput. Vis.* 120 194–214

Leith E, Chen H, Chen Y, Dilworth D, Lopez J, Masri R, Rudd H and Valdmanis J 1991. Electronic holography and speckle methods for imaging through tissue using femtosecond gated pulses *App. Opt.* 30 4204-4210.

Lee N and Cichocki A 2018 Fundamental tensor operations for large-scale data analysis using tensor network formats *Multidim Syst Sign Process.* 29 921-960

Lewis AS and Sendov H 2005 Nonsmooth analysis of singular values. Part I: theory. *Set Valued Anal.* 13 213-241

Li M, Idoughi R, Choudhury B and Heidrich W 2017 Statistical model for OCT image denoising *Biomed. Opt. Express* 8 3903-3917

Liu C, Tang J, Yan S and Lin Z 2016 Nonconvex Nonsmooth Low Rank Minimization via Iteratively Reweighted Nuclear Norm *IEEE Trans. Image Proc.* 25 829-839

Luan F and Wu Y 2013 Application of RPCA in optical coherence tomography for speckle noise reduction *Laser Phys. Lett.* 10 035603

Łojasiewicz S 1963 Une propriété topologique des sous-ensembles analytiques réels Colloques internationaux du C.N.R.S 117 Les Équations aux Dérivées Partielles 87-89.

Ma Y, Chen X, Zhu W, Cheng X, Xiang D and Shi F 2018 Speckle noise reduction in optical coherence tomography images base on edge-sensitive cGAN *Biomed. Opt. Express* 9 5129-5146

Malek-Mohammadi M, Koochakzadeh A, Babaie-Zadeh M, Jansson M and Rojas CR 2016 Successive Concave Sparsity Approximation for Compressed Sensing *IEEE Trans. Sig. Proc.* 64 5657-5671

Oh TH, Tai YW, Bazin JC, Kim H and Kweon IS 2016 Partial sum minimization of singular values in robust PCA: algorithms and applications *IEEE Trans. Patt. Anal. Mach. Int.* 38 744-758

Oseledets IV 2011 Tensor-Train Decomposition *SIAM J. Sci. Comput.* 33 2295-2317

Ozcan A, Bilenca A, Desjardins AE, Bouma BE and Tearney GJ 2007 Speckle reduction in optical coherence tomography images using digital filtering *J. Opt. Soc. Am. A.* 24 1901-1910

Ozdemir A, Iwen MA and Aviyente S 2018 Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA *Proc. IEEE*. 106 1341-1358

Phan AH, Cichocki A, Uschmajew A, Tichavský P, Luta G and Mandić D 2020 Tensor Networks for Latent Variable Analysis: Novel Algorithms for Tensor Train Approximation *IEEE Trans. Neural Net. Learn. Syst.* 31 2174-2188

Parikh N and Boyd S 2013 Proximal algorithms. *Foundations and Trends in Optimization* 1 123-231

Recht B, Fazel M and Parillo PA 2010 Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization *SIAM Rev.* 52 471-501

Said A and Pearlman WA 1996 A new fast and efficient image codec based on set partitioning in hierarchical trees *IEEE Trans. Circuits Syst. Video Technol.* 6 243–250

Shi F, Chen X, Zhao H, W Zhu, Xiang D, Gao E, Sonka M and Chen H 2015 Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images with Serous Pigment Epithelial Detachments *IEEE Trans. Med. Imag.* 34 441 - 452

van Soest G, Villiger M, Regar E, Tearney GJ, Bouma BE, and van der Steen AFW 2012 Frequency domain multiplexing for speckle reduction in optical coherence tomography. *J. Biomed. Opt*. 17 076018

Tucker LR 1966 Some mathematical notes on three-mode factor analysis *Psychometrika* 31 279-311

Wang Y, Yin W and Zeng J 2019 Global convergence of ADMM in nonconvex nonsmooth optimization *J. Sci. Comput.* 78 29-63

Xu Z, Chang X, Xu F and Zhang H 201 L-1/2 regularization: a thresholding representation theory and a fast solver *IEEE Trans. Neural Net. Learn. Syst*. 23 1013–1027

Zhang X, Xu C, Sun X and Baciu G 2016 Schatten-q regularizer for low rank subspace clustering model *Neurocomputing* 182 36-47