

# A Natural-Scene Gradient Distribution Prior and its Application in Light-Microscopy Image Processing

Yuanhao Gong, *Student Member, IEEE*, and Ivo F. Sbalzarini

**Abstract**—Signal processing in light-microscopy and cell imaging is concerned with reconstructing latent ground truth from imperfect images. This typically requires assuming prior knowledge about the latent ground truth. While this assumption regularizes the problem to an extent where it can be solved, it also biases the result toward the expected. It thus often remains unclear what prior to use for a given practical problem. We argue here that the gradient distribution of natural-scene images may provide a versatile and well-founded prior for light-microscopy images that does not impose assumptions about the geometry of the ground-truth signal, but only about its gradient spectrum. We provide motivation for this choice from different points of view, and we illustrate the resulting regularizer for use on light-microscopy images. We provide a simple parametric model for the resulting prior, leading to efficiently solvable variational problems. We demonstrate the use of these models and solvers in a variety of common image-processing tasks, including contrast enhancement, noise-level estimation, denoising, blind deconvolution, and dehazing. We conclude by discussing the limitations and possible interpretations of the prior.

**Index Terms**—Deconvolution, dehazing, denoising, gradient distribution, naturalization, noise-level estimation, parametric prior, variational method.

## I. INTRODUCTION

VIRTUALLY all light-microscopy modalities rely on signal processing and computational image analysis. Image enhancement, including deconvolution and denoising, is often a necessary first step in an analysis pipeline, preceding higher-level analyses like image segmentation [1] and motion tracking [2]. These image-enhancement tasks are often generic to a wide range of applications and imaging modalities. They are, however, inverse problems, as one attempts to reconstruct a latent “perfect image” (i.e., ground truth) from the given imperfect (noisy, blurry, hazy, etc.) observation.

Inverse problems are almost always ill-posed or ill-conditioned, especially if the transformation between ground truth

and observation is nonlinear or unknown. In order to be able to solve such problems, additional knowledge about the unknown ground truth has to be assumed.

Conceptually, there are two approaches to including additional knowledge: interpolation (smoothing or filtering) and model fitting (Bayesian inference). In the former approach, the additionally assumed prior knowledge is encoded in the choice of the interpolation basis, or in the filter kernels used. These choices typically impose certain geometric properties of the perfect image, such as connectivity, smoothness, sparsity, or curvature. In the Bayesian approach, one attempts to reconstruct a latent image such that it resembles as much as possible the observed image when run through the imaging (blurring, noise, etc.) transformation. Bayesian inference requires prior knowledge in the form of a *prior probability distribution* that sufficiently constrains the reconstruction problem to render it well-posed. Frequently used priors in light-microscopy image processing include sparsity in the spatial and/or frequency domain [3], total variation (TV) [4]–[6], mean curvature (MC) [7]–[9], Gaussian curvature (GC) [8], [10]–[12], and hybrid priors [13], [14].

While prior knowledge can regularize the inverse problem to an extent where it can be solved, it also biases the result toward the expected. Inappropriate priors may obscure features in the image, or lead to reconstruction artifacts like fringes or ringing. Choosing the “right” prior, however, is as hard as solving the original problem, since the latent ground-truth image is unknown. The main drawback of frequently used priors is that they are not adaptive to the image contents, and often entirely unrelated to it. They merely postulate geometric properties of the latent image. The popular TV prior [4]–[6], for example, presupposes that the latent image be a collection of uniformly bright regions, i.e., to be piece-wise constant. Imposing this prior leads to removal of image detail and processing artifacts if this presumption is not justified.

Spectral priors have been introduced in order to relax the geometric constraints. They do not directly impose knowledge about a property of the latent image, but only about the histogram (or distribution, i.e., the spectrum) of that property. As such, they are weaker priors. A particularly popular spectral prior is the Gradient Distribution Prior (GDP), which presupposes a certain statistical distribution of the gradients in the image, i.e., a certain gradient histogram. It has been shown to lead to better results than the TV prior in many image-processing tasks [15]–[21], where the weaker regularization is sufficient. This is because the GDP does not constrain the solution to be piecewise constant in this case, but allows more flexible reconstructions. A review about signal processing in the gradient

Manuscript received March 12, 2015; revised August 16, 2015; accepted December 02, 2015. Date of publication December 04, 2015; date of current version January 21, 2016. This work was supported in part by the German Federal Ministry of Research and Education (BMBF) under funding code 031A099 and in part by the Swiss National Science Foundation under Grant CRSII3-132396/1. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Charles Kervrann.

The authors are with the MOSAIC Group, Chair of Scientific Computing for Systems Biology, Faculty of Computer Science, TU Dresden, 01307 Dresden, Germany, and also with the Center for Systems Biology Dresden (CSBD), Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany (e-mail: ivos@mpi-cbg.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2015.2506122

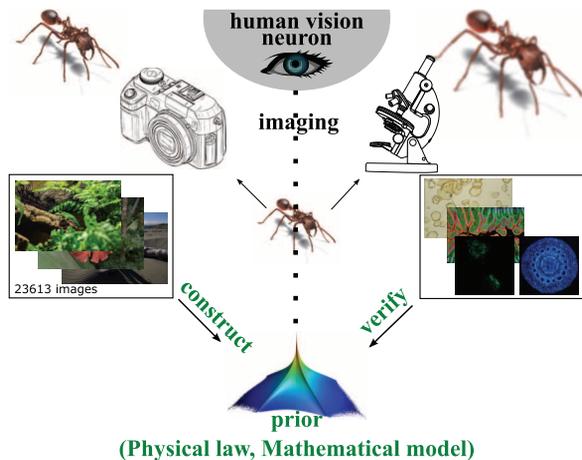


Fig. 1. Concept of using natural-scene GDPs in light-microscopy imaging. The prior is learned from natural-scene images, which are largely aberration-free because the imaged objects are much larger than the wavelength of the light. Since they obey the same physical laws as light-microscopy images, we propose to use this prior also in that case. We hence learn the GDP from natural-scene images and then use it on light-microscopy images. The human vision system links the two, as it has evolved to process natural-scene images, but is also used to look at microscopy images.

domain can be found in [22]. Some recent advances in this area are described in [23]–[27].

Conceptually, GDPs are related to histogram equalization. However, while the latter presupposes a uniform distribution in the intensity domain, the former operates in the gradient domain, and the presupposed distribution is not uniform, but learned from examples.

Despite their flexibility and success in signal processing, GDPs learned from natural-scene images have to the best of our knowledge never been adopted in light-microscopy image processing. Natural-scene images are images of nature-made objects, such as landscapes, animals, and plants, where the size of the imaged objects is much larger than the wavelength of the imaging light. Large collections of natural-scene images are readily available from public sources. When taken under good conditions, these images can be largely free of the aberrations often found in light microscopy. Since both image natural objects according to the same physical laws, albeit on different length scales, we propose using GDPs learned from high-quality natural-scene images as priors in light-microscopy image processing. This philosophy is illustrated in Fig. 1.

We motivate this proposition from different points of view. We further provide a parametric model for the resulting GDP and design efficient algorithms for variational problems including the corresponding regularizer. We illustrate the use of the natural-scene GDP on light-microscopy images in a range of applications, comparing with state-of-the-art methods.

#### A. Why not Directly Learn the GDP From Microscopy Images?

A natural question to ask is why one should process light-microscopy images using a GDP learned from natural-scene images. This seems unintuitive at first. Why not directly learn a GDP from light-microscopy images, maybe even of a certain imaging modality, and then use that? The first answer is that (1)

learning the GDP from natural-scene images avoids learning unwanted aberrations, like diffraction blur, noise, etc., which are present in light-microscopy images. Since removing these aberrations is often the aim of image processing, the prior assumption about the latent ground truth must not contain their signatures. Natural-scene images are near-perfect in the sense of very high SNR, low diffraction (the imaged objects are much larger than the wavelength of the light), and low aberrations.

Additional reasons are: (2) The ground truth of light-microscopy images is unknown. Even when using noise-robust gradient estimation, it would remain unclear whether the learned prior is dominated by the objects, the imaging noise, or the gradient estimation scheme. (3) Human vision is adapted to processing natural-scene images (Fig. 1). Therefore, learning a GDP from natural-scene images likely improves perceived image quality.

This is confirmed for denoising in Section VII-C. As the results in Table VII show, using a GDP learned directly from microscopy images gives the worst results. Using the GDP learned from natural-scene images, however, yields state-of-the-art performance.

#### B. Our Contribution

Our contribution is three-fold: (1) We test the natural-scene GDP on light-microscopy images, illustrating its use as a default regularizer in the absence of more specific prior knowledge. (2) We provide a novel parametric model for GDPs, which includes for the first time the correlation between the  $x$  and  $y$  components of the gradient. This correlation has been ignored in the well-known hyper-Laplace model [20], [21]. As we show here, the gradient components in an image are correlated. Ignoring these correlations may not only lead to artifacts in the result, but also complicates solving the resulting variational problem, frequently requiring alternating optimization over the gradient components. (3) We provide efficient solver algorithms for variational models containing the new parametric model as a regularizer.

#### C. Organization of This Paper

Before presenting our results, we formalize the problem in Section II and provide a motivation for our proposal in Section III. Variability of the GDP on light-microscopy images is studied in Section IV. In Section V, we provide novel parametric models for the GDP in one and two dimensions. In Section VI, we demonstrate how to use the present GDP as a regularizer in a variational framework. In Section VII, we illustrate several applications from light-microscopy image processing. We conclude and discuss this work in Section VIII.

## II. MATHEMATICAL FRAMEWORK

We aim at computing an estimate  $\hat{U}(\vec{x})$  ( $\vec{x}$  is the spatial coordinate) of the unknown, latent perfect image  $\tilde{U}(\vec{x})$  from the observed discrete samples  $I(i, j) = F(\tilde{U}(\vec{x}_{i,j}))$ , which are the pixels  $(i, j) \in \mathbb{Z}_0^+ \times \mathbb{Z}_0^+$  of the data image. The data image  $I$  has been generated from the underlying truth  $\tilde{U}$  by the imaging process  $F$ , introducing aberrations, such as blur, noise, scattering, down-sampling, etc.

The reconstruction problem can be expressed in a Bayesian framework as MAP estimation:

$$\begin{aligned} \hat{U} &= \arg \max_U \{ p(U|I) \propto p(I|U)p(U) \} \\ &\Leftrightarrow \arg \min_U \{ -\log(p(I|U)) - \lambda \log(p(U)) \}, \end{aligned} \quad (1)$$

where the scalar regularization parameter  $\lambda$  is introduced to balance the likelihood  $p(I|U)$  and the prior  $p(U)$ . From this MAP formulation, it is clear that the estimate  $\hat{U}$  only converges toward  $\tilde{U}$  if the prior characterizes the latter.

In variational form, this leads to a minimization problem over an energy  $\mathcal{E}(U)$  where the prior becomes a regularizer:

$$\hat{U} = \arg \min_{U \in F_s} \left\{ \mathcal{E}(U) = \int_{\vec{x} \in \Omega} \Phi_1(U, I) d\vec{x} + \lambda \int_{\vec{x} \in \Omega} \Phi_2(U) d\vec{x} \right\}, \quad (2)$$

where  $\Phi_1$  is a data-fitting cost function, which models the (generally unknown) imaging transformation  $F$  in order to quantify how well a certain hypothetical reconstruction  $U$  fits the data  $I$ . Assuming a Gaussian distribution for the measurement errors, e.g., leads to an  $\ell_2$  norm in the data fitting term.  $\Phi_2$  is the regularization function acting on  $U$ . This term imposes prior knowledge (sparsity, smoothness, etc.) about the unknown perfect image  $\tilde{U}$ .  $\Omega$  is the image domain, and  $F_s$  is the postulated function space in which  $U$  lives.  $\Phi_1$  has a close relationship with  $-\log(p(I|U))$  and  $\Phi_2$  with  $-\log(p(U))$ .

When using a spectrally regularized model, the regularization term does not directly act on  $U$ , but on a distribution or histogram  $p(\cdot)$  of some features of  $U$ :

$$\begin{aligned} \hat{U} &= \arg \min_{U \in F_s} \left\{ \int_{\vec{x} \in \Omega} \Phi_1(U, I) d\vec{x} \right\} \\ &\quad \text{s.t. } p(J(U)) = p_J^{\text{pr}}, \end{aligned} \quad (3)$$

where  $J$  is a filter (map, feature, differential operator, etc.) and  $p_J^{\text{pr}}$  is the corresponding spectral prior. In GDPs, the filter  $J(\cdot) = \nabla(\cdot)$ , and  $p(\cdot)$  is the gradient distribution.

The hard spectral constraint can be relaxed by introducing an auxiliary variable  $\omega$  for decoupling:

$$\begin{aligned} \hat{U} &= \arg \min_{U \in F_s} \left\{ \int_{\vec{x} \in \Omega} [\Phi_1(U, I) + \lambda \Phi_2(U, \omega)] d\vec{x} \right\}, \\ &\quad \text{s.t. } p(J(\omega)) = p_J^{\text{pr}}. \end{aligned} \quad (4)$$

This decoupling technique is generic to variational models with hard constraints. It has previously been used, e.g., in split-Bregman [28], TGV [14], and hyper-Laplacian [20] models.

### III. MOTIVATION: WHY THE GRADIENT DISTRIBUTION?

Spectral priors are typically learned or estimated from image collections. Given a sufficiently diverse collection of images, the histogram or probability distribution of a spectral prior is estimated by averaging  $p(J(I))$  over all images  $I$ . There are many features that can be computed, including color and texture features, but the image gradient  $J(I) = \nabla I$  is particularly interesting. This is first because it is remarkably invariant across images [29]. Second, it is easy to compute and can hence be

TABLE I  
NATURAL-SCENE IMAGE DATASETS USED TO LEARN THE PRIOR. SOURCE URLs ARE GIVEN IN THE FOOTNOTES.<sup>1,2,3,4,5,6,7</sup>

Footnote	1	2	3	4	5	6	7	all
#images	1005	1000	5063	832	1491	6033	8189	23613

learned from large image collections. Third, the gradient has a simple intuitive meaning as the first-order approximation to  $\tilde{U}$ .

In the following, we use the term *gradient field* to denote the gradient image, i.e., an image  $\vec{G}$  that has the same size as the original data image  $I$ , but where each pixel stores two values that are the two components of the gradient of  $I$  at that location. The *gradient distribution* is the histogram or probability distribution of these values across all pixels, and/or across multiple images. We restrict our discussion to two-dimensional images where the gradient has two components. Extensions to higher-dimensional images are possible by adding additional gradient components.

GDPs have been used in Bayesian frameworks for image denoising [15], deblurring [17], restoration [21], super resolution [30], and others [16]–[21]. Deblurring in the gradient domain is more efficient than working with the original pixel values [17], [18]. This can be explained by the reduced correlation in the gradient domain [31].

Reconstructing an image from its gradient field amounts to integration with one point constraint [22], [23], [32]–[37]. One way of doing this is by solving a Poisson equation. With proper boundary conditions, the solution is unique, and there exists a wealth of stable, efficient, and accurate numerical solvers for this equation. We hence prefer this way of performing the numerical integration.

### IV. THE GRADIENT DISTRIBUTION PRIOR (GDP)

In order to learn the GDP from natural-scene images, we collect 23613 images of natural scenes. We compute the resulting gradient distributions and assess the variability of the training images around the average distribution.

#### A. Training Dataset and Gradient Approximation

We collected seven datasets of natural-scene images as shown in Table I. Each image  $I(i, j)$  was converted to 8-bit gray-scale. The gradient field is defined as:

$$\vec{G}(i, j) = \nabla I(i, j), \quad (5)$$

where here we use the first-order forward finite-difference approximations  $\nabla I \approx (I(i+1, j) - I(i, j), I(i, j+1) - I(i, j))$ . Certain features of the resulting gradient histogram are artifacts of this discretization scheme used to compute the gradient. The high-probability lines at 0, 45, and 90 degrees in Fig. 2 are artifacts of using forward differences. When using backward differences, the 45-degree line rotates to  $-45$  degrees. Using central differences, the 0 and 90-degree lines disappear, but smoothing artifacts at high gradients appear, since central differences cannot capture edges in the image. However, as long as the discretization scheme used to learn the GDP is the same as

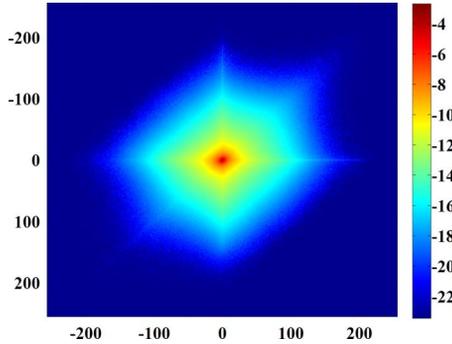


Fig. 2. Average gradient distribution from the training set of 23613 natural-scene images used here. Color codes normalized probability in log-scale.

the one later used to apply the model, the method is consistent, as the artifacts cancel out.<sup>1-7</sup>

We choose the first-order finite-difference scheme for our computations because it has the highest spatial resolution. We compute the gradient at interior pixels, using the outermost pixel layers as boundary condition. Due to the use of 8-bit gray-scale images, possible gradients are in the discrete domain  $[-255, 255] \times [-255, 255]$ , where we can construct the 2D histogram of  $\vec{G}$ . We use  $G^x$  and  $G^y$  to denote respective components of  $\vec{G}$ .

In order to normalize the histogram, we divide all bins by the total number of pixels in the image, i.e., by  $mn$  where  $m$  and  $n$  are the number of pixels along the  $x$  and  $y$  edges of the image. After aggregating data from all images in the training set, we further normalize by the total number of images in the dataset. The resulting empirical distribution  $p^{\text{pr}}$  is shown in Fig. 2. It has previously been shown that this GDP is stable on the training dataset, in the sense that the gradient histograms of individual training images cluster around the GDP learned across the entire dataset [31].

### B. Variability of the Prior

The first step in studying the natural-scene GDP on light-microscopy images is to check that good microscopy images cluster around the GDP. For this, we manually collect a test dataset of 40 high-quality microscopy images from public internet sources. Some examples are shown in Fig. 3.

We compute the RMS distance from the GDP for each image's gradient distribution. The distance histogram is shown in Fig. 4. It confirms that the gradient distributions tested cluster around the GDP learned from natural-scene images.

### C. Correlation With Image Quality

The second step in studying the natural-scene GDP is to check that it is correlated with perceived image quality. Only then,

<sup>1</sup>[HTTP://WWW.VISION.EE.ETHZ.CH/SHOWROOM/ZUBUD/](http://www.vision.ee.ethz.ch/showroom/zubud/)

<sup>2</sup>[HTTP://SEE.XIDIAN.EDU.CN/FACULTY/WSDONG/DATA/FICKR\\_IMAGES.RAR](http://see.xidian.edu.cn/faculty/wsdong/data/flickr_images.rar)

<sup>3</sup>[HTTP://WWW.ROBOTS.OX.AC.UK/~VGG/DATA/OXBUILDINGS/](http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/)

<sup>4</sup>[HTTP://WWW.COMP.LEEDS.AC.UK/SCS6JWKS/DATASET/LEEDSBUTTERFLY/](http://www.comp.leeds.ac.uk/scs6jwks/dataset/leedsbutterfly/)

<sup>5</sup>[HTTP://LEAR.INRIALPES.FR/~JEGOU/DATA.PHP](http://lear.inrialpes.fr/~jegou/data.php)

<sup>6</sup>[HTTP://WWW.VISION.CALTECH.EDU/VISIPEDIA/CUB-200.HTML](http://www.vision.caltech.edu/visipedia/CUB-200.html)

<sup>7</sup>[HTTP://WWW.ROBOTS.OX.AC.UK/~VGG/DATA/FLOWERS/102/INDEX.HTML](http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html)

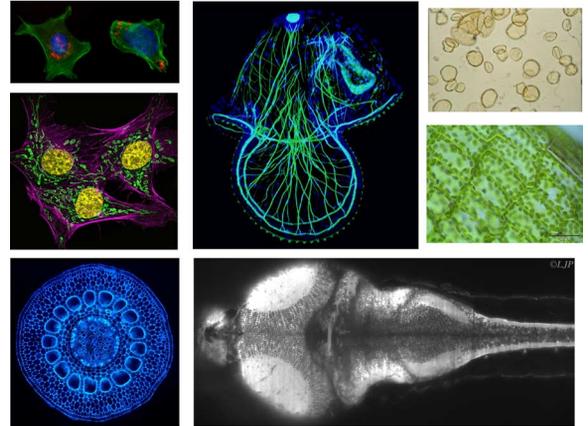


Fig. 3. Seven examples from the 40 microscopy test images.

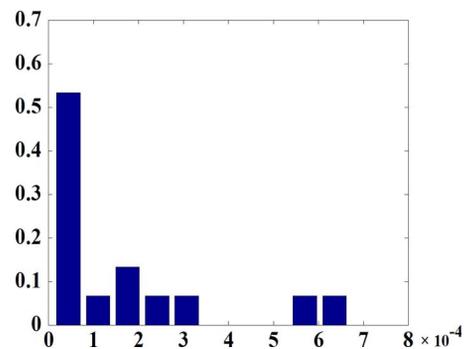


Fig. 4. Histogram of the RMS distances between the natural-scene GDP and the gradient distributions of the 40 test images.

imposing the prior in light-microscopy images is expected to improve the result. As we have shown before, the distance between the gradient distribution of any given image and the GDP is highly correlated with subjectively perceived image quality [31]. We use the standard LIVE benchmark dataset for image quality assessment [38] in order to show that the Hellinger distance is a good metric to use in this context. For this, we test the  $\ell_2$ -norm,  $\ell_1$ -norm, cosine distance, the Earth Mover Distance (EMD),  $\chi^2$  distance, and the Hellinger distance. Using these distances, we form the score:

$$\text{score} = \text{Distance}(p(\nabla I^{\text{true}}), p(\nabla I^{\text{distorted}})). \quad (6)$$

If the ground-truth image  $I^{\text{true}}$  is unknown (i.e., in a real-world application rather than a benchmark setting), the score is defined with respect to the GDP:

$$\text{score}_{\text{pr}} = \text{Distance}(p^{\text{pr}}, p(\nabla I^{\text{distorted}})). \quad (7)$$

A measure of subjectively perceived image quality is provided by the LIVE benchmark's DMOS (difference mean opinion score). Different correlations between DMOS and our objective score are reported in Table II. In all cases, the Hellinger distance between the gradient distributions shows the best correlation with DMOS. Together with its stability, this renders the natural-scene GDP a good default prior for light-microscopy applications, in the absence of more specific prior knowledge.

TABLE II

CORRELATIONS BETWEEN SUBJECTIVELY PERCEIVED IMAGE QUALITY (DMOS FROM LIVE BENCHMARK [38]) AND OUR OBJECTIVE SCORE USING DIFFERENT DISTANCE METRICS. THE FOLLOWING CORRELATIONS ARE REPORTED: PEARSON'S LINEAR CORRELATION COEFFICIENT (PCC), SPEARMAN'S RANK-ORDER CORRELATION COEFFICIENT (SCC), AND KENDALL'S RANK-ORDER CORRELATION COEFFICIENT (KCC).

	$\ell_2$	$\ell_1$	cos	EMD	$\chi^2$	Hellinger
PCC	0.6193	0.7926	0.6277	0.5172	0.7662	<b>0.8687</b>
SCC	0.6434	0.7773	0.6114	0.7576	0.7977	<b>0.8630</b>
KCC	0.4588	0.5822	0.4355	0.5639	0.6027	<b>0.6745</b>

## V. PARAMETRIC MODEL FOR THE GDP

In order to efficiently use the GDP as a regularization term, a parametric model is desirable. We here provide parametric models for the marginal and joint gradient distributions, and we assess their approximation accuracy. We compare these new models with traditional gradient-distribution models, such as hyper-Laplace models, and with TV in 1D and 2D. We then discuss the convexity, sparsity, and entropy of the new models.

### A. 1D Marginal Model

Traditionally, image gradient histograms have been modeled as component-wise generalized Gaussian distributions:

$$\log(P(G^x)) = -a_1|G^x|^{b_1} + c_1, \quad (8)$$

where  $a_1$ ,  $b_1$ , and  $c_1$  are the model parameters. Here and elsewhere, this model is treated as a function rather than a distribution. Therefore,  $c_1$  is an independent parameter, instead of being the normalization constant. This is inconsequential for the later variational model, since the normalization constant becomes an additive offset to the total energy, which has no impact on the location of energy minima. We do, however, use the normalized models and histograms whenever we compute probabilistic quantities, such as entropies. The model in (8) includes Gaussian ( $b_1 = 2$ ), Laplacian ( $b_1 = 1$ ), and hyper-Laplacian distributions ( $b_1 = 0.6$ ) as special cases [20].

Combining considerations of model simplicity, computational efficiency, and solvability of the resulting variational optimization problem, we propose the following new model for the 1D marginal gradient histogram:

$$\log(P(G^x)) = -a_2(G^x)^2 - \log(b_2 + |G^x|^2) + c_2, \quad (9)$$

where  $a_2$ ,  $b_2$ , and  $c_2$  are the parameters (again, the fitting function need not be normalized to a probability distribution, for the reasons outlined above). The results of fitting this model to the training data are shown in Table III, compared with other models. As shown in Table III and Fig. 5, our model outperforms the previously used Laplacian, Gaussian, and hyper-Laplacian models in terms of fitting quality.

In addition to the increased fitting accuracy, our model has several other advantages:

- **Integrability:** Our model is integrable, which is convenient for use in optimization algorithms and for analytically computing the cumulative distribution function (CDF). As shown in Fig. 5(b), the CDF of our model is still meaningful when other models become invalid (Gaussian,

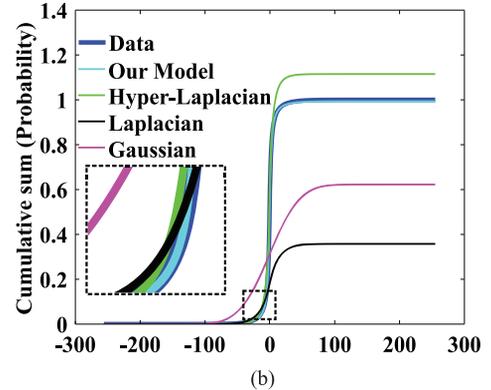
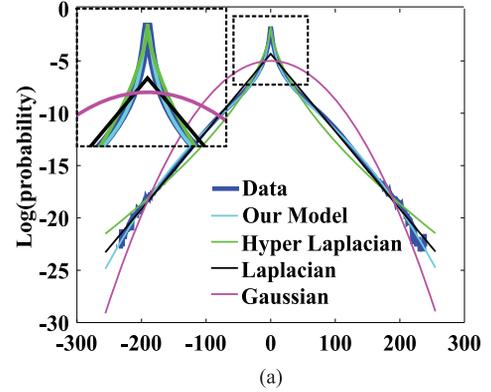


Fig. 5. Comparison of (a) marginal models (log scale) and (b) their cumulative sums (linear scale). Optimal parameters are used for each model. The quantitative fits are given in Table III. Lines are distinguished by color, as reproduced in the electronic version.

TABLE III  
GOODNESS OF FIT COMPARISON FOR ALL MODELS: OUR MODEL (9), HYPER-LAPLACIAN, LAPLACIAN, GAUSSIAN.

	Image set	1	2	3	4	5	6	7
our	SSE	271	324	266	44.4	38.2	62.8	30.7
	$R^2$	0.96	0.93	0.96	0.99	0.99	0.98	0.99
hyp	SSE	576	301	537	45.4	389	70.5	250
	$R^2$	0.92	0.93	0.91	0.98	0.96	0.98	0.97
Lap	$SSE \times 10^{-3}$	1.86	3.01	3.02	3.95	2.34	3.90	3.95
	$R^2$	0.74	0.30	0.52	0.13	0.81	0.10	0.57
Ga- uss	$SSE \times 10^{-4}$	0.83	1.02	1.10	1.24	1.32	1.23	1.64
	$R^2$	-0.12	-1.3	-0.72	-2.6	-0.046	-2.5	-0.75

Laplacian, and hyper-Laplacian have asymptotes  $\neq 1$ ) or hard to integrate (hyper-Laplacian).

- **Computational efficiency:** our model has a simple mathematical form that can efficiently be evaluated. The gain is substantial, as shown in Suppl. Fig. 1.
- **Optimization efficiency:** the variational problem resulting from our model can be written as the difference of two convex functions. Optimization problems involving our model can hence efficiently be solved using D.C. programming (short for: difference of convex).

### B. 2D Joint Model

In previous works, the 2D gradient distribution has been assumed to be the product of two independent 1D marginals along  $x$  and  $y$  [17]–[21]. This is not necessarily the case, as there can be correlations between the gradient components. We hence estimate the joint 2D gradient distribution from the training dataset and provide a 2D parametric model for it.

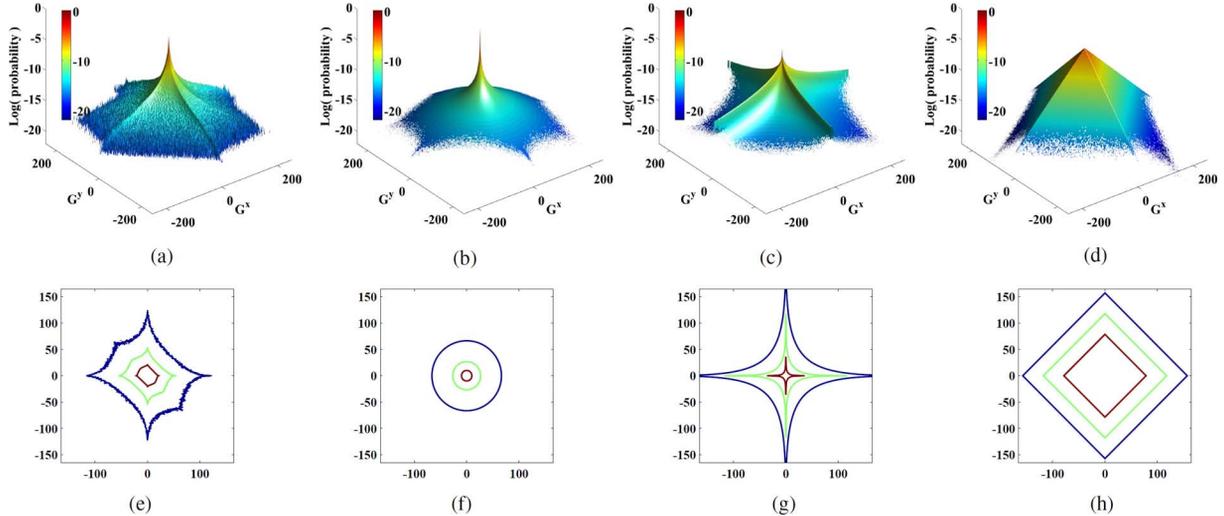


Fig. 6. Visual comparison of model fits for 2D joint gradient distributions in log scale. (a,e) Empirical 2D gradient distribution of the training data. (b,f) Best fit of our new model from (11). (c,d,g,h) Best fits of the hyper-Laplacian and Laplacian models. For each model, we show contour lines of values  $-13$ ,  $-11$ , and  $-9$ . The area included by isoline  $-13$  of our model is 3% of the whole domain, but the total probability mass in that area is 99%. Moreover, our new model is rotationally symmetric, not imposing any *a priori* preferred gradient orientation. (a) data (b) our model (c) hyper-Laplacian (d) Laplacian (e) data iso lines (f) our model's iso lines (g) hypLap iso lines (h) Laplace iso lines.

TABLE IV  
CORRELATION BETWEEN  $G^x$  AND  $G^y$ .

Image set	1	2	3	4	5	6	7
correlation	-0.12	-0.23	-0.19	-0.22	-0.12	-0.25	-0.11
(log scale)	0.37	0.31	0.28	0.37	0.17	0.39	0.18

As shown in Table IV, the two gradient components are weakly negatively correlated in the training dataset (from edges in the images). This weak correlation between the gradient components explains why alternating minimization had to be used in previous works that considered the marginal models independently, and why the results were still good even though the prior was not strictly correct.

The traditional model (8) can easily be extended to 2D:

$$\log(P(\vec{G})) = -a_1(|G^x|^{b_1} + |G^y|^{b_1}) + c_1, \quad (10)$$

where  $a_1$ ,  $b_1$ , and  $c_1$  are the parameters. This model, including the hyper-Laplacian as a special case, treats the  $x$  and  $y$  components of the gradient as independent and identically distributed.

Considering that a correlation between the gradient components may exist, we instead propose the following model (derived from (9)) for the 2D joint gradient histogram:

$$\log(P) = -a_2(|G^x|^2 + |G^y|^2) - \log(b_2 + |G^x|^2 + |G^y|^2) + c_2. \quad (11)$$

This model is based on the weaker assumption that the differences between gradients at neighboring pixels are uniformly randomly distributed.

The fitted parameters are shown in Table V. Fig. 6 compares the model with previous models. The gradient histogram of the training data is shown in Figs. 6(a), 6(e) whereas the best-fit parametric models are plotted in the remaining panels. The area included by isoline  $-13$  (blue isoline) of our model is only 3% of the whole domain, but the total probability mass in that area is 99%, indicating a high sparsity of our model (see also Section V-D). Moreover, the new model is rotationally symmetric, which is desirable for two reasons: (1) It renders the

TABLE V  
PARAMETERS AND GOODNESS OF FIT OF THE 2D MODEL.

Image set	1	2	3	4	5	6	7	all
$a_2 \times 10^5$	10.9	4.81	8.27	4.42	16.5	5.21	11.0	6.21
$b_2 \times 10^2$	4.56	6.67	4.22	2.60	1.85	3.62	1.01	2.39
$c_2$	-4.74	-4.38	-4.60	-4.99	-5.79	-4.91	-6.03	-5.24
$SSE \times 10^{-5}$	1.13	0.892	1.32	1.47	1.72	1.26	1.17	1.48
$R^2$	0.83	0.85	0.90	0.82	0.86	0.89	0.89	0.90

model robust against discretization artifacts of how the gradient has been computed (cf. Section IV). (2) Except for special images (e.g., architecture) there is no *a priori* reason why certain gradient directions should be preferred.

The cumulative distribution function (CDF) of the 2D model

$$C(\vec{G}) = \int_{-\infty}^{G^y} \int_{-\infty}^{G^x} P((u,v)) du dv \quad (12)$$

is sensitive to image transformations (see Suppl. Fig. 2), rendering the parameters identifiable and the model useful for detecting and undoing such transformations. This inspires us to define a single CDF parameter that quantifies how close the gradient distribution of an image is to the GDP expected for a natural-scene image.

### C. The Naturalness Factor

We define a scalar number measuring the distance between an image's gradient histogram and the natural-scene GDP. For this, however, we need to simplify the model to a single parameter. Since the model is most sensitive w.r.t. parameter  $a_2$  (see Suppl. Fig. 3), we define  $T = \sqrt{a_2}$  and set  $b_2 = c_2 = 0$ , leading to:

$$P(G^x) = |G^x|^{-2} e^{-(TG^x)^2}. \quad (13)$$

Using integration by parts, the CDF of this simplified model can be analytically computed:

$$\tilde{C}(G^x) = -\frac{e^{-(TG^x)^2}}{G^x} - T\sqrt{\pi}\text{erf}(TG^x) + H(G^x), \quad (14)$$

where  $H$  is the Heaviside and  $\text{erf}$  the error function.

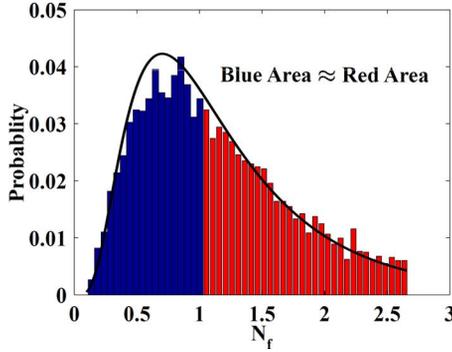


Fig. 7. Naturalness factor ( $N_f$ ) distribution for the natural-scene images of the training set. Blue bars indicate ( $N_f < 1$ ), red bars  $N_f > 1$ . The black line is a log-normal distribution with parameters  $\mu = 0.039$  and  $\sigma = 0.613$ . The median is 1.

This model has only one single scalar parameter,  $T$ , which can be determined by solving the convex problem:

$$\min_T \int (\log(p(G^x)) + T^2(G^x)^2 + 2\log(|G^x|))^2 dG^x, \quad (15)$$

which has the unique analytical solution:

$$T = \left( \frac{-\int (2\log(|G^x|) + \log(p(G^x)))(G^x)^2 dG^x}{\int (G^x)^4 dG^x} \right)^{1/2}, \quad (16)$$

where  $p(G^x)$  is the marginal distribution. Therefore, the parameter  $T$  can explicitly be computed for any given image.

Comparing an image's parameter  $T$  with the expected value  $T_{pr}$  from natural-scene images, i.e., from the GDP, we define:

**Definition:** For any image  $I$ , the **naturalness factor**  $N_f$  is defined as  $N_f = T/T_{pr}$  and the image  $I_n$  generated from  $I$  such that  $T_n \approx T_{pr}$  is called the **naturalized image**.

Since  $T_{pr}$  is obtained as an average over natural-scene images, and natural-scene images cluster around the GDP, the  $N_f$  of a natural-scene image is expected to be distributed around 1, as confirmed in Fig. 7. The range of values in our training set is  $N_f \in [0.2, 2.7]$ , and the naturalness factors of the training images satisfy a log-normal distribution, as expected from the definition of  $T$ .

#### D. Convexity, Sparsity, and Entropy of the GDP

1) **Convexity:** The TV (Laplacian) prior is popular because it leads to convex variational models. While the hyper-Laplacian would fit the data better, it leads to a non-convex variational model, which is hard to solve. We show here that our model (9) and its 2D variant (11) are quasi-concave, which means that all iso-sets are convex, simplifying optimization while still fitting the data better.

**Lemma V.1:** Equations (9) and (11) are quasi-concave.

*Proof:* For (9), we have:

$$\log(P(G_1^x)) < \log(P(G_2^x)) \text{ when } G_1^x > G_2^x > 0 \quad (17)$$

$$\log(P(G_1^x)) < \log(P(G_2^x)) \text{ when } G_1^x < G_2^x < 0. \quad (18)$$

This monotonicity property with respect to 0 ensures that (9) is quasi-concave. Equation (11) is a rotation of (9) with respect to the  $y$  axis. Therefore, the set  $\{\vec{G} : \log(P(\vec{G})) \geq h\}$  is convex  $\forall h$ .  $\square$

TABLE VI  
COMPARISON OF DIFFERENT MODELS.

Model	Convexity	Fitting accuracy
Eq. 9	quasi	high
hyper-Laplacian	quasi	medium
Laplacian	yes	low
Gaussian	yes	very low
Eq. 11	quasi	high
2D hyper-Laplacian	no	medium
2D Laplacian	yes	low
2D Gaussian	yes	very low

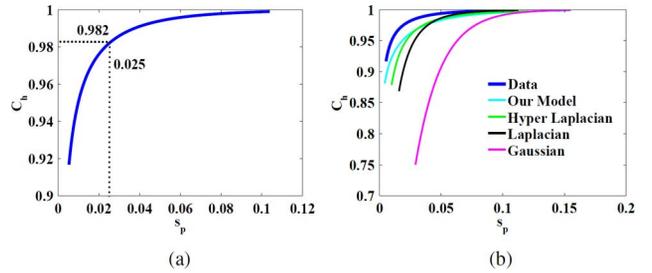


Fig. 8. Sparsity of the GDP. (a) 98.2% of the information can be encoded with only 2.5% of the dictionary at a cutoff level of  $h = 3.6 \times 10^{-6}$ . (b) The present model is almost as sparse as the data, and sparser than previous models. Lines are distinguished by color, as reproduced in the electronic version.

The resulting variational energy functional when using this model as a regularizer, however, is not quasi-concave. Nevertheless it can be written as the difference of two convex functions. Such optimization problems are known as D.C. problems (short for: difference of convex), and efficient solvers are available for them. The present model hence leads to efficiently solvable variational problems, while fitting the data best. Table VI qualitatively compares different models.

2) **Sparsity:** The gradient distribution balances sparsity and signal encoding. We quantify the sparsity of  $p(\vec{G})$  using the classical  $\ell_h^0$ -norm sparsity measure:

$$s_p(h) = \frac{\int \int \chi_{p(\vec{G}) > h} dG^x dG^y}{\int \int dG^x dG^y}, \quad (19)$$

where  $\chi : \mathbb{R} \rightarrow \{0, 1\}$  is an indicator function. Further:

$$C_h(h) = \int \int \chi_{p(\vec{G}) > h} p(\vec{G}) dG^x dG^y, \quad (20)$$

which is the total probability mass on levels larger than  $h$ . The sparsity  $s_p(h)$  measures how many words (of some dictionary) are needed to encode the information in  $C_h(h)$  with an accuracy or tolerance  $h$ . The relationship between  $s_p$  and  $C_h$  is shown in Fig. 8(a) for the images from our training set. We observe that the gradient signal is sparse already at a low cutoff  $h$ . Fig. 8(b) shows the sparsity/information curves for different parametric models of the GDP. Our model is about as sparse as the data, whereas all other models are less sparse and throw away more information.

3) **Entropy:** The entropy of a 2D distribution is defined as:

$$E(P) = - \int \int P(\vec{G}) \log(P(\vec{G})) dG^x dG^y. \quad (21)$$

Since the entropy is entirely determined by the distribution, imposing a GDP implies imposing an entropy prior (rather than

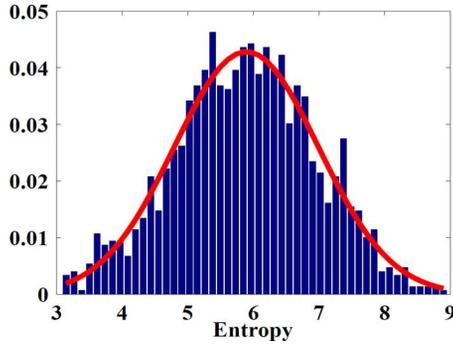


Fig. 9. Entropy distribution of the natural-scene images from the training set. The average is 5.88. The red line is the Gaussian  $0.043 \exp[-(E - 5.88/1.56)^2]$ .

simply minimizing or maximizing entropy). The entropy distribution of the natural-scene images from the training dataset is shown in Fig. 9. It is normally distributed with mean 5.88. This suggests that neither a maximum-entropy nor a minimum-entropy estimator is appropriate, but entropy should follow a prior distribution similar to the one shown in Fig. 9.

## VI. IMPOSING THE GDP IN VARIATIONAL PROBLEMS

In a variational framework, there are two ways of imposing a prior: as a *hard constraint* and as a *soft constraint*. Both are possible for the GDP. For a hard constraint, the GDP is imposed by gradient remapping. The mapped gradient field is then used to reconstruct the output image by solving a Poisson equation with Dirichlet boundary conditions. For a soft constraint, the GDP can be imposed as a regularization term, leading to a D.C. minimization problem. As shown in Section VII, the decision between using a soft or hard constraint depends on the specific application.

### A. As a Hard Constraint

We impose the GDP as a hard constraint by gradient-field remapping. The idea is to map the original gradient field, using a nonlinear mapping function, into a new gradient field that exactly satisfies the GDP. From this remapped gradient field, the output image is reconstructed by solving a Poisson equation. In the special case of a linear mapping function, the reconstruction simplifies to rescaling the image pixel values.

In order to guarantee integrability of the result and a well-posed reconstruction problem, we propose the use of parametric mapping functions, determined by exact histogram specification [39], [40].

1) *Gradient Field Remapping*: Let  $\text{Map}$  remap the gradient field to a new field  $\vec{G}_n$  that satisfies the GDP:

$$\vec{G}_n = \text{Map}(\vec{G}), \text{ s.t. } p(\vec{G}_n) = p^{\text{pr}}. \quad (22)$$

This mapping is nonlinear in general, but can be approximated by a linear mapping. While bi-variate mapping is non-trivial in general, using the parametric models introduced here renders it straightforward.

Fig. 10 illustrates the effects of linear vs. nonlinear remapping. The nonlinear remapping uses exact histogram specification [39] in order to make the gradient histogram of the resulting

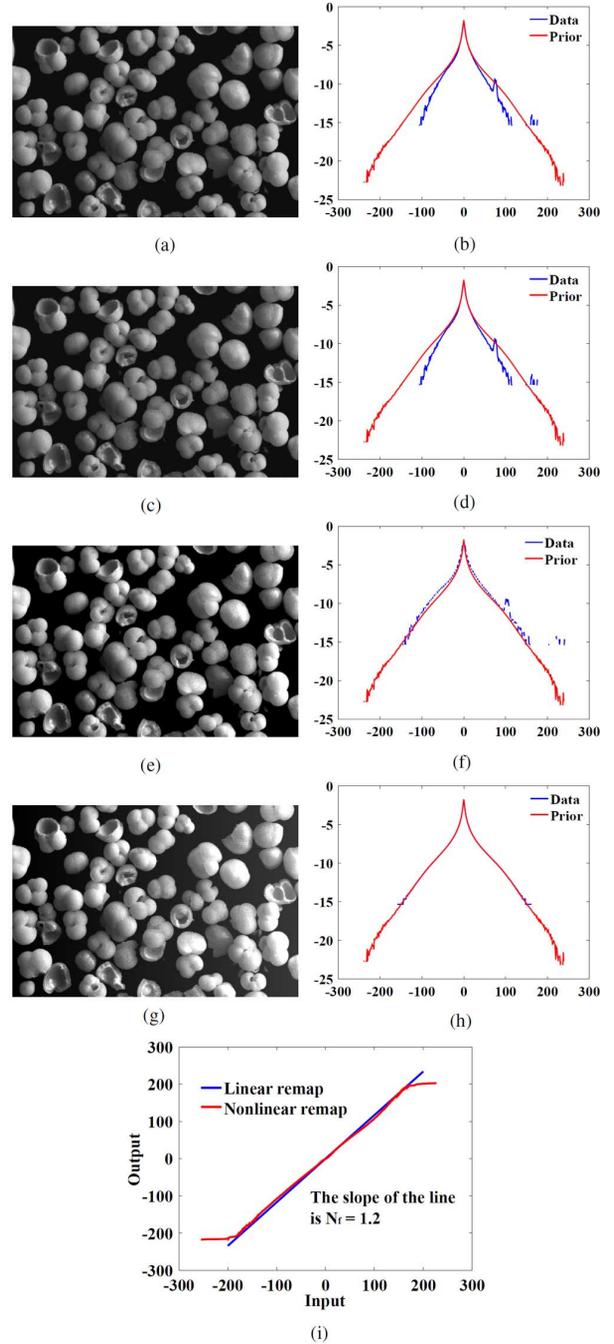


Fig. 10. Comparison of different gradient field remapping methods: original image (converted to grayscale at  $N_f = 1.2$ ) and its gradient distribution (a,b), image reconstructed from the original gradient field without any remapping (c,d), with linear remapping (e,f), and with nonlinear remapping (g,h). The absolute RMS errors of the reconstructions are 0.0, 1.9, 20, and 38, respectively, with respect to the original image. The corresponding gradient distributions after remapping are shown to the right of the images. The linear and nonlinear remapping functions used are shown in (i), as determined by exact histogram specification [39]. Lines are distinguished by color, as reproduced in the electronic version.

image match the GDP exactly (Fig. 10(h)). The linear approximation leads to an approximate fit (Fig. 10(f))<sup>8</sup>.

<sup>8</sup>Figs. 10(b), 10(d), 10(f), 10(h) shows plots of the average of the two marginals for better visualization. The actual remapping is done on the full 2D joint distribution.

Linear remapping amounts to a simple rescaling of the pixel intensities, such that the gradient distribution fits the GDP in average. This is reminiscent of histogram equalization, with the important difference that the scaling parameter is not user-adjustable, but automatically determined such that the GDP is best fit. The output image is reconstructed from the remapped gradient field, as described next.

2) *Image Reconstruction*: Reconstructing the output image from the remapped gradient field amounts to minimizing the following  $q$ -Dirichlet energy:

$$\arg \min_{I_n} \left\{ \|\nabla I_n - \vec{G}_n\|_q \right\} \\ \text{s.t. } I_n \in \text{Lip}(\Omega), \quad (23)$$

where  $q > 1$ ,  $\|\cdot\|_q$  is the standard  $\ell_q$ -norm, and  $\text{Lip}(\Omega)$  is the space of Lipschitz-continuous functions on domain  $\Omega$ .

Existence and uniqueness of the solution of (23) with Dirichlet boundary conditions have been proven [41]. We use Dirichlet boundary conditions by solving only on the internal pixels of the image, fixing the boundary pixels as a boundary condition. Commonly used norms are  $\ell_2$  ( $q = 2$ ) and  $\ell_1$  ( $q = 1$ ), which correspond to reducing measurement errors (unspecific) and gross errors (outliers), respectively.

Taking the  $\ell_2$  norm in (23), we recover the output image  $I_n$  from the remapped gradient field  $\vec{G}_n$  by solving the Poisson equation

$$\Delta I_n = \nabla \cdot \vec{G}_n \quad (24)$$

with Dirichlet boundary conditions given by the boundary pixels of the original image. Using different values of  $q$  leads to fractional Poisson equations with Caputo-type fractional derivatives.

The standard  $\ell_2$  Poisson equation can be solved efficiently, e.g., by FFT-based algorithms or wavelet [42] solvers. A short summary of available Poisson solvers is given in Suppl. Table 1.

Reconstructing an image from its gradient field is accurate. An example is shown in Fig. 10(c). The original image (a) is an 8-bit grayscale image from the 2015 Nikon Photomicrography Competition (C. Maolagain, Wellington, New Zealand) showing a reflected-light microscopy image of *foraminifera* shells from the sea (40x). The relative RMS error of the reconstruction without remapping (c) is  $10^{-4}$ . Reconstruction using the wavelet Poisson solver [42] in Matlab takes 3.5 seconds on an Apple MacBook Pro (early 2011). When reconstructing after linear or nonlinear remapping to the natural-scene GDP (e-h). The mapping functions as determined by exact histogram specification [39] are shown in Fig. 10(i).

### B. As a Soft Constraint

Imposing the GDP as a soft constraint is done by using it as a regularization term. For a variational energy  $\mathcal{E}(U)$  this can be done by evolving the Partial Differential Equation (PDE)

$$\frac{\partial U}{\partial t} = -\frac{\partial \mathcal{E}(U)}{\partial U} \quad (25)$$

over pseudo-time  $t$  (i.e., the iterations of the algorithm). Since the energy  $\mathcal{E}$  is non-convex in general, minimization should be performed in a multi-scale space in order to avoid local minima and accelerate the computation. This can, e.g., be done using

multi-scale anisotropic diffusion, similar to the Perona-Malik model [43].

When using our parametric model for the GDP, the minimization problem simplifies. In this case, the variational energy is the difference of two convex functions, and the minimization problem can efficiently be solved using algorithms based on D.C. programming [44], eliminating the need for PDE pseudo-time evolution. For the present GDP model, the following decomposition holds:

$$\mathcal{E}(U) = \mathcal{E}_1(U) - \mathcal{E}_2(U), \quad (26)$$

where

$$\mathcal{E}_1(U) = \int_{\vec{x} \in \Omega} \left( \frac{1}{2} |U - I|_*^2 + \frac{\lambda}{2} a_2 |\nabla U|^2 \right) d\vec{x} \quad (27)$$

$$\mathcal{E}_2(U) = -\frac{\lambda}{2} \int_{\vec{x} \in \Omega} \log(b_2 + |\nabla U|^2) d\vec{x} \quad (28)$$

are differentiable convex functions,  $|\cdot|_*$  is the norm used in the data-fitting term, and  $a_2, b_2$  are the GDP model parameters.

In the rest of this paper, we use  $\mathcal{R}$  denote the regularization term derived from our model when using the natural-scene GDP as a soft constraint:

$$\mathcal{R}(U) = \int_{\vec{x} \in \Omega} (a_{\text{pr}} |\nabla U|^2 + \log(b_{\text{pr}} + |\nabla U|^2)) d\vec{x}, \quad (29)$$

where  $a_{\text{pr}} = 6.21 \cdot 10^{-5}$  and  $b_{\text{pr}} = 2.39 \cdot 10^{-2}$  are the parameters  $a_2$  and  $b_2$  of the natural-scene GDP prior (see Table V). The parameter  $c_2$  is an additive constant and plays no role in the regularization. The integration over the image domain is valid under the assumption that gradients at neighboring pixels are statistically independent, which is the assumption underlying our model. The functional form  $\Phi_2 = a|\nabla U|^2 + \log(b + |\nabla U|^2)$  is a generalization of many well-known priors, including the TV prior for  $\Phi_2 = |\nabla U|$ .

For the above GDP regularizer, we have the functional derivative (Euler-Lagrange equation):

$$-\frac{\partial \mathcal{R}(U)}{\partial U} = \left( \frac{d}{d\vec{x}} \frac{\partial}{\partial \nabla U} - \frac{\partial}{\partial U} \right) (a_{\text{pr}} |\nabla U|^2 + \log(b_{\text{pr}} + |\nabla U|^2)) \\ = \left( a_{\text{pr}} + \frac{b_{\text{pr}} - |\nabla U|^2}{(b_{\text{pr}} + |\nabla U|^2)^2} \right) \Delta U. \quad (30)$$

The resulting model can efficiently be solved using D.C. programming [44]. In this case, one needs to choose a convex Bregman function  $\phi$ , the choice of which however does not matter much to the algorithm performance [44]. Then, (26) can be minimized using Algorithm 1. The convergence proof can be found in Refs. [44], [45]. In the special case when  $\phi$  is chosen to be a quadratic function (corresponding to an  $\ell_2$  data-fitting energy), Algorithm 1 reduces to the standard proximal point algorithm.

---

#### Algorithm 1 Minimization using D.C. programming

---

**Require:**  $\mathcal{E}_1, \mathcal{E}_2, \phi$ , step size  $\delta t > 0, \varepsilon > 0$

- 1: **while**  $\|(U_i - U_{i-1})\|_\infty > \varepsilon$  **do**
  - 2:  $U_{i+1} = (\nabla \phi + \delta t \nabla \mathcal{E}_1)^{-1} (\nabla \phi(U_i) + \delta t \nabla \mathcal{E}_2(U_i))$
  - 3: **end while**
-

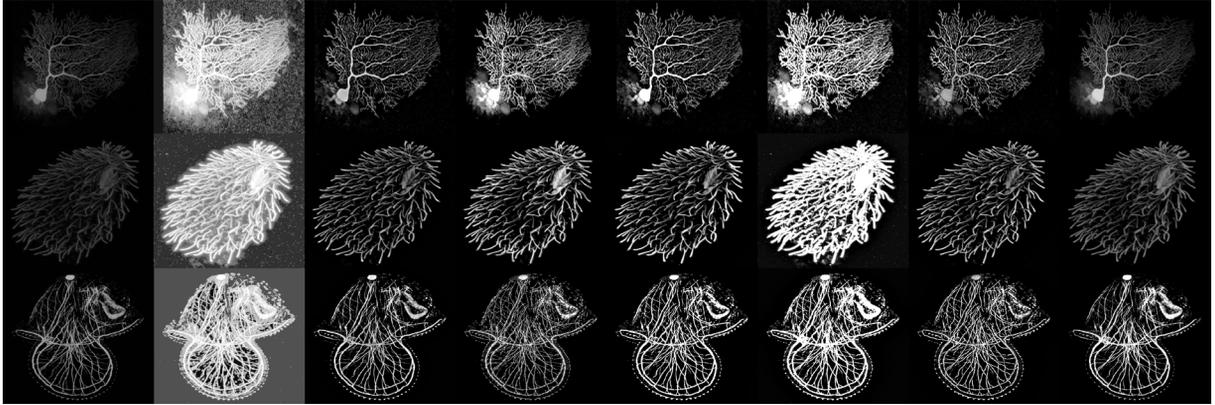


Fig. 11. Comparison of image naturalization with other image-enhancement methods for fluorescence microscopy. From left to right: original, histogram equalized, [48], [23], [49], [50], [24], and naturalized. Naturalization preserves the original signal while increasing the contrast without artifacts. The naturalness factors  $N_f$  for the three input images are 1.5, 1.8, and 2. Image sources from top to bottom: (Maryann Martone, CCDB), (Lee & Matus, U Hawaii, amicros.org), (Gaertig lab, U Georgia, bmc.uga.edu).

### C. Implementation Details

For hard GDP constraints, we implement nonlinear remapping with a map determined by exact histogram specification [39] based on our parametric 2D GDP model. For image reconstruction, we implement FFT-based and wavelet-based [42] Poisson solvers.

For soft GDP constraints, we implement the general diffusion solver, which is valid for all priors, and the D.C. solver in Algorithm 1 for our parametric GDP model.

The source code is available from our website as Matlab code, C++ code included in the OpenCV library, and Java code as an ImageJ/Fiji plugin. All applications shown below were done using the C++ implementation.

## VII. EXAMPLE APPLICATIONS

We exemplify the use of the present GDP, the parametric models, and the solvers in a variety of image-processing tasks, ranging from contrast enhancement, to noise-level estimation, denoising, deconvolution, and dehazing. We do not compete with specialist methods for every single case, like state-of-the-art denoising methods that were especially designed for biomedical images [46], [47]. However, we achieve comparable results across all applications using one and the same simple method. This indicates that the natural-scene GDP might be a good standard choice when better-founded prior knowledge is unavailable.

### A. Image Naturalization

Image naturalization is an image enhancement method that we have proposed earlier [31]. It solves the variational model:

$$\mathcal{E}(U) = \int_{\vec{x} \in \Omega} |U - I|_2^2 d\vec{x} \text{ such that } p(\nabla U) = p^{\text{PF}}. \quad (31)$$

Imposing the hard GDP constraint is done, as detailed in Section VI-A, by remapping the gradient field of the input image to match the GDP, and then reconstructing the output image. This output image looks more “natural”.

Since the natural-scene GDP correlates with image quality, this makes the image look more appealing. We hence propose

to use image naturalization when displaying light-microscopy images to a human observer.

Some examples of fluorescence microscopy images are shown in Fig. 11 along with their naturalized versions (last column) and results from six other image-enhancement methods. In all cases, the naturalized image looks more appealing than the results from the other image-enhancement methods. This extends to microscopy images our previous observation that image naturalization provides better results than many other known image-enhancement methods [31].

When using a linear remapping function, naturalization can also be used to automatically determine a good scaling parameter for standard histogram equalization. Naturalization then amounts to a simple rescaling of pixel intensities, albeit with a scaling factor that renders the naturalness factor of the result as close to 1 as possible.

### B. Noise-Level Estimation

Denoising methods often rely on having an estimate of the noise level in order to adjust algorithm parameters. In practical applications, however, the true noise level is unknown. We show how the GDP can be used to estimate the noise level of an image. As shown in Suppl. Fig. 2, the CDF model is sensitive to noise. This can be exploited to estimate the noise level by relating the fitted parameter  $T$  of the simplified 1D CDF (see Section V-C) of a given image to noise level through a calibration curve. We construct such a calibration curve ( $T$  vs. true noise level) by randomly choosing images from our natural-scene training dataset and adding to them Gaussian noise of varying  $\sigma = [0.02 : 0.02 : 0.8]$ . The dependence of  $T$  on  $\sigma$  shows a distinct characteristic, which is almost independent of image content (Fig. 12(a)). We fit this dependence using the mixture of exponentials:

$$\tilde{\sigma} = \sum_{i=0}^{i=N} q_i \exp\{s_i T\}, \quad (32)$$

where  $N > 0$ ,  $q_i > 0$ , and  $s_i < 0$  are parameters to be determined. For our dataset, we find the best fit  $N = 2$ ,  $q_{\{1,2\}} = \{772.6, 0.9538\}$ ,  $s_{\{1,2\}} = \{-5321, -931.2\}$ . The goodness of

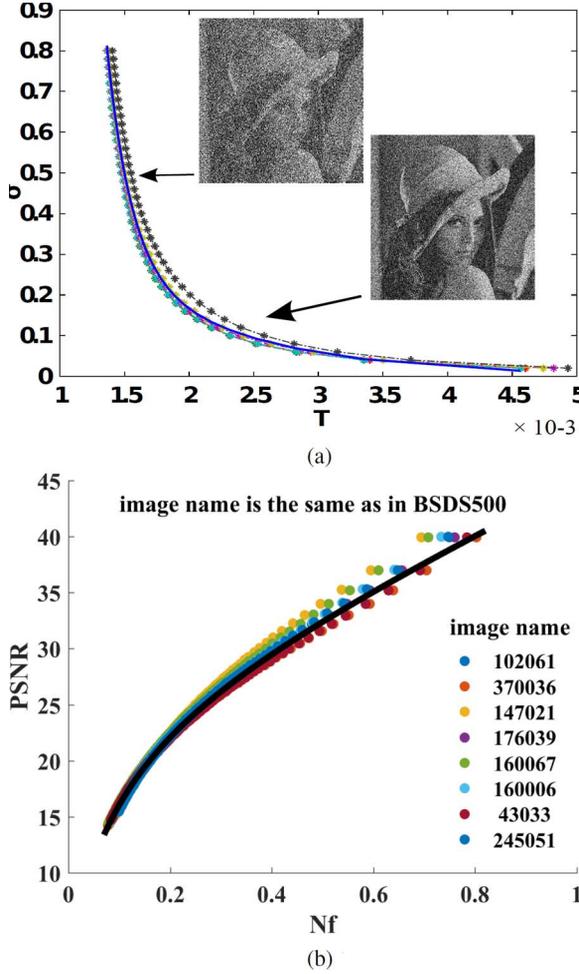


Fig. 12. (a) We build the noise-level estimation model (solid blue line) from seven natural-scene images (individual lines with symbols), including *Lena*, for which two insets are shown to illustrate noise appearance. (b) We test this model on ten images from the BSDS500 benchmark [51]. The black line is the prediction from our model, while the colored dots are the true noise levels for the images named in the inset legend. (a) build model (b) test model. Lines are distinguished by color, as reproduced in the electronic version.

fit is:  $SSE = 0.2741$ ,  $RMSE = 0.034$ ,  $R^2 = 0.979$ . This model is shown by the solid blue line in Fig. 12(a).

We test the model by adding Gaussian noise of different, known  $\sigma$  to a disjoint randomly selected set of 10 test images from the standard BSDS500 benchmark [51]. The differences between the noise levels  $\tilde{\sigma}$  estimated by our model (solid black line) and the ground truth (colored dots) are shown in Fig. 12(b). 87% of the predictions have an accuracy  $|\tilde{\sigma} - \sigma| < 0.04$ . We find similar results also for other random image sets tested.

This suggests that the parameter  $T$  can provide accurate and robust noise-level estimation. A particularly favorable property of this estimator is its high sensitivity to changes in  $\sigma$  for low noise levels ( $\sigma < 0.2$ ). Correctly estimating low noise levels is particularly hard for traditional, pixel-based estimators.

### C. Denoising

The high sensitivity of the cumulative gradient distribution to small amounts of noise makes the GDP a good prior for image denoising. Small non-zero gradients play a key role to recovering image detail. Traditional denoising methods with spatial

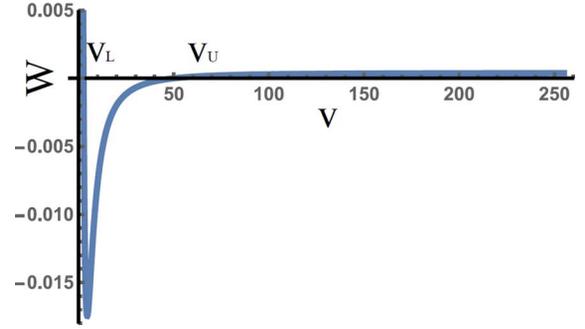


Fig. 13. Behavior of the diffusion coefficient  $W$  vs. the square gradient magnitude  $v = |\nabla U|^2$  for the  $a_{pr}$  and  $b_{pr}$  values of the natural-scene GDP.

regularization (such as TV and its variants, GC, etc.) remove both noise and small signal gradients. In contrast, the spectral GDP can distinguish between noise and small signal gradients based on their frequency of occurrence.

This is compatible with many researchers' observation that split-Bregman solvers for TV- $\ell_1$  models achieve better results in the sense of PSNR [1], [28]. This is because the auxiliary variable introduced in Bregman splitting changes the model to allow for small gradients. These small gradients improve the result. Another example is non-local TV, using spatially repeated patterns to allow for small signal gradients [52]. A third example is stochastic (Monte Carlo) denoising [53]. While all of these methods allow for small image gradients, distinguishing them from noise is mostly *ad hoc*.

Here, the GDP can provide additional information. This has recently been demonstrated in a MAP Bayesian framework [21], which also has the capability of recovering image details. Using our novel parametric GDP model as a soft constraint, we have:

$$\mathcal{E}(U) = \int_{\vec{x} \in \Omega} |U - I|_2^2 d\vec{x} + \lambda \mathcal{R}(U). \quad (33)$$

This denoising model is differentiable with respect to  $U$  and can be efficiently solved by gradient descent (Algorithm 2). From (30), this model can be interpreted as a combination of anisotropic and inverse diffusion.

---

#### Algorithm 2 Denoising with GDP

---

**Require**  $I$ ,  $\lambda$ , step size  $\delta t$ ,  $a_{pr}$ ,  $b_{pr}$ ,  $\varepsilon$

- 1:  $U_0 = I$
  - 2: **while**  $\|(U_i - U_{i-1})\|_\infty > \varepsilon$  **do**
  - 3:  $W = a_{pr} + (b_{pr} - |\nabla U_i|^2) / (b_{pr} + |\nabla U_i|^2)^2$
  - 4:  $U_{i+1} = U_i / (1 + \delta t) + \delta t / (1 + \delta t) I + \delta t / (1 + \delta t) \lambda W \Delta U_i$
  - 5:  $i = i + 1$
  - 6: **end while**
- 

The regularization term is a hybrid of diffusion and inverse diffusion, which is different from traditional approaches that only depend on one of them. For example, the traditional anisotropic diffusion  $(1 + |\nabla U|_2^2)^{-1}$  only tries to smooth the image, while inverse diffusion only enhances the image [54]. The behavior of the diffusion coefficient  $W$  in Algorithm 2 is illustrated in Fig. 13.

TABLE VII

QUALITY COMPARISON OF DENOISING RESULTS FROM DIFFERENT MODELS. USING THE GDP DIRECTLY LEARNED FROM MICROSCOPY IMAGES YIELDS THE WORST RESULT. THE PRESENT METHOD USING THE NATURAL-SCENE GDP YIELDS THE BEST RESULT.

	input	Perona-Malik	TV	Present	microscopy GDP
PSNR	17.12	26.13	26.39	26.39	18.2
SSIM	0.21	0.69	0.64	0.72	0.24

*Lemma VII.1:* In Algorithm 2, let  $v = |\nabla U|^2$ , if  $a_{\text{pr}} b_{\text{pr}} \geq 1/8$ , then  $W \geq 0$ . If  $a_{\text{pr}} b_{\text{pr}} < 1/8$ , then there are two fixed points  $v_L$  and  $v_U$  ( $v_L < v_U$ ) such that

$$\begin{cases} W \leq 0 & \text{if } v_L \leq v \leq v_U \\ W > 0 & \text{else.} \end{cases} \quad (34)$$

The proof is given in the Supplement. For the values of  $a_{\text{pr}} = 6.21 \cdot 10^{-5}$  and  $b_{\text{pr}} = 2.39 \cdot 10^{-2}$  (see Table V),  $W$  has two fixed points. This means that  $U$  gets enhanced (inverse diffusion,  $W < 0$ ) or smoothed (diffusion,  $W > 0$ ) depending on the local gradient magnitude  $\sqrt{v}$ . Even though this behavior is similar to forward-backward diffusion [55], the difference is that our model is derived from a distribution prior, rather than from the gradient itself. As a result, the parameters  $a_{\text{pr}}$  and  $b_{\text{pr}}$  are learned from datasets and not manually adjusted, as in forward-backward diffusion [55].

Results from using Algorithm 2 on an example test image are shown in Table VII. The test image was synthetically corrupted by additive Gaussian noise of the indicated input SNR. The present model achieves state-of-the-art PSNR with better image quality, as quantified by the SSIM [38] quality measure. For comparison, we also use the present method with a GDP learned from the 40 high-quality microscopy test images. When using natural-scene GDPs learned from random subsets of 40 training images, the PSNR are between 25.4 and 26.4, the SSIM between 0.60 to 0.72. This confirms that learning the GDP directly from (good) microscopy images is undesirable (see Section I-A).

#### D. Blind Deconvolution

Light-microscopy images are often blurred. In deconvolution problems, the assumption is that the blur can be described by a convolution of the latent ground-truth image  $\tilde{U}$  with a blur kernel  $K$ . The task of *deconvolution* is to estimate  $\tilde{U}$  from the observed blurred image  $I$ . If the blur kernel  $K$  is unknown and to be estimated along, the problem is referred to as *blind deconvolution*. This is an ill-posed inverse problem.

$K$  and  $U$  can be estimated either in the spatial and/or the gradient domain. We provide here an algorithm for blind deconvolution using the natural-scene GDP. The algorithm is inspired by the observation that pixel auto-correlation is significantly reduced in the gradient domain [31], which is a favorable property for blur-kernel estimation. The latent image, however, is better estimated in the spatial domain, where the auto-correlation signal can be exploited. Different combinations of spatial/gradient-domain deconvolution have already been previously presented (see Table VIII). The present algorithm, however, is the first one to combine gradient-domain kernel estimation with spatial-domain image estimation.

Besides the working domain, the regularizer used is of key importance. Mostly, sparsity of the kernel and minimum-TV of

TABLE VIII

SUMMARY OF BLIND DECONVOLUTION ALGORITHMS.

	Kernel $K$	Image $U$	Typical Method
domain	spatial	spatial	[56]
	spatial	gradient	[57]
	gradient	gradient	[19]
	gradient	spatial	present

the latent image are imposed in deconvolution [58]–[61]. However, a GDP on the latent image can provide a better choice, removing less image detail than TV [17]–[20], [57].

Here, we use the present parametric natural-scene GDP model as a soft constraint for the latent image, but impose no prior on the kernel other than a hard non-negativity constraint. This renders our method generic to a wide variety of different blur kernels that do not have to be previously known.

We use alternating minimization to estimate the kernel  $K$  and the latent image  $U$  by minimizing:

$$\begin{aligned} \mathcal{E}_k(K) &= \|\nabla U_i * K - \nabla I\|_2^2 \\ \text{s.t. } \|K\|_2 &= 1, K \geq 0, \end{aligned} \quad (35)$$

$$\mathcal{E}_u(U) = \frac{1}{2} \|U * K_{i+1} - I\|_2^2 + \frac{\lambda}{2} \mathcal{R}(U), \quad (36)$$

where  $i$  is the iteration number of the alternating minimization scheme and  $*$  is the convolution operator. Equation (35) is a convex function with convex constraints, guaranteeing a globally optimal solution. In Algorithm 3 we hence solve this part of the problem analytically by projection. Equation (36) is not convex, but can be solved by the diffusion process in Algorithm 2. Algorithm 3 summarizes the resulting overall blind deconvolution process, which is performed in a multi-scale fashion to avoid local minima and accelerate computation ( $\circ$  denotes point-wise multiplication).

#### Algorithm 3 Blind Deconvolution with GDP

**Require:**  $I, \lambda, a_{\text{pr}}, b_{\text{pr}}, \varepsilon$

- 1:  $\mathbf{F}_I = FFT(\nabla I)$
- 2: **while**  $\|\nabla(U_i - U_{i-1})\|_\infty > \varepsilon$  **do**
- 3:  $\mathbf{F}_U = FFT(\nabla U_i)$
- 4:  $\hat{K}_{i+1} = FFT^{-1} \left( (\overline{\mathbf{F}_U}^T \circ \mathbf{F}_I) / (\overline{\mathbf{F}_U}^T \circ \mathbf{F}_U) \right)$
- 5:  $K_{i+1} \leftarrow \hat{K}_{i+1} > 0, \int_{\vec{x} \in W_K} \hat{K}_{i+1} d\vec{x} = 1$
- 6:  $U_{i+1} \leftarrow$  run Algorithm 2 with input  $U_i$
- 7: **end while**

An example with a complicated blur kernel is shown in Fig. 14 and Table IX. The ground-truth image (Fig. 14(a)) is blurred with a known kernel (Fig. 14(b)). Figs. 14(c), 14(d) show the reconstructed images using two different *non-blind* deconvolution methods with the ground-truth kernel provided to them. Figs. 14(e), 14(f) show the results of two blind deconvolution methods along with the estimated kernels (insets). Fig. 14(g) shows the estimated  $\hat{K}$  at different scales of the multi-scale process used in the present method. As evident from Table IX and visual comparison, the result from the present GDP method is closest to the ground-truth image and achieves high image quality (measured by the quality metrics

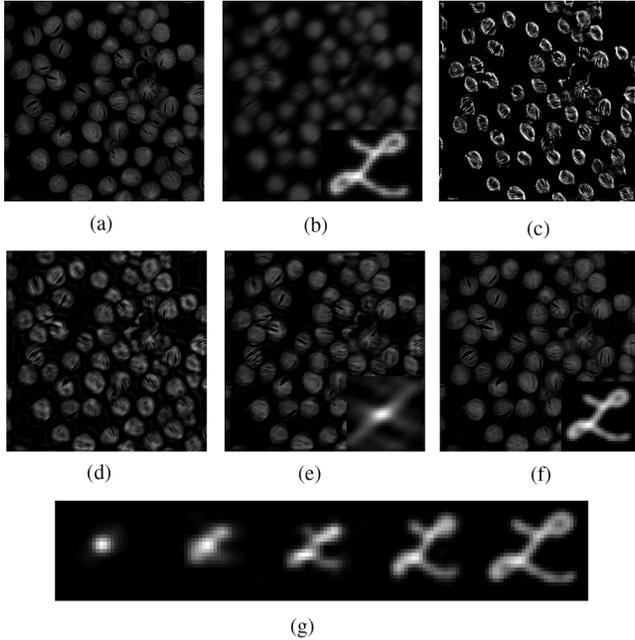


Fig. 14. Deconvolution example. (a) Ground truth image (Fluorescence confocal microscopy image of peace lily pollen, 63x, M. Guervos, University of Oviedo, 2015 Nikon Photomicrography Competition). (b) Input image to the deconvolution methods, obtained by blurring the image in (a) with the kernel shown in the inset. (c,d) Results from two non-blind deconvolution methods with the ground truth blur kernel given; the classical Lucy-Richardson algorithm [62] and the hyper-Laplace method [20]. (e,f) Results from two blind deconvolution methods ([60] and our Algorithm 3) along with the estimated blur kernels (insets). (g) Blur kernel estimated by our method on different levels of the multi-scale process. (a) ground truth image (b) blurred image (c) Lucy-Richardson (d) hyper-Laplace prior (e) normalized sparsity (f) present GDP result (g) estimated  $\hat{K}$  at different scales of our method.

TABLE IX  
QUALITY COMPARISON OF DECONVOLUTION RESULTS.

	PSNR	SSIM
Blurred input image (Fig. 14(b))	24.46	0.49
Lucy-Richardson (Fig. 14(c))	15.93	0.53
Hyper-Laplacian (Fig. 14(d))	24.83	0.76
Normalized sparsity (Fig. 14(e))	23.65	0.55
Present method (Fig. 14(f))	29.46	0.75

PSNR and SSIM [38]) combined with good estimation of the hidden blur kernel.

As shown in the Supplement, the same model and algorithm can also be used for image zooming when fixing the blur kernel.

### E. Scatter Light Removal and Dehazing

Scatter light is a common nuisance in light microscopy when imaging thick samples. The light propagating through the sample is scattered, similarly to how fog and haze scatter light in photographs. The resulting image is the superposition of the scatter light and the latent image. *Dehazing* aims at reducing the scatter light in the observed image. In classical dehazing methods, the observed image  $I$  is modeled as [63]:

$$I(\vec{x}) = U(\vec{x})t(\vec{x}) + A(1 - t(\vec{x})), \quad (37)$$

where  $U$  is the latent image,  $t(\vec{x}) = e^{-\beta d(\vec{x})}$  is the unknown scattering map, and  $A$  is the environment light constant. The unknown scattering coefficient  $\beta$  is a material property of the medium, and  $d(\vec{x})$  is the distance from the object to the camera.

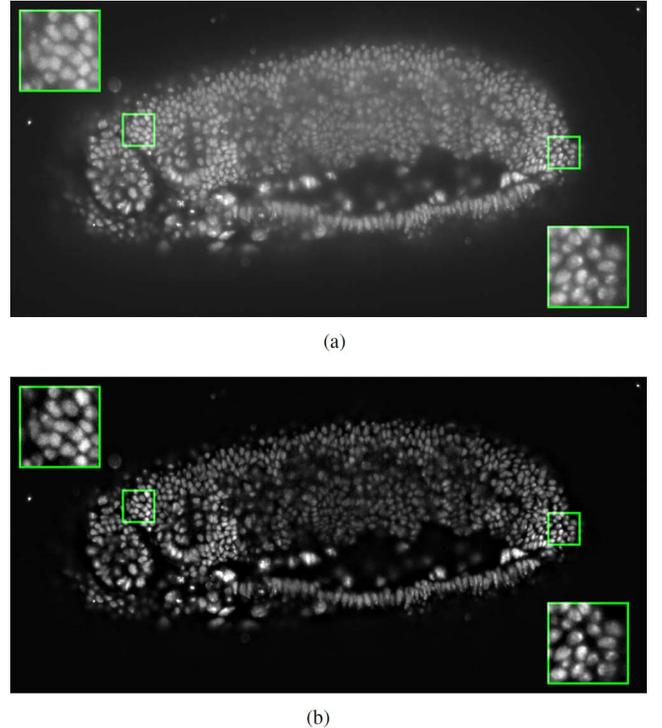


Fig. 15. Scatter light removal in a fluorescence microscopy image of a whole *Drosophila melanogaster* embryo with the nuclei labeled by fluorescence. (a) Original image as recorded by SPIM (single-plane illumination microscopy; source: Tomancak lab, MPI-CBG). Due to the thickness of the sample, there is significant scatter light. (b) Result from the present dehazing method, significantly reducing scatter light.

Solving this model for  $U$  is ill-posed. A popular prior to regularize the problem in the spatial domain is the *dark-channel prior* [63]. Alternatively, the problem can be regularized in a Bayesian framework [64]. Here, we impose the natural-scene GDP for the latent image as a hard constraint, while using TV regularization for the scattering map  $t$ :

$$\mathcal{E}(U, t) = \|U(\vec{x})t(\vec{x}) + A(1 - t(\vec{x})) - I(\vec{x})\|_2^2 + \lambda \int_{\Omega} |\nabla t(\vec{x})| d\vec{x} \\ s.t. p(U(\vec{x})) = p^{PF}. \quad (38)$$

This hard-constrained model, however, is difficult to solve. Moreover, (37) does not hold anymore in this model. Therefore, we relax the model to a soft constraint and solve:

$$\mathcal{E}(U, t) = \frac{1}{2} \|U(\vec{x})t(\vec{x}) + A(1 - t(\vec{x})) - I(\vec{x})\|_2^2 + \frac{\lambda}{2} \left[ \mathcal{R}(U) + \alpha \left\| \frac{\nabla t(\vec{x})}{t(\vec{x})} \right\|_1 \right] \quad (39)$$

with the parameter  $\alpha = 1$  by default, balancing the priors for  $U$  and  $t$ . We use alternating minimization over  $U$  and  $t$  to obtain the final result.

An example is shown in Fig. 15. After dehazing using the present method, the image appears largely free of scatter haze. This facilitates visual inspection, as well as further downstream processing and analysis.

## VIII. CONCLUSION AND DISCUSSION

We proposed learning a gradient distribution prior (GDP) for light-microscopy images from natural-scene images. Both types

of images use the same physical laws to image objects of nature, albeit at different length scales. While light-microscopy images are often corrupted by significant aberrations, such aberrations can be absent in natural-scene images taken under good conditions. The latter may hence provide a good standard GDP in the absence of more specific prior knowledge. We have confirmed that the resulting GDP is stable and correlated with image quality. We have provided a novel parametric model for natural-scene GDPs in 1D and 2D. This model fits the data better than previous models, is sparse, and leads to efficiently solvable quasi-convex optimization problems when used as a regularizer in a variational framework.

We have provided the variational frameworks to impose the model either as a soft or a hard constraint. When imposed as a hard constraint, the natural-scene GDP model amounts to nonlinear gradient-field remapping. The resulting image can be reconstructed by solving a Poisson equation. When a linear remapping function is used, standard histogram rescaling is recovered, albeit with automatically determined parameter. When imposed as a soft constraint, the present GDP model leads to difference-of-convex (D.C.) optimization problems with analytically known energy gradient. They can efficiently be solved using D.C. programming.

Based on a simplified 1D Cumulative Distribution Function (CDF) model of the natural-scene GDP, we defined the naturalness factor of an image in order to quantify how far its gradient histogram is from the one expected for a natural-scene image. This can be used as an image feature or an image-quality indicator, for example when calibrating microscopes.

We have illustrated the use of the new GDP model in various applications, ranging from contrast enhancement to denoising, blind deconvolution, and dehazing. This has illustrated that the GDP is versatile. However, it may not be the optimal choice in every given application, and it was not our goal to beat specialist solutions for individual tasks. We do believe, though, that the natural-scene GDP is a good choice in the absence of better-founded priors.

An important advantage of the presented methods is their low number of parameters. There are no user-adjustable scaling factors, thresholds, or noise levels. The only parameter is the regularization constant  $\lambda$  when using the GDP as a soft constraint. When using it as a hard constraint, such as in image naturalization, the method is entirely parameter-free.

While the present natural-scene GDP can help achieve better results in light-microscopy image processing, the resulting images are biased by the prior to look more like natural-scene images. If quantitative fluorometry or single-molecule quantification are the goal, this prior should not be used, since the intensities in the resulting image are altered, and no longer reflect chemical concentrations of fluorophores.

Moreover, we have only considered 8-bit grayscale images here. Naturally, the GDP for other bit depths looks different, and the model parameters have to be re-estimated. The functional shape of the model, however, remains unchanged. It is important, though, that the discrete operator used to approximate the gradient in the training data is the same that is also used later when using the GDP. This ensures that the unavoidable discretization artifacts cancel out. Unfortunately, a continuous

model seems infeasible, since the gradient magnitude cannot be bounded.

The GDP model presented here is based on a number of assumptions: First, we assumed gradients at neighboring pixels to be independent, in order for the integration over the image domain to be valid. This is a standard assumption. Second, we assumed that the GDP models are un-normalized fitting functions and not proper probability distributions. This makes the parameter  $c$  a true free parameter, giving more freedom to the fit. At the same time, it does not harm, since the normalization becomes irrelevant when going from the Bayesian to the variational formulation. Third, we assumed the gradient components to be identically distributed, but not necessarily independent. This assumption seems reasonable for microscopy images, where we do not expect any preferential gradient orientation. Fourth, our parametric GDP model is rotationally symmetric, assuming that there are no preferred gradient orientations. While this seems reasonable for natural and microscopy images, it may not hold, e.g., for architecture images or modern art.

Certainly, the gradient distribution also varies *within* the set of natural-scene images. Images of trees have a different gradient distribution than images of beaches. Nevertheless, the difference between tree and beach GDPs is less than the difference between natural-scene and non-natural-scene GDPs. It is a matter of scale, which differences one considers important and hence interesting. Our aim was to make light-microscopy images look more natural, in order for the human eye and automatic image analysis methods to be able to better process them, since both the eye and the methods of computer vision have developed for natural-scene images. Our goal was not to claim invariance of gradient distributions across natural-scene images, nor to make microscopy images look more like images of trees, or beaches. This defines the choice of scale at which we attribute importance to differences.

While we have exclusively focused on the image gradient, the same work could also be done for higher-order derivatives, like the Laplace operator. Statistics of higher-order differential operators can provide additional regularization in the same framework [31]. Of special interest could also be the Gaussian curvature (GC) distribution [12], as it directly relates to the geometry of cell membranes through the Willmore membrane energy. As shown in Suppl. Fig. 4(a), these second-order quantities satisfy similar distributions as the gradient. An interesting observation is that the naturalness factor derived from the distribution of the Laplacian is highly correlated with the one derived from the gradient distribution (Suppl. Fig. 4(b)). This suggests that the naturalness factor is a universal image feature that does not depend on the order of the statistic over which it is defined. Confirming this, however, is still outstanding.

Another potential extension would be to exploit the sparsity of the presented GDP model for compressive sensing [65], where it could be interesting that the present GDP model is sparser than the hyper-Laplace prior, while still leading to quasi-convex variational problems. The present work can also be extended to higher-dimensional images. Constructing and using, e.g., a GDP for 3D light-microscopy images is straightforward. The parametric models presented here are sums or products of 1D models and can easily be extended to higher dimensions.

However, it will not be easy to estimate the parameters, because there exist almost no 3D natural-scene images.

Notwithstanding these open questions and limitations, GDPs have repeatedly proven useful and competitive in image processing. The results presented here confirm many of the known favorable properties of spectral regularizers in image processing, and extend the use of GDPs to light-microscopy images. We believe, for the arguments set out here, that efficient parametric models of GDP learned from natural-scene images are a powerful and well-founded tool for light-microscopy image processing. In order to make this available to the community, we provide open-source codes of all algorithms presented here on the MOSAIC Group web page.

## REFERENCES

- [1] G. Paul, J. Cardinale, and I. F. Sbalzarini, "Coupling image restoration and segmentation: A generalized linear model/Bregman perspective," *Int. J. Comput. Vis.*, 2013.
- [2] N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemsse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. O. de Solórzano, J.-C. Olivo-Marin, and E. Meijering, "Objective comparison of particle tracking methods," *Nat. Methods*, vol. 11, no. 3, pp. 281–289, 2014.
- [3] N. Pustelnik, C. Chaux, and J. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.
- [4] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.
- [5] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, no. 2, pp. 503–516, 2000.
- [6] G. Chantas, N. P. Galatsanos, R. Molina, and A. K. Katsaggelos, "Variational Bayesian image restoration with a product of spatially weighted total variation image priors," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 351–362, Feb. 2010.
- [7] A. I. El-Fallah and G. E. Ford, "Mean curvature evolution and surface area scaling in image filtering," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 750–753, May 1997.
- [8] H. Zhu, H. Shu, J. Zhou, X. Bao, and L. Luo, "Bayesian algorithms for PET image reconstruction with mean curvature and Gauss curvature diffusion regularizations," *Comput. Biol. Med.*, vol. 37, no. 6, pp. 793–804, 2007.
- [9] X. Liu, Z. Ying, and S. Qiu, "A fourth-order partial differential equations method of noise removal," in *Proc. 4th Int. Congr. Image and Signal Process. (CISP)*, Oct. 2011, vol. 2, pp. 641–645.
- [10] S.-H. Lee and J. K. Seo, "Noise removal with Gauss curvature-driven diffusion," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 904–909, Jul. 2005.
- [11] B. Lu, H. Wang, and Z. Lin, "High order Gaussian curvature flow for image smoothing," in *Proc. Int. Conf. Multimedia Technol. (ICMT)*, Jul. 2011, pp. 5888–5891.
- [12] Y. Gong and I. F. Sbalzarini, "Local weighted Gaussian curvature for image processing," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 534–538.
- [13] S. Kim, "PDE-based image restoration: A hybrid model and color image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1163–1170, May 2006.
- [14] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.
- [15] S. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 11, pp. 1236–1250, Nov. 1997.
- [16] Y. Weiss and W. T. Freeman, "What makes a good model of natural images?," in *Proc. CVPR*, 2007, pp. 1–8.
- [17] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, p. 73, 2008.
- [18] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, p. 145, 2009.
- [19] X. Chen, J. Yang, and Q. Wu, "Image deblur in gradient domain," *Opt. Eng.*, vol. 49, p. 117003, 2010.
- [20] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," *Adv. Neural Inf. Process. Syst.*, vol. 22, pp. 1–9, 2009.
- [21] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman, "Image restoration by matching gradient distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 683–694, Apr. 2012.
- [22] A. Agrawal and R. Raskar, "Gradient domain manipulation techniques in vision and graphics," *ICCV Short Course*, 2007.
- [23] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 174:1–174:12, Dec. 2011.
- [24] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 139:1–139:10, Nov. 2012.
- [25] R. Fattal, "Edge-avoiding wavelets and their applications," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–10, 2009.
- [26] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1–67:10, Aug. 2008.
- [27] J. McCann and N. S. Pollard, "Real-time gradient-domain painting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 93:1–93:7, Aug. 2008.
- [28] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
- [29] F. Long and D. Purves, "Natural scene statistics as the universal basis of color context effects," *Proc. Nat. Acad. Sci.*, vol. 100, no. 25, pp. 15190–15193, 2003.
- [30] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative Bayesian image super resolution with natural image prior," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4054–4067, Sep. 2012.
- [31] Y. Gong and I. F. Sbalzarini, "Image enhancement by gradient distribution specification," in *Comput. Vision—ACCV Workshops, Revised Select. Papers, Part II, Volume 9009 of Lect. Notes Comput. Sci.*, Singapore, Nov. 2014.
- [32] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, 2002.
- [33] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, pp. 313–318, 2003.
- [34] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless, "Gradientshop: A gradient-domain optimization framework for image and video filtering," *ACM Trans. Graph.*, vol. 29, no. 2, p. 10, 2010.
- [35] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proc. 4th Eurographics Symp. Geometry Process. Eurographics Assoc.*, 2006, pp. 61–70.
- [36] Y. Gong, G. Paul, and I. F. Sbalzarini, "Coupled signed-distance functions for implicit surface reconstruction," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2012, pp. 1000–1003.
- [37] T. Guillemot, A. Almans, and T. Boubekeur, "Non local point set surfaces," in *Proc. 2nd Int. Conf. 3D Imag., Modeling, Process., Visualizat., Transmiss. (3DIMPVT)*, Oct. 2012, pp. 324–331.
- [38] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [39] D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1143–1152, May 2006.
- [40] M. Nikolova, Y.-W. Wen, and R. H. Chan, "Exact histogram specification for digital images using a variational approach," *J. Math. Imag. Vis.*, vol. 46, no. 3, pp. 309–325, 2013.
- [41] L. Boccardo, T. Gallouët, and L. Orsina, "Existence and uniqueness of entropy solutions for nonlinear elliptic equations with measure data," *Ann. de l'Inst. Henri Poincaré-Nonlinear Anal.*, vol. 13, no. 5, pp. 539–552, 1996.
- [42] R.-Q. Jia and S.-T. Liu, "Wavelet bases of Hermite cubic splines on the interval," *Adv. Comput. Math.*, vol. 25, no. 1-3, pp. 23–39, 2006.
- [43] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [44] A. Hamdi, "A modified Bregman proximal scheme to minimize the difference of two convex functions," *Appl. Math. E-Notes, Electron. Only* vol. 6, pp. 132–140, 2006 [Online]. Available: <http://eudml.org/doc/55428>

- [45] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.
- [46] S. P. Awate and R. T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 364–376, Mar. 2006.
- [47] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salameo, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 442–454, Feb. 2010.
- [48] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. ECCV'10*, 2010, pp. 1–14.
- [49] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph. Proc. SIGGRAPH'11*, vol. 30, no. 4, pp. 69:1–69:12, 2011.
- [50] E. S. L. Gastal and M. M. Oliveira, "Adaptive manifolds for real-time high-dimensional filtering," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 33:1–33:13, 2012.
- [51] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Comput. Vis. ICCV*, Vancouver, BC, Canada, 2001, pp. 416–423.
- [52] X. Liu and L. Huang, "A new nonlocal total variation regularization algorithm for image denoising," *Math. Comput. Simulat.*, vol. 97, no. 0, pp. 224–233, 2014.
- [53] A. Wong, A. Mishra, W. Zhang, P. Fieguth, and D. A. Clausi, "Stochastic image denoising based on Markov-chain Monte Carlo sampling," *Signal Process.*, vol. 91, pp. 2112–2120, 2011.
- [54] J. Calder and A.-R. Mansouri, "Anisotropic image sharpening via well-posed Sobolev gradient flows," *SIAM J. Math. Anal.*, vol. 43, no. 4, pp. 1536–1556, 2011.
- [55] G. Gilboa, N. A. Sochen, and Y. Y. Zeevi, "Forward-and-backward diffusion processes for adaptive image enhancement and denoising," *IEEE Trans. Image Process.*, vol. 11, no. 7, pp. 689–703, Jul. 2002.
- [56] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, *Deconvolution Using Natural Image Priors*. Cambridge, MA, USA: Mass. Inst. of Technol., Comput. Sci. and Artif. Intell. Lab., 2007.
- [57] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, pp. 787–794, 2006.
- [58] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.
- [59] A. Marquina, "Nonlinear inverse scale space methods for total variation blind deconvolution," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 64–83, 2009.
- [60] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE CVPR*, 2011, pp. 233–240.
- [61] W. Li, Q. Li, W. Gong, and S. Tang, "Total variation blind deconvolution employing split Bregman iteration," *J. Vis. Commun. Image Represent.*, vol. 23, no. 3, pp. 409–417, 2012.
- [62] D. S. C. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms," *Appl. Opt.*, vol. 36, no. 8, pp. 1766–1775, 1997.
- [63] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[64] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, pp. 263–278, 2012.

[65] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.



**Yuanhao Gong** was a doctoral student with the MO-SAIC Group from January 2011 until March 2015 and earned his doctoral degree from ETH Zurich in computer science. He earned his Bachelor degree in mathematics and applied mathematics from Tsinghua University (Beijing, China) in 2007. From 2007 to 2010, he studied computer science at Xiamen University (Fujian, China). His Master project focused on microscopy image processing, reverse multi-scale spaces, and local symmetry detection. He obtained his Master degree in computer science in June 2010.

Since 2015, he has been a Postdoctoral Researcher in the Computational Brain Imaging group of Prof. Thomas Yeo at the SINAPSE Institute of the National University of Singapore.

He won the Gold Medal in the National Mathematics Olympics in 1999. He won a National Support Scholarship in 2006. In 2012, he received the Best Student Paper Award from the IEEE International Symposium on Biomedical Imaging (ISBI).



**Ivo F. Sbalzarini** received his degree in mechanical engineering, with majors in control theory and applied mathematics, from ETH Zurich (Switzerland) in 2002, which was awarded the Willi Studer Prize for the best graduation of that year. In 2006 he obtained his doctoral degree (Dr. sc. techn.) in Computer Science from ETH Zurich under the supervision of Prof. Petros Koumoutsakos. His dissertation was awarded the Dimitri N. Chorafas Award for 2006. He is the founder and head of the MOSAIC Group.

From 2006 until 2012, he was Assistant Professor of Computational Science at ETH Zurich. He was also invited Professor of Biology at École Normale Supérieure in Paris (France) and Research Group Leader in Bioinformatics at the Mediterranean Institute for Life Sciences (MedILS) in Split (Croatia). Since 2012 he is a tenured Senior Research Group Leader with the Center for Systems Biology Dresden (CSBD) at the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) in Dresden (Germany). Since 2014 he holds the Chair of Scientific Computing for Systems Biology as a Full Professor on the Faculty of Computer Science of TU Dresden.

He is a life-long honorary member of the Technical Society of Zurich, a path co-leader of the Center for Advancing Electronics Dresden, an Associate Editor for BMC Bioinformatics, a member of the Program Committee of the IEEE Symposium on Biomedical Imaging (ISBI) since 2010, and serves on various conference, program, and reviewer boards. His research interests include biological image analysis and image processing, adaptive discretization schemes for PDEs, randomized algorithms, and parallel algorithms.