2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

CONTINUOUS LOCALIZATION USING SPARSITY CONSTRAINTS FOR HIGH-DENSITY SUPER-RESOLUTION MICROSCOPY

Junhong Min[a], Cédric Vonesch[b], Nicolas Olivier[c], Hagai Kirshner[b], Suliana Manley[c], Jong Chul Ye[a], Michael Unser[b]

[a]Bio-Imaging & Signal Processing Lab, KAIST, Republic of Korea
[b]Biomedical Imaging Group, [c] Laboratory of Experimental Biophysics, EPFL, Switzerland
Email:[minimok, jong.ye]@kaist.ac.kr, [cedric.vonesch, n.olivier, hagai.kirshner, suliana.manley, michael.unser]@epfl.ch

ABSTRACT

Super-resolution localization microscopy relies on sparse activation of photo-switchable probes. Such activation, however, introduces limited temporal resolution. High-density imaging overcomes this limitation by allowing several neighboring probes to be activated simultaneously. In this work, we propose an algorithm that incorporates a continuous-domain sparsity prior into the high-density localization problem. We use a Taylor approximation of the PSF, and rely on a fast proximal gradient optimization procedure. Unlike currently available methods that use discrete-domain sparsity priors, our approach does not restrict the estimated locations to a pre-defined sampling grid. Experimental results of simulated and real data demonstrate significant improvement over these methods in terms of accuracy, molecular identification and computational complexity.

Index Terms— Super resolution, High-density imaging, Localization, Proximal gradient

1. INTRODUCTION

Fluorescence microscopy is an invaluable tool in cell biology that has undergone many developments over the past decade. In particular, far-field techniques such as stochastic optical reconstruction microscopy (STORM) [1] and photo-activated localization microscopy (PALM) [2] can achieve nano-scale resolution by means of photo-switchable probes and single-molecule localization algorithms. These methods rely on sparse activation of fluorophore in both the spatial and temporal domains. As a result, the Point-Spread Functions (PSFs) of activated fluorophores do not overlap,they can be localized individually. Common localization methods are based on a least-squares [1, 2] or maximum-likelihood[3] criteria. In the above context, reconstructing a sub-cellular structure typically requires several thousands of frames . This implies acquisition times of an order of minutes, which is a serious limitation for live-cell imaging.

One possible approach for overcoming this limitation is high-density imaging. This means that the acquired data frequently contains overlapping PSFs. There exist few algorithms that can address such a case. For example, the DAOSTORM algorithm[4] fits multiple overlapping spots in an greedy manner, adding more point-sources at each iteration for improving the data fit in regions of overlapping sources. The CSSTORM (compressed sensing STORM) method [5] uses a sparsity-promoting prior for reconstructing the biological structure on a finer sub-pixel grid. The latter approach has been shown to be more adapted to high-density data. Specifically, the CSSTORM method formulates the localization as a convex optimization problem, which is then solved by linear programming. As linear programming is computationally demanding, the CSSSTORM algorithm takes a local approach in which every camera image is divided into several blocks of 7×7 pixels that are processed individually. This non-global approach can potentially degrade the localization accuracy.

In this paper, we present a new localization algorithm for high-density imaging. We combine a global sparsity-promoting variational formulation with a Taylor approximation of the shifted PSF which is valid within a small neighborhood of the finer sub-pixel grid. Instead of being tied to the s ub-pixel grid, this allows us to estimate the spatial locations of the particles on the continuum, demonstrating better accuracy.

2. PROBLEM FORMULATION AND EXISTING APPROACHES

At a given time instant, we can describe the activated fluorophores by a collection of infinitesimal sources:

$$f = \sum_{k=1}^{K} c[\boldsymbol{k}]\, \delta(\cdot - \mathbf{x}_{\boldsymbol{k}}).$$

Here, K is the number of point sources, $\mathbf{x}_{\boldsymbol{k}}$ is the position of the k-th source and $c[\boldsymbol{k}]$ is its brightness in terms of the number of emitted photons. In a space-invariant model, an

This research was supported by the National Research Foundation of Korea(NRF) grant funded in part by the Korean government (MEST) (No.2011-0030933) and the Swiss National Science foundation under grant No.200020-144355

image is obtained by convolving with the PSF h of the microscope and then sampled on a Cartesian grid. Furthermore, the measurements are perturbed by a background signal b and a random noise contribution represented by $\mathcal{N}$:

$$g[\boldsymbol{n}] \sim \mathcal{N}\left(\sum_{k=1}^{K} c[\boldsymbol{k}]\, h(\boldsymbol{n}-\mathbf{x}_{\boldsymbol{k}}) + b\right).$$

The goal is then to retrieve the coefficients $c[\boldsymbol{k}]$ and the locations $\mathbf{x}_{\boldsymbol{k}}$ from the measurements $g[\boldsymbol{n}]$. This can be formulated as a non-linear least-squares problem, where the goal is to minimize the cost functional

$$J(c,\mathbf{x},b) = \sum_{\boldsymbol{n}}\left(g[\boldsymbol{n}] - \sum_{k=1}^{K} c[\boldsymbol{k}]\, h(\boldsymbol{n}-\mathbf{x}_{\boldsymbol{k}}) - b\right)^2. \quad (1)$$

Under the assumption that there is no overlap, the number of point sources K can easily be estimated and each source can be localized independently (e.g., using a minimization procedure that alternates between estimating c and estimating $\mathbf{x}$). However, in high-density imaging, K cannot be estimated accurately due to overlapping PSFs. Thus, one needs to make additional assumptions with respect to the spatial density of activated molecules. An alternative is to reconstruct an image on a finer sub-pixel grid while promoting a sparsity constraint on the brightness coefficients, for example by penalizing their ℓ_1 norm. The cost function is the following form

$$J(c,\mathbf{x},b) = \sum_{\boldsymbol{n}}\left(g(\boldsymbol{n}) - \sum_{\boldsymbol{m}} c[\boldsymbol{m}]\, h(\boldsymbol{n}-\tfrac{\boldsymbol{m}}{M}) - b\right)^2 + \lambda\sum_{\boldsymbol{m}}|c[\boldsymbol{m}]|, \quad (2)$$

where M is an oversampling factor and λ is the regularization parameter. An equivalent formulation (used for example in [5]) is to minimize $\sum_{\boldsymbol{m}}|c[\boldsymbol{m}]|$ subject to

$$\sum_{\boldsymbol{n}}\left(g(\boldsymbol{n}) - \sum_{m} c[\boldsymbol{m}]\, h(\boldsymbol{n}-\tfrac{\boldsymbol{m}}{M}) - b\right)^2 \le \boldsymbol{\beta}^2, \quad (3)$$

where β^2 is noise variance. In the unconstrained form (2), λ is not only related to the noise variance $\boldsymbol{\beta}^2$, but also has a direct interpretation in terms of the minimal brightness that a point source must have in order to be considered active. However, both formulations are still based on the sub-pixel grid. Moreover, we have observed experimentally that imposing sparsity often results in an underestimation of the coefficient value $c[\boldsymbol{k}]$ and their distances among closely spaced particles, which implies a spatial bias of localization.

3. CONTINUOUS-SPACE LOCALIZATION WITH SPARSITY CONSTRAINTS

Our approach combines continuous domain localization with the sparsity paradigm. Going back to (1), we can describe every point source position $\mathbf{x}_{\boldsymbol{k}}$ by the closest grid point $\frac{\boldsymbol{m}_{\boldsymbol{k}}}{M}$:

$$\mathbf{x}_{\boldsymbol{k}} = \frac{\boldsymbol{m}_{\boldsymbol{k}}}{\boldsymbol{M}} + \boldsymbol{\epsilon}_{\boldsymbol{k}}.$$

We then use a first-order Taylor-series approximation

$$h(\boldsymbol{n}-\mathbf{x}_{\boldsymbol{k}}) \simeq h(\boldsymbol{n}-\frac{\boldsymbol{m}_{\boldsymbol{k}}}{\boldsymbol{M}}) - \boldsymbol{\epsilon}_{\boldsymbol{k}}^T\nabla h(\boldsymbol{n}-\frac{\boldsymbol{m}_{\boldsymbol{k}}}{\boldsymbol{M}})$$

and rewrite (1) as

$$J(c,\boldsymbol{\epsilon},b) = \sum_{\boldsymbol{n}}\left(g[\boldsymbol{n}] - \sum_{\boldsymbol{k}}^{K} c[\boldsymbol{k}]\left[h(\boldsymbol{n}-\tfrac{\boldsymbol{m}_{\boldsymbol{k}}}{M}) - \boldsymbol{\epsilon}_{\boldsymbol{k}}^T\nabla h(\boldsymbol{n}-\tfrac{\boldsymbol{m}_{\boldsymbol{k}}}{M})\right] - b\right)^2. \quad (4)$$

Finally, we impose sparsity promoting constraints on (4), by carrying out following three subsequent phases:

$$\begin{aligned}
&\text{Phase1} \quad : \\
&\quad \min_{\boldsymbol{c},b}\left(\|\boldsymbol{g}-\boldsymbol{D}\boldsymbol{H}\boldsymbol{c}-\mathbf{1}b\|^2 + \lambda\textstyle\sum_{\boldsymbol{m}}\boldsymbol{c}[\boldsymbol{m}] + \mathcal{I}_{\mathfrak{R}_+}(\boldsymbol{c})\right) \\
&\text{Phase3} \quad : \\
&\quad \min_{\boldsymbol{c},b}\left(\|\boldsymbol{g}-\boldsymbol{D}\boldsymbol{H}\boldsymbol{c}-\mathbf{1}b\|^2 + \mathcal{I}_S(\boldsymbol{c})\right) \\
&\text{Phase3} \quad : \\
&\quad \min_{\boldsymbol{\epsilon}}\left(\|(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{H}\boldsymbol{c}-\mathbf{1}b) - \boldsymbol{D}\nabla\boldsymbol{H}\boldsymbol{C}\boldsymbol{\epsilon}\|^2 + \mathcal{I}_{\tilde{\epsilon}}(\boldsymbol{\epsilon})\right)
\end{aligned} \quad (5)$$

where $\mathbf{1}$ is constant vector, $\boldsymbol{D}$ is a downsampling matrix, $\boldsymbol{H}$ and $\nabla\boldsymbol{H}$ are convolution matrices corresponding to h and ∇h respectively, S is indices of non-zero $\boldsymbol{c}[\boldsymbol{m}]$ from phase 1 and $\boldsymbol{C}$ = diag($\boldsymbol{c}$). The indicator functions $\mathcal{I}_{\mathfrak{R}_+}(\boldsymbol{c})$, $\mathcal{I}_S(\boldsymbol{c})$ and $\mathcal{I}_{\tilde{\epsilon}}(\boldsymbol{\epsilon})$ are as follows:

$$\begin{aligned}
\mathcal{I}_{\mathfrak{R}_+}(\boldsymbol{c}) &= \begin{cases} 0 & \text{if } \forall m,\ \boldsymbol{c}[m] \ge 0; \\ \infty & \text{otherwise;} \end{cases} \\
\mathcal{I}_S(\boldsymbol{c}) &= \begin{cases} 0 & \text{if } \forall \boldsymbol{c}[m] > 0,\ \ m \in S \\ \infty & \text{otherwise;} \end{cases} \\
\mathcal{I}_{\tilde{\epsilon}}(\boldsymbol{\epsilon}) &= \begin{cases} 0 & \text{if } \forall m,\ |\boldsymbol{\epsilon}[m]| \le \tilde{\epsilon}; \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Conceptually, we first estimate the locations and the brightness coefficients on a fine uniform grid using the sparsity constraints of Phase 1. Then, we refine these quantities while restricting their support to the set S in Phase 2. Finally we introduce the displacements ϵ in Phase 3, hence allowing for continuous domain location. We found it necessary to introduce Phase 2 and Phase 3 to compensate for the estimation bias introduced by the sparsity constraint in Phase 1. Thus, the proposed problem formulation (5) is more robust than (2) with respect to localization accuracy.

In general, (5) is a large-scale problem involving millions of variables. Instead of linear programming, we use a proximal gradient method where cost function is minimized iteratively via proximal mapping such as soft-thresholding or projection operations. We optimize (5) as follows.

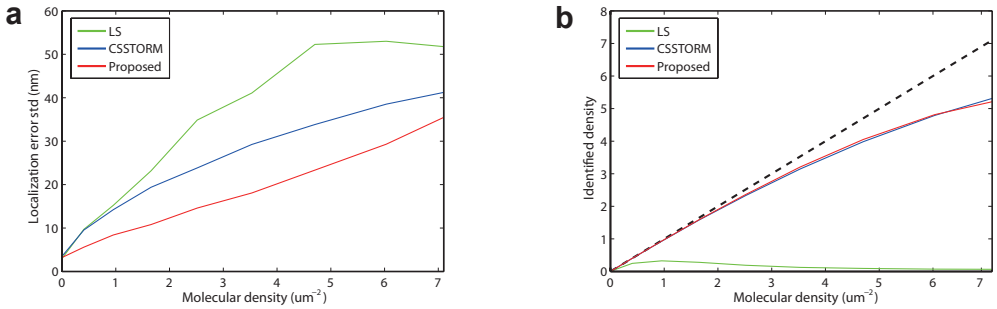

Fig. 1. Monte Carlo simulation. (a) Localization error. (b) Identified molecular density.

Simplified Algorithm 1

1. Initialization : Initialize variables.

2. Phase 1: Repeat until stop criterion

$$\begin{aligned} \boldsymbol{c}_{k+1} &= \mathcal{T}_{\lambda/L_c}(\boldsymbol{c}_k - \tfrac{2}{L_c}(\boldsymbol{DH})^T(\boldsymbol{DHc}_k + \mathbf{1}b_k - \boldsymbol{g})) \\ b_{k+1} &= b_k - \tfrac{2}{L_c}\mathbf{1}^T(\boldsymbol{DHc}_k + b_k - \boldsymbol{g})) \end{aligned}$$

3. Phase 2: Repeat until stop criterion.

$$\begin{aligned} \boldsymbol{c}_{k+1} &= \mathcal{P}_S(c_k - \tfrac{2}{L_c}(\boldsymbol{DH})^T(\boldsymbol{DHc}_k + \mathbf{1}b_k - \boldsymbol{g})) \\ b_{k+1} &= b_k - \tfrac{2}{L_c}\mathbf{1}^T(\boldsymbol{DHc}_k + b_k - \boldsymbol{g})) \end{aligned}$$

4. Phase 3: Repeat until stop criterion.

$$\begin{aligned} \boldsymbol{C\epsilon}_{k+1} &= \mathcal{P}_{C\tilde{\epsilon}}(\boldsymbol{C}\epsilon_k - \tfrac{2}{L_\epsilon}\nabla\boldsymbol{H}^T\boldsymbol{D}^T(\boldsymbol{D}\nabla\boldsymbol{HC}\boldsymbol{\epsilon}_k \\ &\quad -(\boldsymbol{DHc} + \mathbf{1}b - \boldsymbol{g})) \end{aligned}$$

Here, $L_c \geq 2\lambda_{max}(A^TA)$ and $L_\epsilon \geq \lambda_{max}(\nabla A^T \nabla A)$ denote Lipschitz constants where $A = [H \;\; \mathbf{1}]$ and $\nabla A = [\nabla H \;\; \mathbf{1}]$. $\mathcal{P}_S(x)$ and $\mathcal{P}_{C\tilde{\epsilon}}(x)$ are projection operators on the sets defined by the characteristic functions $\mathcal{I}_S(c)$ and $\mathcal{I}_{\tilde{\epsilon}}$ respectively. The soft-thresholding operator $\mathcal{T}_\alpha(x)$ is defined as

$$\mathcal{T}_\alpha(x)_i = \begin{cases} x_i - \alpha & if \;\; x_i > \alpha, \\ 0 & otherwise. \end{cases}$$

The simplified Algorithm 1 has linear convergence rate, which is relatively slow. Thus, the proposed algorithm uses the faster proximal gradient method called MFISTA [6] which improves the convergence rate by one order. We can easily modify the described form by using two-step updating rules for next estimate (see [6] for details). Due to the limited space, we omit further details.

The proposed algorithm typically returns several non-zero valued elements of the c and corresponding $\boldsymbol{\epsilon}[m]$ near the true position of a fluorophore because of the high coherence of convolution matrix $\boldsymbol{H}$ and $\nabla\boldsymbol{H}$. This effect was also reported in [5]. We follow the post-processing method used in [5] in order to retrieve number of localized particles and their locations. In the method, neighboring non-zero $\boldsymbol{c}[\boldsymbol{m}]$ are counted as a cluster. We then calculates each center of mass with $\boldsymbol{c}[\boldsymbol{m}]$ and $\boldsymbol{\epsilon}[m]$ as the localized molecular position.

4. EXPERIMENTS

We performed experiments both on synthetic data (Monte Carlo simulations) and on real high-density STORM data. In every experiment, the regularization parameter is set to $\lambda = background + \frac{\sqrt{\sum \boldsymbol{g}(\boldsymbol{n})}}{4}$; here $background$ is estimated by averaging 4×4 pixels in boundary region and $\sum \boldsymbol{g}(\boldsymbol{n})$ represents an estimate of the noise variance, assuming that the CCD shot noise is Poisson distributed. We observed that this choice worked well with real data too. The oversampling factor M is equal to 7 and the displacement of ϵ is bounded by half of the finer grid size. The initial guess is obtained by Wiener-type filtering: $\boldsymbol{c}_0 = (\boldsymbol{H}^T\boldsymbol{D}^T\boldsymbol{DH} + \sigma^2 I)\boldsymbol{H}^T\boldsymbol{D}^T\boldsymbol{g}$. After Phase 1 and Phase 2, coefficients $\boldsymbol{c}[\boldsymbol{m}]$ that are below 10% of the maximum of c are discarded as spurious points.

In the simulated data, each molecule corresponds to a Gaussian PSF of 340nm full width half maximum (FWHM) and 70 background photons are added to each CCD pixel of size 160 nm. In addition, we introduced Poisson shot noise and a small Gaussian readout noise with unit variance. The number of emitted photons is 4000 per molecule. We compared our algorithm with two other methods: conventional single-molecule least-square fitting and CSSTORM[5]. The least-squares method works by fitting an elliptical Gaussian PSF to local maxima of the image [2]. CSSTORM divides the image into several blocks of 7×7 pixels, excluding outermost pixel-wide region in each block for the localization. To quantify the error, we find the closest true molecular position to each localized position and then calculate the standard deviation of the localization errors and number of identified true molecules.

For the Monte Carlo simulation, we generated a wide range of molecular densities, from $0.02\mu m^{-2}$ to $7\mu m^{-2}$. We used 100 realizations for every density. In all of our experiments, the proposed algorithm demonstrated better results than the least square fitting method and CSSTORM. The localization of the proposed algorithm is comparable to the least square fitting method in the very sparse density case ($0.02\mu m^{-2}$). For the high-density cases, the least square method is not effective, as expected. We also found that our global approach is more accurate than CSSTROM over the whole range of molecular densities. This improvement mainly originates from our global approach in phase 1. Additional improvement of about 10% is achieved by the

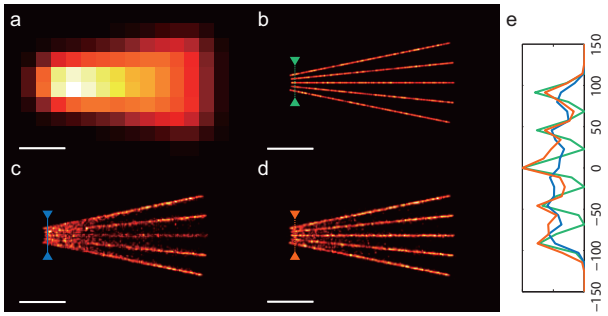

Fig. 2. Synthetic resolution phantom. (a) Diffraction limited image. (b) True phantom image. (c) CSSTORM image. (d) Image obtained with the proposed algorithm. (e) line plot in (b-d). Scale-bars in (a-d) are 500nm.

refinement steps in Phases 2 and 3 thereby reducing the effect of the discrete grid and the spatial bias induced by the sparsity constraint of Phase 1. Therefore, the proposed algorithm can identify over 10 times more molecules than least square fitting and twice more than CSSTORM at the same localization accuracy. In addition, the proposed algorithm is twice as faster as CSSTORM in spite of the additional fine-tuning.

Next, we generated 1000 frames of synthetic data with high activation density, representing 5 lines that are diverging equi-angularly. The minimum distance between two adjacent lines is $40nm$. Super-resolution images obtained with CSSTORM and the proposed algorithm are shown in Figure 2. Only the proposed algorithm clearly resolves the minimum gap between the lines2(e).

Last, we compared the algorithms on 200 real STORM images of microtubules labeled with Alexa 647. The exposure time for each image was 100 ms. In figure3(b-d), CSSTORM and the proposed algorithm both identified significantly more molecules than the least-squares method. However, it appears that the proposed algorithm also outperforms CSSTORM in configurations with crossing-over microtubules, see Fig. 3(e,f).

5. CONCLUSION

We presented a new high-density localization algorithm that is based on a space-domain sparsity constraint without restricting the particle positions to the sub-pixel grid. Our approach yields better localization accuracy and higher molecular identification rates than other currently available methods over a wide range of experimental conditions. Moreover, it has reduced computation times compared to CSSTORM, which is also based on a sparsity-promoting prior. Our experimental results indicate that the proposed approach could reduce the number of images that is required for obtaining super-resolution image reconstruction. This is crucial for studying biological interactions at the nanometer scale.

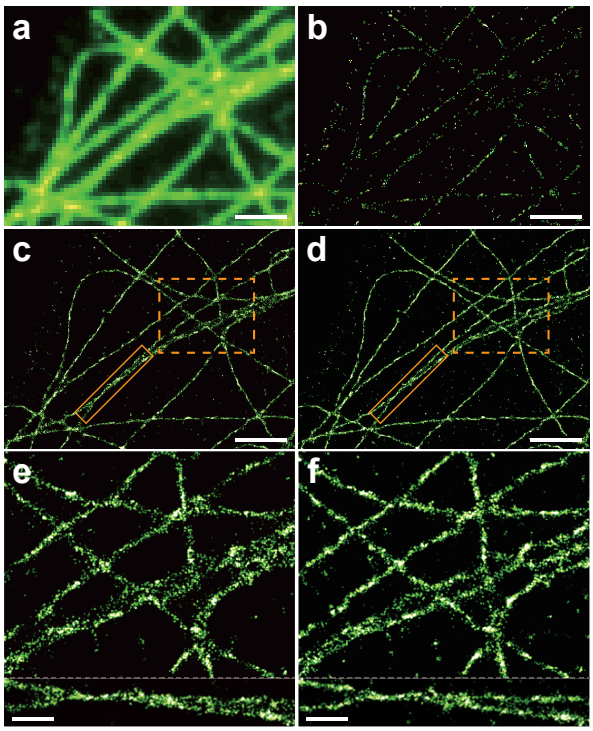

Fig. 3. Microtubules stained with Alexa647. (a) Diffraction limited image. (b) Image obtained by least-squares fitting. (c) CSSTORM image. (d) Image obtained with the proposed algorithm. (e-f) Close-up images of solid and dotted line boxes in (c,d) Scale-bars are $2\mu m$ in (a-d) and 300nm in (e-f).

6. REFERENCES

[1] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, 2006.

[2] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J.S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642, 2006.

[3] Carlas S. Smith., Nikolai Joseph., Bernd Rieger., and Keith A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nature Methods*, vol. 7, no. 5, pp. 373–375, 2010.

[4] Seamus J Holden., Stephan Uphoff., and Achillefs N Kapanidis, "Daostorm: an algorithm for high-density super-resolution microscopy," *Nature Methods*, vol. 8, no. 4, pp. 279–280, 2011.

[5] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster storm using compressed sensing," *Nature Methods*, vol. 9, no. 7, pp. 721–723, 2012.

[6] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2419–2434, 2009.