Diffuse Optical Tomography Enhanced by Clustered Sparsity for Functional Brain Imaging

Chen Chen, *Student Member, IEEE*, Fenghua Tian, Hanli Liu, and Junzhou Huang*, *Member, IEEE*

Abstract—Diffuse optical tomography (DOT) is a noninvasive technique which measures hemodynamic changes in the tissue with near infrared light, which has been increasingly used to study brain functions. Due to the nature of light propagation in the tissue, the reconstruction problem is severely ill-posed. For linearized DOT problems, sparsity regularization has achieved promising results over conventional Tikhonov regularization in recent experimental research. As extensions to standard sparsity, it is widely known that structured sparsity based methods are often superior in terms of reconstruction accuracy, when the data follows some structures. In this paper, we exploit the structured sparsity of diffuse optical images. Based on the functional specialization of the brain, it is observed that the *in vivo* absorption changes caused by a specific brain function would be clustered in certain region(s) and not randomly distributed. Thus, a new algorithm is proposed for this clustered sparsity reconstruction (CSR). Results of numerical simulations and phantom experiments have demonstrated the superiority of the proposed method over the state-of-the-art methods. An example from human *in vivo* measurements further confirmed the advantages of the proposed CSR method.

Index Terms—Clustered sparsity, diffuse optical tomography (DOT), functional brain imaging, structured sparsity.

I. INTRODUCTION

DIFFUSE optical tomography (DOT) is an emerging technique used to study brain functions, which is quickly gaining favorable recognition because of its noninvasive manner and relatively low cost [1], [2]. This technique uses near infrared light in a range of 650–900 nm, which is sensitive to the absorptions of oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (Hb). The light sources and detectors are arranged on the scalp. The diffused light from the cortical layer of the brain is acquired to form an image of activation. Compared with other neuroimaging modalities, such as functional magnetic resonance imaging (fMRI), DOT can provide more comprehensive information of cerebral hemodynamics, while fMRI has better spatial resolution. A recent review of DOT is given in [3].

One of the main challenges in DOT is the image reconstruction (or inverse problem). Due to the diffusive nature of light and limited numbers of sources and detectors, the inverse problem is severely ill-posed. In order to make the problem more tractable, it is necessary to make linearization approximation [4], [5], e.g., Rytov approximation. Regularization is often applied to the linear inverse problem to obtain a unique solution. Conventionally, the ℓ_2-norm regularization (also known as Tikhonov regularization) is the most commonly used method because it can be easily implemented [6]. The drawback is its tendency to over-smooth the image by penalizing large values. Thus, sharp boundaries for the reconstructed images are very difficult to obtain by the ℓ_2-norm regularization.

Since the perturbation from a homogeneous background or reference medium is relatively small in volume and contrast, sparsity of the reconstructed image is generally assumed [7]–[11]. Guided by compressive sensing theory [12], a sparse signal or image can be recovered from fewer measurements than that dictated by Shannon–Nyquist theorem under mild conditions. Sparsity inducing methods have been shown to be repeatedly successful in many real-world applications [13]–[16]. In DOT, ℓ_1 norm is first used in [7], [9], [11] to induce sparsity. A more recent work shows that regularization with ℓ_p ($0 < p < 1$) and smooth-ℓ_0 norms can improve the results of ℓ_1 norm regularization [10], while the inverse problem is more difficult to solve due to the nonconvexity and nonsmoothness of such norms.

While promising results have been obtained in these sparsity-inducing methods over the conventional ℓ_2 norm regularization, accurate reconstruction of diffuse optical images is still challenging. First, experiments often involve noise, while *standard sparsity* based methods are often not robust to noise, e.g., those in our later simulations. True signals are difficult to distinguish from significant noise, as the noise may also satisfy the sparsity assumption. Second, based on compressive sensing theory, the minimal number of measurements for successful recovery is required to be $\mathcal{O}(K + K\log(N/K))$, where K is the number of nonzero components and N is the length of the signal. Limited by the number of measurements in DOT, these standard sparsity based methods may fail when the image is less sparse (i.e., K is larger). Fortunately, such limitations of standard sparsity have been overcome in advanced sparsity techniques called structured sparsity [17], [18]. According to structured sparsity theories [17]–[19], fewer measurements are required for signals with structured sparsity than those with standard sparsity, or the recovery accuracy can be improved with the same number of measurements. Also, structured sparsity based methods are often more robust to noise. However, for different

types of data, discovering the underlying structures of the data and developing an efficient method to solve the corresponding problem is still open to questions. In this study, we aim to improve DOT based on structured sparsity.

The diffuse optical images do have some special structures if we look at the brain features in biology. It is widely known that human actions correspond to certain regions of brain activation. For example, brain state changes specifically in the dorsal medial prefrontal area during Vipassana meditation [20]. These changes of brain state only take place in a region or regions but are not randomly distributed over the whole brain, which will make the change of absorption have a clustered appearance. In contrast to previous works that use no priori information other than sparsity, we propose a new method to improve the reconstruction by exploiting this clustered structure. We call this method clustered sparsity reconstruction (CSR). The clustered sparsity problem is modeled with convex programming and solved by a new algorithm based on the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) framework [21]. Comprehensive experimental results have demonstrated significant improvements achieved by our CSR method when compared with previous works.

II. THEORY

A. Regularization for DOT

The relative change in optical density is measured by each source-detector (S-D) pair. Photon propagation in human tissue is mathematically described by the Boltzmann transport equation [7]. Following previous work [22], we assume the background optical parameters are known in this study. Therefore, the reconstruction problem in DOT can be simplified by the linear Rytov approximation [7], [9]

$$Ax = b \tag{1}$$

where $b \in \mathbb{R}^{m \times 1}$ is the vector of measured relative light density changes; $x \in \mathbb{R}^{n \times 1}$ represents an image (after vectorizing) of $\triangle \mu_a$ (i.e., the change of absorption coefficient); $A \in \mathbb{R}^{m \times n}$ is the forward sensing matrix referring to the sensitivity of pixels with different S-D pairs. Due to the limitation on the number of S-D pairs and the diffusive nature of light, this problem is severely ill-posed, i.e., $m \ll n$. To obtain a unique solution, the ℓ_2-norm regularization is widely used and the objective function becomes

$$\hat{x} = \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_2^2 \right\} \tag{2}$$

where λ is a positive parameter and can be selected by the L-curve method [23]. To overcome the over-smoothing by ℓ_2-norm regularization, many recent methods have been proposed to exploit the sparsity of the reconstructed image [7], [9], [10], [24]. Sparsity-inducing norms, e.g., ℓ_1, ℓ_p $(0 < p < 1)$ and smooth-ℓ_0 norms, are used for regularization instead of ℓ_2 norm

$$\hat{x} = \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_p^p \right\} \tag{3}$$
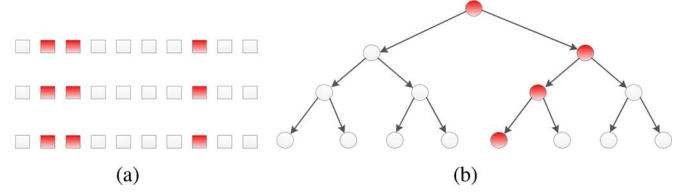

Fig. 1. Examples of structured sparse data. Red denotes nonzeros components and the white denotes zero components. (a) Group sparse data, where each column is a group. (b) Tree sparse data. Ancestors of a nonzero node to the root are all nonzeros.

$$\hat{x} = \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_0 \right\} \tag{4}$$

where the ℓ_0 norm is approximated with the Gaussian function. Promising results have been obtained using these methods. However, these standard sparsity based methods only exploit the sparseness of the reconstructed image, while the correlations or structures of the nonzero values have not been utilized.

B. Structured Sparsity

Compressive sensing theory [12] provides a theoretical guarantee for robust recovery with standard sparsity (e.g., ℓ_1 norm). Under mild conditions, it has been proven that $\mathcal{O}(K + K\log(N/K))$ measurements are required for successful recovery with high probability, where K is the number of nonzero components and N is the total number of components. However, for DOT problems, the number of measurements is limited by physical reasons, e.g., the diffusive nature of light and the size of source-detectors. In some cases where the images are less sparse, i.e., K is relatively large to N, the performance of standard sparsity based methods (e.g., those mentioned above) cannot be guaranteed.

To overcome this limitation, structured sparsity theories have been developed recently [17], [18]. These theories suggest that, if we could exploit more prior information than sparsity, the measurement bound can be reduced [25]. For example, the nonzero components may have a group structure, where the components in the same group are simultaneously zeros or nonzeros [26], [27]. Mixed $\ell_{2,1}$ norm can be used to model group sparsity, which is also feasible for overlapping groups [28]. Another common structure is the tree structure [16], [29]. Fig. 1 shows these two examples. If the group structure and tree structure are exploited, the required measurements for successful recovery can be reduced to $\mathcal{O}(K + Q\log(G/Q))$ and $\mathcal{O}(K + \log(N/K))$, respectively, where $Q \leq K$ is the number of nonzero groups and G denotes the total number of groups. The group structure has already been successfully utilized for the support estimation of absorption and scattering coefficients in tumor imaging [22], [30]. However, the reconstruction is still performed with standard sparsity. To the best of our knowledge, the structured sparsity has rarely been used in DOT.

Although structured sparsity provides advantages over standard sparsity in reconstruction problems, the structures of the data is often much more difficult to be observed than sparseness. In addition, unlike standard sparsity that has been studied for a couple of decades, there is much fewer algorithms for structured sparsity. For some complex structures, developing an efficient
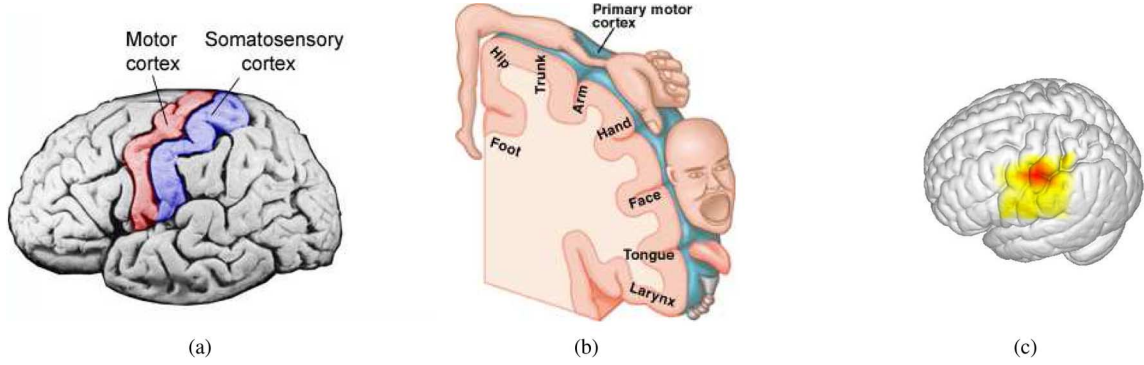
Fig. 2. (a) Motor cortex in human brain. (b) Anatomical cross section of the motor cortex, where different regions correspond to different actions. Images come from Google Images. (c) Reconstructed diffuse optical image of brain activity caused by a finger-tapping task. Yellow color represents the activated brain region, while red color indicates the maximal activation within the region.

algorithm is still challenging. In this article, we aim to study the inherent structures in diffuse optical images and develop an efficient algorithm for fast reconstruction.

III. METHOD

A. Clustered Appearance

The diffuse optical image is often sparse as the change of absorption is relatively small compared to the whole imaging area. This prior information has been utilized in many existing methods [7]–[10]. We further observe that diffuse optical images have a clustered appearance, i.e., the change of absorption is often clustered in certain regions but not randomly distributed. In biology, it is because human actions correspond to certain regions of brain activation, which is widely known. Besides the example discussed above, Fig. 2 shows the motor cortex of human brain and the diffuse optical tomography caused by a finger-tapping task. In this example, the brain activation clusters in the motor cortex area.

By structured sparsity theories [17]–[19], it has been proven that only $\mathcal{O}(K + C\log(N/C))$ measurements are required to recover clustered sparse signals instead of $\mathcal{O}(K + K\log(N/K))$ for standard sparse signals. Here C denotes the number of clusters with $C \ll K$. When the number of measurement is not sufficient for standard sparsity, it is still possible to achieve successful recovery with clustered sparsity. When using the same number of measurements, significant improvement can be gained by clustered sparsity.

In contrast to previous works, one of our contributions is to utilize this prior information to boost reconstruction. If we take a close look, the nonzero pixels are mutually connected and cluster together, which can be modeled as a undirect 2-D graph (e.g., Fig. 3). Each pixel is a vertex and only a cluster or a few clusters are nonzeros. In previous works involving clustered sparsity [17]–[19], greedy algorithms were used to solve the corresponding problem. However, for such greedy algorithms, the sparsity number K should be approximately known before running the algorithm, which is not available for DOT. Instead, we efficiently solve this problem via convex programming. We only assume the image has a clustered tendency, while other information such as the sparsity number K, the size and locations of the cluster do not need to be known.
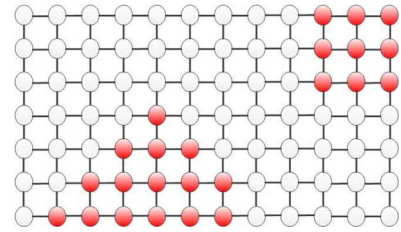

Fig. 3. Clustered sparsity data. Red nodes denotes nonzeros components and the white ones denotes zero components. Nonzero nodes tend to be mutually connected but not randomly distributed.

B. Efficient Algorithm

In this study, the clustered sparsity is approximated by overlapping group sparsity (OGS) [28]. We assign each pixel with its neighbor pixels (e.g., 4, 8) into a group. With this group setting, nonzero pixels of the reconstructed image will be only in the same groups, leading to the clustered structure of the nonzero pixels. The optimal group size (e.g., four neighbors or eight neighbors) may differ depending on the data, but whether we use the structured sparsity prior information or not is what matters the most. The experiments in the latter sections demonstrate that reconstruction accuracy can be significantly improved with the eight neighbors group setting. Now, we can summarize our model for structured sparsity based DOT

$$\hat{x} = \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_{2,1} \right\} \quad (5)$$

where $\|x\|_{2,1} = \sum_g \|x_g\|_2$, g denotes one of the group described above and x_g denotes the components in this group. The geometric view of $\ell_{2,1}$ norm with overlapping groups and more discussions can be found in [28].

Optimizing the overlapping group sparsity based problems is not an easy task, due to nonseparability of the mixed $\ell_{2,1}$ norm. A few software packages could solve this problem (5), e.g., [31], [32]. However, such algorithms duplicate the overlapped components as the original work [28]. The computational complexity will increase significantly due to the highly overlapped structure, e.g., 5 and 9 times higher when four and eight neighbors are grouped together. We would like to solve it more efficiently.

To solve (5), we propose a new algorithm based on the FISTA [21] framework, which has been proven to reach the optimal convergence rate for first order gradient methods. The whole algorithm is summarized in Algorithm 1. We call it as clustered sparsity reconstruction (CSR). For the first step, $f(x) = (1)/(2)\|Ax - b\|_2^2$, and $\nabla f(x) = A^T(Ax - b)$ denotes its gradient which has Lipschitz constant L. The smallest Lipschitz constant L can be selected based on the maximum eigenvalue of $A^T A$. A^T denotes the transpose of A. In the original FISTA algorithm for ℓ_1 norm regularization, the second step has a closed form solution by soft-thresholding. However, due to the nonsmoothness and nonseparability of the overlapped $\ell_{2,1}$ norm, there is no closed form solution for the OGS thresholding/ denosing problem in the second step. We apply the reweighted least squares algorithm [33]–[35] to solve it. Finally, each x^k is updated by the results in the previous two iterations to accelerate the convergence.

Algorithm 1 Clustered Sparsity Reconstruction (CSR)

Purpose: $\min_x\{(1)/(2)\|Ax - b\|_2^2 + \lambda\|x\|_{2,1}\}$

Input: $\rho = (1)/(L)$, λ, $t^1 = 1$, $y^0 = x^0$

for $k = 1$ to N **do**
  1) $s = x^k - \rho\nabla f(x^k)$
  2) $y^k = \arg\min_y\{(1)/(2\rho)\|y - s\|^2 + \lambda\|y\|_{2,1}\}$
  3) $t^{k+1} = [1 + \sqrt{1 + 4(t^k)^2}]/2$
  4) $x^{k+1} = y^k + (t^k - 1)/(t^{k+1})(y^k - y^{k-1})$

end for

The OGS thresholding algorithm is listed in Algorithm 2. $\text{conv}(x, J)$ denotes the convolution operation for x with template J and "·" denotes the element-wise operations. J depends on our group setting that has been discussed before. Note that both x and b need to be reshaped in 2-D in order to apply convolutions. Compared with standard OGS solvers [31], [32] with $\mathcal{O}(NS)$ complexity, this algorithm only costs $\mathcal{O}(N\log S)$, where S is the size of each group.

Algorithm 2 OGS Thresholding

Purpose: $\min_x\{(1)/(2)\|x - b\|_2^2 + \lambda\|x\|_{2,1}\}$

Input: λ, $x^0 = b$, $J = [1, 1, 1; 1, 1, 1; 1, 1, 1]$

for $k = 1$ to N **do**
  1) $r = \sqrt{\text{conv}(x.\times x, J)}$
  2) $x = b \cdot /[1 + \lambda * conv(r, J)]$

end for

IV. EXPERIMENT

A. Simulation

Simulations are conducted using the PMI Toolbox [36]. The probe geometry of these simulations and later phantom experiments is the same as that in previous work [37]. The number of measurements is $m = 188$, and the field of view (FOV) is
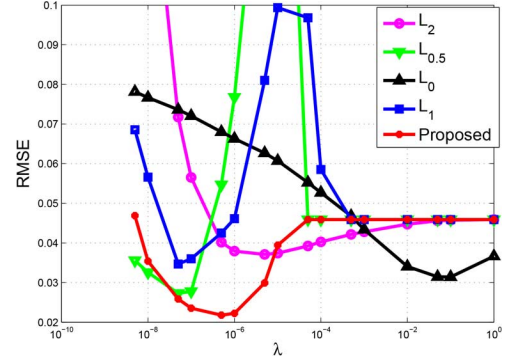

Fig. 4. Reconstruction performance of different algorithms when $\sigma = 0.0004$ (SNR = 20.79 dB).

6 cm $\times$ 6 cm with resolution 61×61 pixels. The optical properties of the medium are absorption coefficient $\mu_a = 0.08\cdot$ cm^{-1} and reduced scattering coefficient $\mu'_s = 8.8\cdot$ cm^{-1}. The objects (two spheres of 1 cm diameter) have the same scattering coefficient as the medium, and a higher absorption coefficient $\mu_a = 0.3\cdot$ cm^{-1}. The sensitivity matrix A is generated by Rytov approximation [9]. Random Gaussian noise with standard derivation σ is added into the measurement vector b. Root mean square error (RMSE) and contrast-to-noise ratio (CNR) are used as metrics for evaluation. From the definition, the reconstructed image with larger CNR means better performance. As suggested in [10], p in the ℓ_p norm regularization is set as 0.5. There is an additional ℓ_2 term combined in the software of smooth ℓ_0 method. We set this parameter as the best value of that in the ℓ_2 norm regularization method.

The RMSEs for different λ are presented in Fig. 4 when $\sigma = 0.0004$. This corresponds to a signal-to-noise ratio (SNR) of 20.79 dB. Compared with previous methods, smaller errors are achieved by the proposed method with a proper λ, which coincides with the structured sparsity theories. Both the ℓ_2 norm regularization [38] and the proposed method are less sensitive to the parameter setting.

Fig. 5(a) presents the reconstruction results in terms of absorption change ($\triangle\mu_a$, cm^{-1}) at the optimal parameters for each algorithm when $\sigma = 0.0004$. All the reconstructed images are shown at the same scale. The absorbers reconstructed by ℓ_2 norm regularization have a low contrast to the background due to over-smoothing. Sparse results are achieved by smooth ℓ_0, $\ell_{0.5}$ [10] and ℓ_1 norm regularization [9], which have much higher contrasts compared with the previous one. However, these norms only encourage sparsity and have no other constraints on the locations of the nonzero values. The absorbers on the images are slightly distorted and distributed, e.g., those by the ℓ_1 norm and $\ell_{0.5}$ norm regularization. Our method not only induces sparsity, but also encourage the nonzero values to be clustered. The cross sections of these recovered images at $x = 0$ are shown in Fig. 5(b). Compared to the ground-truth image, the result obtained by the ℓ_2 norm has smaller pixel values, while those reconstructed by ℓ_1 norm and smooth ℓ_0 norm have significantly larger pixel values. It is consistent with our visual observations of Fig. 5(a).

We also validate these algorithms with increased noise. When we gradually increase σ, the reconstructed absorbers
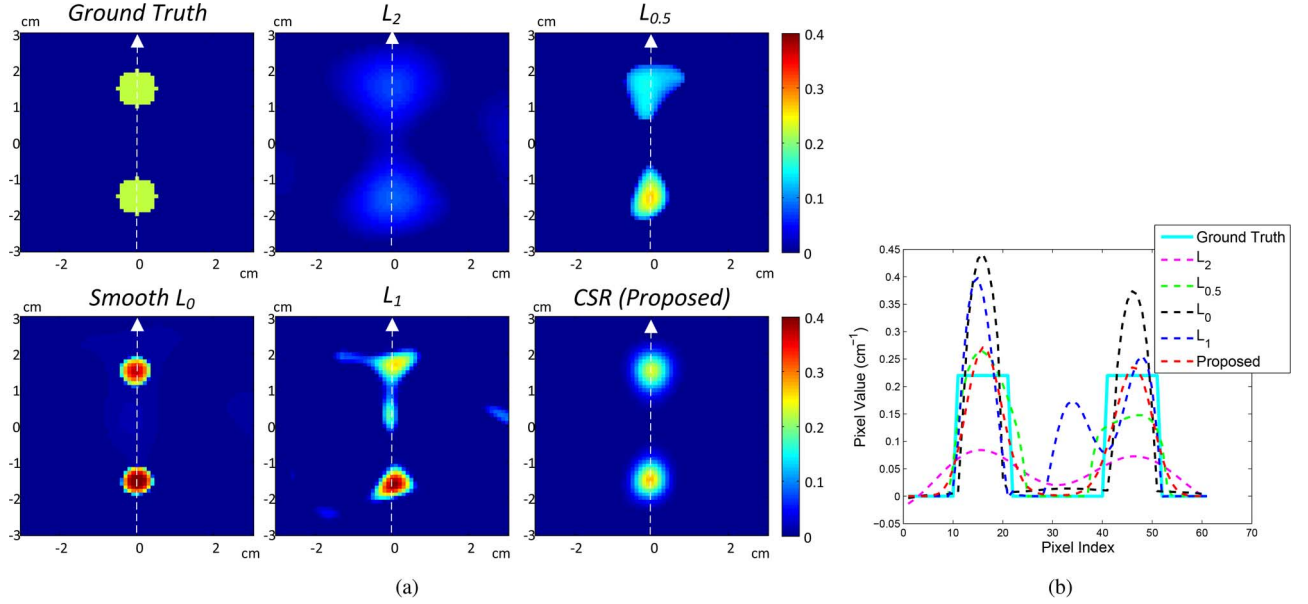
Fig. 5. (a) Ground-truth image and the reconstructed images of $\triangle\mu_a$ in cm^{-1} based on simulations, where $\sigma = 0.0004$ (SNR $= 20.79$ dB). (b) Cross sections of the absorbers at $x = 0$, which are indicated by the white arrows in (a).
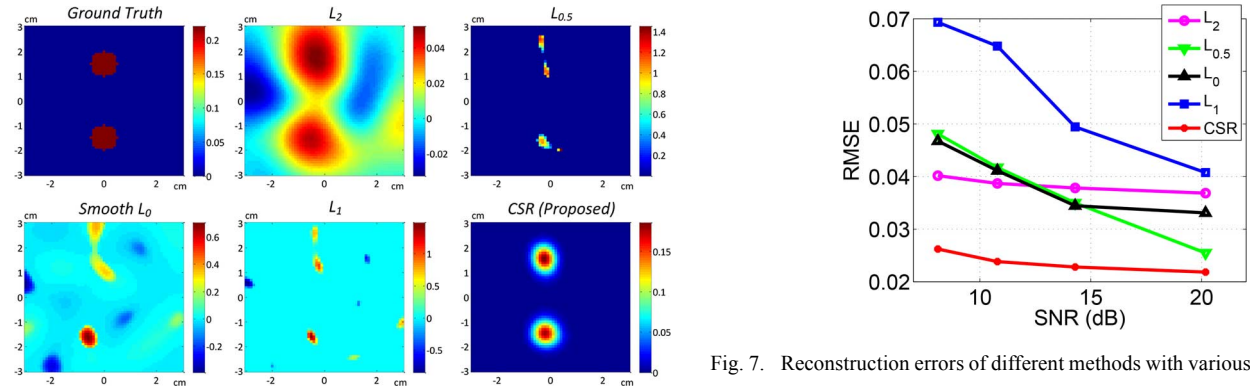

Fig. 6. Ground-truth image and the reconstructed images from the simulation ($\triangle\mu_a$ in cm^{-1}), where $\sigma = 0.002$ (SNR $= 7.66$ dB).

Fig. 7. Reconstruction errors of different methods with various levels of noise.

TABLE I
PARAMETER SETTING FOR THE DIFFERENT METHODS

	ℓ_2	$\ell_{0.5}$	smooth ℓ_0	ℓ_1	CSR
λ	6.7×10^{-6}	10^{-7}	0.05	6×10^{-7}	10^{-6}

of the existing sparsity based methods tend to be severely distorted, while that by the conventional ℓ_2 norm regularization has a very low contrast. Fig. 6 presents the results when $\sigma = 0.002$ (SNR $= 7.66$ dB). We use different color bars to show these results. Even with big noise, the result given by our CSR method is only sightly distorted and the shapes of the absorbers can be clearly observed. As the higher accuracy of our method can be clearly observed in this figure, we do not compare the cross sections of different images here. Comparing results with a different power of noise in Figs. 5 and 6, CSR is more robust to noise than the standard sparsity based methods ($\ell_{0.5}$, ℓ_1, and ℓ_0 norm regularization).

To quantitatively validate the above conclusion, the reconstruction errors of different methods are presented in Fig. 7 with different levels of noise. To reduce randomness, we run each method at each setting 20 times, and the average results are reported here. It can be clearly observed that the conventional ℓ_2 norm method and the proposed method are more robust to noise, while the standard sparsity based methods using $\ell_{0.5}$ norm and

ℓ_1 norm are sensitive to noise. The proposed CSR method consistently outperforms all of the other methods. These results further confirm the advantages of the proposed CSR method.

B. Phantom

We further conducted experiments to validate our method using laboratory tissue phantoms. The experiment environment was the same as that in previous work [37]. A large tank (approximately $15 \times 10 \times 10$ cm^3) was used to contain the phantom. The walls of this tank were covered by black tape so that no light was reflected. The phantom had an absorption coefficient of $\mu_a = 0.08\cdot$ cm^{-1} and a reduced scattering coefficient of $\mu'_s = 8.8\cdot$ cm^{-1}. A 1-cm-diameter spherical absorber with $\mu_a = 0.3\cdot$ cm^{-1} and the same reduced scattering coefficient was placed around the x-axis and 3 cm below the surface of the phantom. The measured data with and without the absorber were respectively acquired by all 188 channels.
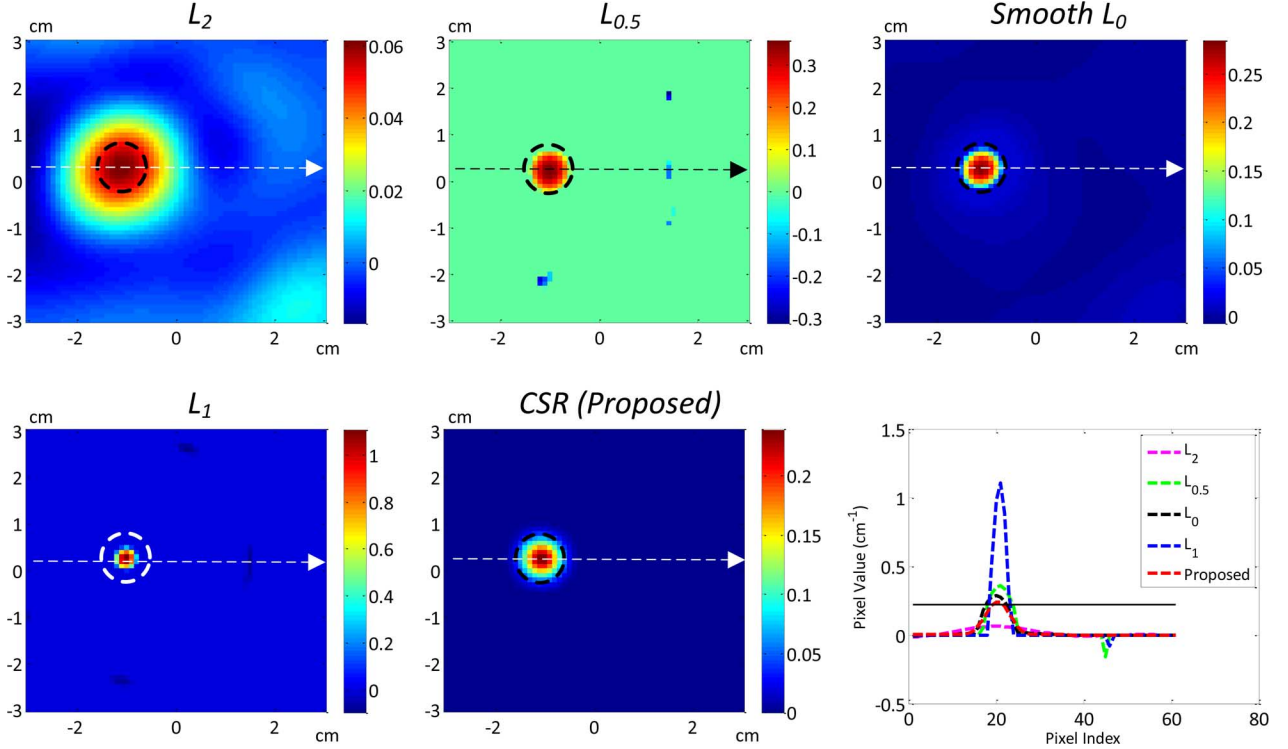
Fig. 8. Reconstructed images of a single absorber ($\triangle \mu_a$ in cm^{-1}). Dashed circles indicate the actual size of the object. Right bottom panel shows the cross sections of different images at the maximum pixel value, which are indicated by the arrows in the reconstructed images. Black solid line in the right bottom panel indicates the maximum $\triangle \mu_a$ caused by the actual absorber.

Keeping the experimental setup, we then place two spherical absorbers with 1 cm diameter around the x-axis at the same depth to conduct another experiment.

In previous methods [9], [38], the L-curve method [23] was used to select the parameters. However, such an approach often does not lead to the optimal parameters [39], [40]. In this study, we first selected a small range of the parameters by the L-curve method, and then the final parameter was selected in this range with user's adjustment. A sparser solution with fewer clusters is preferred. We list the parameters of different methods in Table I, which are used for both phantom experiments. An adaptive method proposed recently may alleviate this parameter tuning process [41]. Interestingly, we find these parameter settings are consistent with those in Fig. 4. It indicates another way to tune the parameters in clinic applications. The parameters tuned in some known tasks may be used for other tasks with the same experimental environment.

Figs. 8 and 9 present the reconstruction results for these two phantom experiments. Due to their significantly different reconstructed values, separate color bars are used. Since we know the actual size of the absorbers, the absorbers reconstructed by the ℓ_2 norm look dispersive, with larger areas and smaller intensity. We could find that the results obtained by the $\ell_{0.5}$ norm, smooth ℓ_0 norm, and ℓ_1 norm tend to be smaller than the ground-truth. Reconstruction of the second data (Fig. 9) is quite difficult as it is less sparse. The true absorbers are hard to be distinguished by $\ell_{0.5}$ and ℓ_1 norm methods. If we take a close look, the images recovered by the ℓ_2 norm and smooth ℓ_0 norm contain pixels of negative values. This may result in cross-talk in actual functional brain study [42]. Quantitative comparisons of these exper-

iments are listed in Table II, which are consistent with our visual observations. Based on the simulations results and the above analysis, the absorbers recovered by our CSR method should be the closest one to the ground-truth in terms of object area and intensity. One of the reasons is that our structured sparsity based method is less sensitive to noise. The random noise often does not follow the structures of the true signal, while it is hard to distinguish using the standard sparsity based methods.

C. Functional Human Brain Imaging

We have validated that the proposed CSR outperforms previous methods on the simulated data and phantom data. Our next task is to show how to apply the proposed method on functional brain imaging. A well-known motor task [43] (i.e., finger-tapping) is used here to evoke motor cortex activation. We follow the same protocol as that in [44] and the measurements were acquired by a multichannel, continuous-wave NIRS system (CW-5, Techen Inc., Milford, MA, USA) [45]. Following a fixed rhythm, the subjects were instructed to simultaneously tap four fingers (except the thumb) up and down without moving the wrist and arm. Light of 690 nm and 830 nm wavelengths were emitted from sources to measure the changes in concentrations of oxy- and deoxy- hemoglobin. The reconstructed images (with 41×13 pixels) were sliced at $Z = -1.5$ cm depth.

We randomly selected one of the eight subjects that was reported in [44]. With the same measurements, we reconstructed the images with different methods, which contain ℓ_2 norm, $\ell_{0.5}$ norm [10], smooth ℓ_0 norm [10], ℓ_1 norm [9] regularization
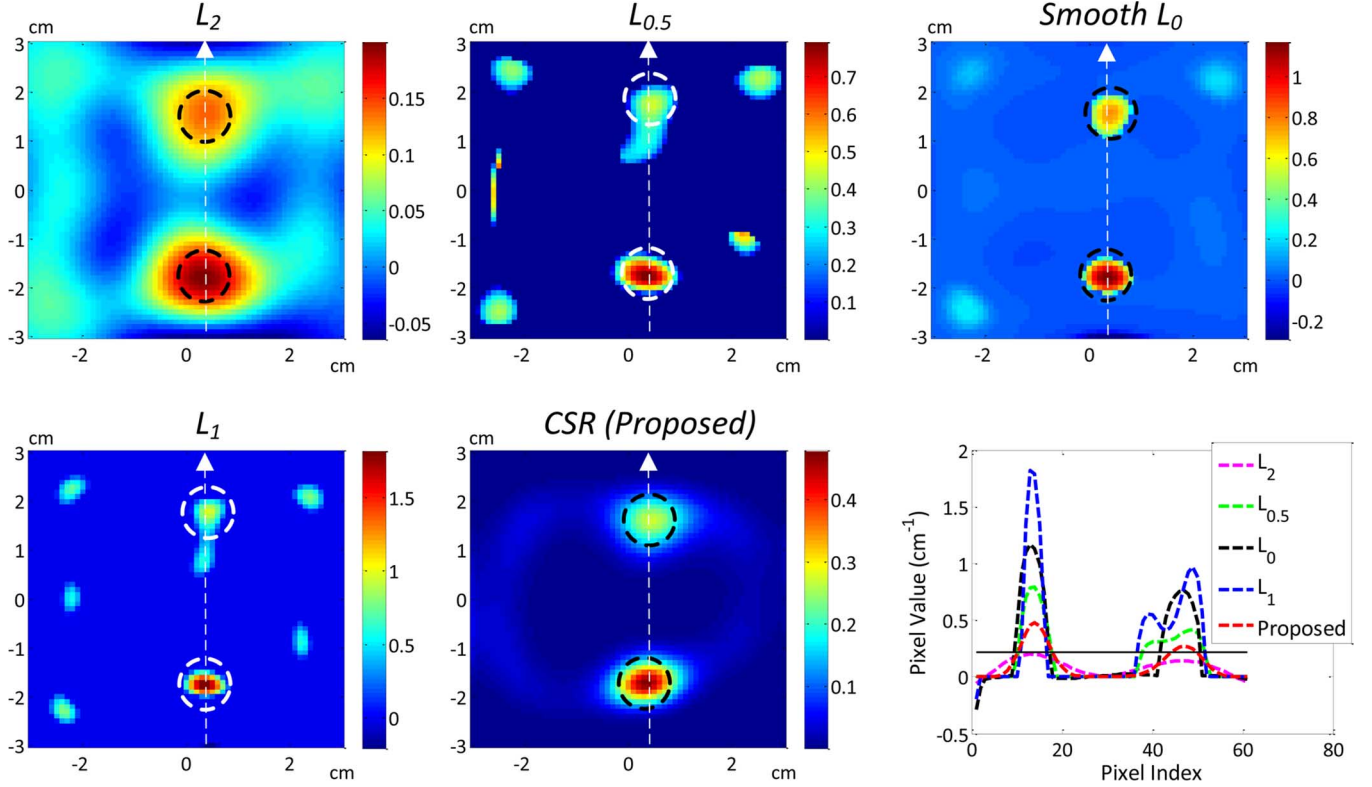
Fig. 9. Reconstructed images of two separated absorbers ($\triangle \mu_a$ in cm^{-1}). Dashed circles indicate the actual sizes of the objects. Right bottom panel shows the cross sections of different images at the maximum pixel value, which are indicated by the white arrows in the reconstructed images. Black solid line in the right bottom panel indicates the maximum $\triangle \mu_a$ caused by the actual absorbers.

TABLE II
CNRs OF DIFFERENT PHANTOM RECONSTRUCTION RESULTS

	ℓ_2	$\ell_{0.5}$	smooth ℓ_0	ℓ_1	CSR
Fig. 8	4.69	8.34	10.71	3.66	**14.03**
Fig. 9	3.36	4.14	5.32	3.07	**6.70**

and the proposed CSR. The λ in our method was selected as 7.9×10^{-4}. Those results are presented in Fig. 10 on the same scale.

The images reconstructed by all methods had brain activation on the lower right side. This is expected due to the left hand finger tapping (contra-lateral activation) of the subject. Enlarged activation areas were obtained by the ℓ_2 norm and smooth ℓ_0 norm regularization (the first row and the third row of Fig. 10), while the remaining images showed smaller activation areas. If comparing the sparsity inducing methods (i.e., $\ell_{0.5}$ norm, smooth ℓ_0 norm, ℓ_1 norm with the proposed method), the proposed method and ℓ_0 norm produced more localized images. In addition, it is obvious that the images obtained by our CSR method (the last row of Fig. 10) had much higher contrasts than those obtained by ℓ_0 norm. Considering the shapes of different results, our method provided a more concentrated and accurate result. Comparing the reconstruction results of oxyhemoglobin and deoxyhemoglobin (i.e., the first column and the second column), it seems that our method can potentially reduce the cross-talk [42]. These results confirms the benefit of our method in functional human brain studies.

V. DISCUSSION AND CONCLUSION

In this study, we proposed to use structured sparsity to improve the reconstruction accuracy of DOT. More precisely, the clustered prior information was utilized by the mixed $\ell_{2,1}$ norm regularization. This was motivated by the fact that functional brain activation is often localized in some special region(s) but not randomly distributed. Before this study, the clustered sparsity had already been successfully used in compressed sensing and computer vision [46]. It leads to several advantages: 1) improving the reconstruction accuracy with the same number of measurements; 2) maintaining stable recovery when the measurements are not sufficient for standard sparsity (e.g., by ℓ_1 norm); 3) enhancing the robustness to noise and preventing artifacts in the background. By structured sparsity theories [17]–[19], it has been proved that only $\mathcal{O}(K + C \log(N/C))$ measurements are required to recover clustered sparse signals instead of $\mathcal{O}(K + K \log(N/K))$ for standard sparse signals. Here C denotes the number of clusters, which is significantly smaller than the number of nonzero pixels K. There is no additional information (e.g., shape, size, location of the absorbers) required for the proposed algorithm. These are why it could facilitate diffuse optical imaging with high accuracy.

Numerical simulation and phantom experiments have validated the effectiveness of our method when compared with conventional and recent algorithms. Qualitative analysis demonstrated that out method can outperform existing approaches up to 30% in terms of CNR. The superior performance of our method was further confirmed on *in vivo* data.
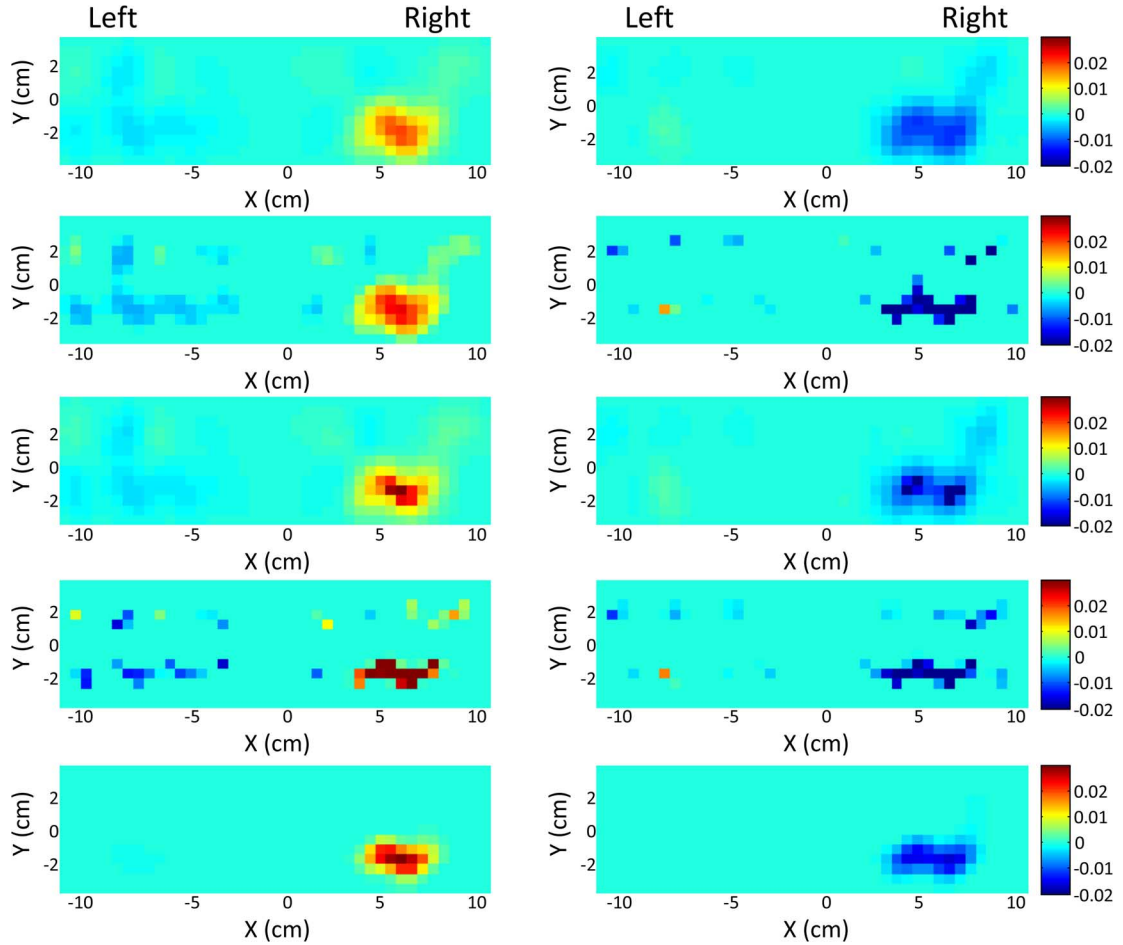
Fig. 10. 2-D slices (1.5 cm below the scalp surface) of reconstructed human brain images induced by a finger tapping task. Left column: the images reconstructed for increased oxy-hemoglobin concentration (arbitrary unit). Right column: the images reconstructed for decreased deoxy-hemoglobin concentration (arbitrary unit). From the first row to the last row, the images are reconstructed using ℓ_2 norm, $\ell_{0.5}$ norm, smooth ℓ_0 norm, ℓ_1 norm, and the proposed method, respectively. Figure is best viewed on screen, rather than in print.

Our method can recover images with the fewest artifacts and best contrasts in the expected region. Practical applications can benefit from the proposed CSR method with little or no revision on the hardware.

Currently, parameter selection is still an open active research area for DOT. To the best of our knowledge, there is no efficient way to accurately select the optimal regularization parameter or parameters for each algorithm. For this reason, an algorithm that has fewer parameters and is not sensitive to the parameters is preferred. The ℓ_1 norm and $\ell_{0.5}$ norm based methods are very sensitive to parameter settings, as illustrated by their sharp curves in Fig. 4. We used the original FISTA algorithm to solve the ℓ_1 norm regularization problem and ran sufficient iterations until the algorithm converged. The human brain imaging results seem to be better if the stopping criteria are controlled manually [11], e.g., by setting the number of Newton iterations. However, multiple parameter selection is a drawback of the algorithm. A similar issue is also in the method of smooth ℓ_0 norm regularization, where an additional ℓ_2 norm regularization term is included in the software. By contrast, there is only one nonsensitive parameter λ in our method. We have made our best effort to select the optimal parameters for different methods for fair comparisons. Although the parameters for the previous methods may not be exactly optimal, the experimental results are sufficient to demonstrate the benefit of our method, which can achieve substantial improvement in image reconstruction with only one nonsensitive parameter.

All experiments are conducted using MATLAB on a desktop with 3.4-GHz Intel core i7 3770 CPU. The reconstruction speed of our CSR method is slightly slower than the ℓ_1 norm regularization method, due to the more difficult problem with overlapping groups. In the phantom experiments, the reconstruction times of the ℓ_2 norm, $\ell_{0.5}$ norm, smooth ℓ_0 norm, ℓ_1 norm based methods and the proposed method are around 3 s, 26 s, 2 s, 14 s, and 23 s, respectively. Due to the linear approximation, such reconstruction speed is quite acceptable.

Although the proposed CSR method has achieved promising results, some applications in DOT is still very challenging. Different from functional brain imaging, tumor imaging involves the problem to reconstruct both scattering and absorption coefficients [22]. The current work cannot be directly applied to such case. In some scenarios, such as breast imaging [47], the image may not be sparse. Some sparsity transformation may be required, e.g., the total variation [22]. Future work will focus on extending the proposed method in such cases.

REFERENCES

[1] D. A. Boas, A. M. Dale, and M. A. Franceschini, "Diffuse optical imaging of brain activation: Approaches to optimizing image sensitivity, resolution, and accuracy," *Neuroimage*, vol. 23, pp. S275–S288, 2004.

[2] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.*, vol. 20, no. 10, pp. 435–442, 1997.

[3] T. Durduran, R. Choe, W. Baker, and A. Yodh, "Diffuse optics for tissue monitoring and tomography," *Rep. Prog. Phys.*, vol. 73, no. 7, p. 076701, 2010.

[4] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Piscataway, NJ: IEEE Press, 1999.

[5] S. R. Arridge and J. C. Schotland, "Optical tomography: Forward and inverse problems," *Inverse Probl.*, vol. 25, no. 12, p. 123010, 2009.

[6] M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with a priori anatomical information," *Phys. Med. Biol.*, vol. 50, no. 12, p. 2837, 2005.

[7] N. Cao, A. Nehorai, and M. Jacobs, "Image reconstruction for diffuse optical tomography using sparsity regularization and expectationmaximization algorithm," *Opt. Exp.*, vol. 15, pp. 13 695–13 708, 2007.

[8] J. Ye, S. Lee, and Y. Bresler, "Exact reconstruction formula for diffuse optical tomography using simultaneous sparse representation," in *Proc IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 1621–1624.

[9] M. Süzen, A. Giannoula, and T. Durduran, "Compressed sensing in diffuse optical tomography," *Opt. Exp.*, vol. 18, no. 23, pp. 23 676–23 690, 2010.

[10] J. Prakash, C. Shaw, R. Manjappa, R. Kanhirodan, and P. K. Yalavarthy, "Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 2, p. 6800609, Mar./Apr. 2014.

[11] V. C. Kavuri, Z.-J. Lin, F. Tian, and H. Liu, "Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography," *Biomed. Opt. Exp.*, vol. 3, no. 5, p. 943, 2012.

[12] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[13] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[14] Y. Zheng, E. Daniel, A. A. Hunter, III, R. Xiao, J. Gao, H. Li, M. G. Maguire, D. H. Brainard, and J. C. Gee, "Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix," *Med. Image Anal.*, vol. 18, no. 6, pp. 903–913, 2014.

[15] J. Huang, S. Zhang, H. Li, and D. Metaxas, "Composite splitting algorithms for convex optimization," *Comput. Vis. Image Und.*, vol. 115, no. 12, pp. 1610–1622, 2011.

[16] C. Chen and J. Huang, "Compressive sensing MRI with wavelet tree sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1124–1132.

[17] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, 2011.

[18] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[19] V. Cevher, P. Indyk, C. Hegde, and R. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," presented at the Int. Conf. Sampl. Theory Appl., Marseille, France, 2009.

[20] B. K. Hölzel, U. Ott, H. Hempel, A. Hackl, K. Wolf, R. Stark, and D. Vaitl, "Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators," *Neurosci. Lett.*, vol. 421, no. 1, pp. 16–21, 2007.

[21] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[22] O. Lee and J. C. Ye, "Joint sparsity-driven non-iterative simultaneous reconstruction of absorption and scattering in diffuse optical tomography," *Opt. Exp.*, vol. 21, no. 22, pp. 26 589–26 604, 2013.

[23] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.

[24] S. Okawa, Y. Hoshi, and Y. Yamada, "Improvement of image quality of time-domain diffuse optical tomography with lp sparsity regularization," *Biomed. Opt. Exp.*, vol. 2, no. 12, pp. 3334–3348, 2011.

[25] C. Chen, Y. Li, and J. Huang, "Forest sparsity for multi-channel compressive sensing," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2803–2813, Nov. 2014.

[26] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2005.

[27] J. Huang and T. Zhang, "The benefit of group sparsity," *Ann. Stat.*, vol. 38, no. 4, pp. 1978–2004, 2010.

[28] L. Jacob, G. Obozinski, and J. Vert, "Group lasso with overlap and graph lasso," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Montreal, Canada, 2009, pp. 433–440.

[29] C. Chen and J. Huang, "The benefit of tree sparsity in accelerated MRI," *Med. Image Anal.*, vol. 18, no. 6, pp. 834–842, 2014.

[30] O. Lee, J. M. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: Noniterative exact reconstruction using joint sparsity," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1129–1142, May 2011.

[31] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning With Efficient Projections*. Phoenix, AZ: Arizona State Univ., 2009.

[32] W. Deng, W. Yin, and Y. Zhang, Group sparse optimization by alternating direction method Tech. Rep., 2011.

[33] P. Chen and I. Selesnick, "Translation-invariant shrinkage/thresholding of group sparse signals," *Signal Process.*, vol. 94, pp. 476–489, 2014.

[34] C. Chen, J. Huang, L. He, and H. Li, "Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Columbus, OH, 2014, pp. 2713–2720.

[35] C. Chen, Z. Peng, and J. Huang, "O(1) algorithms for overlapping group sparsity," in *Int. Conf. Pattern Recognit.*, Stockholm, Sweden, 2014.

[36] D. Boas, D. Brooks, R. Gaudette, T. Gaudette, E. Miller, and Q. Zhang, Photon Migration Imaging (PMI) Toolbox [Online]. Available: http://www.nmr.mgh.harvard.edu/PMI/toolbox/

[37] H. Niu, F. Tian, Z. Lin, and H. Liu, "Development of a compensation algorithm for accurate depth localization in diffuse optical tomography," *Opt. Lett.*, vol. 35, no. 3, pp. 429–431, 2010.

[38] F. Tian, G. Alexandrakis, and H. Liu, "Optimization of probe geometry for diffuse optical brain imaging based on measurement density and distribution," *Appl. Opt.*, vol. 48, no. 13, pp. 2496–2504, 2009.

[39] J. Culver, R. Choe, M. Holboke, L. Zubkov, T. Durduran, A. Slemp, V. Ntziachristos, B. Chance, and A. Yodh, "Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: Evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging," *Med. Phys.*, vol. 30, no. 2, pp. 235–247, 2003.

[40] J. Prakash and P. K. Yalavarthy, "A lsqr-type method provides a computationally efficient automated optimal choice of regularization parameter in diffuse optical tomography," *Med. Phys.*, vol. 40, no. 3, p. 033101, 2013.

[41] J. Feng, C. Qin, K. Jia, D. Han, K. Liu, S. Zhu, X. Yang, and J. Tian, "An adaptive regularization parameter choice strategy for multispectral bioluminescence tomography," *Med. Phys.*, vol. 38, no. 11, pp. 5933–5944, 2011.

[42] Y. Zhan, A. T. Eggebrecht, J. P. Culver, and H. Dehghani, "Singular value decomposition based regularization prior to spectral mixing improves crosstalk in dynamic imaging using spectral diffuse optical tomography," *Biomed. Opt. Exp.*, vol. 3, no. 9, pp. 2036–2049, 2012.

[43] D. R. Leff, F. Orihuela-Espina, C. E. Elwell, T. Athanasiou, D. T. Delpy, A. W. Darzi, and G.-Z. Yang, "Assessment of the cerebral cortex during motor task behaviours in adults: A systematic review of functional near infrared spectroscopy (FNIRS) studies," *NeuroImage*, vol. 54, no. 4, pp. 2922–2936, 2011.

[44] F. Tian, M. R. Delgado, S. C. Dhamne, B. Khan, G. Alexandrakis, M. I. Romero, L. Smith, D. Reid, N. J. Clegg, and H. Liu, "Quantification of functional near infrared spectroscopy to assess cortical reorganization in children with cerebral palsy," *Opt. Exp.*, vol. 18, no. 25, p. 25973, 2010.

[45] M. A. Franceschini, D. K. Joseph, T. J. Huppert, S. G. Diamond, and D. A. Boas, "Diffuse optical imaging of the whole head," *J. Biomed. Opt.*, vol. 11, no. 5, pp. 054 007–054 007, 2006.

[46] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Proc. Int. Conf. Comput. Vis.*, Kyoto, Japan, 2009, pp. 64–71.

[47] M. L. Flexman, M. A. Khalil, R. Al Abdi, H. K. Kim, C. J. Fong, E. Desperito, D. L. Hershman, R. L. Barbour, and A. H. Hielscher, "Digital optical tomography system for dynamic breast imaging," *J. Biomed. Opt.*, vol. 16, no. 7, pp. 076 014–076 014, 2011.