MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Multi-Layer Terahertz Imaging of Non-Overlapping Contents

Wang, P.; Fu, H.; Koike-Akino, T.; Orlik, P.V.

TR2018-098 July 13, 2018

Abstract

This paper considers terahertz (THz) imaging of multi-layer nonoverlapping contents with compressed measurements. One issue here is the shadow effect from front layers to deep layers due to the non-uniform penetrating illumination. In the case of nonoverlapping contents in layered structures, the shadow effect can be utilized to improve recovery performance and reduce the number of measurements. To this end, we propose several approaches based on the total variation (TV) minimization principle and take into account individual-layer sparsity, group sparsity over layers, and hierarchical group sparsity over layers to reduce the number of measurements. Numerical evaluation confirms the effectiveness of the proposed approaches.

IEEE Sensor Array and Multi-Channel Signal Processing Workshop (IEEE SAM)

MULTI-LAYER TERAHERTZ IMAGING OF NON-OVERLAPPING CONTENTS

Pu Wang[1]*, Haoyu Fu*[2]*, Toshiaki Koike-Akino*[1]*, Philip V. Orlik*[1]

[1] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
E-mails: {pwang, koike, porlik}@merl.com
[2] Ohio State University, Columbus, OH 43210, USA
E-mails: fu.436@osu.edu

ABSTRACT

This paper considers terahertz (THz) imaging of multi-layer non-overlapping contents with compressed measurements. One issue here is the shadow effect from front layers to deep layers due to the non-uniform penetrating illumination. In the case of non-overlapping contents in layered structures, the shadow effect can be utilized to improve recovery performance and reduce the number of measurements. To this end, we propose several approaches based on the total variation (TV) minimization principle and take into account individual-layer sparsity, group sparsity over layers, and hierarchical group sparsity over layers to reduce the number of measurements. Numerical evaluation confirms the effectiveness of the proposed approaches.

Index Terms— Terahertz, multi-layer structure, content extraction, compressed measurements.

1. INTRODUCTION

Over the past two decades, there have been increased interests in terahertz (THz) sensing using the time-domain spectroscopy (TDS) in either a reflection or transmission mode, due to the broad applications in gas sensing, moisture analysis, non-destructive evaluation, biomedical diagnosis, package inspection, and security screening [1]. By sending an ultra-short pulse (e.g., 1-2 picoseconds), the THz-TDS system is able to inspect not only the top surface of the sample but also its internal structure, either a defect underneath the top layer or a multi-layer structure, due to its capability of penetrating a wide range of non-conducting materials.

The THz-TDS can operate in a *raster* or *compressed* scanning mode [2, 3]. In the raster scanning mode, the sample under inspection is illuminated by a THz-TDS point source and a programmable mechanical raster moves the sample in the plane perpendicular to the incidental waveform in order to measure the two-dimensional surface of the sample. Single-layer and multi-layer content extractions have been tested in [4–6]. One issue here is, due to either irregular sample surfaces or vibration from the mechanical raster move, to deal with depth variation and its induced delay variation from one pixel to another.

On the other hand, the compressed scanning mode aims to remove the mechanical raster move and reduce the total acquisition time. As shown in Fig. 1, a terahertz transmitter sends a terahertz ultra-short pulse beam to a collimating lens. The collimated beam passes through layered samples and is then modulated by a random pattern with the help of a THz-band spatial light modulator (SLM), followed by a focusing lens and a single-pixel photoconductive detector [7, 8]. In other words, only one measurement is formed for a mask at a time. The compressed scanning process repeats with different realizations of random masks and collects multiple sequential measurements. The sample image can then be recovered by sparsity-driven minimization methods, e.g., the total variation (TV) minimization method [7].

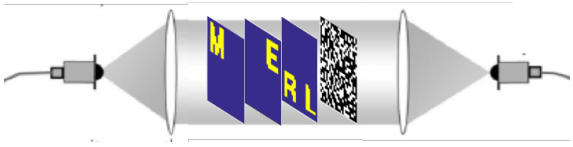

Fig. 1. THz compressed scanning setup for multi-layer imaging of non-overlapping contents.

In this paper, we consider the use of compressed scanning mode to extract non-overlapping contents from layered structures with THz measurements. One issue here is the *shadow* effect due to the non-uniform penetrating illumination from front layers to deep layers. In the case of non-overlapping contents over multiple layers, the shadow effect can be utilized to improve the recovery performance and reduce the number of measurements. We build the proposed approaches on the TV minimization principle as a regularization approach for noise reduction while preserving image edges, to promote the sparsity on the image gradient domain for the multi-layered non-overlapping content extraction with compressed measurements. Furthermore, we take into account *individual-layer sparsity*, *group sparsity over layers*, and *hierarchical group sparsity over layers* to reduce the number of measurements while maintaining the recovery performance.

The remainder of this paper is organized as follows. Section 2 introduces the signal model for the multi-layer THz imaging. In Section 3, several approaches based on the TV minimization are proposed. Numerical results are provided in Section 4, followed by the conclusion in Section 5.

2. SIGNAL MODEL

Let $\mathbf{x}_l = \big[x_l(1), x_l(2), \ldots, x_l(N)\big]^T$ denote a reflectance vector by stacking the columns of the two-dimensional reflectance matrix at the l-th layer of the sample, where N is the number of total pixels at each layer of the sample. As the THz source illuminates the sample from a spatially encoded mask, the received measurements can be described as

$$\mathbf{y}_l = \mathbf{A}_l \mathbf{x}_l + \mathbf{v}_l, \tag{1}$$

The work of H. Fu was performed during his internship at MERL.

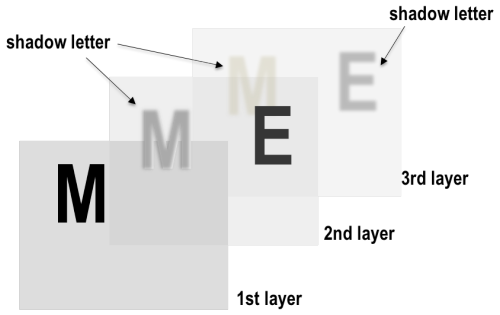

Fig. 2. The shadow effect of non-overlapping content for front layers to deep layers.

where $\mathbf{A}_l = \left[\mathbf{a}_{l,1}, \ldots, \mathbf{a}_{l,M}\right]^T$ is the measurement matrix corresponding to the l-th layer, $\mathbf{v}_l = \left[v_l(1), \ldots, v_l(M)\right]^T$ is the Gaussian distributed noise with zero mean and an unknown variance σ_l^2, i.e., $\mathbf{v}_l \sim \mathcal{N}(\mathbf{0}, \sigma_l^2 \mathbf{I}_M)$, $\mathbf{y}_l = \left[y_l(1), \ldots, y_l(M)\right]^T$, and M is the number of measurements.

In practice, the THz-band SLM likely remains the same during the electro-optic sampling process which leads to the same measurement matrix $\mathbf{A}$ over all layers. Here, we assume the measurement matrix $\mathbf{A}_l$ is a function of the layer index as the measurement matrix can absorb layer-dependent inter-reflections and surface irregularity.

It is noticed that the signal model of (1) can, in fact, describe both raster and compressed scanning acquisitions:

- In the case of the *raster scanning*, i.e., each pixel is illuminated and measured individually, we have $M = N$ and $\mathbf{A}$ reduces to a diagonal matrix with diagonal elements responsible for the depth variation [6, Section III.I.4].
- In the case of the *compressed scanning*, e.g., the single-pixel THz camera [7], we have $M < N$ and each row of the measurement matrix $\mathbf{A}$ corresponds to one random mask pattern used to form one measurement y_m.

In the case of layered structures, the shadow effect is present due to the non-uniform penetrating illumination from front layers to deeper layers. For instance, as shown in Fig. 2, two letters 'M' and 'E' are present at two non-overlapping areas in the first and second layers, respectively. Due to different reflection coefficients and reflection index between the letter 'M' and the rest area of the first layer, the penetrating illumination towards the second layer is hence different, which causes a shadow letter 'M' in the second layer and, subsequently, in the third layer. This shadow effect is also applied to the letter 'E' in the second layer with its shadow letter in the third layer.

3. PROPOSED APPROACHES

In this section, we utilize the TV, originally proposed in [9] as a regularization approach for noise reduction while preserving image edges, to promote the sparsity on the image gradient domain for the multi-layered non-overlapping content extraction with compressed measurements. Particularly, we consider several variants of the TV over the multi-layer structure to enforce either *individual-layer sparsity*, *group sparsity over layers*, or *hierarchical group sparsity over layers* to reduce the number of measurements while maintaining the recovery performance.

3.1. TV Minimization over Individual Layers

First, a straightforward solution is, similar to [7], to apply the TV-regularized minimization independently over each individual layer. This solution can be formulated as follows

$$\hat{\mathbf{x}}_l = \arg\min_{\mathbf{x}_l} \frac{1}{2}\|\mathbf{y}_l - \mathbf{A}_l\mathbf{x}_l\|_2^2 + \lambda_l \|\mathbf{x}_l\|_{\text{TV}}, \tag{2}$$

where λ_l is the regularization parameter for the l-th layer, and $\|\mathbf{x}\|_{\text{TV}}$ is a discrete TV (semi)-norm with two popular choices of 1) the isotropic TV

$$\|\mathbf{x}\|_{\text{TV}} = \sum_{n=1}^{N} \text{TV}_I(x_n) = \sum_{n=1}^{N} \sqrt{(\Delta_n^h(\mathbf{x}))^2 + (\Delta_n^v(\mathbf{x}))^2}, \tag{3}$$

and 2) the anisotropic TV

$$\|\mathbf{x}\|_{\text{TV}} = \sum_{n=1}^{N} \text{TV}_A(x_n) = \sum_{n=1}^{N} \left[|\Delta_n^h(\mathbf{x})| + |\Delta_n^v(\mathbf{x})|\right], \tag{4}$$

with the operators $\Delta_n^h(\mathbf{x})$ and $\Delta_n^v(\mathbf{x})$ correspond to, respectively, the horizontal and vertical first order differences at pixel n. Specifically, $\Delta_n^h(\mathbf{x}) = x_n - x_{h(n)}$ and $\Delta_n^v(\mathbf{x}) = x_n - x_{v(n)}$ with $h(n)$ and $v(n)$ denoting the nearest horizontal and vertical neighbors of pixel n, respectively. Fast algorithms such as iterative shrinkage/thresholding algorithms (ISTA) and its accelerated version (FISTA) have been proposed to circumvent the non-smoothness of the TV regularization term [10].

3.2. Group TV Minimization over Multiple Layers

It is seen that the solutions from (2) do not explore the shadow effect and require more measurements for deep layers as the sparsity decreases over layers. To utilize the shadow effect, one can enforce the group sparsity over layers such that the content in front layers always appear in deep layers. One way to formulate this group sparsity over layers can be described as follows

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \sum_{l} \frac{1}{2}\|\mathbf{y}_l - \mathbf{A}_l\mathbf{x}_l\|_2^2 + \lambda \|\mathbf{X}\|_{\text{GTV}} \tag{5}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_L] \in \mathbb{R}^{N \times L}$ groups all images over L layers, λ is the regularization parameter, $\|\mathbf{X}\|_{\text{GTV}}$ is the group TV over multiple layers defined as

$$\begin{aligned}\|\mathbf{X}\|_{\text{GTV}} &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} [\text{TV}_I(x_n^l)]^2} \\ &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} (\Delta_n^h(\mathbf{x}_l))^2 + (\Delta_n^v(\mathbf{x}_l))^2},\end{aligned} \tag{6}$$

for the isotropic TV and

$$\begin{aligned}\|\mathbf{X}\|_{\text{GTV}} &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} [\text{TV}_A(x_n^l)]^2} \\ &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} \left[|\Delta_n^h(\mathbf{x})| + |\Delta_n^v(\mathbf{x})|\right]^2},\end{aligned} \tag{7}$$

for the anisotropic TV. It is easy to see that both group TV definitions are the $\ell_{2,1}$ norm, i.e., $\|\mathbf{D}\|_{2,1} = \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} D_{n,l}^2}$, of the (isotropic or anisotropic) TV map $\mathbf{D}$, i.e., $D_{n,l} = \text{TV}_I(\mathbf{x}_l(n))$ or $D_{n,l} = \text{TV}_A(\mathbf{x}_l(n))$. In other words, (5) encourages solutions which yield layer-wise sparsity on the TV map of $\mathbf{X}$.

It is also noted that, for a given pixel n, instead of 2 (horizontal and vertical) gradients in the single-layer case, there are $2L$ gradients over L layers. The isotropic group TV of (6) simply uses the ℓ_2 norm over the expanded $2L$ gradients followed by the ℓ_1 norm over all pixels. Inspired by this motivation, we introduce another version of the anisotropic TV

$$\begin{aligned}\|\mathbf{X}\|_{\text{TV}} &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} [\text{TV}_A(x_n^l)]} \\ &= \sum_{n=1}^{N} \sqrt{\sum_{l=1}^{L} [|\Delta_n^h(\mathbf{x})| + |\Delta_n^v(\mathbf{x})|]},\end{aligned} \tag{8}$$

which uses the anisotropic definition, i.e., ℓ_1 norm, over the $2L$ gradients followed by the ℓ_1 norm over pixels.

3.3. Hierarchical Group TV Minimization over Multiple Layers

In addition to the shadow effect, the non-overlapping assumption imposes an additional feature over layers, i.e., the TV-domain sparsity extent increases over layers, i.e., $S(\text{TV}(\mathbf{x}_l)) \subset S(\text{TV}(\mathbf{x}_{l+1}))$, where $S(\mathbf{x})$ is the set of indices where the elements of $\mathbf{x}$ are non-zeros. In fact, the combined shadow effect and non-overlapping assumption introduces a hierarchical layer-wise structure in the TV domain.

As an illustrative example of $L = 3$ in Fig. 3 where non-zero and zero gradient denoted as color and gray arrows, respectively, two pixels with non-zero (horizontal and vertical) gradients are present in the upper left corner of the first layer and, due to the shadow effect, the gradients of the same pixels at the second and third layers are also non-zeros. At the second layer, two more pixels at the upper right corner have non-zero gradients which results in the same pixels at the third layer have non-zero gradients. Finally, at the third layer, two more pixels at the lower right corner have non-zero gradients. From this illustrative example, one can clearly see that the hierarchical structure over layers. To be more precise, it is a *nested group sparsity* pattern over layers in the TV domain.

The nested group structure can be mathematically formulated as follows. Denote the l-th layer image as $\mathbf{X}_l \in \mathbb{R}^{N_r \times N_c}$ where $N = N_r \times N_c$, where $\mathbf{x}_l$ is the vectorized version of $\mathbf{X}_l$. We can define the matrix $\mathbf{I}_{\text{TV}}$ below indicating the smallest layer index where a pixel has non-zero gradients

$$[\mathbf{I}_{\text{TV}}]_{n_r,n_c} = \begin{cases} \min\limits_{l}\{\text{TV}([\mathbf{X}_l]_{n_r,n_c}) \neq 0\} \\ 0, \text{ if } \text{TV}([\mathbf{X}_l]_{n_r,n_c}) = 0, l = 1, \cdots, L \end{cases} \tag{9}$$

For the illustrative example in Fig. 3, the corresponding matrix $\mathbf{I}_{\text{TV}}$ is given at the right upper corner. For a given pixel at $[n_r, n_c]$, the nested group structure implies that

$$\text{TV}([\mathbf{X}_l]_{n_r,n_c} = 0, \text{ for all } l < [\mathbf{I}_{\text{TV}}]_{n_r,n_c}. \tag{10}$$

In other words, all gradients at layers before the layer $[\mathbf{I}_{\text{TV}}]_{n_r,n_c}$ are all zeros. On the other hand, the number of non-zero gradients

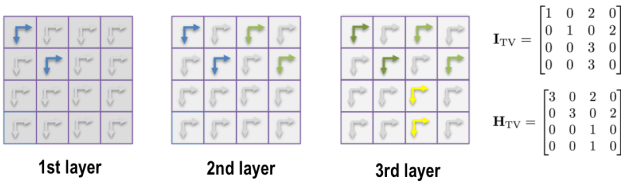

Fig. 3. Hierarchical group sparsity in the TV domain where colored arrows represent non-zero (horizontal/vertical) gradients and grey arrows denote zero gradients. The matrix $\mathbf{I}_{\text{TV}}$ indicates the smallest layer index where the pixel has non-zero gradients, and the matrix $\mathbf{H}_{\text{TV}}$ indicates the number of non-zero gradients over layers.

for the pixel at $[n_r, n_c]$ over layers is equal to the number of layers minus the corresponding smallest layer index in $\mathbf{I}_{\text{TV}}$.

$$[\mathbf{H}_{\text{TV}}]_{n_r,n_c} = L - [\mathbf{I}_{\text{TV}}]_{n_r,n_c}, \text{ if } [\mathbf{I}_{\text{TV}}]_{n_r,n_c} > 0. \tag{11}$$

The corresponding $\mathbf{H}_{\text{TV}}$ is also given in Fig. 3.

The hierarchical nested group structure has been utilized in the hierarchical group LASSO [11] which uses hierarchical sparsity constraints directly in selecting *solution variables* and the hierarchical vector auto-regressive modeling [12] that imposes hierarchical sparse *correlation matrices* over multiple lags. For the problem of interest here, we introduce similar hierarchical sparsity constraints to the TV domain. To describe the hierarchical group structure, we define the following variables

$$\mathbf{X}_l = [\mathbf{x}_1, \cdots, \mathbf{x}_l] \in \mathbb{R}^{N \times l} \tag{12}$$

$$\mathbf{D}(\mathbf{X}_l) = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_l] \in \mathbb{R}^{N \times l} \tag{13}$$

where $\mathbf{d}_l = \text{TV}(\mathbf{x}_l) = [\text{TV}(\mathbf{x}_l(1)), \cdots, \text{TV}(\mathbf{x}_l(N))]^T$ denotes the (isotropic or anisotropic) TV vector of $\mathbf{x}_l$ at the l-th layer. Essentially, $\mathbf{D}(\mathbf{X}_l)$ groups the first l TV vectors for the first l layers. Note that $\mathbf{D}_1 = \mathbf{d}_1$. Then the hierarchical group TV minimization over multiple layers is defined as follows

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \sum_{l} \frac{1}{2} \|\mathbf{y}_l - \mathbf{A}_l \mathbf{x}_l\|_2^2 + \lambda \|\mathbf{X}\|_{\text{HGTV}} \tag{14}$$

where

$$\|\mathbf{X}\|_{\text{HGTV}} = \sum_{n=1}^{N} \sum_{l=1}^{L} \|\mathbf{D}(\mathbf{X}_l)\|_2 \tag{15}$$

with each $\|\mathbf{D}(\mathbf{X}_l)\|_2$ imposed for attaining the group structure from the first layer to the l-th layer. We can rewrite

$$\|\mathbf{X}\|_{\text{HGTV}} = \sum_{n=1}^{N} \|\mathbf{D}(\mathbf{X}_L)\|_2 + \sum_{n=1}^{N} \sum_{l=1}^{L-1} \|\mathbf{D}(\mathbf{X}_l)\|_2 \tag{16}$$

which shows the first item on the right is the same as the penalty term in (6) or (7) used in the group TV minimization. Then additional penalty terms $\sum_{n=1}^{N} \sum_{l=1}^{L-1} \|\mathbf{D}(\mathbf{X}_l)\|_2$ enforce extra sparsity constraints when the layer index l becomes smaller. As a result, it encourages more sparse solutions at the front layers than the deep layers and, therefore, gives a solution with the nested group sparse pattern in the TV domain.

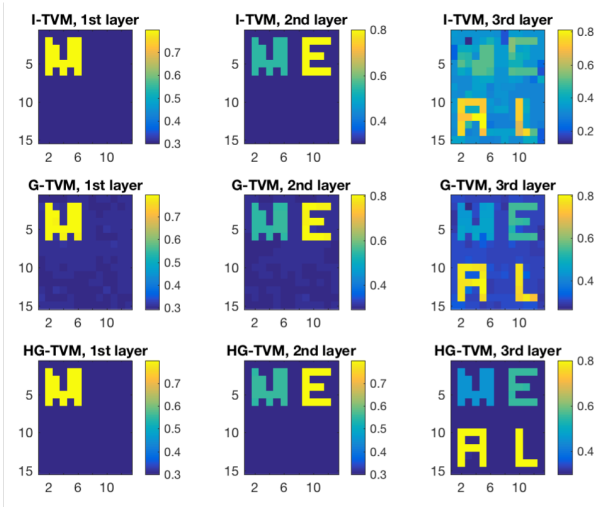

Fig. 4. The recovered four letters along with shadow letters in $L = 3$ layers with proposed approaches when $M/N = 0.5$.

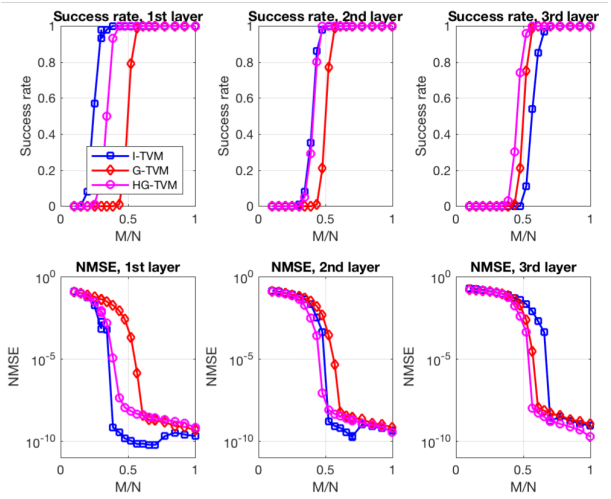

Fig. 5. Performance comparison in terms of the success rate (top row) and NMSE (bottom row) as a function of the compression ratio M/N for all three layers.

4. NUMERICAL RESULTS

In this section, numerical results are provided to evaluate all three proposed approaches in terms of the success rate and normalized mean squared error (NMSE) as a function of the compression ratio M/N. Denote the three approaches, sequentially, as 1) I-TVM, 2) G-TVM and 3) HG-TVM. To obtain the results, we run all three approaches with several choices of regularization parameter λ over a broad interval and report the best result with the regularization parameter giving the smallest the estimation error with respect to the ground truth.

Fig. 4 shows the recovery performance of $L = 3$ layers with four non-overlapping letters when the number of measurements is half of the number of pixels in each layer, i.e., $M/N = 0.5$. The first row shows the performance of the individual TV minimization (I-TVM) in Section. 3.1. One can observe good performance at the first two layers along the shadow letter of 'M' in the second layer. However, for the third layer, the shadow letters 'M' and 'E' are not recovered and the two letters 'A' and 'L' are visible with significant background artifacts. As explained previously, with both shadow letters from front layers and new letters in the third layer, the number of non-zero gradients in the third layer is significantly increased, which in turns requires more measurements for improved performance at deep layers. The second row gives the performance of the group TV minimization (G-TVM) approach with the same number of measurements. Compared with the I-TVM approach, the G-TVM shows improved performance in the third layer by imposing layer-wise group constraints. Nevertheless, layer-wise group constraints introduce visible background artifacts in the first and second layers. The best performance across all three layers is obtained by the hierarchical group TV minimization (HG-TVM) approach shown in the bottom row where all four letters and shadow letters are fully recovered with the minimum artifacts.

The performance is further quantified in terms of the success rate and NMSE by using the Monte-Carlo simulation. The NMSE for each layer is defined as $\|\hat{\mathbf{x}}_l - \mathbf{x}_l\|_2^2/\|\mathbf{x}_l\|_2^2$ and the Monte-Carlo trail is considered to be a success if the layer-wise NMSE is less than a threshold, i.e., 10^{-5} in our simulation. In each Monte-Carlo run, the measurement matrix $\mathbf{A}_l$ is generated as independent Gaussian matrices with zero mean and unit variance. Fig. 5 shows the two performance metrics for the proposed approaches over all three layers. For the I-TVM, the success rate curve (denoted as blue curves with dots) moves towards higher compression ratios from the first layer to the third layer. This observation is in line with the qualitative results shown in Fig. 4, where the deeper layers with more shadow letters requires higher compression ratios and larger number of measurements. The G-TVM (denoted as red curves with diamonds) shows the worst performance for the first two layers as more background artifacts are introduced due to the group structure constraints. At the third layer, the G-TVM is better than the I-TVM. Finally, the HG-TVM shows worse performance than the I-TVM at the first layer but the best performance over the second and third layers.

5. CONCLUSION

This paper investigated THz imaging of multi-layer non-overlapping contents and, particularly, utilized the shadow effect to improve the recovery performance for deep layers with increased number of shadow letters from front layers. Our approach to the problem of interest was built on the total variation minimization principle and exploited the hierarchical (nested) group structure in the total variation domain over layers. The proposed hierarchical group total variation minimization approach achieved the best trade-off performance across all three layers in the considered examples.

6. REFERENCES

[1] N. Horiuchi, "Terahertz technology: Endless applications," *Nature Photonics*, vol. 4, no. 4, pp. 140, Sept. 2010.

[2] E. P. Parrott, S. M. Sy, T. Blu, V. P. Wallace, and E. Pickwell-MacPherson, "Terahertz pulsed imaging in vivo: measurements and processing methods," *Journal of Biomedical Optics*, vol. 16, no. 10, pp. 106010–1–106010–8, 2011.

[3] G. C. Walker, J. W. Bowen, J. Labaune, J-B. Jackson, S. Hadjiloucas, J. Roberts, G. Mourou, and M. Menu, "Terahertz

deconvolution," *Optics Express*, vol. 20, no. 25, pp. 27230–27241, Dec. 2012.

[4] N. Sunaguchi, Y. Sasaki, N. Maikusa, M. Kawai, T. Yuasa, and C. Otani, "Depth-resolving THz imaging with tomosynthesis," *Optics Express*, vol. 17, no. 12, pp. 9558–9570, June 2009.

[5] A. Redo-Sanchez, B. Heshmat, A. Aghasi, S. Naqvi, M. Zhang, J. Romberg, and R. Raskar, "Terahertz time-gated spectral imaging for content extraction through layered structures," *Nature Communications*, vol. 7, pp. 1–7, Sept. 2016.

[6] A. Aghasi, B. Heshmat, A. Redo-Sanchez, J. Romberg, and R. Raskar, "Sweep distortion removal from terahertz images via blind demodulation," *Optica*, vol. 3, no. 7, pp. 754–762, July 2016.

[7] W.L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Applied Physics Letters*, vol. 93, no. 121105, Sept. 2008.

[8] C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla, "Terahertz compressive imaging with metamaterial spatial light modulators," *Nature Photonics*, vol. 8, no. 8, pp. 605–609, Aug. 2014.

[9] L. I. Rudin, S.Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

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

[11] P. Zhao, G. Rocha, and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *The Annals of Statistics*, pp. 3468–3497, 2009.

[12] W. Nicholson, J. Bien, and D. Matteson, "High dimensional forecasting via interpretable vector autoregression," *ArXiv e-prints*, 2016.