https://doi.org/10.2352/ISSN.2470-1173.2020.14.COIMG-006

A Splitting-Based Iterative Algorithm for GPU-Accelerated Statistical Dual-Energy X-Ray CT Reconstruction

Fangda Li, Ankit Manerikar, Tanmay Prakash, Avinash C. Kak; School of Electrical and Computer Engineering, Purdue University; West Lafayette, Indiana, USA. [1]

Abstract

When dealing with material classification in baggage at airports, Dual-Energy Computed Tomography (DECT) allows characterization of any given material with coefficients based on two attenuative effects: Compton scattering and photoelectric absorption. However, straightforward projection-domain decomposition methods for this characterization often yield poor reconstructions due to the high dynamic range of material properties encountered in an actual luggage scan. Hence, for better reconstruction quality under a timing constraint, we propose a splitting-based, GPU-accelerated, statistical DECT reconstruction algorithm. Compared to prior art, our main contribution lies in the significant acceleration made possible by separating reconstruction and decomposition within an Alternating Direction Method of Multipliers (ADMM) framework. Experimental results, on both synthetic and real-world baggage phantoms, demonstrate a significant reduction in time required for convergence.

Introduction

X-ray Computed Tomography (CT) is a widely deployed technique for threat detection in baggage at airport security checkpoints. However, using a single X-ray spectrum limits its reconstruction to only that of LAC (linear attenuation coefficient) or HU (Hounsfield Unit) images which are, at best, an approximation to the underlying energy-dependent characteristics of the materials. While LAC might suffice for discriminating between different types of tissues in medical imaging, in the adversarial case of airport baggage scanning, materials used commonly for homemade explosives can possess LAC values that are nearly the same as for benign materials. As a result, the technique of DECT, where projections from two X-ray spectra are collected simultaneously at each angle, has been proposed for material discrimination because it allows for the material property to be recovered in an additional dimension by using an energy-dependent attenuation model.

The most commonly used DECT model represents X-ray attenuation as the combined effect of Compton scattering and photoelectric absorption, which can be written as:

$$\mu(E,x) = x_c f_{\mathrm{KN}}(E) + x_p f_p(E), \tag{1}$$

where $f_{\mathrm{KN}}(E)$ and $f_p(E)$ denote the energy-dependent multipliers[2] to the Compton coefficient x_c and the photoelectric (PE) coefficient x_p, respectively. Therefore, the goal of DECT is to recover both the Compton and the PE coefficients of the object using the projections $m_h, m_l \in \mathbb{R}^M$ measured at two different X-ray spectra. This amounts to solving simultaneously for $x_c, x_p \in \mathbb{R}^N$ from the following two equations that can be shown compactly by:

$$m_{h/l} = -\ln \int S_{h/l}(E) e^{-R\mu(E)} dE + \ln \int S_{h/l}(E) dE, \tag{2}$$

where R denotes the forward projection matrix, and $S_h(E)$ and $S_l(E)$ denote, respectively, the photon distribution across energy levels in the high- or low-energy spectra.

Solving the two equations shown above is made challenging by the fact the two phenomena that contribute to X-ray attenuation — Compton and PE, occur at grossly different scales. At most applicable energy levels, Compton scattering is the dominant contributor to attenuation and this disparity between the two phenomena becomes worse with increasing energy since $f_p(E)$ has a cubic decay with E. As a result, stable recovery of PE coefficients requires more sophisticated inversion algorithms, such as those described in [2, 3, 4, 5].

Unfortunately, the algorithms cited above tend to be iterative and, with run-of-the-mill computing hardware, take a long time to return the results. Therefore, a straightforward implementation of these algorithms is not appropriate for the end-goals that motivate our research — high-throughput baggage screening at airports. The focus of the concepts presented in this paper is on *improving the computational efficiency* of an ADMM-based statistical inversion algorithm.

Related Work

DECT involves two key steps: dual-energy decomposition and tomographic reconstruction. The two tasks can be done either sequentially, as in projection-wise decomposition, or in a unified step using iterative statistical approaches.

Projection-wise Decomposition: One of the earliest approaches for DECT decomposition, the Constrained Decomposition Method (CDM) [1] involves directly decomposing the dual energy projections to Compton and PE line integrals followed by the FBP reconstructions of the two. In [6], CDM was also extended to operate for Multi-Energy CT. A major disadvantage of this method is that it guards itself poorly against artifacts, especially in PE coefficients but it is still a preferred approach as it enables parallel implementation.

Iterative Statistical Approaches: The statistical methods for DECT solve for the MAP estimates, finding the Compton and PE coefficients that best correspond to the measurements and any prior knowledge. The literature on Multi-Energy CT has focused largely on designing the models and the priors that best leverage the structural similarity across bases. In [4], Compton/PE images are reconstructed on a set of exponential basis functions with an

[1]This research was funded by the BAA 17-03 AATR contract with Department of Homeland Security's Science and Technology Directorate.

[2]See [1] for a detailed formulation of $f_{\mathrm{KN}}(E)$ and $f_p(E)$.

edge-correlation penalty term. This idea of encouraging geometric similarity between the more stable Compton image and the PE image is further explored in [5]. For this purpose, the authors have proposed a new Non-Local Mean (NLM) regularizer on the PE image and laid out an ADMM formation that scaled up to a problem size of practical interest. By treating the energy level as a third dimension, the contributions in [2, 3] adopt a tensor-based model for reconstruction using sparse priors.

Despite being able to produce high-quality DECT, statistical reconstructions generally fail drastically with regard to the timing constraints of practical applications. Compared to LAC reconstructions, DECT has to deal with the added computational burden associated with the decomposition step. As a result, in existing approaches, solving decomposition and reconstruction in one step becomes inefficient due to the combined complexity and high-dimensionality of the problem.

We therefore propose a statistical DECT approach that combines the best of the projection-wise decomposition methods and the iterative statistical methods. More specifically, we employ a splitting-based MAP estimation method embedded in an ADMM framework. As we will show in Section , our new splitting scheme not only provides a better convergence rate, but also allows for powerful hardware-enabled acceleration.

Proposed Method

Problem Formulation

To describe our proposed method, we first define the forward model for X-ray attenuation, i.e., the nonlinear transform $f(\cdot)$ from Compton/PE coefficients to logarithmic projections:

$$f_{h/l}(\widetilde{R}x) = C_{h/l} - \ln \int S_{h/l} e^{-f_{\mathrm{KN}} R x_c - f_p R x_p} dE, \tag{3}$$

where $C_{h/l}$ are constants, $x = \left[x_c^T; x_p^T\right]^T$ and $\widetilde{R} = \left[R; R\right]$. Given a pair of line-integral measurements $m = \left[m_h^T; m_l^T\right]^T$ corrupted by Poisson photon noise and possibly other artifacts, we construct a MAP estimate with the Total Variation (TV) term and the non-negativity prior:

$$\widehat{x}_{\mathrm{MAP}} = \underset{x}{\operatorname{argmin}} \frac{1}{2} \left\| f(\widetilde{R}x) - m \right\|_{\Sigma}^2 + \lambda |x|_{\mathrm{TV}} + g(x), \tag{4}$$

where Σ is a diagonal matrix with photon counts as trace elements as proposed in [7], $g(x_i) = \begin{cases} 0, x_i \in \mathbb{R}^+ \\ \infty, x_i \notin \mathbb{R}^+ \end{cases}$ and λ is the regularization parameter. While the unconstrained optimization in (4) is highly inefficient to solve directly, the ADMM method provides a flexible splitting-based framework for both optimality and convergence [5].

Formulation for ADMM

ADMM begins with the conversion of (4) to its constrained equivalent. By introducing two new auxiliary variables y and z for the TV and the non-negativity terms, respectively, and by posing x as the primal variable, the MAP minimization can be transformed into:

$$\begin{aligned} \widehat{x}_{\mathrm{MAP}} = &\underset{x,y,z}{\operatorname{argmin}} \frac{1}{2} \left\| f(\widetilde{R}x) - m \right\|_{\Sigma}^2 + \lambda |y| + g(z), \\ &\text{s.t.} \begin{bmatrix} y_{c/p} \\ z_{c/p} \end{bmatrix} = \begin{bmatrix} D \\ I \end{bmatrix} x_{c/p} = C x_{c/p}, \end{aligned} \tag{5}$$

where D denotes the finite difference operator. Subsequently, the corresponding Augmented Lagrangian (AL) can be written as:

$$\begin{aligned} L(x,y,z) = &\frac{1}{2} \left\| f(\widetilde{R}x) - m \right\|_{\Sigma}^2 + \lambda |y| + g(z) \\ &+ \sum_{\beta \in \{c,p\}} \frac{\rho_\beta}{2} \left\| \begin{bmatrix} y_\beta \\ z_\beta \end{bmatrix} - C x_\beta + u_\beta \right\|^2, \end{aligned} \tag{6}$$

where u denotes the dual variable, or the scaled Lagrangian multiplier, and ρ denotes the penalty parameter. ADMM splits the minimization of AL into subproblems that are solved separately in an iterative framework. One such intuitive splitting scheme proposed in [5] is for the iterative updates to be carried out according to (7) through (11) shown below. First, the primal variables are updated using:

$$\begin{aligned} \widehat{x}_c = &\underset{x_c}{\operatorname{argmin}} \frac{1}{2} \left\| f(R x_c, R x_p) - m \right\|_{\Sigma}^2 \\ &+ \frac{\rho_c}{2} \left\| \begin{bmatrix} y_c \\ z_c \end{bmatrix} - C x_c + u_c \right\|^2, \end{aligned} \tag{7}$$

$$\begin{aligned} \widehat{x}_p = &\underset{x_p}{\operatorname{argmin}} \frac{1}{2} \left\| f(R \widehat{x}_c, R x_p) - m \right\|_{\Sigma}^2 \\ &+ \frac{\rho_p}{2} \left\| \begin{bmatrix} y_p \\ z_p \end{bmatrix} - C x_p + u_p \right\|^2. \end{aligned} \tag{8}$$

Note that the update to x_p is made subsequent to $\widehat{x}_c$ since we can expect $\widehat{x}_c$ to stabilize the recovery of the more noise-prone x_p. Secondly, we update the auxiliary variables y and z corresponding to the TV and non-negativity terms, respectively, by using

$$\widehat{y}_{c/p} = \mathit{shrinkage}(D \widehat{x}_{c/p} - u^y_{c/p}, \frac{\lambda_{c/p}}{\rho_{c/p}}), \tag{9}$$

$$\widehat{z}_{c/p} = \max(0, \widehat{x}_{c/p} - u^z_{c/p}), \tag{10}$$

where the shrinkage function is defined in [8, Eq. 9.35]. Then the dual variable u is updated with

$$\widehat{u}_{c/p} = u_{c/p} + \begin{bmatrix} \widehat{y}_{c/p} \\ \widehat{z}_{c/p} \end{bmatrix} - C \widehat{x}_{c/p}. \tag{11}$$

Lastly, we update ρ adaptively using a method described in [9] that is based on the primal and dual residual. While (9)-(11) can be realized by straightforward element-wise operations, solving (7)-(8) requires a nonlinear least squares algorithm such as Levenberg-Marquardt (LM). Despite its robustness, LM, is a sequential algorithm, offering little room for parallelization. Intuitively, the overall computational inefficiency can be attributed to (7) or (8) which at once addresses two problems of very different nature: nonlinear dual-energy decomposition and regular linear tomographic reconstruction.

Proposed Splitting Scheme

For the new sped-up implementation presented in this paper, we further decompose (7)-(8) into two simpler subproblems: tomographic reconstruction followed by dual-energy decomposition. This is achieved by viewing measurement fitting as an additional constraint: $a = \widetilde{R}x$, where a is a new auxiliary variable for that purpose. Incorporating this constraint with the others, for the first subproblem we pose tomographic reconstruction as an unweighted unconstrained least squares problem:

$$\widehat{x}_{c/p} = \underset{x_{c/p}}{\operatorname{argmin}} \frac{\rho_{c/p}}{2} \left\| \begin{bmatrix} a_{c/p} \\ y_{c/p} \\ z_{c/p} \end{bmatrix} - \begin{bmatrix} R \\ C \end{bmatrix} x_{c/p} + \begin{bmatrix} u^a_{c/p} \\ u^y_{c/p} \\ u^z_{c/p} \end{bmatrix} \right\|^2 . \quad (12)$$

and we can now separately deal with the decomposition, i.e. solving for a, in the following subproblem. We have chosen to use a CG solver for the minimization shown above because our system is huge and sparse. Additionally, since the projection-wise weights Σ are now absorbed by the decomposition step to be described later, the linear system corresponding to (12) is shift-invariant. Therefore, for improved convergence rate, each CG iteration lends itself well to preconditioning using a high-passing filter [10]. In our implementation, we experimented with the ramp filter as the preconditioner.

The second subproblem, that of dual-energy decomposition, can be solved by finding the pair of Compton/PE coefficients that minimizes the cost function at each ray projection independently. Therefore, the decomposition is achieved by using the Unconstrained Decomposition Method (UDM):

$$\begin{aligned} \begin{bmatrix} \widehat{a}_c \\ \widehat{a}_p \end{bmatrix} = \underset{a_c, a_p}{\operatorname{argmin}} & \frac{1}{2} \left\| f\left(\begin{bmatrix} a_c \\ a_p \end{bmatrix} \right) - m \right\|^2_{\Sigma} \\ & + \sum_{\beta \in \{c,p\}} \frac{\rho_\beta}{2} \left\| a_\beta - R\widehat{x}_\beta + u^a_\beta \right\|^2 . \end{aligned} \quad (13)$$

Compared to (7) and (8), the update formulas in (12) and (13) have the following advantages: First, treating reconstruction and decomposition separately allows us to use two entirely different algorithms, with each tailored for the corresponding subproblem, which in our case are Preconditioned CG (PCG) and UDM. As an additional benefit, while LM is still used in UDM, backtracking of the damping parameter, which is necessary for dealing with non-linear optimization, no longer involves expensive tomographic operations. Lastly, computational efficiency is improved in terms of both the number of operations required, as listed in Table 1, and parallelization. We can now not only independently execute (12) for both bases, but also leverage massively parallelized hardware such as GPUs to solve (13).

Table 1. *Minimum number of operations per ADMM iteration. In* ***LM(m)×CG(n)****, n CG iterations are taken to compute the update per LM iteration. Note that LM may require more operations than listed to tune the damping parameter.*

	R	R^T	$f(\cdot)$
LM(m)×CG(n)	$2m(n+1)$	$2m(n+1)$	$4m$
UDM(m)-PCG(n)	$2n$	$2(n+1)$	$4m$

Experimental Results

We now present qualitative and quantitative results on three phantoms: a simulated phantom (`sim18`), a real-world water bottle phantom (`Water`), and an actual baggage phantom (`Clutter`). We have implemented the following algorithms and evaluated them on these three phantoms: (i) **CDM-FBP**: CDM [1] with 16 CPU threads; (ii) **LM(m)×CG(n)**: ADMM as in [5]; and (iii) **UDM(m)-(P)CG(n)**: proposed ADMM method with m UDM iterations and n (P)CG iterations.

Our Python implementations use the Astra toolbox [11] for GPU-based calculations of forward and backward projections. Additionally, as a key factor for acceleration, we use Gpufit [12] to parallelize the decomposition step. All experiments are carried out on a single computing cluster node with 16 cores, 20GB RAM and an Nvidia 1080Ti GPU. Regarding initialization, in our experiment we have chosen the CDM output as the initial Compton estimate, since it is generally stable. For initial PE coefficients, we use a scaled version of the Compton estimate, similar to what is done in [5]. For both Compton and PE, we used $\lambda = 10^{-5}$ as the TV regularization parameter and $\rho = 10^{-3}$ as the initial penalty parameter. Furthermore, for scale-invariant evaluation, quantitative reconstruction quality is assessed with the normalized ℓ_2-distance between x and ground truth x^*:

$$\xi(x) = 20 \log_{10} \left(\frac{\|x - x^*\|_2}{\|x^*\|_2} \right). \quad (14)$$

The first phantom, `sim18`, contains seven circular regions of different materials. The reconstructed images, of size 512×512, are constructed from 720 angles, each containg 725 parallel ray projections. The first row in Figure 1 shows the reconstructions by **CDM-FBP** and the proposed **UDM-PCG** algorithm. In Figure 2a, we display the average computational time per iteration inside parentheses in the inset box and plot $\xi(x)$ versus iteration to compare the convergence rates. In general, the ADMM algorithm with the proposed splitting scheme results in significantly shorter total execution time than **LM×CG** to reach the same error.

For results on the two real-world phantoms, the parallel beam projection data for the two was collected on the Imatron C300 CT scanner. The X-ray source emits 1.8×10^5 and 1.7×10^5 photons per ray with two energy spectra at 95keV and 130keV, respectively. The high- and the low-energy sinograms are subsampled by two, resulting in 360 angles for each and with 512 bins for each angle. The reconstruction results for the `Water` phantom are shown in the second row of Figure 1. For quantitative evaluation, in Figure 2b we plot $\xi(x)$ versus iteration within the ROI – the central circular region that is occupied by distilled water. Our proposed **UDM-PCG** algorithm is significantly more efficient not only in terms of the number of the iterations needed, but also in terms of the time per iteration as compared to **LM×CG**.

The real-world baggage phantom `Clutter` is particularly challenging because it contains metallic objects, and illustrates well the need for a statistical reconstruction. As shown in the third row of Figure 1, while the **CDM-FBP** PE reconstruction is completely overshadowed by streaking artifacts, **UDM1-PCG1** is able to recover object shapes to a reasonable degree.

Conclusions and Future Work

In this paper, we have proposed a new splitting scheme for implementing ADMM to reconstruct Compton and PE coefficient images using dual-energy projection data. By separating the the reconstruction and decomposition steps, the proposed GPU-accelerated ADMM algorithm achieves a significant speedup in time when compared to the prior state-of-the-art. Future work will be aimed at generalization to 3D reconstruction and improving initialization heuristics.

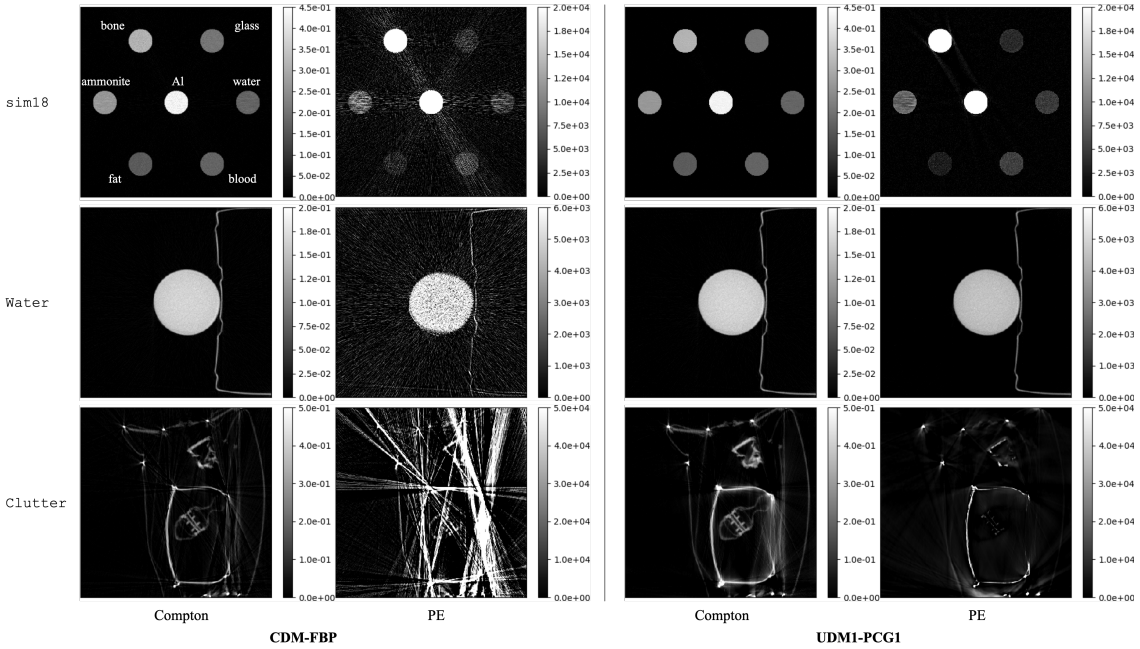


Figure 1. *Compton and PE reconstructions by **CDM-FBP** and **UDM1-PCG1**. Note that we clipped the values in PE images of `sim18` to show the streaking artifacts in the **CDM-FBP** reconstruction. Such artifacts, appearing also in the **CDM-FBP** reconstructions of `Water` and `Clutter`, are significantly suppressed by the **UDM1-PCG1** method.*

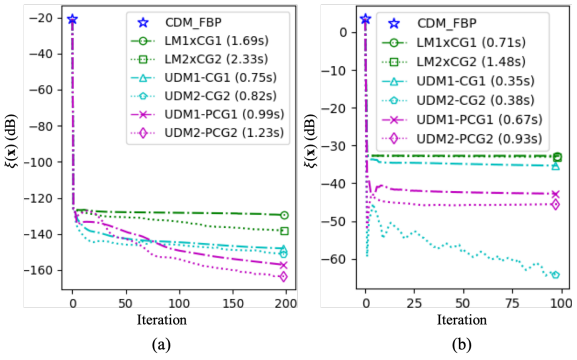


Figure 2. *$\xi(x)$ vs. iteration for PE and average seconds per iteration for (a) `sim18` and (b) `Water`. The same plots for Compton are omitted since they exhibit similar trends.*

References

[1] Z. Ying, R. Naidu and C.R. Crawford, Dual energy computed tomography for explosive detection, *Journal of X-ray Science and Technology* **14**(4) (2006), 235–256.

[2] O. Semerci, N. Hao, M.E. Kilmer and E.L. Miller, Tensor-based formulation and nuclear norm regularization for multienergy computed tomography, *IEEE Transactions on Image Processing* **23**(4) (2014), 1678–1693.

[3] Y. Zhang, X. Mou, G. Wang and H. Yu, Tensor-based dictionary learning for spectral CT reconstruction, *IEEE transactions on medical imaging* **36**(1) (2017), 142–154.

[4] O. Semerci and E.L. Miller, A parametric level-set approach to simultaneous object identification and background reconstruction for dual-energy computed tomography, *IEEE transactions on image processing* **21**(5) (2012), 2719–2734.

[5] B.H. Tracey and E.L. Miller, Stabilizing dual-energy x-ray computed tomography reconstructions using patch-based regularization, *Inverse Problems* **31**(10) (2015), 105004.

[6] Y. Yuan, B. Tracey and E. Miller, Robust x-ray based material identification using multi-energy sinogram decomposition, in: *Anomaly Detection and Imaging with X-Rays (ADIX)*, Vol. 9847, International Society for Optics and Photonics, 2016, p. 98470V.

[7] C.A. Bouman and K. Sauer, A unified approach to statistical tomography using coordinate descent optimization, *IEEE Transactions on image processing* **5**(3) (1996), 480–492.

[8] C.A. Bouman, Model based image processing (2013).

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine learning* **3**(1) (2011), 1–122.

[10] J.A. Fessler and S.D. Booth, Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction, *IEEE transactions on image processing* **8**(5) (1999), 688–699.

[11] W. van Aarle, W.J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K.J. Batenburg and J. Sijbers, Fast and flexible X-ray tomography using the ASTRA toolbox, *Optics express* **24**(22) (2016), 25129–25147.

[12] A. Przybylski, B. Thiel, J. Keller-Findeisen, B. Stock and M. Bates, Gpufit: An open-source toolkit for GPU-accelerated curve fitting, *Scientific reports* **7**(1) (2017), 15722.

Author Biography

Fangda Li is a Ph.D. student working with Prof. Avinash Kak at the Robot Vision Lab in the School of Electrical and Computer Engineering at Purdue University, where he also received his B.Sc. degree in Electrical Engineering in 2015 and M.Sc. degree in Computer Engineering in 2017. His research interests span computed tomography, computer vision, machine learning, robotics and visualization.

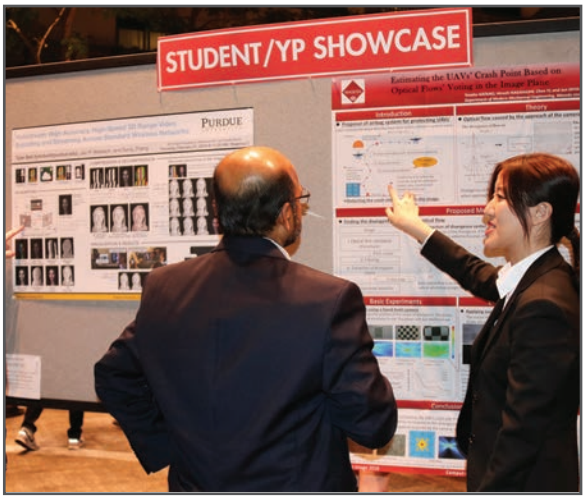
JOIN US AT THE NEXT EI!

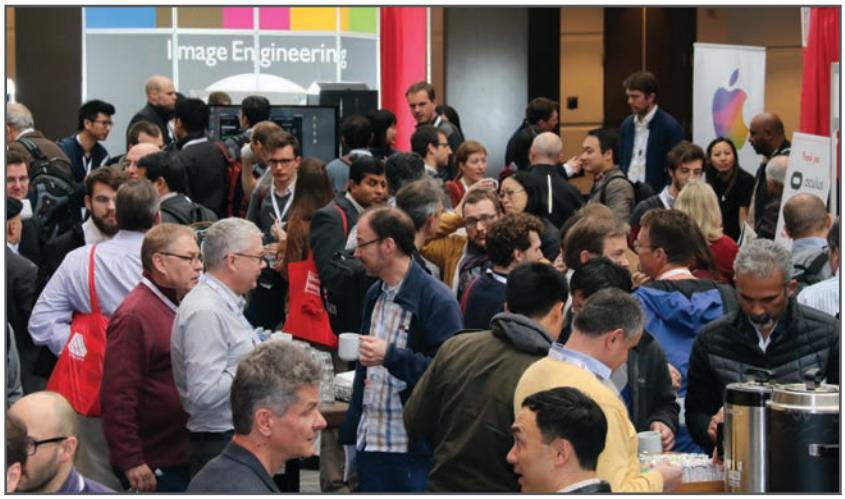
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org