Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Computed tomography imaging system design for shape threat detection

Ahmad Masoudi
Ratchaneekorn Thamvichai
Mark A. Neifeld

SPIE.

Ahmad Masoudi, Ratchaneekorn Thamvichai, Mark A. Neifeld, "Computed tomography imaging system design for shape threat detection," *Opt. Eng.* **56**(4), 041308 (2016), doi: 10.1117/1.OE.56.4.041308.

Computed tomography imaging system design for shape threat detection

Ahmad Masoudi,^{a,*} Ratchaneekorn Thamvichai,^a and Mark A. Neifeld^{a,b}

^aUniversity of Arizona, Electrical and Computer Engineering Department, 1230 East Speedway Boulevard, Tucson, Arizona 85719, United States

^bUniversity of Arizona, College of Optical Sciences, 1630 East University Boulevard, Tucson, Arizona 85721, United States

Abstract. In the first part of this work, we present two methods for improving the shape-threat detection performance of x-ray computed tomography. Our work uses a fixed-gantry system employing 25 x-ray sources. We first utilize Kullback–Leibler divergence and Mahalanobis distance to determine the optimal single-source single-exposure measurement. The second method employs gradient search on Bhattacharyya bound on error rate (P_e) to determine an optimal multiplexed measurement that simultaneously utilizes all available sources in a single exposure. With limited total resources of 10^6 photons, the multiplexed measurement provides a $41.8\times$ reduction in P_e relative to the single-source measurement. In the second part, we consider multiple exposures and develop an adaptive measurement strategy for x-ray threat detection. Using the adaptive strategy, we design the next measurement based on information retrieved from previous measurements. We determine both optimal “next measurement” and stopping criterion to insure a target P_e using sequential hypothesis testing framework. With adaptive single-source measurements, we can reduce P_e by a factor of $40\times$ relative to the measurements employing all sources in sequence. We also observe that there is a trade-off between measurement SNR and number of detectors when we study the performance of systems with reduced detector numbers. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.OE.56.4.041308]

Keywords: computational imaging; x-ray; detection; computed tomography.

Paper 161333SS received Aug. 27, 2016; accepted for publication Nov. 15, 2016; published online Dec. 8, 2016.

1 Introduction

X-ray computed tomography (CT) is a powerful imaging methodology because it is capable of providing (1) high spatial resolution, (2) material specificity, and (3) three-dimensional information. These properties make x-ray CT popular in a wide variety of medical, security, and industrial applications. X-ray CT has been used for airport baggage screening purposes since December 1994 when InVision Technologies introduced CTX 5000, an explosive detection system (EDS) machine based on x-ray CT.¹ The important performance requirements of an EDS machine are high probability of detection of specific types and masses of explosive materials and low false alarm rates.² Today there are a large number of transportation security laboratory certified EDS machines in use throughout the country.³ In order to detect potential threats, all of these EDS machines employ tomographic measurement followed by image reconstruction and pattern recognition algorithms. We refer to this approach as a “conventional system.” Conventional systems employ a “rotating gantry” to produce a collection of projection angles through the object via physical motion. An x-ray source and a detector array are rotated around the object to produce a diverse collection of tomographic slices. For example, ~ 1000 projections at different angles are needed⁴ in order to reconstruct a two-dimensional (2-D) image of size 1000×1000 pixels by a system with a detector array of size 1024. Since this approach can be slow due to its physical motion requirement, an alternative “rectangular fixed gantry” (RFG) system employs 25 x-ray sources arranged around the object with

associated detector arrays to collect the transmitted photons.⁵ Measurement diversity is now obtained via the sequence of sources used to illuminate the object. Note that a fixed-gantry system can offer additional flexibility in terms of nonstandard sequencing of x-ray sources and/or opportunities for illuminating with multiple sources simultaneously. Owing to the speed and flexibility of this approach, we have based our study on the fixed-gantry architecture. To establish a baseline performance, we study the detection performance of our rectangular fixed-gantry system when it collects measurement employing using all sources in sequence. We refer to this approach as RFG approach.

The detection of threat objects in a piece of luggage using either the conventional system or the fixed-gantry system typically consists of three steps: image reconstruction, feature extraction, and classification.⁶ In this work, we will show that bypassing the image reconstruction and other image processing steps and performing the classification directly on projections can increase the system classification performance.

Our goal will be to design a measurement strategy that minimizes the classification error rate between probability distributions of two classes which are threat bags and nonthreat bags. We will model each class using a Gaussian mixture model (GMM) probability density function (PDF). The system that minimizes the similarity between two probability distributions yields the minimum misclassification rate. The most common metric to measure PDF similarity is Kullback–Leibler (KL) divergence. The method by Jensen et al.⁷ computes the KL divergence between two GMMs

*Address all correspondence to: Ahmad Masoudi, E-mail: masoudi@email.arizona.edu

using a numerical integration which requires a large number of samples in order to get accurate value. Stochastic integration can provide a more accurate result than simple numerical integration, but still requires a large number of samples when the dimension is large.⁸ The computation of KL divergence using a matching-based approximation and unscented transform approximation is proposed, but it also requires numerical integration.⁹ All of the aforementioned approaches employ sampling to compute KL divergence which causes the computation time to greatly increase when dimensionality of data increases. Another approach to measure the similarity between two PDFs is the Cauchy–Schwartz (CS) divergence. It is shown that there is a closed analytical form of CS divergence for two GMMs¹⁰ and CS divergence outperforms KL divergence based on numerical and stochastic integration in real object classification. In this work, we employ two methods to design optimal static measurement. In the first method, we use a lower-bound on KL divergence which has closed analytical form and in the second method instead of measuring the similarity directly, we use Bhattacharyya distance to determine an upper-bound for classification error because of its closed analytical form for GMMs. We design a system such that it minimizes the error upper-bound.

Before studying our main problem, we establish baseline performance in the next three subsections. In Sec. 1.1 we describe the physical system model of a rectangular fixed-gantry x-ray CT that will be employed throughout the paper, Sec. 1.2 presents an RFG system which employs reconstruction, feature extraction, and classification, and Sec. 1.3 demonstrates that it is possible to decrease the classification error rate via implementing a detection algorithm directly on the measurement vector (vector of projections).

1.1 System Model

We study the 2-D fixed-gantry x-ray CT system model shown in Fig. 1. This rectangular fixed-gantry system consists of 25 sources and a U-shaped photon-counting array containing 2200 detectors. Each source generates a 45-deg fan-beam and a broad photon energy spectrum that is typical of a tungsten anode ranging from 20 to 150 keV. The tunnel

Fig. 1 Fixed-gantry x-ray CT system model.

size is consistent with current systems and provides sufficient space for a 50 cm × 50 cm piece of luggage (i.e., the object is a bag containing various threat and/or nonthreat items). The objects used in our studies are 500 × 500 pixels synthetic bags generated by a stochastic bag generator (SBG) described further in Sec. 1.3.¹¹

In order to quantify the operation of the system shown in Fig. 1, we first consider the case in which only one source, s 'th source, is turned on. Based on Beer's law, we can compute the mean number of photons at the detectors by $\bar{\mathbf{r}} = I_s e^{-\mathbf{H}_s \mathbf{f}}$ where the vector $\mathbf{r}$ represents the random signal vector whose components indicate the number of photons collected at each detector in the detector array, we represent the mean of $\mathbf{r}$ by $\bar{\mathbf{r}}$, I_s is the number of photons generated by the s 'th source, $\mathbf{H}_s$ is the projection matrix for the s 'th source, and $\mathbf{f}$ is the bag attenuation coefficients which are lexicographically ordered into a vector such that if the bag has the size of $N \times N$ pixels then $\mathbf{f}$ is an $N^2 \times 1$ vector. For the case that a combination of sources are simultaneously turned on, the equation of $\bar{\mathbf{r}}$ becomes $\bar{\mathbf{r}} = \sum_{s=1}^{25} I_s e^{-\mathbf{H}_s \mathbf{f}}$. Since the dominant noise in the system is Poisson, then $\mathbf{r} \sim \text{Poiss}(\bar{\mathbf{r}})$. We can accurately approximate $\mathbf{r}$ by a Gaussian distribution $\mathbf{r} \sim N[\bar{\mathbf{r}}, \Lambda(\bar{\mathbf{r}})]$, if all the elements of $\bar{\mathbf{r}}$ are greater than 20. $\Lambda(\mathbf{x})$ is a diagonal matrix with vector $\mathbf{x}$ on the diagonal. In this work, we will use the Gaussian approximation to simplify our optimization problems.

1.2 Rectangular Fixed-Gantry Detection System

Our goal in this section is to briefly introduce the operation of an RFG system by use of a simple example. Suppose that we have two classes of bags: threat bags and nonthreat bags, and that each class contains only two bags. We can see that the threat bags have threat items defined to be a wrench and gun in Fig. 2(b). These two items are absent in nonthreat bags as shown in Fig. 2(a). For detecting the threats, the RFG system uses the steps shown in Fig. 3 which are (1) obtain projections (measurement data) by employing all 25 sources in sequence where each source uses an equal number of photons; (2) reconstruct the image using the algebraic reconstruction technique; (3) segment the possible threat from the reconstructed image; and (4) make a decision.

The notation introduced above captures any measurement consisting of a single exposure. In order to capture the outcome of a sequence of such measurements, we use $\mathbf{R}$ to represent the concatenation of a collection of signal vectors. An RFG system will use the same source brightness for each exposure and will, therefore, produce $\mathbf{R} = [\mathbf{r}_1 \dots \mathbf{r}_{25}]$ where $\mathbf{r}_s$ is the measurement vector or signal vector from s 'th source as explained in Sec. 1.1. We define the output of the segmentation step as $\mathbf{g} = \text{Seg}[\text{Recon}(\mathbf{R})]$, where $\text{Recon}()$ is the reconstruction function and $\text{Seg}()$ is the segmentation function in the RFG system. With each class containing two bags where each bag is modeled using Gaussian PDF, we use GMM with two Gaussian components/mixtures to model $\mathbf{g}$. In this simple case, we can use the optimal decision algorithm, the likelihood ratio test. Since there are two classes of bags, our problem has two hypotheses

$$H^c: \mathbf{g} = p_1^c * N(\boldsymbol{\mu}_1^c, \boldsymbol{\Sigma}_1^c) + p_2^c * N(\boldsymbol{\mu}_2^c, \boldsymbol{\Sigma}_2^c), \quad (1)$$

$$H^t: \mathbf{g} = p_1^t * N(\boldsymbol{\mu}_1^t, \boldsymbol{\Sigma}_1^t) + p_2^t * N(\boldsymbol{\mu}_2^t, \boldsymbol{\Sigma}_2^t), \quad (2)$$

Fig. 2 (a) Two nonthreat bags and (b) two bags containing shape threat items (gun and wrench).

Fig. 3 RFG system flowchart containing steps: (1) measurement, (2) reconstruction, (3) segmentation, and (4) decision making.

where superscripts c and t indicate that the variable belongs to nonthreat/clutter and threat classes, respectively. p_1 and p_2 are mixture coefficients/weights (i.e., each mixture has an equal weight of 0.5) and μ s and Σ s are means and covariance matrices (i.e., which we estimate by running a large number of simulation experiments) and $p_c = 0.5$ and $p_t = 0.5$ are prior probabilities for nonthreat and threat classes, respectively. The optimal decision rule is

$$\frac{Pr(\mathbf{g}|c)^{H^c \text{ or } H^t}}{Pr(\mathbf{g}|t)} \geq \frac{p_t}{p_c}. \quad (3)$$

The probability of error (P_e) or error rate is defined as follows:

$$P_e = p_t * Pr\{H_c|H_t\} + p_c * Pr\{H_t|H_c\},$$

where $Pr\{H_t|H_c\}$ is probability that we decide there is a threat while the bag contains no threat and $Pr\{H_c|H_t\}$ is probability that we miss the threat in the bag. Figure 4 shows the RFG system error rate (red solid curve) as a function of total number of photons used to generate $\mathbf{R}$. As we expected, the error rate decreases as the number of photons increases.

1.3 Classification in Measurement Space

In an RFG system, it is possible that we may lose information during the reconstruction and segmentation steps since they can be imperfect. One way to overcome this imperfection will be to undertake the classification step directly on the measurement data. Within this framework, our problem is modified to

$$H^c: \mathbf{R} = p_1^c * \text{Poiss}(\overline{\mathbf{R}}_1^c) + p_2^c * \text{Poiss}(\overline{\mathbf{R}}_2^c), \quad (4)$$

$$H^t: \mathbf{R} = p_1^t * \text{Poiss}(\overline{\mathbf{R}}_1^t) + p_2^t * \text{Poiss}(\overline{\mathbf{R}}_2^t), \quad (5)$$

where once again superscripts c and t indicate that each variable belongs to the nonthreat or threat classes, respectively, $\overline{\mathbf{R}}_1^c$ and $\overline{\mathbf{R}}_2^c$ represent the sequences of means associated with the signal vectors for the first and second nonthreat bags,

Fig. 4 Error rate of the RFG system with reconstruction and segmentation (red solid line) as a function of number of photons per measurement. Error rate of the RFG system without reconstruction and segmentation (blue dashed line) as a function of number of photons per measurement.

respectively, and $\overline{\mathbf{R}}_1^c$ and $\overline{\mathbf{R}}_2^c$ represent the sequences of means associated with signal vectors for two threat bags. Making one of these definitions more explicit, we have $\overline{\mathbf{R}}_1^c = [I_1 e^{-\mathbf{H}_1 \mathbf{f}_1^c}; \dots; I_{25} e^{-\mathbf{H}_{25} \mathbf{f}_1^c}]$ where I_s is number of photons at s 'th source, $\mathbf{H}_s$ is the projection matrix for s 'th source, and $\mathbf{f}_1^c$ is the first nonthreat bag. $\overline{\mathbf{R}}_2^c$, $\overline{\mathbf{R}}_1^t$, and $\overline{\mathbf{R}}_2^t$ can be defined similarly. Analogous to our earlier definitions p_1^c , p_2^c , p_1^t , and p_2^t are mixture weights of nonthreat and threat classes, respectively. With this "measurement space" notation we modify the likelihood ratio testing as well

$$\frac{Pr(\mathbf{R}|c)}{Pr(\mathbf{R}|t)} \stackrel{H^c \text{ or } H'^?}{\gtrless} \frac{p_t}{p_c}. \quad (6)$$

The performance of the measurement space classifier is shown via the blue dashed curve in Fig. 4. From this result, we can see that it is indeed possible to reduce the classification error rate by performing the classification directly in measurement space because information that is useful for classification may be lost since reconstruction and segmentation algorithms are not perfect. For example, when the total available photon budget is 10^5 photons, we observe an $1000\times$ reduction in P_e for this idealized example. Motivated by this example and the more general statement of the information processing inequality, all subsequent results in this paper will be based on classification directly in the measurement space.¹²

In the simple example presented above, each class contained only two bags. This was sufficient to demonstrate plausibility; however, henceforth we will generate statistical results based on a larger-scale problem. We will retain the two classes (i.e., threat and nonthreat), but each class now contains 100 bags generated by our SBG. The SBG is a software tool that mimics the packing of luggage to generate virtual bags of the type shown in Fig. 2. The SBG includes a large library of shapes and materials, any subset of which can be designated as threat or nonthreat. For the work reported here, we identify three threat shapes: wrench, gun, and knife, each of which can appear at a variety of positions, orientations, and scales. The remainder of the shape library is designated as nonthreat. The SBG generates a nonthreat bag by uniformly selecting at random a collection of nonthreat shapes from the library and placing them in a bag. Next, each shape is "filled" with one of materials from a library of materials. For this paper, only nonthreat materials are used since we consider only shape-based threats. Each nonthreat bag generated by the SBG is used to generate a companion threat bag. This is done by selecting one of the shapes from the nonthreat bag and replacing it with a threat shape without changing the associated fill material. Upon completion of this process we have 100 pairs of nonthreat and threat bags such that in each pair, the threat bag differs from the nonthreat bag by the shape of only one item. This process generates a very challenging classification problem that we can write as follows:

$$H^c: \mathbf{R} = \sum_{i=1}^{100} p_i^c \text{Poiss}(\overline{\mathbf{R}}_i^c), \quad (7)$$

$$H^t: \mathbf{R} = \sum_{i=1}^{100} p_i^t \text{Poiss}(\overline{\mathbf{R}}_i^t), \quad (8)$$

where $p_i^c = p_i^t = 1/100$ for $i = 1, \dots, 100$ and other parameters are defined as before. We continue to employ the optimal decision rule via the likelihood ratio test

$$\frac{Pr(\mathbf{R}|c)}{Pr(\mathbf{R}|t)} \stackrel{H^c \text{ or } H'^?}{\gtrless} \frac{p_t}{p_c}, \quad (9)$$

where p_t and p_c are class priors for threat class and nonthreat class, respectively, and they are both equal to 0.5. As a reminder, a single measurement is represented by lowercase $\mathbf{r}$ and sequence of measurements is represented by uppercase $\mathbf{R}$ throughout this paper and when we mention RFG system, we refer to our fixed-gantry x-ray CT using all 25 sources in sequence.

The remainder of this article is organized as follow. In Sec. 2, we consider nonadaptive tomographic measurements for shape-threat detection. First we find the optimal source for minimizing probability of error (P_e) in a single-exposure measurement and we then determine the optimal multiplexed measurement (i.e., using multiple sources simultaneously in a single exposure). The optimal multiplexing strategy is found by using a gradient search on the upper-bound on probability of error (P_e). Both optimal single source and optimal multiplexed measurements are nonadaptive because they are designed to achieve minimum P_e based on only a single exposure. In Sec. 3, we extend the aforementioned nonadaptive single-exposure systems to the adaptive multi-exposure systems by applying the sequential hypothesis testing (SHT) algorithm.^{13–16} By using SHT, instead of employing a single measurement with high signal-to-noise ratio (SNR) measurement, we can use a sequence of lower SNR measurements that benefit from accumulated knowledge. Section 4 studies the behavior of both adaptive and nonadaptive systems when the number of photon-counting detectors is reduced. Section 5 provides conclusion and possible future work.

2 Nonadaptive Systems

In the two sections that follow we will develop the optimal single-source measurement and the optimal multiplexed measurement systems, respectively.

2.1 Single-Source Solution

Considering the problem stated in Eqs. (7) and (8) with the modification that we make a single measurement, the signal vector $\mathbf{r}$ is, therefore, used in our hypotheses below:

$$H^c: \mathbf{r} = \sum_{i=1}^{100} p_i^c \text{Poiss}(\overline{\mathbf{r}}_i^c), \quad (10)$$

$$H^t: \mathbf{r} = \sum_{i=1}^{100} p_i^t \text{Poiss}(\overline{\mathbf{r}}_i^t). \quad (11)$$

The likelihood ratio test is given by

$$\frac{Pr(\mathbf{r}|c)}{Pr(\mathbf{r}|t)} \stackrel{H^c \text{ or } H'^?}{\gtrless} \frac{p_t}{p_c},$$

where, for example, $\overline{\mathbf{r}}_i^c = I_s e^{-\mathbf{H}_s \mathbf{f}_i^c}$ and s is the source position and I_s is the number of photons generated by the s 'th

source. The signal vector mean for each bag is, therefore, a function of s 'th source position. Note that in order to simplify the optimization problem, we use the Gaussian approximation described in Sec. 1.1 which converts Eqs. (10) and (11) into GMMs (instead of Poisson mixture models).

We are interested in determining the optimal source position which minimizes P_e . One possible approach is to estimate the P_e for all 25 source positions by Monte–Carlo simulation and select the one that yields the minimum. Unfortunately, for Monte–Carlo estimation to be accurate, we must run a very large number of simulations. This complexity cost is further exacerbated when the parameter of interest (i.e., P_e) is small. So instead of doing Monte–Carlo simulation, one may prefer to use other metrics such as symmetric KL divergence since the source that yields the maximum symmetric KL divergence is expected to result in minimum P_e . However, there is no closed analytical form for the KL divergence between two GMMs and it must be estimated by Monte–Carlo simulation which is expensive and time-consuming. To overcome the difficulty of computing KL divergence between GMMs, we instead compute the KL divergence between every threat/nonthreat bag-pair (i.e., 10^4 pairs in our ensemble of 100 bag pairs). Since each bag can be described by a Gaussian PDF, this KL metric is available in closed form. By selecting the source position that yields the maximum of the smallest KL divergence, we aim to minimize P_e . Using this approach, we compute a lower-bound for symmetrical KL divergence between two GMMs. Mathematically, the optimal source position is given as follows:

$$s_b^{kl} = \underset{s=1 \dots 25}{\operatorname{argmax}} \min_{i,j} \frac{1}{2} D_{kl} \{ N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)] \| N[\bar{\mathbf{r}}_j^t, \Lambda(\bar{\mathbf{r}}_j^t)] \} + \frac{1}{2} D_{kl} \{ N[\bar{\mathbf{r}}_j^t, \Lambda(\bar{\mathbf{r}}_j^t)] \| N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)] \}. \quad (12)$$

Considering $\bar{\mathbf{r}}_i^c = I_s e^{-H_s f_i^c}$ and $\bar{\mathbf{r}}_j^t = I_s e^{-H_s f_j^t}$, we can then rewrite Eq. (12) as Eq. (13),

$$s_b^{kl} = \underset{s=1 \dots 25}{\operatorname{argmax}} \min_{i,j} \frac{1}{2} D_{kl} \{ N[I_s e^{-H_s f_i^c}, \Lambda(I_s e^{-H_s f_i^c})] \| N[I_s e^{-H_s f_j^t}, \Lambda(I_s e^{-H_s f_j^t})] \} + \frac{1}{2} D_{kl} \{ N[I_s e^{-H_s f_j^t}, \Lambda(I_s e^{-H_s f_j^t})] \| N[I_s e^{-H_s f_i^c}, \Lambda(I_s e^{-H_s f_i^c})] \}, \quad (13)$$

where s_b^{kl} is the optimal source position, $i, j = 1, \dots, 100$ (number of bags) and D_{kl} stands for KL divergence. We know that if $p \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $q \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ then $D_{kl}(p \| q) = \frac{1}{2} (\operatorname{tr}(\boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) + \ln \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|}) + K$ where K is number of elements in $\boldsymbol{\mu}_0$ or $\boldsymbol{\mu}_1$. An analogous approach can be used to compute a bound on the Mahalanobis distance between a pair of GMMs. Once again we compute the pairwise Mahalanobis distance between threat/nonthreat bag pairs and select the source position that yields the maximum of the smallest Mahalanobis distance. Mathematically we can write

Fig. 5 Error rate of nonadaptive systems based on pairwise Mahalanobis distance [Eq. (14)] and symmetrical KL divergence [Eq. (12)] and RFG system as a function of number of photons.

$$s_b^{mhl} = \underset{s=1 \dots 25}{\operatorname{argmax}} \min_{i,j} (\mathbf{r}_i^c - \mathbf{r}_j^t)^T \Lambda(\mathbf{r}_i^c + \mathbf{r}_j^t)^{-1} (\mathbf{r}_i^c - \mathbf{r}_j^t), \quad (14)$$

where the parameters are defined as in Eq. (12).

We study the classification performance of single-source measurement systems based on both metrics [i.e., KL divergence from Eq. (12) and Mahalanobis distance from Eq. (14)], and quantify the error probability as a function of total photon count. Note that for single-exposure systems, photon count refers to the photons produced by the source during a single exposure; whereas, for multiexposure systems (e.g., an RFG systems employing all 25 sources in sequence), it refers to the sum of the photons used by all sources during the sequence. Our results are based on the following values of total photon counts: 10^4 , 5×10^4 , 10^5 , 5×10^5 , and 10^6 . Both metrics result in the optimal source number 12 for all values of total photon count, therefore, we expect that both metrics will yield comparable P_e . Figure 5 shows that both KL and Mahalanobis metrics yield nearly identical P_e . Also as expected, we see that the error rate decreases as the number of photons increases. In addition to the optimal single-source error rates described above, the error rate of an RFG system is also plotted in Fig. 5. We can see that the RFG system outperforms the nonadaptive single source system because of the spatial diversity provided by use of all 25 sources. This observation motivates us to consider the potential benefit of multiplexed single-exposure measurements which can provide spatial diversity.

2.2 Multiplex Solution

The drawbacks of the approaches discussed in the previous section are that we use only one pair of threat and nonthreat bags (i.e., the worst-case bag pair) to represent the complete GMM for each class. By ignoring other bag pairs which could potentially be used to contribute to making a correct decision, we are overestimating P_e . As mentioned above, we are also sacrificing spatial diversity in our measurement by limiting our attention to only a single source. In order to address these two problems, we need to define a new metric that considers all the sources and all the bags and develop an

Fig. 6 Optimal multiplexed measurement $\mathbf{I}^{\text{opt}}$ for (a) 10^4 photons and (b) $\mathbf{I}^{\text{opt}}$ for 10^5 photons.

algorithm that finds the best combination of sources that yields the best projection (i.e., the optimal multiplexed measurement). For multiplexed measurement, the mean of the signal vector is $\bar{\mathbf{r}} = \sum_{s=1}^{25} I_s e^{-\mathbf{H}_s \mathbf{f}}$ and the signal vector has Poisson distribution $\mathbf{r} \sim \text{Pois}(\bar{\mathbf{r}})$ that can be approximated by $N[\bar{\mathbf{r}}, \Lambda(\bar{\mathbf{r}})]$. We also wish to retain our constraint on the total photons so that $\sum_{s=1}^{25} I_s = I_0$ and I_0 is the total number of photons.

Bhattacharyya distance is a special case of Chernoff distance, and is the new metric that we use because the upper-bound classification error for two Gaussian distributions; $N(\omega_1, \Omega_1)$ and $N(\omega_2, \Omega_2)$ is $P_e \leq e^{-d_B[N(\omega_1, \Omega_1), N(\omega_2, \Omega_2)]}$ where d_B is the Bhattacharyya distance between two Gaussians given in Eq. (15). Even though Bhattacharyya distance does not have a closed analytical form between two GMMs, it has been shown that the classification between two GMMs can be upper-bounded by Eq. (16) which is the summation of all of the pairwise error upper-bounds.¹⁷

$$d_B[N(\omega_1, \Omega_1), N(\omega_2, \Omega_2)] = \frac{1}{8} (\omega_1 - \omega_2)^T \Omega^{-1} [(\omega_1 - \omega_2)] + \frac{1}{2} \ln \left(\frac{|\Omega|}{|\Omega_1| + |\Omega_2|} \right),$$

where $\Omega = \frac{\Omega_1 + \Omega_2}{2}$ (15)

$$P_e^{\text{GMM}} \leq \sqrt{p_i p_c} \sum_{i=1}^{100} \sum_{j=1}^{100} \sqrt{p_i^c p_j^c} e^{-d_B\{N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)], N[\bar{\mathbf{r}}_j^c, \Lambda(\bar{\mathbf{r}}_j^c)]\}}.$$

(16)

This overcomes both shortcomings identified above and because Eq. (16) is a smooth function of the photon number vector $\mathbf{I} = [I_1 \dots I_{25}]$, it can be used as a convenient metric for a gradient search to find the optimal multiplexed solution. Our multiplexed solution, therefore, is obtained by minimizing Eq. (16) subjected to $\sum_{s=1}^{25} I_s = I_0$, where I_0 is a constant representing total number of photons in the system.

$$\mathbf{I}^{\text{opt}} = \underset{\mathbf{I}}{\text{argmin}} \sum_{i=1}^{100} \sum_{j=1}^{100} \sqrt{p_i^c p_j^c} e^{-d_B\{N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)], N[\bar{\mathbf{r}}_j^c, \Lambda(\bar{\mathbf{r}}_j^c)]\}}$$

subject to $\sum_{s=1}^{25} I_s = I_0$ and $\forall s I_s \geq 0$. (17)

We solve the above optimization problem by gradient descent (i.e., the derivation of the gradient can be found in the Appendix). The optimal photon allocation among sources for the multiplexed exposure (i.e., the optimal photon number vector $\mathbf{I}$) obtained for $I_0 = 10^4$ and $I_0 = 10^5$ can be seen in Figs. 6(a) and 6(b), respectively. We observe that for low photon count, the optimal photon allocation is concentrated on a single source; whereas, when the total photon count is large, the optimal measurement employs multiple sources and obtains the associated spatial diversity. This result makes sense because for small I_0 if the system distributes the photons to several sources then the signal vector from each source will be deteriorated by the noise, hence the system chooses only one source trading diversity for SNR. On the other hand, the system with large I_0 can distribute the photons among several sources and the signal vector from each selected source still has high SNR. Therefore, the system is able to perform multiplex measurement. Another observation is that if we use this metric which is based on Bhattacharyya distance, to find the optimal single source, we get the same results as previous approaches discussed in the previous section. In other words, this approach also gives us source number 12 as the optimal single projection for the range of I_0 values studied.

Figure 7(a) compares the value of P_e^{GMM} (error rate upper-bound) for the optimal single source (green square) and multiplex (blue diamond) measurement as a function of total photon count. As expected, the error bound for a multiplexed measurement is substantially improved relative to the single-source measurement. This is the result of increased spatial diversity. Figure 7(b) shows the actual error rate (P_e) obtained in simulation for an optimal single source (green square), optimal multiplexed (blue diamond), and RFG

Fig. 7 (a) The upper error bound Eq. (16) for the optimal single-source and multiplexed measurement for different numbers of photons. (b) Error-rate of the optimal single-source and the multiplexed measurement systems.

(black circle) measurement strategies. We can see that the optimal multiplexed measurement yields substantially smaller P_e compared with the optimal single-source measurement. For example, at $I_0 = 10^6$ photons we observe a 41.8 times reduction in P_e via use of multiplexing. Equally important is the observation that an optimally multiplexed solution can provide P_e approximately equal to that of an RFG measurement strategy at comparable total photon count.

3 Adaptive System

We have thus far studied the performance of three types of nonadaptive system: RFG, optimal single source, and optimal multiplexed. In this section, we are going to enhance the performance of the two optimal nonadaptive systems by making them adaptive. By adaptive we mean that, instead of making one exposure yield a single-optimal measurement, the system performs a sequence of exposures where, at each exposure, the optimal source or combination of the sources are selected based on the information obtained from previous exposures. The SHT approach defines the measurement sequence. We compare the classification performances of adaptive, nonadaptive, and RFG systems.

3.1 Single-Source Solution

Before describing the adaptive single-source strategy, we summarize the decision-making process of SHT. After each iteration (i.e., an iteration yields a measurement from a single exposure of the sequential measurement process), the SHT algorithm executes one of the following two actions: (a) it chooses one of the hypotheses as true and rejects the other one or (b) it cannot make a decision and instead performs the next measurement. At each iteration, SHT divides the decision space into three mutually exclusive regions based on the desired error-rate P_e^* . The decision at each iteration is based on a likelihood ratio test. Mathematically the SHT at k 'th iteration makes a decision based on

$$\frac{Pr(\mathbf{R}^k|c)}{Pr(\mathbf{R}^k|t)} \times \frac{p^c}{p^t} > A, \quad (18)$$

$$\frac{Pr(\mathbf{R}^k|c)}{Pr(\mathbf{R}^k|t)} \times \frac{p^c}{p^t} < B, \quad (19)$$

where $A = \frac{1-P_e^*}{P_e^*}$ and $B = \frac{P_e^*}{1-P_e^*}$ are the upper threshold and lower threshold, respectively, and $\mathbf{R}^k = [\mathbf{r}^1 \dots \mathbf{r}^k]$ represents the accumulated measurement vector which is created by concatenating the k previous measurements. SHT chooses H^c if the inequality in Eq. (18) is satisfied and chooses H^t if the inequality stated in Eq. (19) is true. If none of the inequalities in Eqs. (19) and (18) is satisfied, then SHT makes the next measurement.

We can explain the adaptive single-source algorithm based on SHT as follows. Figure 8 also shows the following algorithm steps:

1. We first specify a value for P_e^* and compute the thresholds A and B . We set the iteration counter k to one and set the belief states for each class to 0.5, so $P_c^1 = P_t^1 = 0.5$. We then find the optimal single source for the first measurement based on one of the approaches in Eqs. (12) or (14)—name that source s_1 . Using the optimal source we make a first measurement and name it $\mathbf{r}^1$. We set $\mathbf{R}^1 = \mathbf{r}^1$ and go to step 2.
2. By checking the inequalities in Eqs. (18) and (19), we either make a threat/nonthreat decision or continue to step 3 to update the current belief states. The belief state is the updated probability of threat or nonthreat class. After each iteration, SHT should provide more information about the class leading to higher belief in one class over the other.
3. We update the belief state for each class using Bayes rule

Fig. 8 Flowchart of adaptive system based on SHT.

$$\begin{aligned}
 P_c^{k+1} &= \frac{Pr(\mathbf{R}^k|c)P_c^k}{Pr(\mathbf{R}^k|c)P_c^k + Pr(\mathbf{R}^k|t)P_t^k}, \\
 P_t^{k+1} &= \frac{Pr(\mathbf{R}^k|t)P_t^k}{Pr(\mathbf{R}^k|c)P_c^k + Pr(\mathbf{R}^k|t)P_t^k}.
 \end{aligned} \quad (20)$$

Then we go to next step.

- In this step, we are going to find the optimal source for the next measurement. Figure 9 schematically represents the space of all possible signal vector means that could have been produced by the sequence of sources through the k 'th measurement. Black circles represent the collection of possible mean vectors of nonthreat bags and the red triangles represent the threat bags. The actual accumulated measurement vector obtained via the specific sequence of k measurements is depicted by the blue square. Using our earlier insight from the nonadaptive design strategy, we consider the worst case P_e in order to select the next optimal source s_{k+1} . Since in each class we know that the classification is harder for the closest bag to the k 'th measurement, we find the closest bags to the current measurement in nonthreat and threat classes by computing the Mahalanobis distance of every bag in each class to the current measurement and choosing the bag that has the smallest Mahalanobis distance. As shown in Fig. 9, we name the closest bag in nonthreat class and threat class with C and T , respectively. To find the next optimal source s_{k+1} , we compute the Mahalanobis distance between the mean of the signal vector related to C and T for all 25 sources using Eq. (21) and choose the one maximizing the distance between C and T . After finding the next optimal source s_{k+1} , we update the iteration counter ($k \leftarrow k + 1$) and go back to step 2.

$$\begin{aligned}
 s_{k+1} = \operatorname{argmax}_{s=1 \dots 25} & (I_s e^{-H_s f_C} - I_s e^{-H_s f_T})^* [P_C^k \Lambda(I_s e^{-H_s f_C}) \\
 & + P_T^k \Lambda(I_s e^{-H_s f_T})]^{-1} \times (I_s e^{-H_s f_C} - I_s e^{-H_s f_T}).
 \end{aligned} \quad (21)$$

In Eq. (21), f_C is the mean of the signal vector representing bag C and f_T is the corresponding quantity for bag T in Fig. 9. Note that step 4 is a critical aspect of the SHT process because we use information obtained from previous measurements to find the next optimal source in this step. The

updated belief states from step 3 represent this information from previous measurements and must be used in step 4. We accomplish this by multiplying the covariance matrices of closest bags by their related belief states as shown in Eq. (21). By this approach, we can control the randomness of each Gaussian related to f_C or f_T via its current belief state. In other words, the updated belief states are considered as weights for covariance matrices in Eq. (21), therefore, the more probable Gaussian has the larger weight (belief state). We will see below that this approach works well.

Next we present the classification performance of the adaptive single-source system. SHT will be used to find the next “best source” (i.e., it will not simply sequence through all 25 sources) in an effort to minimize the number of photons required to achieve the desired probability of error P_e^* . Recall that each class has 100 bags and $p_t = p_c = 0.5$. We choose the desired probability of error $P_e^* = 0.01$ which results in $A = 99$ and $B = 0.0101$ in step 1 of SHT. Figure 10(a) shows the evolution of the belief states for one experiment when the measurement apparatus is presented with a threat bag selected at random. We can see that SHT pushes the true belief state to one and the wrong belief state to zero. It is important to note that the number of measurements for each experiment is a random number. This is due to the random nature of the bag selection as well as the stochastic nature of the noisy measurement process. Figure 10(b) confirms the expectation that the mean number of measurements $E[k]$ is a monotonic decreasing function of number of photons per measurement. As the measurement

Fig. 9 Determining the closest pair of bags (C, T) to the current measurement (blue circle).

Fig. 10 (a) The evolution of belief states for threat and nonthreat classes when one threat bag is selected at random and (b) mean number of measurements $E[k]$ as a function of number of photons per measurement.

SNR increases, SHT requires fewer measurements on average to achieve a reliable classification result.

Figure 11 compares the error rate P_e of adaptive, nonadaptive, and RFG systems as functions of total photon budget. For adaptive systems, we define the total photon budget as $E[k] \times$ number of photons per measurement. For example, consider a system that employs 10^4 photons per measurement (i.e., per each exposure in the sequence), Fig. 10(b) shows that $E[k]$ for this exposure level is 1.68 resulting in a required total photon budget of 16,800 photons in order to guarantee with high probability that the P_e value remains below our target P_e^* . We can see from Fig. 11 that the classification error is a monotonic decreasing function of total photon budget for all methods that we have studied. The adaptive single-source system (red diamond) outperforms the nonadaptive single-source system (blue square) as well

as the RFG system (black x) and its error rate is always below the desired error rate as expected (green line at 0.01). For example, for a total number of photons of 10^6 , we can reduce P_e by factor of 40 relative to the RFG system. Compared with the nonadaptive single-source solution, the adaptive system benefits from a spatial diversity advantage as a result of utilizing more than a single exposure where each exposure may use a different source location chosen by Eq. (21). In contrast with the RFG multiexposure system, the adaptive system uses relatively few exposures [i.e., note that Fig. 10(b) indicates that fewer than two exposures is often sufficient] thus providing higher measurement SNR for a fixed total photon budget.

3.2 Multiplex Solution

In this section, we extend the adaptive single-source solution described above to the case of multiplexed exposures. By employing source multiplexing to provide spatial diversity to each measurement in the sequence, we expect to further reduce the required number of exposures and thereby improve the P_e performance obtained under the total photon count constraint. The SHT procedure in this case is nearly identical to the algorithm described in the previous section with two modifications. In the first step, instead of finding the optimal single source, we find the best combination of sources $\mathbf{I}^{\text{opt}}$ by solving the optimization problem stated in Eq. (17). Also we need to modify step 4 to incorporate the current belief state into the optimization of the source brightness vector. There are two natural ways to do this. First we can embed the belief states into the problem stated in Eq. (17) by multiplying the belief states by their associated covariance matrices and then solve the resulting optimization problem. Unfortunately this approach is very time-consuming because we estimate $E[k]$ and P_e by Monte-Carlo simulations. The second way to modify step 4 is to consider the closest bag pair to the current measurement similar to the approach we used for the single-source adaptive system.

Fig. 11 Error rates of nonadaptive single-source system, adaptive single-source, and RFG system.

Fig. 12 (a) Mean number of measurements $E[k]$ for adaptive single-source and adaptive multiplexed measurement systems. (b) Error rates of nonadaptive single, nonadaptive multiplexed measurement, RFG, adaptive single-source and adaptive multiplexed measurement systems.

Therefore, we can say that we are looking for the combination of sources that maximizes the Mahalanobis distance between the closest bag pair and we multiply each class covariance matrix by its relative belief state to use the retrieved information from previous measurements. Because we have limited the total number of photons per measurement in the system, this maximization problem can be written as

$$\begin{aligned} \mathbf{I}_{k+1}^{\text{opt}} = \underset{\mathbf{I}}{\text{argmax}} & \left(\sum_{s=1}^{25} I_s e^{-H_s f_c} - \sum_{s=1}^{25} I_s e^{-H_s f_T} \right)^T \\ & \times \left[P_c^k \Lambda \left(\sum_{s=1}^{25} I_s e^{-H_s f_c} \right) + P_t^k \Lambda \left(\sum_{s=1}^{25} I_s e^{-H_s f_T} \right) \right]^{-1} \\ & \times \left(\sum_{s=1}^{25} I_s e^{-H_s f_c} - \sum_{s=1}^{25} I_s e^{-H_s f_T} \right). \end{aligned}$$

subject to $\sum_{s=1}^{25} I_s = I_0$. (22)

Note that Eq. (22) can be very quickly solved using gradient descent compared to solving Eq. (17). We use Eq. (22) to find the optimal combination of sources for the next measurement.

Figure 12(a) illustrates the mean number of measurements, $E[k]$, required for both adaptive single source (green square) and adaptive multiplexed (blue diamond) approaches. As expected, the adaptive multiplexed solution provides spatial diversity within each exposure and, therefore, requires slightly fewer measurements on average compared to the adaptive single-source system. In Fig. 12(b), we plot the classification error obtained from all of the methods described herein. We note that the performance of the adaptive solutions (dashed) is consistently superior to the nonadaptive and RFG solutions. The adaptive multiplexed strategy offers only a slight advantage compared with the adaptive single-source

solution because, for this range of total photon count, the required number of measurements is already small.

4 Reducing the Number of Detectors

To reduce the cost of building an x-ray CT, one may reduce the number of detectors and use wider detectors instead as shown in Fig. 13, a reduction of the number of detectors from M to $M/2$. We may choose to double the detector width in order to cover the same total detector area and avoid losing photons. Using wider detectors increases the signal vector SNR since each detector can collect more photons. Hence there is a trade-off between signal vector SNR and its dimensionality (number of detectors in the detector array).

In this section, we are interested in studying the performance of both adaptive single-source and adaptive multiplexed measurement systems under dimensionality reduction of detector array. In previous sections, we observed that SHT keeps the actual probability of error far below the desired probability of error P_e^* . This means that the detection performance can remain acceptable by SHT algorithm even if there is slight increase in probability of error. To prevent this phenomenon causing any trouble and have a fair comparison between systems with different numbers of detectors, we repeat the experiments for different values of P_e^* from 0.08 to 3×10^{-4} and record the probability of error (P_e) for each system with a different number of detectors. Based on the recorded data points (P_e), we create a scatter plot and find the lower-bound on P_e for each system. We will use these lower-bounds to compare detection performance of different systems because our goal here is to minimize the P_e .

Figure 14 shows error-rate scatter plots of the single-source adaptive systems with 2200 (circle), 1100 (star), 220 (diamond), or 44 (square) detectors in the detector array as well as the lower-bounds (solid curves) on the error rates for each detector number. Each data point in the plot is P_e value estimated by Monte-Carlo simulations. We can see that, for a low photon budget, the system with 220 detectors

Fig. 13 Reducing the number of detectors by factor of 2, from M to $M/2$.

Fig. 14 The probability of error for single-source adaptive system with different numbers of detectors.

outperforms other systems. As the photon budget increases, the system with 2200 detectors has a better performance compared to other systems. This happens because there is a trade-off between signal-vector SNR and spatial diversity (number of detectors in the detector array). For low photon count, the system sacrifices the spatial diversity for an increase in SNR by using the smaller number of detectors in the detector array. Therefore, it prevents the noise from deteriorating the signal vector severely. For high photon count, there is no need to reduce number of detectors in order to increase the SNR. The system, therefore, benefits from spatial diversity as shown in Fig. 14 where the system with 2200 detectors has the best performance in the high photon count region.

Figure 15 shows error-rate scatter plots of multiplexed adaptive systems with 2200, 1100, 220, or 44 numbers of detectors in the detector array with the lower-bounds on the error rates. Like the single-source adaptive system, for a multiplexed adaptive system, there is a trade-off between SNR and spatial diversity in the multiplexed adaptive system. Both single-source and multiplexed sources adaptive systems yield the same trend such that the 220-detector system outperforms others in the low-photon count region and the 2200-detector system outperforms others in the high-photon count region. Figures 14 and 15 show that adaptive

Fig. 15 The probability of error for multiplexed measurement adaptive system with different number of detectors.

single source and adaptive multiplexed systems have comparable performances under dimensionality reduction.

5 Conclusion

In this work, we first discussed two nonadaptive methods for improving the shape-threat detection performance of x-ray CT. In the first method, we found the optimal source for single-exposure measurement based on KL divergence and Mahalanobis distance. In the second method, we determined the optimal combination of sources for a single-exposure measurement by minimizing the Bhattacharyya bound on probability of error. The RFG system outperformed the single-source system whereas the multiplexed sources system showed comparable performance to the RFG system. We then discussed the adaptive systems by applying SHT algorithm and witness that the adaptive systems outperformed the nonadaptive systems in term of probability of error and both adaptive single-source and multiplexed measurement systems had comparable performances. At the end, we study the effect of reducing the number of detectors in the detector array on the performance of both adaptive systems. We observed that, for a low photon count, reducing the number of detectors generally enhances the system performance whereas, in a high photon there is no need for the reduction.

Appendix

In this section, we determine the gradient of Eq. (17) respective to $\mathbf{I} = [I_1 \dots I_{25}]$:

$$\nabla_{\mathbf{I}} \left(\sum_{i=1}^{100} \sum_{j=1}^{100} \sqrt{p_i^c p_j^c} e^{-d_B \{N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)], N[\bar{\mathbf{r}}_j^c, \Lambda(\bar{\mathbf{r}}_j^c)]\}} \right), \quad (23)$$

where $\bar{\mathbf{r}}_i^c$ and $\bar{\mathbf{r}}_j^c$ are functions of I , e.g., $\bar{\mathbf{r}}_i^c = \sum_{s=1}^{25} I_s e^{-\mathbf{H}_s \mathbf{r}_i^c}$ and $\mathbf{r}_i^c$ is the i 'th nonthreat bag. We can move the gradient into the sum in Eq. (23)

$$\sum_{i=1}^{100} \sum_{j=1}^{100} \sqrt{p_i^c p_j^t} e^{-d_B\{N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)], N[\bar{\mathbf{r}}_j^t, \Lambda(\bar{\mathbf{r}}_j^t)]\}} \times \nabla_{\mathbf{I}}(-d_B\{N[\bar{\mathbf{r}}_i^c, \Lambda(\bar{\mathbf{r}}_i^c)], N[\bar{\mathbf{r}}_j^t, \Lambda(\bar{\mathbf{r}}_j^t)]\}), \quad (24)$$

where

$$d_B(N[\bar{\mathbf{r}}_i^c, \bar{\mathbf{r}}_i^c], N[\bar{\mathbf{r}}_j^t, \bar{\mathbf{r}}_j^t]) = \frac{1}{8} (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t)^T \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t) + \frac{1}{2} \ln \left[\frac{|\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)|}{|\Lambda(\bar{\mathbf{r}}_i^c)| + |\Lambda(\bar{\mathbf{r}}_j^t)|} \right]. \quad (25)$$

As we can see, we need to only compute the gradient of Eq. (25). The gradient of the first term on the right-hand side of Eq. (25) is

$$\begin{aligned} & \frac{1}{8} [\nabla_{\mathbf{I}}(\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t)^T] \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t) \\ & + \frac{1}{8} (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t)^T [\nabla_{\mathbf{I}} \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1}] (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t) \\ & + \frac{1}{8} (\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t)^T \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} [\nabla_{\mathbf{I}}(\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t)], \end{aligned} \quad (26)$$

where

$$\nabla_{\mathbf{I}}(\bar{\mathbf{r}}_i^c - \bar{\mathbf{r}}_j^t) = \begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_i^c} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_i^c} \end{pmatrix} - \begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_j^t} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_j^t} \end{pmatrix}, \quad (27)$$

$$\begin{aligned} \nabla_{\mathbf{I}} \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} &= -\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} \Lambda \left[\begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_i^c} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_i^c} \end{pmatrix} \right. \\ &\quad \left. + \begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_j^t} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_j^t} \end{pmatrix} \right] \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1}. \end{aligned} \quad (28)$$

By substituting Eq. (28) and Eq. (27) into Eq. (26), we compute the derivative of the first term on the right-hand side of Eq. (25)

The gradient of the second term on the right-hand side of Eq. (25) is

$$\begin{aligned} & \nabla_{\mathbf{I}} \left\{ \frac{1}{2} \ln \left[\frac{|\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)|}{|\Lambda(\bar{\mathbf{r}}_i^c)| + |\Lambda(\bar{\mathbf{r}}_j^t)|} \right] \right\} \\ &= \frac{\nabla_{\mathbf{I}}[|\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)|]}{2|\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)|} - \frac{\nabla_{\mathbf{I}}[|\Lambda(\bar{\mathbf{r}}_i^c)| + |\Lambda(\bar{\mathbf{r}}_j^t)|]}{2[|\Lambda(\bar{\mathbf{r}}_i^c)| + |\Lambda(\bar{\mathbf{r}}_j^t)|]}, \end{aligned} \quad (29)$$

where

$$\begin{aligned} \nabla_{\mathbf{I}}[|\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)|] &= |\Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)| \\ &\times \text{Tr} \left\{ \Lambda(\bar{\mathbf{r}}_i^c + \bar{\mathbf{r}}_j^t)^{-1} \Lambda \left[\begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_i^c} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_i^c} \end{pmatrix} + \begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_j^t} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_j^t} \end{pmatrix} \right] \right\}, \end{aligned} \quad (30)$$

$$\begin{aligned} \nabla_{\mathbf{I}}[|\Lambda(\bar{\mathbf{r}}_i^c)| + |\Lambda(\bar{\mathbf{r}}_j^t)|] &= |\Lambda(\bar{\mathbf{r}}_i^c)| \text{Tr} \left\{ \Lambda(\bar{\mathbf{r}}_i^c)^{-1} \Lambda \left[\begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_i^c} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_i^c} \end{pmatrix} \right] \right\} \\ &+ |\Lambda(\bar{\mathbf{r}}_j^t)| \text{Tr} \left\{ \Lambda(\bar{\mathbf{r}}_j^t)^{-1} \Lambda \left[\begin{pmatrix} e^{-\mathbf{H}_1 \mathbf{f}_j^t} \\ \vdots \\ e^{-\mathbf{H}_{25} \mathbf{f}_j^t} \end{pmatrix} \right] \right\}, \end{aligned} \quad (31)$$

where $\text{Tr}()$ is trace of a matrix. By substituting Eqs. (31) and Eq. (30) into Eq. (29) which is the derivative of the second term in Eq. (25), we are now able to compute the derivative of Eq. (25) for all the i and j terms and use it to determine Eq. (23).

References

1. K. A. Heiskanen and F. L. Roder, "Explosives detection systems: the real-world experience," *Proc. SPIE* **2936**, 142–155 (1997).
2. J. D. Baldeschwieler, *Detection of Explosives for Commercial Aviation Security*, Vol. **471**, National Academies Press, New York (1993).
3. Y. White, "Automated threat detection for baggage screening," in *Mercury White Paper*, Mercury Computer Systems, United Kingdom (2009).
4. S. Singh and M. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Process.* **83**(1), 31–55 (2003).
5. B. Gonzales et al., "Rectangular fixed-gantry CT prototype: combining CNT x-ray sources and accelerated compressed sensing-based reconstruction," *IEEE Access* **2**, 971–981 (2014).
6. T. Feather et al., "CAXSS: an intelligent threat-detection system," *Proc. SPIE* **1824**, 153–161 (1993).
7. J. H. Jensen et al., "Evaluation of distance measures between Gaussian mixture models of MFCCs," in *Int. Conf. on Music Information Retrieval*, pp. 107–108 (2007).
8. E. Pampalk, "Speeding up music similarity," in *2nd Annual Music Information Retrieval eXchange*, London (2005).
9. J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. Ninth IEEE Int. Conf. on Computer Vision*, pp. 487–493, IEEE (2003).
10. K. Kampa, E. Hasanbelliu, and J. C. Principe, "Closed-form Cauchy-Schwarz PDF divergence for mixture of Gaussians," in *Int. Joint Conf. on Neural Networks (IJCNN '11)*, pp. 2578–2585, IEEE (2011).
11. J. Huang and A. Ashok, "Information optimal compressive x-ray threat detection," *Proc. SPIE* **9847**, 98470T (2016).
12. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, New Jersey (2012).
13. A. Wald, "Sequential tests of statistical hypotheses," in *Breakthroughs in Statistics*, pp. 256–298, Springer, New York (1992).
14. P. K. Baheti and M. A. Neifeld, "Adaptive feature-specific imaging: a face recognition example," *Appl. Opt.* **47**(10), B21–B31 (2008).
15. A. Masoudi, M. A. Neifeld, and R. Thamvichai, "Target detection via adaptive tomography," in *Computational Optical Sensing and Imaging, CW2F-5*, Optical Society of America (2015).
16. A. Masoudi, R. Thamvichai, and M. A. Neifeld, "Shape threat detection via adaptive computed tomography," *Proc. SPIE* **9847**, 98470H (2016).
17. M. Thangavelu and R. Raich, "On linear dimension reduction for multi-class classification of Gaussian mixtures," in *IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP '09)*, pp. 1–6, IEEE (2009).

Biographies for the authors are not available.