

On the Relevance of Denoising and Artefact Reduction in 3D Segmentation and Classification within Complex Computed Tomography Imagery

Andre Mouton¹, Toby P. Breckon²

*School of Engineering, Cranfield University, Bedfordshire, UK¹
Department of {Computer Science — Engineering}, Durham University, Durham, UK²
contact: toby.breckon@durham.ac.uk*

Abstract

We evaluate the impact of denoising and Metal Artefact Reduction (MAR) on 3D object segmentation and classification in low-resolution, cluttered dual-energy Computed Tomography (CT). To this end, we present a novel 3D materials-based segmentation technique based on the Dual-Energy Index (DEI) to automatically generate subvolumes for classification. Subvolume classification is performed using an extension of Extremely Randomised Clustering (ERC) forest codebooks, constructed using dense feature-point sampling and multiscale Density Histogram (DH) descriptors. Within this experimental framework, we evaluate the impact on classification accuracy and computational expense of pre-processing by intensity thresholding, Non-Local Means (NLM) filtering, Linear Interpolation-based MAR (LIMar) and Distance-Driven MAR (DDMar) in the domain of 3D baggage security screening. We demonstrate that basic NLM filtering, although removing fewer artefacts, produces state-of-the-art classification results comparable to the more complex DDMar but at a significant reduction in computational cost - bringing into question the importance (in terms of automated CT analysis) of computationally expensive artefact reduction techniques. Overall, it was found that the use of MAR pre-processing approaches produced only a marginal improvement in classification performance ($< 1\%$) at considerable additional computational cost ($> 10\times$) when compared to NLM pre-processing.

Keywords: Denoising, Metal artefact reduction, Classification, Segmentation, Dual-energy Computed Tomography, DECT

1. Introduction

The use of 3D X-ray Computed Tomography (CT) has seen a massive surge in popularity in recent years, beyond the traditional domain of medical imaging. Current industrial applications of CT range from micro-CT for non-invasive imaging of wood anatomy to automated baggage and parcel screening within the aviation security infrastructure [1, 2, 3]. The use of Dual-energy Computed Tomography (DECT), whereby objects are scanned at two distinct energy levels, has been especially widespread. DECT provides an effective means for performing materials-based discrimination of objects and has thus been used for a variety of tasks such as the automated detection of explosives in baggage-screening [4, 5] and mineral and rock characterisation in Non-Destructive Testing (NDT) applications in the geoscience domain [6, 7]. In this work, we consider the dual-energy baggage-CT imagery obtained in the aviation security screening domain.

The main non-object recognition based objective of dual-energy baggage-CT scanners [5], combined with the demand for throughput in the aviation-security domain, lead to compromises in image quality in terms of noise, artefacts and voxel resolutions [5, 3]. While medical-CT imagery typically presents with sub-millimetre isotropic voxel resolutions [8, 9] and relatively low quantities of background clutter, baggage-CT data contains dense collections of man-made objects scanned at low millimetre scale, anisotropic resolutions with high levels of noise and metal streaking artefacts [10, 3]. This characteristically complex nature of security-screening CT imagery has complicated the implementation of traditional 3D object segmentation and classification techniques in this domain [11, 2, 12].

Image noise and metal streaking artefacts in particular, have been highlighted as the predominant factors limiting the efficacy of object classification and segmentation techniques in such complex volumetric imagery. Megherbi *et al.* [11] investigated the effectiveness of classical medical-segmentation techniques when applied directly to low-quality baggage-CT scans. Successful segmentation is shown to be dependent on careful parameter tuning. Furthermore, even after parameter optimisation, the effects of image noise, metal streaking artefacts and background clutter result in significantly poorer segmentation than those observed in the medical domain. Although several recent studies have demonstrated successful, unsupervised 3D object segmentation in cluttered baggage-CT data [13, 14, 15], these were performed using medical-grade CT imagery. Despite high near-isotropic voxel resolutions and relatively low

levels of noise and artefacts, each of the studies again cited metal streaking artefacts as limiting factor on performance.

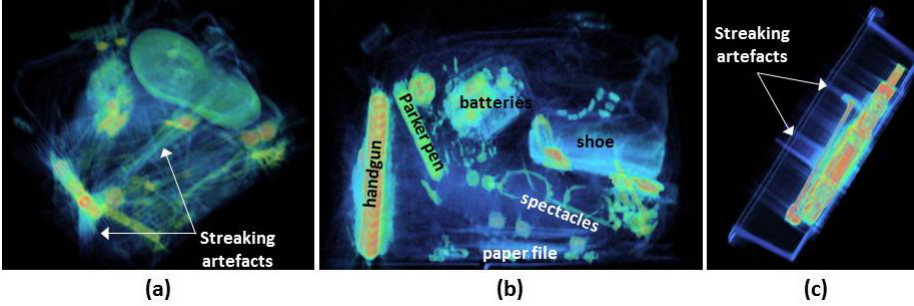


Figure 1: Baggage-CT scans illustrating poor image quality, low resolution, artefacts and clutter (obtained on a Reveal CT80-DR baggage scanner).

Owing to these limitations in the state-of-the-art in unsupervised 3D object segmentation techniques for low-resolution, complex volumetric imagery, prior classification studies in this domain have relied largely on the manual segmentation of images [16, 17, 2, 12, 10].

Megherbi *et al.* [18, 17] present a comparison of classifier-based approaches using volumetric shape characteristics for the classification of pre-segmented objects in cluttered volumetric baggage-CT imagery. The performance of combinations of three shaped-based feature descriptors (3D Zernike descriptors [19]; the Histogram-of-Shape Index (HSI) [20] and a combination of both) and five classifiers (Support Vector Machines (SVM) [21]; neural networks [22]; decision trees [23]; boosted decision trees [24] and random forests [24]) are considered for the classification of pre-segmented bottles. Correct classification rates in excess of 98.0% for pre-segmented bottles are presented using the HSI descriptor in conjunction with an SVM or random forest classifier. The study, however, uses a limited data-set and does not consider the effects of image noise or artefacts.

Extending upon their earlier work [25], Flitton *et al.* [12] present an experimental comparison to investigate the suitability of the Bag-of-Words (BoW) model [26] for the detection of threat items in both manually-segmented as well as unsegmented baggage-CT imagery. Combinations of four 3D interest-point descriptors (Density Histograms (DH) [25]; Density Gradient Histograms (DGH) [25]; the 3D Scale-Invariant Feature Transform (SIFT) [16] and the 3D Rotationally-Invariant Feature Transform (RIFT) [27]) and three codebook assignment methodologies (hard, kernel and uncertainty) are considered. In the classification of handguns and bottles in manually segmented subvolumes, op-

timal correct classification rates ($\sim 89\%$ for bottles; $\sim 97\%$ for handguns) are obtained using an uncertainty assignment protocol [28] in conjunction with simple density-based descriptors [25] sampled at 3D SIFT [16] keypoint locations. The impact of the classifier type, the clustering method and the keypoint detection protocol are, however, not considered. Further experimentation, using unsegmented whole volumes, results in a significant decline in performance, particularly in terms of false-positive rates ($\sim 16\%$ for whole volumes vs. $< 3\%$ for subvolumes). Poor resolution, image noise and metal-streaking artefacts (characteristic to baggage-CT imagery) are again shown to negatively impact the efficacy of the 3D descriptors and subsequently classification performance in both the segmented and unsegmented volumes. Image denoising and metal artefact reduction are not considered. Despite the high false-positive rates for the whole volumes, this work currently represents the only fully-automated (end-to-end) 3D object classification approach for low-quality baggage-CT imagery [12].

Flitton *et al.* [2] present a novel 3D extension to the hierarchical visual cortex model for 3D object classification in pre-segmented baggage-CT subvolumes. The approach outperforms a traditional BoW approach with correct classification rates in excess of 95% and low false-positive rates. Classification performance is again hindered by image noise, metal streaking artefacts and clutter (denoising and metal artefact reduction are not considered). Furthermore, an extremely high computational cost is associated with the construction of the cortex model.

With the exception of the whole-volume approach of Flitton *et al.* [12], the aforementioned techniques all operate on manually segmented subvolumes and none of them consider image denoising and/or metal artefact reduction. Nonetheless, prior studies investigating noise and Metal Artefact Reduction (MAR) in low-resolution, cluttered CT imagery have demonstrated improved image quality, despite the complex nature of the data [10, 29, 30, 3]. Although Mouton *et al.* [10, 29] have presented a novel 3D SIFT-based [16] quantitative performance evaluation metric, which gives an indication of the impact of denoising/MAR on 3D object classification, a true system-level evaluation, whereby the impact on overall system performance (e.g. classification accuracy) is quantified, has not previously been performed.

Given the challenges of complex CT imagery, pre-classification denoising and MAR would initially appear to be a natural choice in any end-to-end processing pipeline. Here, we examine the relevance of this assumption, given its computa-

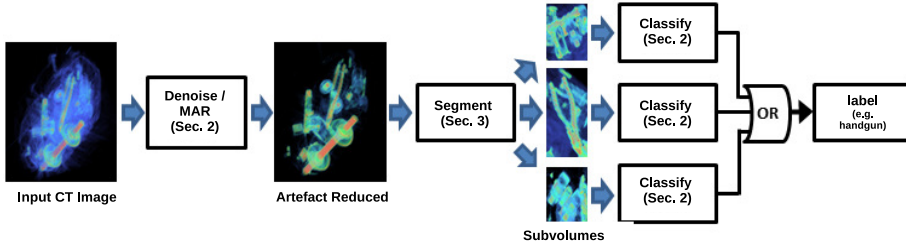


Figure 2: An overview of the joint 3D object segmentation and classification processing pipeline used to evaluate the impact of the denoising and MAR pre-processing.

tional overhead and impact on overall classification performance. To these ends we present novel methodologies for the segmentation and classification of 3D objects in complex dual-energy CT imagery and evaluate the impact of four denoising and artefact reduction techniques on overall performance. This work is the first to perform an experimental evaluation of this nature on complex dual-energy baggage-CT imagery, whereby the relevance of denoising and artefact reduction are addressed.

Here, we address these limitations in prior work in the following ways. First, we present a dual-energy-based segmentation technique to automatically generate subvolumes for classification (Section 3). The approach is based on three novel contributions: 1) a materials-based coarse segmentation technique using the Dual-Energy Index (DEI) [31] and connected component analysis; 2) a random-forest-based model for measuring the quality of individual object segments, which is used to guide the segmentation process and 3) an efficient segmentation-refinement procedure for splitting fused objects. Secondly, we present a codebook 3D object classification approach based on Extremely Randomised Clustering (ERC) forests [32], a dense feature sampling strategy [33] and an Support Vector Machine (SVM) classifier [34] (Section 4). Finally, we perform an overall system-level performance evaluation of four denoising/artefact reduction techniques: 1) simple intensity thresholding; 2) Non-Local Means (NLM) filtering [35, 36]; 3) Linear Interpolation MAR (LIMar) [37] and 3) Distance-Driven MAR (DDMar) [30] (Section 5). The overall outcome of this work is to show the efficacy of such denoising and artefact reduction techniques within the context of a fully automated, yet computationally efficient joint 3D object segmentation and classification processing pipeline for low-quality, complex baggage-CT imagery. An overview of this processing pipeline is shown in Figure 2 within which each the operations of denoising/MAR, seg-

mentation and classification are each associated to corresponding section of this study (Figure 2 / Sections 2 - 4).

2. Noise and Artefact Reduction

We extend the prior work of Mouton *et al.* [10, 29], where the impact of denoising/MAR on object class recognition is quantified using the ratio of object to noise 3D SIFT interest points, by performing a classification-task-driven performance evaluation of four denoising and MAR techniques.

A comprehensive review and evaluation of noise and artefact reduction in low-quality cluttered CT imagery has been presented in [29, 10, 3, 30]. Most notably, it is shown that the superiority of the state-of-the-art techniques from the medical literature, over simpler techniques is reduced as a result of the complexity and poor quality of non-medical CT imagery. It is thus not obvious if the performance gains (if any) of such techniques merit their additional computational overhead. We thus compare the classification-impact of four noise and artefact reduction techniques (of varying complexities): 1) simple intensity thresholding; 2) Non-Local Means (NLM) filtering [35, 36]; 3) MAR by linear interpolation [37] (denoted LIMar) and 4) Distance-Driven MAR [30] (denoted DDMar). Note that DDMar was denoted Mou_{β} in the original study [30]. Although significant advancement within MAR techniques continue, recent developments continue to focus on the medical domain [38, 39, 40] whereas extensive comparative experimental studies [3] have already clearly identified the discrete and varied challenges of baggage-CT when compared to the medical domain. Furthermore, MAR approaches targeting the baggage-CT domain consistently show superior performance when compared to leading medical domain MAR approaches [10, 30, 3]. For a further overview of work in this area the reader is directed to the recent reviews of [41, 42].

Denoising by NLM filtering involves computing the weighted mean of all points whose Gaussian neighbourhood is similar to the neighbourhood of the current pixel [36]. The NLM algorithm (in its original form) is computationally demanding, making it ill-suited for practical applications [43, 44]. We thus use the linear-complexity NLM implementation of Mahmoudi and Sapiro [43], to optimise computational efficiency and reduce processing times. NLM filtering is accelerated by pre-classifying neighbourhoods using the mean neighbourhood intensities and local gradients as measures of similarity (i.e. by computing their responses to a mean intensity-based filter and a local gradient-based filter).

The technique is shown to reduce the quadratic complexity of the original NLM algorithm to a linear complexity.

DDMar [30] is a sinogram completion-based MAR approach [3] that imposes a weighted limit on the intensities of the MAR-corrected pixels where weights are based on the spatial locations of the pixels relative to the metal objects within the image. In particular, pixels falling within the straight-line regions connecting multiple metal objects are subjected to less intensive intensity-limiting, thereby compensating for the characteristic dark bands occurring in these regions [3]. These filtering and MAR methods are applied on a per-slice basis.

3. 3D Object Segmentation

In prior work [25, 2, 45], 3D object classification has been performed on manually generated subvolumes, containing at most a single object of interest. To address this, we present a dual-energy based unsupervised 3D object segmentation technique. In a related study [46], the proposed approach has demonstrated comparable performance to the state-of-the-art Isoperimetric Distance Tree (IDT) technique [15] at significant reduction in computational cost. In view of the high classification rates achieved using such volumes ($> 98\%$ [45]), the output label maps generated using the proposed segmentation method, are used to create a set of single-object subvolumes for each given input volume. These subvolumes are then passed to the chosen classifier.

Dual-Energy CT (DECT): Conventional, single-energy CT scanners produce reconstructions representative of the Linear Attenuation Coefficients (LAC) of the object under investigation such that voxel intensities are dependent on the LAC of the scanned object [47]. It is thus difficult to distinguish between materials that share similar LAC. In contrast, DECT, whereby attenuation data is captured using two distinct X-ray spectra, offers a means for characterising the chemical composition (e.g. atomic number and/or electron density) of the scan based on its response under these different spectral conditions. The physical basis of DECT relies on the energy dependence of the interaction of X-ray photons with matter. Within a photon energy range of approximately 30 keV to 200 keV, these interactions are known to be dominated by the photoelectric effect and Compton scattering [48]. Dual-energy decomposition involves determining the Compton scatter and the photoelectric absorption coefficients of the material from the measured high and low-energy projections. Decomposed dual-energy data may subsequently be used to compute effective atomic num-

bers and electron densities (this is the basis for traditional explosives detection systems in aviation security-screening [5, 4]).

DECT techniques typically fall into one of three categories [49]: 1) post-reconstruction techniques; 2) pre-reconstruction techniques and 3) iterative-reconstruction techniques. Pre-reconstruction [50] and iterative reconstruction [51] techniques are concerned with solving the material decomposition problem and determining the effective atomic number and electron density images. While a variety of DECT-decomposition techniques exist (e.g. direct approximation [48]; iso-transmission lines [52]), they typically require access to raw projection data and rely on a physical calibration procedure [50]. A more rudimentary estimation of the chemical characteristics of a scan may be obtained via post-reconstruction (or image-based) DECT techniques. Post-reconstruction DECT does not require a calibration procedure or the availability of raw-data. The Dual-Energy Index (DEI) [31] is a post-reconstruction DECT measure that represents an estimate of the effective atomic numbers of the materials in a scan. For a material within air (i.e. not dissolved in water) the DEI is given by:

$$DEI = \frac{x_L - x_H}{x_L + x_H + 2000} \quad (1)$$

where x_L and x_H are the pixel values, in Hounsfield Units (HU), for the low and high-energy scans respectively [31]. In contrast to the true effective atomic number [4], the DEI does not rely on the photoelectric cross-section characteristics of the material (which are not precisely known) [31].

We generate initial coarse segmentations using the DEI, simple thresholding operations and connected component analysis. More particularly, based on the assumption that different materials in a scan are relatively well separated by their DEI, material-specific images are generated by thresholding:

$$I_{\tau_i} = \begin{cases} 1 & \text{if } \tau_{i-1} \leq I_{\text{dei}} \leq \tau_i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, N_\tau \quad (2)$$

where I_{dei} is the DEI image and the number of thresholds N_τ is a user specified parameter. Connected component analysis [53] is then performed on each of the N_τ thresholded images. Those connected components larger than a predefined minimum object size are assigned individual labels and represent the image foreground (or coarse segmentation) to be passed to the segmentation-

refinement algorithm.

The quality of the individual components of the coarse segmentations are determined using a random-forest-based evaluation metric (denoted the Random Forest Score (RFS)), which is trained to recognise high-quality, single-object segments. Each foreground object in the coarse segmentation is characterised by a 42D descriptor comprising feature attributes falling into one of five categories [54]: 1) unweighted geometric features (quantifying the size and regularity of the segmentation); 2) weighted geometric features (locally emphasising the geometric features when intensity values are similar to each other); 3) intensity features (measuring absolute intensity and intensity distributions within segmentations); 4) gradient features and 5) ratio features (computed as ratios of previously computed features). Individual trees in the random forest are constructed in a top-down recursive manner using a simple thresholding function as the node split function for all internal nodes:

$$f(v_i, \theta_j) = \begin{cases} 0 & v_i < \theta_j \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

where $v_i, i = 1, \dots, D$ is a single feature attribute selected from a D -dimensional descriptor vector $\mathbf{v} \in \mathbb{R}^D$ and θ_j is a scalar valued threshold ($D = 42$). The optimality criterion used for node splitting is the classical Information Gain (IG) [55]. Randomness is injected into the trees via random node optimisation, whereby a random subset of the available node test parameter values is considered at each node. Trees are grown to a maximum depth D and leaf nodes are generated if the IG falls below a minimum threshold IG_{\min} . The RFS of a given segmentation is computed by averaging the corresponding posterior probabilities of each of the leaf nodes reached in the forest:

$$\text{RFS} = p(c|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^T p_t(c|\mathbf{v}) \quad (4)$$

where T is the number of trees in the forest; $p(c|\mathbf{v})$ is the estimated conditional probability that a given test point \mathbf{v} belongs to the class c and c is a discrete class label (i.e. $(0, 1) \rightarrow (\text{bad}, \text{good})$).

The RFS is computed for each component (or object) in a given coarse segmentation. Those components yielding an RFS below a given threshold τ_{RFS} are considered to be composed of multiple objects and are partitioned at the

estimated intersection (or touching) points of the multiple objects comprising that segment. These points are found by detecting the perimeter voxels of the original object that are likely to be common to two objects. Non-zero (i.e. object) voxels are assumed to lie on the perimeter of the object if they are connected to at least one zero (i.e. background) voxel. The assumption is made that those voxels corresponding to the intersections of multiple objects will be surrounded by higher numbers of object voxels compared to regular (non-intersection) perimeter points [13]. If the total number of object voxels in a predefined local cubic neighbourhood ($11 \times 11 \times 11$) of a given perimeter voxel is greater than a threshold, τ_{HP} , then that perimeter voxel is considered to be an intersection point [13]. A Connected Component Analysis (CCA) [53] is performed on a mask obtained by removal (i.e. setting to zero) of these so-called *hot points*. The two connected components returning the highest RFS (computed with the hot-points reinserted) are retained. When CCA results in only a single connected component, morphological dilation of the zeroed-out region is performed until CCA returns at least two components. If the RFS of one of the regions falls below the RFS of the original region, or if the region is smaller than the minimum permissible size of an object, then the region is assumed to be noise or artefact-induced and is discarded. After CCA, if both components result in a decrease relative to the original RFS, then the original object is retained. Although objects split in this way are not guaranteed to produce segments with $RFS > \tau_{RFS}$, only splits resulting in improved scores are permitted. For objects containing multiple hot-point clusters, the RFS of the final split objects are affected by the order in which the clusters are considered. As the described splitting procedure is fast and the numbers of hot-point clusters per object are generally low (typically ≤ 3), the optimal order (i.e. that which results in the individual objects with the highest RFS) is determined by testing all possible orders.

Finally, for each labelled object in the resulting segmentation, a subvolume is generated by computing the minimum bounding box encompassing that object and then increasing the dimensions of the box by approximately 30mm in all 3 dimensions (as per [2, 45]). The high correct classification rates presented in [2, 45] were achieved without considering noise and/or artefact reduction. This suggests that classification performance is robust to image noise and artefacts, provided a given subvolume is dominated by a single object. The final subvolume used for classification is thus obtained by extracting the entire region corresponding to the expanded bounding box from the pre-processed volume

(as opposed to setting the non-object, background, voxels to zero in the subvolume). This strategy is adopted to ensure that contextual information is not lost in the subsequent feature extraction and description process. A segmentation composed of N labelled objects thus results in N subvolumes, each theoretically containing a single, distinct object (Figure 3).

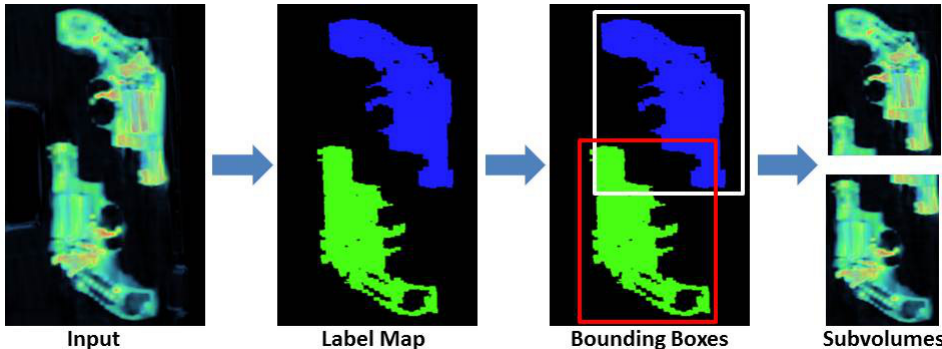


Figure 3: Generation of single-object subvolumes for object classification.

4. 3D Object Classification

We present a codebook sub-volume classification technique to independently classify each of the subvolumes in a given input volume. In a related study [45], the proposed approach has been shown to outperform the state-of-the-art visual cortex approach [2] both in terms of classification accuracy and processing time. The method adheres to the following traditional classification framework [56]: 1) feature detection and description; 2) visual codebook generation and vector quantisation and 3) classification.

Feature detection and description: In accordance with [33], we adopt a dense feature-sampling strategy, whereby 3D keypoints are sampled uniformly and randomly from the baggage-CT sub-volume using the Density Histogram (DH) descriptor of [25]. An invariance to uniform changes in image scale is obtained by sampling interest points from three image scales [57] (as per [16]). At each of the scales a limit of $\tau_N = 0.006N$ on the number of randomly sampled points is enforced (where N is the number of voxels in the Gaussian scale-space image and τ_N is determined empirically). Flitton *et al.* [25] have shown that simple density statistics-based descriptors outperform more complex 3D descriptors (SIFT [16] and RIFT [58]) in object detection within low resolution, baggage-CT imagery. We thus use the Density Histogram (DH) descriptor [25]

here. The descriptor characterises the local density variation at a given interest point as an N -bin histogram defined over a continuous density range. Optimal descriptor parameters are selected in accordance with [25, 59] and result in a 60-dimensional feature vector.

Visual codebook generation: We employ Extremely Randomised Clustering (ERC) forests [32] for feature encoding and codebook generation. ERC forests assign separate codewords to every leaf node in a given forest. A BoW representation for a given image is obtained by accumulating the codeword counts after applying the forest to all the descriptors in the image. The resulting histogram of codewords is then used in subsequent classification in the same way as any standard BoW model. ERC forests are built in the same manner as the random forests used for computing the RFS. Particularly, trees are trained in a top-down recursive fashion [60] using a set of labelled training descriptors, where the labels are obtained from global image annotations (i.e. all descriptors from a given image share the same label). Internal nodes are optimised using the thresholding function in Equation 3 and the information gain optimality criterion [55]. Randomness is injected into the trees by considering a fixed-size random subset of the available node split function parameter values at each node.

Classification is performed using a multi-class (one-versus-the-rest) Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel [34] that is optimized via grid search over the kernel parameter space [45]. Subvolumes are classified individually and the final class label of a given input volume, composed of N segmented objects (i.e. N subvolumes) is computed as the logical ‘OR’ of the class labels of each of its N constituent subvolumes. While the classification of each of the N subvolumes is easily parallelised the processing time of a serial classification of a given N -object input volume may be reduced by only classifying the i^{th} subvolume if the $(i - 1)^{th}$ subvolume (where $i = 2, \dots, N$) has returned a negative class label (i.e. classification is terminated as soon as a positive label is produced).

For further detail, on all aspects of the 3D object classification approach used as a reference within the processing pipeline for this study the reader is directed to [45]. Additional parameter details are presented in Section 5.

5. Results

Classification performance was evaluated for two independent object types (handguns and bottles) in complex 3D baggage-CT imagery using traditional measures (true-positive rate, false-positive rate, precision, accuracy and processing time). Processing times were measured for all experiments performed on an Intel Core i5 machine running a 2.30GHz processor with 6GB of RAM.

The 3D dual-energy baggage-CT imagery was obtained on a CT-80DR dual-energy baggage scanner manufactured by Reveal Imaging Inc, which produces volumes with low anisotropic resolutions of $1.56 \times 1.61 \times 5$ mm. All images were gathered using a set of ~200+ test baggage items, packed with numerous benign items of clutter (e.g. clothing, books, paperwork, electronics) representative of the aviation traveling public. The set of baggage items used conform to the variety and packing profile commonly used for test and evaluation within the baggage-CT security domain. These images were scanned with and without the target object types (handguns and bottles) present.

Within these CT images, all voxels with intensities lower than a predefined threshold of 1000 Modified Hounsfield Units (MHU) were considered to belong to the background and thus set to zero prior to segmentation. The minimum permissible object volume was set to 50 cm^3 .

All non-target objects were considered as clutter and chosen to provide an environment that is comparable to that encountered within a transport security setting. Typical clutter items included both low density items (e.g. clothing, books etc.) and high density items (e.g. belt buckles, batteries, pliers, dumbbells etc.). We used separate training and testing datasets, due to the multi-faceted nature of the proposed approach. From the available baggage CT-imagery, sub-volume training and test sets were extracted manually for the target object types under consideration. The handgun training set consisted of 101 manually segmented positive (handgun) subvolumes and 134 manually segmented negative (clutter) subvolumes, while the bottle training set consisted of 88 and 90 manually segmented positive and negative subvolumes respectively (Figure 4). The handgun and bottle test sets consisted of 208/150 and 146/190 positive/negative samples (whole volumes) respectively (Figure 5). Within an operational context, it is expected that a given baggage-CT image would be exhaustively split into sub-volumes, either uniformly with possible inter-subvolume overlap or via the use of object segmentation (Section 3), such that all the sub-volumes are processed (screened) for detection of the target objects (e.g. handguns and

bottles, Figure 5)

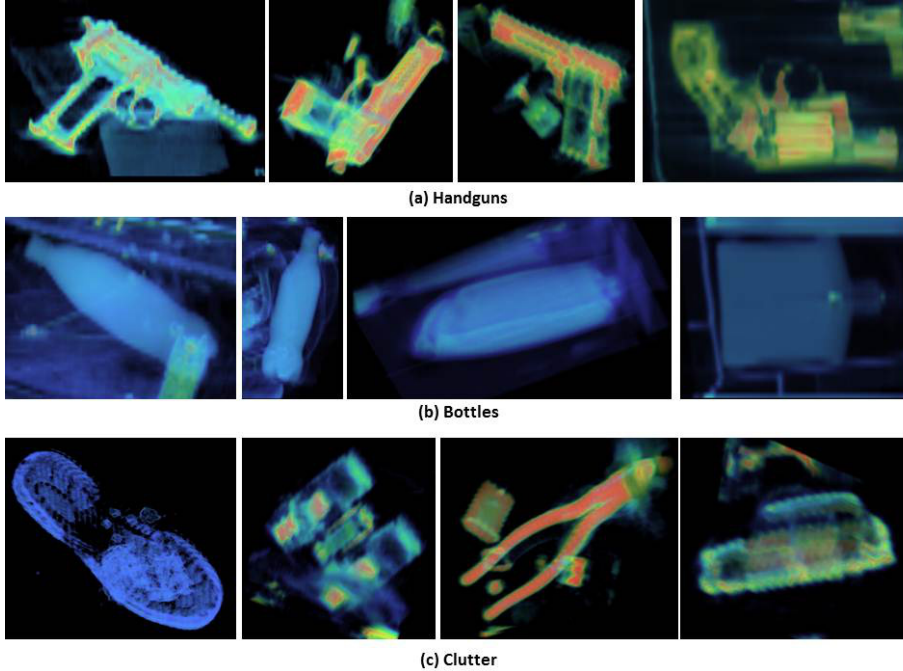


Figure 4: Examples of manually segmented training data.

Optimal algorithm parameters were determined independently for each object class and kept constant for every instance in that experiment (i.e. one set of parameters used for the entire handgun experiment and another set for the entire bottle experiment). Pre-processing parameters were determined using a small set of validation volumes and several different sets of input parameters. Those parameters that produced the most visually satisfying segmentations were chosen. The block-wise linear-time NLM implementation [43], with a block size of 128 pixels, was used to reduce the high processing time of NLM filtering [35, 36]. Due to the efficiency of the random forest-based classification process and the relatively small data-set, the optimal ERC forest and SVM parameters were determined empirically via cross-validation performed over the entire dataset (i.e. those parameters yielding the highest correct classifications rates were chosen). The number of tests performed for each node split in the ERC forest was set to $|\mathcal{T}_j| = 10$ - this value was fixed for all nodes. Trees were grown to a maximum depth of $D = 5$, with a lower bound of $IG_{min} = 10^{-4}$ on the information gain. These settings resulted in tree growth terminating prior to maximum depth

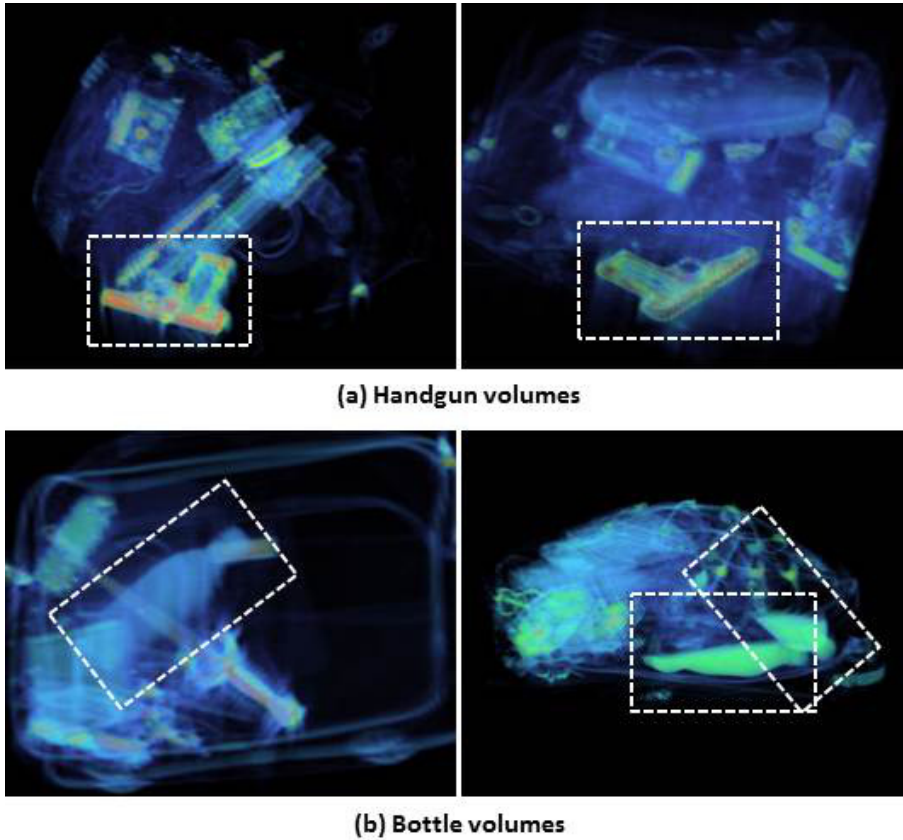


Figure 5: Example test volumes - target object indicated.

and thus no tree pruning was performed. For a forest containing $T = 5$ trees, codebooks typically contained approximately 1500 codewords.

Table 1 summarises the results of the classification experiments. In terms of handgun classification, pre-processing by DDMar (TPR = 97.60%; FPR = 1.33%) and NLM filtering (TPR = 96.63%; FPR = 1.33%) yielded the optimal results. While LIMar (TPR = 95.70%; FPR = 2.67%) outperformed simple intensity thresholding (TPR = 90.87%; FPR = 6.0%), the latter still performed surprisingly well given its simplicity.

In the classification of bottles, NLM filtering correctly classified all positive (bottle) instances (TPR = 100%) while maintaining a low false-positive rate (FPR = 2.11%). DDMar again yielded high correct classification rates (TPR = 98.62%; FPR = 1.58%). The superiority of NLM filtering and DDMar over LIMar (TPR = 89.04%; FPR = 1.58%) was more pronounced compared to the

Table 1: Overall classification performance for tested methods.

Method	Class	TPR (%)	FPR (%)	Prec.	Acc.
Intensity	Handgun	90.87	6.0	0.955	0.922
threshold	Bottle	21.23	3.68	0.816	0.637
NLM filter	Handgun	96.63	1.33	0.990	0.975
[36]	Bottle	100.0	2.11	0.973	0.988
LIMar	Handgun	95.70	2.67	0.980	0.964
[37]	Bottle	89.04	1.58	0.977	0.944
DDMar	Handgun	97.60	1.33	0.990	0.980
[30]	Bottle	98.63	1.58	0.980	0.985

Table 2: Mean per-volume processing times.

Method	Avg. Processing Times (s/volume)		
	Denoising	Segmentation	Total
Intensity threshold	0.23	245.87	246.10
NLM filter [36]	34.01	127.31	161.32
LIMar [37]	49.05	145.67	194.72
DDMar [30]	401.42	129.44	530.86

handgun experiments. Intensity thresholding (TPR = 21.23%; FPR = 3.68%) performed significantly poorer than all three methods, yielding a classification accuracy of only 0.637 (while all other methods yielded accuracies > 0.92).

The mean, per-volume processing times for each of the four pre-processing methods (averaged over both sets of experiments) are shown in Table 2. The use of the average is justified by the fact that similar volumes were used in both experiments (with only the target objects differing) and thus processing times across the two experiments were relatively consistent. As expected, simple intensity thresholding incurred virtually no computational overhead (0.23s per volume). As suggested by the experimentation in [3], the processing times of LIMar (49.05s per volume) and DDMar (401.42s per volume) were higher than that of the NLM filter (34.01s per volume) due to the computational expense associated with the Filtered Back-Projection (FBP) reconstructions. DDMar was, on average, significantly more computationally demanding than the other three methods. As expected, segmentation times were lower for volumes with higher signal-to-noise ratios (i.e. better denoising/artefact reduction). The mean segmentation times for the NLM volumes (127.31s) and the LIMar and DDMar vol-

umes (145.67s and 129.44s respectively) were thus significantly lower than those of the thresholded volumes (245.87s). Although the processing times associated with the final stage of the proposed framework (random forest clustering and SVM classification) are theoretically dependent on the number of objects (i.e. subvolumes) segmented from the original volume, the average times for all four methods were negligible relative to the pre-processing and segmentation stages and are thus not shown in Table 2. The total mean, per-volume processing time was lowest for the NLM-filtered volumes (161.32s).

6. Effects of Denoising and MAR

The relatively high correct handgun classification rates obtained using simple intensity thresholding may be attributed to the predominantly metallic nature of the handguns in the dataset. Their correspondingly high atomic numbers lead to significantly higher intensity values compared to the majority of other commonly encountered, low-density items (e.g. clothing and books) and high-density streaking artefacts. Such high-density items are thus fairly easily isolated, even in cluttered and noisy volumes, using only a single threshold (a property which is exploited in most MAR techniques [3]). The obvious shortcoming of segmentations performed in this manner is the elimination of the majority of the contents of typical baggage scans.

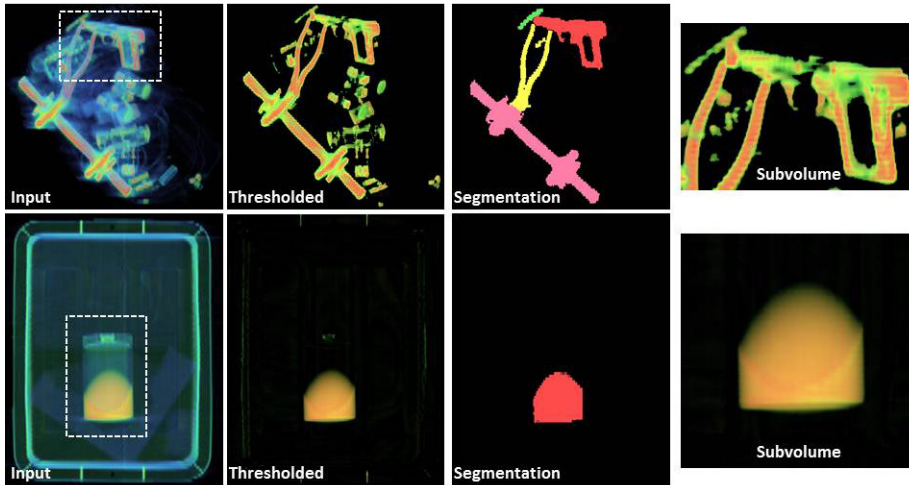


Figure 6: Threshold pre-processing examples. Top row: handgun correctly classified high-density information if not eliminated by the threshold. Bottom row: bottle misclassified as threshold eliminates empty (top) half of bottle.

In the top row of Figure 6 we see an example of a threshold which has been chosen such that the handgun is successfully segmented from a cluttered bag. However, we note that only the high-density objects (the handgun, pliers and dumbbell) in the original scan are accurately depicted in the segmentation map, while the majority of the remaining items are eliminated. This limitation of pre-processing by thresholding is further emphasised by the massive decline in performance when applied to the bottle classification task (where the correct classification rate is significantly lower than random guessing - Table 1). The bottom row in Figure 6 illustrates an example of a bottle-containing volume which was incorrectly classified using intensity thresholding. Note that only the liquid containing region of the bottle is retained in the segmentation while the remaining part of the bottle is eliminated by the threshold. The corresponding subvolume bears little resemblance to a bottle and is thus misclassified. Thresholding is thus only suitable when considering objects with very high densities - it is important to emphasise that these densities need to be higher than the high-density noise and artefacts in the image. In the majority of cases a thresholding approach will not suffice (as illustrated by the poor performance on the bottle dataset). At best, thresholding may be used as an initial screening for high-density threats.

The most important observation that can be made from the results in Section 5 is the high-quality performance of the NLM filtering approach relative to the two MAR-based approaches (both in terms of classification performance as well as processing time). The experimentation presented in [3] suggested that dedicated MAR techniques (e.g. DDMar and LIMar) outperform simple denoising filters (e.g. NLM filter) in the reduction of noise and metal-streaking artefacts in baggage-CT imagery. The results in Section 5, however, bring into question the importance of this superior artefact reduction. Figure 7, for example, shows a handgun-containing volume that was correctly classified by both the DDMar and NLM-filtering approaches. The MAR volume, however, contains notably less streaking than the NLM volume (clearly visible in the final subvolumes) and leads to a handgun segmentation that is superior to the under-segmented NLM handgun - both qualitatively as well as quantitatively (as determined by a higher Random Forest Score (RFS)). The key factor appears to be in the generation of the subsequent subvolumes. Since the original denoised volume information is retained in these subvolumes (as opposed to retaining only the foreground/object information) the errors related to the incorrect labelling in the under-segmented object are largely eliminated - note that both the MAR

and NLM subvolumes contain the object which was incorrectly labelled in the NLM segmentation. In related work [45], it has been shown that classification performance using ERC forests is relatively robust to background noise and clutter, provided the clutter objects are small in relation to the target object (i.e. the target object forms the main part of the subvolume). It is apparent (as illustrated by the additional examples in Figure 8), that for both the handgun and bottle datasets, the NLM segmentations produced sufficiently small under-segmentations to allow for correct classifications. This indicates that the degree of artefact reduction offered by more complex MAR techniques (which comes at a significant increase in computational cost - Table 2) as well as very precise segmentations are not absolutely necessary for successful 3D object classification in complex baggage-CT imagery. This observation is particularly encouraging in the context of security screening, where the demand computational efficiency is paramount to maximise throughput [5]. It is worth emphasising, however, that some degree of denoising is still important, as illustrated by the comparatively poor performance of simple intensity thresholding.

A more detailed analysis of the classification errors produced by the NLM, LIMar and DDMar techniques is necessary. Figure 9 illustrates examples of two subvolumes that resulted in false-positive handgun classifications for each of the three pre-processing techniques. The most obvious trend (evident for all three techniques) is the presence of high-density objects. Beyond this, the subvolumes bear little obvious resemblance to the handguns in the training set (e.g. Figure 4), making it difficult to determine, with any confidence, the root of the missed classifications.

Further investigation of the results have indicated trends in the characteristics of the missed NLM and DDMar handguns (false negatives). In particular, the missed handguns contain relatively large low-density regions in their grips/handles, resulting in over-segmentations of the handguns into separate handle and barrel components. Examples of such handguns (together with their corresponding segmentation maps and relevant subvolumes) are shown in Figure 10 (a) and (b) for the NLM and DDMar techniques respectively. Interestingly, the handgun barrel subvolumes that lead to false negatives are notably similar in appearance to the false-positive generating subvolumes for all three methods (Figure 9) and bear some resemblance to the whole-gun subvolumes. This suggests that the false-positive and false-negative instances lie near to the decision boundary established by the SVM classifier. It is likely that a larger and more diverse training set and a finer tuning of the input parameter space

may alleviate these errors. Note that the fact that handguns with these properties (low-density handles) were included in the classification training data is immaterial, as the errors occur in the segmentation phase of the classification procedure. This highlights the challenge of defining a suitable ‘object philosophy’ in the development of a segmentation algorithm [13]. Hierarchical approaches to segmentation, whereby the relation between the individual parts of composite objects are stored in tree-like structures, allow for multi-part objects to be represented both by their constituent parts as well as single composite objects [14]. The incorporation of such techniques into the proposed framework is likely to be beneficial in the aforementioned scenarios and is left as an area for future work.

We have found that the main cause of false-negative handgun classifications for the LIMar method is an introduction of new streaking artefacts arising from the FBP reconstructions of linearly interpolated sinograms [3], which ultimately leads to under-segmentations (i.e. multiple objects segmented as one). Two such examples are shown in Figure 11. In both cases, the subvolumes containing the handguns also contain the majority of the large items present in the original scans. While the NLM under-segmentations are typically small (Figure 8), the LIMar under-segmentations contain objects similar in size (or larger) than the target objects, thus leading to erroneous classifications.

Figures 12 - 14 show examples of misclassified bottle volumes for each of the three pre-processing techniques (NLM, LIMar, DDMar). The subvolumes that led to false-positive classifications for all three techniques are dominated by objects with circular cross-sections and densities in the range of common liquids (Figure 12). This is a similar observation to that made in related work [12, 2]. It is worth noting that not all such objects resulted in false-positive classifications and the reasons for these particular misclassifications are not clear.

NLM filtering returns a perfect classification of the positive (bottle-containing) test volumes, while DDMar results in only two false negatives. A closer examination of these two cases has indicated that both false positives are caused by half-filled bottles surrounded by high-density objects and hence corrupted (i.e. intersected) by streaking artefacts. One of the two volumes is shown in Figure 13 in addition to the corresponding denoised volume, the segmentation map and the subvolume for NLM filtering. Due to the presence of four high-density objects in the input volume (indicated in Figure 13), the degree of streaking is severe, making it challenging to distinguish the upper border of the bottle. The bottom half of the bottle is more pronounced due to the presence of a higher-

density liquid (relative to the plastic of the bottle). While DDMar successfully removes the artefacts, it also appears to over-compensate and removed valuable image information. Most importantly, the upper border of the bottle is almost entirely eliminated. Consequently, only the liquid-containing region of the bottle is segmented and represented in the corresponding subvolume. In contrast, while NLM filtering removes considerably less streaking (as expected), the entire bottle is retained. The resulting segmentation map, although including background noise in the bottle region, captures all of the relevant information and hence the subvolume is correctly classified. This again illustrates that classification is not affected by inaccurate segmentation provided the error relatively small. Note that, in the case of the DDMar volume, the information related to the upper half of the bottle is eliminated prior to segmentation (since it is the pre-processed volume that is used in the generation of the subvolume). A possible solution may thus be to use the original, unprocessed volume as the input to the subvolume generation.

The false negatives resulting from LIMar again appear to stem from large under-segmentations, caused by the introduction of new streaking artefacts in the MAR procedure. Figure 14 illustrates two such examples: in both instances the post-MAR volumes contain new streaking artefacts, leading to under-segmentations. The limitations of linear interpolation-based MAR in the presence of multiple metal objects, which are already well documented [3], are further substantiated here.

7. Conclusion

We have evaluated the impact of image denoising and Metal Artefact Reduction on 3D object segmentation and classification in low-resolution, cluttered dual-energy CT imagery. To this end, four novel contributions have been made: 1) a materials-based technique for the segmentation of unknown 3D objects from low-resolution, cluttered CT imagery; 2) a high-speed codebook subvolume classification technique using Extremely Randomised Clustering (ERC) forests [32]; 3) a novel, fully automated approach for 3D object classification in cluttered CT imagery (by combining (1) and (2)) which produces correct classification rates in excess of 97% with false positive rates of less than 2% and 4) a full classification-driven performance evaluation of denoising and MAR.

The experimental comparison to investigate the impact on overall performance of the four denoising and Metal Artefact Reduction (MAR) techniques

(intensity thresholding, NLM filtering [35, 36], LIMar [37], DDMar [30]), has demonstrated the superiority of the NLM filtering and DDMar over LIMar and intensity thresholding. NLM filtering is further shown to outperform both MAR techniques in terms of processing time. Considering the comparable classification performance following pre-processing by NLM filtering and DDMar (within 1%), we make the important conclusion that the benefits of superior reduction of artefacts and noise of dedicated MAR do not sufficiently justify the large associated increase in processing time (up to $10\times$).

Future work will thus shift away from denoising and artefact reduction, seeking instead to improve image segmentation by considering a more accurate materials-based characterisation of scans using dual-energy decomposition techniques [4, 51].

Acknowledgments

Parts of this study were originally funded under the Innovative Research Call in Explosives and Weapons Detection (2010), sponsored by the UK Home Office, the UK Department for Transport, the UK Centre for the Protection of National Infrastructure and the Metropolitan Police Service. The authors thank Reveal Imaging Technologies Inc. (USA) for additional support.

- [1] G. van Kaick, S. Delorme, Computed tomography in various fields outside medicine, *European Radiology Supplements* 15 (2005) 74–81.
- [2] G. Flitton, T. Breckon, N. Megherbi, A 3D Extension to Cortex Like Mechanisms for 3D Object Class Recognition, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3634–3641.
- [3] A. Mouton, N. Megherbi, K. van Slambrouk, J. Nuyts, T. Breckon, An experimental survey of metal artefact reduction in computed tomography, *Journal of X-Ray Science and Technology* 21 (2) (2013) 193–226.
- [4] Z. Ying, R. Naidu, C. R. Crawford, Dual energy computed tomography for explosive detection, *Journal of X-ray Science and Technology* 14 (4) (2006) 235–256.
- [5] S. Singh, Explosives detection systems (EDS) for aviation security, *Signal Processing* 83 (1) (2003) 31–55.

- [6] F. Mees, R. Swennen, M. Van Geet, P. Jacobs, Applications of X-ray computed tomography in the geosciences, Geological Society, London, Special Publications 215 (1) (2003) 1–6.
- [7] F. Nachtrab, S. Weis, P. Keßling, F. Sukowski, U. Haßler, T. Fuchs, N. Uhlmann, R. Hanke, Quantitative material analysis by dual-energy computed tomography for industrial NDT applications, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 633 (2011) S159–S162.
- [8] Toshiba Medical Systems Corporation Aquilion 32 <http://www.toshiba-medical.co.uk/ct-systems.asp> ([Jan. 01, 2012]).
- [9] GE Healthcare discovery CT750 HD <http://www.gehealthcare.com/euen/ct/products/> ([Jan. 01, 2012]).
- [10] A. Mouton, N. Megherbi, G. Flitton, T. Breckon, A novel intensity limiting approach to metal artefact reduction in 3D CT baggage imagery, in: Proceedings of the IEEE International Conference on Image Processing, 2012, pp. 2057–2060.
- [11] N. Megherbi, T. Breckon, G. Flitton, A. Mouton, Investigating existing medical CT segmentation techniques within automated baggage and package inspection, in: Proceedings SPIE Security and Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence, Vol. 8901, 2013, pp. 1–8.
- [12] G. Flitton, A. Mouton, T. Breckon, Object classification in 3d baggage security computed tomography imagery using visual codebooks, Pattern Recognition 48 (8) (2015) 2489–2499. doi:10.1016/j.patcog.2015.02.006.
- [13] C. Crawford, H. Martz, H. Piena, Segmentation of objects from volumetric CT data, northeastern University, Boston. This report can be found at: https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf (2013).
- [14] D. F. Wiley, D. Ghosh, C. Woodhouse, Automatic segmentation of CT scans of checked baggage, in: Proceedings of the 2nd International Meeting on Image Formation in X-ray CT, 2012, pp. 310–313.

- [15] L. Grady, V. Singh, T. Kohlberger, C. Alvino, C. Bahlmann, Automatic segmentation of unknown objects, with application to baggage security, European Conference on Computer Vision, Springer, 2012, pp. 430–444.
- [16] G. Flitton, T. Breckon, N. Megherbi, Object recognition using 3D SIFT in complex CT volumes, in: Proceedings British Machine Vision Conference, 2010, pp. 11.1–11.12.
- [17] N. Megherbi, J. Han, T. P. Breckon, G. T. Flitton, A comparison of classification approaches for threat detection in CT based baggage screening, in: Proceedings of the IEEE International Conference on Image Processing, 2012, pp. 3109–3112.
- [18] N. Megherbi, G. T. Flitton, T. P. Breckon, A classifier based approach for the detection of potential threats in CT based baggage screening, in: Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 1833–1836.
- [19] M. Novotni, R. Klein, Shape retrieval using 3D zernike descriptors, Computer-Aided Design 36 (11) (2004) 1047–1062.
- [20] C. Dorai, A. K. Jain, COSMOS-a representation scheme for free-form surfaces, in: Proceedings of the IEEE International Conference on Computer Vision, 1995, pp. 1024–1029.
- [21] A. Ben-Hur, J. Weston, A user’s guide to support vector machines, Methods in Molecular Biology 609 (2010) 223–239.
- [22] E. A. Wan, Neural network classification: A Bayesian interpretation, IEEE Transactions on Neural Networks 1 (4) (1990) 303–305.
- [23] S. R. Safavian, D. Landgrebe, A survey of decision tree classifier methodology, IEEE Transactions on Systems, Man and Cybernetics 21 (3) (1991) 660–674.
- [24] A. Criminisi, J. Shotton, Decision Forests for Computer Vision and Medical Image Analysis, Springer Publishing Company, Incorporated, 2013.
- [25] G. Flitton, T. Breckon, N. Megherbi, A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery, Pattern Recognition 46 (9) (2013) 2420–2436. doi:10.1016/j.patcog.2013.02.008.

- [26] J. Sivic, A. Zisserman, Video Google: A text retrieval approach to object matching in videos, in: Proceedings of the IEEE International Conference on Computer Vision, 2003, pp. 1470–1477.
- [27] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using local affine regions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (8) (2005) 1265–1278.
- [28] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, J. M. Geusebroek, Visual word ambiguity, *IEEE transactions on pattern analysis and machine intelligence* 32 (7) (2010) 1271–1283.
- [29] A. Mouton, N. Megherbi, G. Flitton, T. Breckon, An evaluation of CT image denoising techniques applied to baggage imagery screening, in: Proceedings IEEE International Conference on Industrial Technology, 2013, pp. 1063–1068.
- [30] A. Mouton, N. Megherbi, T. Breckon, K. Van Slambrouck, J. Nuyts, A distance driven method for metal artefact reduction in computed tomography, in: Proceedings IEEE International Conference on Image Processing, 2013, pp. 2334–2338.
- [31] T. R. Johnson, *Medical radiology/diagnostic imaging: dual energy CT in clinical practice*, Springer, 2011.
- [32] F. Moosmann, B. Triggs, F. Jurie, Fast discriminative visual codebooks using randomized clustering forests, *Advances in Neural Information Processing Systems* 19 (2007) 985–992.
- [33] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: *European Conference on Computer Vision*, 2006, pp. 490–503.
- [34] V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York Inc, 2000.
- [35] A. Buades, B. Coll, J. M. Morel, A non-local algorithm for image denoising, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2005, pp. 60–65.

- [36] A. Buades, B. Coll, J. M. Morel, A review of image denoising algorithms, with a new one, *SIAM Multiscale Modeling and Simulation* 4 (2) (2005) 490–530.
- [37] W. A. Kalender, R. Hebel, J. Ebersberger, Reduction of CT artifacts caused by metallic implants., *Radiology* 164 (2) (1987) 576.
- [38] T. Kano, M. Koseki, A new metal artifact reduction algorithm based on a deteriorated ct image, *Journal of X-ray Science and Technology* 24 (6) (2016) 901–912.
- [39] M. Yazdi, Z. Mansourian, Metal artifact reduction in spiral fan-beam ct using a new sinogram segmentation scheme, *Journal of X-ray Science and Technology* 25 (5) (2017) 737–749.
- [40] A. Jaiswal, M. A. Williams, A. Bhalerao, M. K. Tiwari, J. M. Warnett, Markov random field segmentation for industrial computed tomography with metal artefacts, *Journal of X-ray Science and Technology* (2018) 1–19.
- [41] A. Mouton, T. Breckon, A review of automated image understanding within 3d baggage computed tomography security screening, *Journal of X-Ray Science and Technology* 23 (5) (2015) 531–555. doi:10.3233/XST-150508.
- [42] T. W. Rogers, N. Jaccard, E. J. Morton, L. D. Griffin, Automated x-ray image analysis for cargo security: Critical review and future promise, *Journal of X-ray science and technology* 25 (1) (2017) 33–56.
- [43] M. Mahmoudi, G. Sapiro, Fast image and video denoising via non-local means of similar neighborhoods, *Signal Processing Letters, IEEE* 12 (12) (2005) 839–842.
- [44] A. Buades, B. Coll, J.-M. Morel, Image denoising methods. a new nonlocal principle, *SIAM Review* 52 (1) (2010) 113–147.
- [45] A. Mouton, T. Breckon, G. Flitton, N. Megherbi, 3d object classification in baggage computed tomography imagery using randomised clustering forests, in: *Proc. International Conference on Image Processing, IEEE*, 2014, pp. 5202–5206. doi:10.1109/ICIP.2014.7026053.

- [46] A. Mouton, T. Breckon, Materials-based 3d segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening, *Pattern Recognition* 48 (6) (2015) 1961–1978. doi:10.1016/j.patcog.2015.01.010.
- [47] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*, SPIE and John Wiley and Sons, 2003.
- [48] R. E. Alvarez, A. Macovski, Energy-selective reconstructions in X-ray computerised tomography, *Physics in Medicine and Biology* 21 (5) (1976) 733–744.
- [49] Y. Jin, *Implementation and optimization of dual energy computed tomography* [PhD Thesis], University of Erlangen-Nuremberg (2011).
- [50] Z. Ying, R. Naidu, S. Simanovsky, C. R. Crawford, Method of and system for computing effective atomic number images in multi-energy computed tomography, US Patent 7,190,757.
- [51] O. Semerci, E. 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.
- [52] K. Chuang, H. Huang, A fast dual-energy computational method using isotransmission lines and table lookup, *Medical Physics* 14 (1987) 186–192.
- [53] C. Solomon, T. Breckon, *Fundamentals of digital image processing: a practical approach with examples in Matlab*, Wiley-Blackwell, 2010.
- [54] T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, L. Grady, Evaluating segmentation error without ground truth, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, 2012, pp. 528–536.
- [55] A. Criminisi, Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning, *Foundations and Trends in Computer Graphics and Vision* 7 (2-3) (2011) 81–227.
- [56] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: *Proceedings European Conference on Computer Vision*, 2006, pp. 490–503.

- [57] T. Lindeberg, Scale-space theory in computer vision, Kluwer Academic Publishers Netherlands, 1994.
- [58] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using affine-invariant regions, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Vol. 2, 2003, pp. II-319 – II-324.
- [59] G. Flitton, Extending computer vision techniques to recognition problems in 3D volumetric baggage imagery [PhD Thesis], Cranfield University UK (2012).
- [60] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

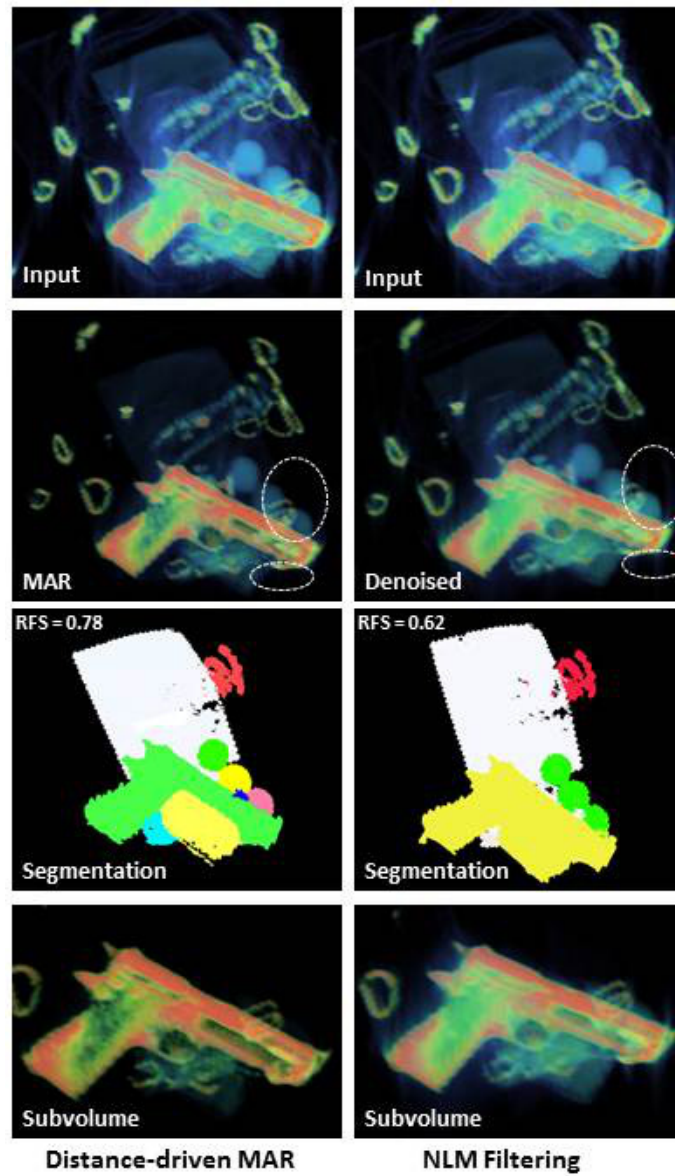


Figure 7: Comparison of DDMar and NLM pre-processed handgun segmentations. Superior artefact removal by MAR indicated in MAR/denoised volumes. Although NLM leads to an under-segmentation of the handgun, the error is small and both methods return correct classifications.

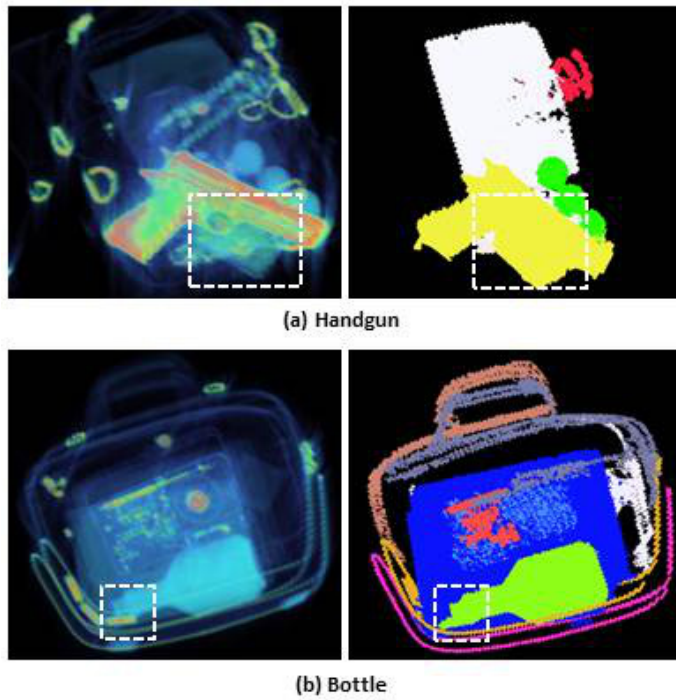


Figure 8: Under-segmentations (outlined regions) resulting from NLM filtering. Since errors are small, classification is not affected.

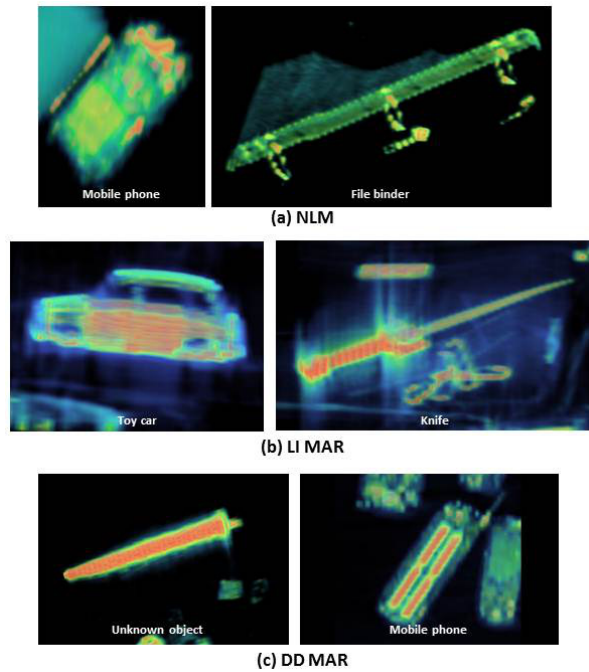


Figure 9: Handgun false-positive examples caused by high-density objects.

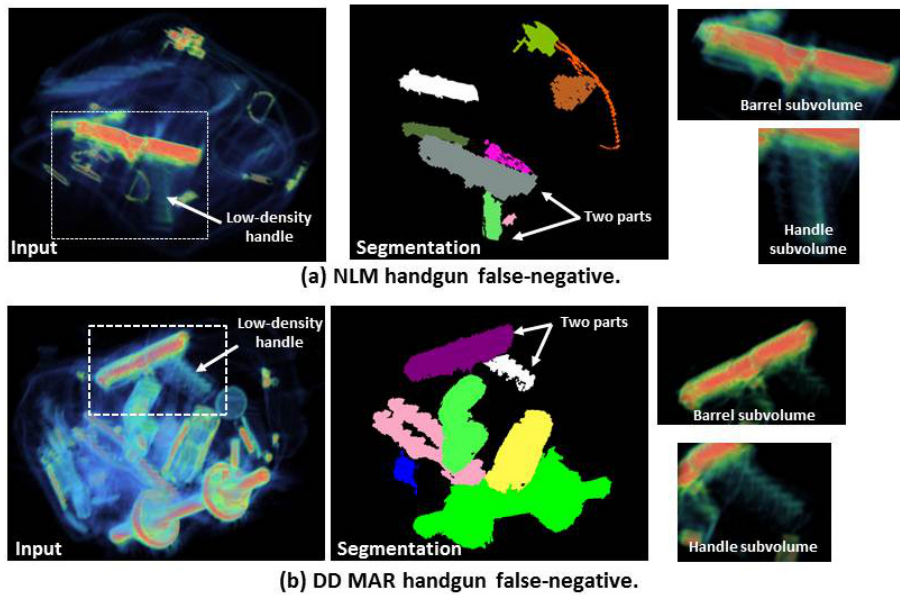


Figure 10: NLM and DDMar handgun false-negative examples. Handguns (outlined) are over-segmented into barrels and handles due to uncharacteristically low-density handles. Resulting subvolumes classified as clutter.

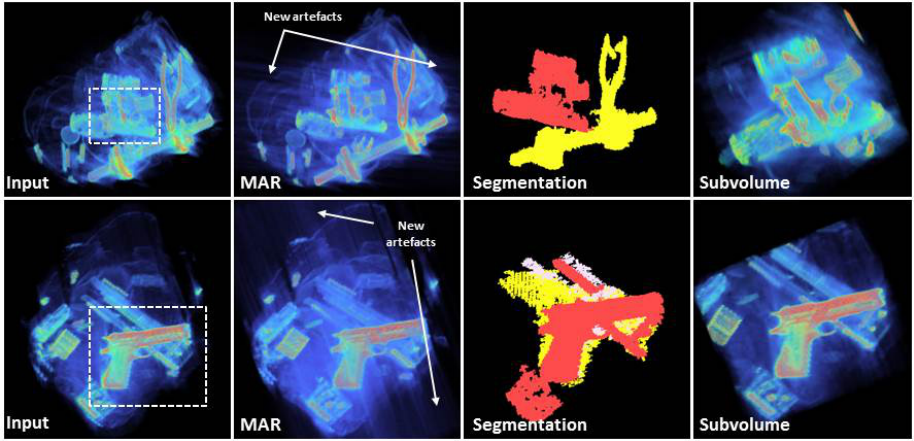


Figure 11: LIMar handgun false-negative examples. New streaking introduced in MAR procedure leads to over-segmentations of the handguns and hence incorrect classifications. Handguns outlined in inputs.

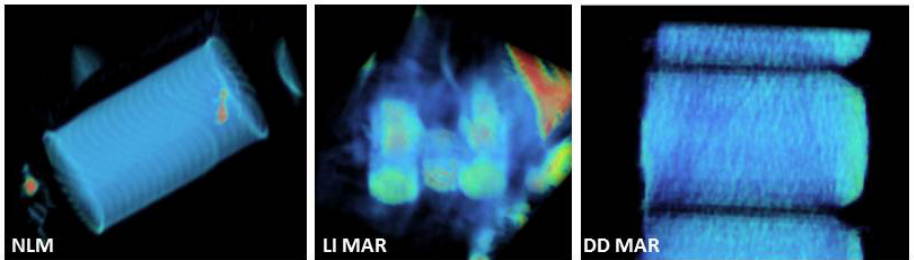


Figure 12: Bottle false-positive examples. Caused by objects with circular cross-sections and similar densities to liquids.

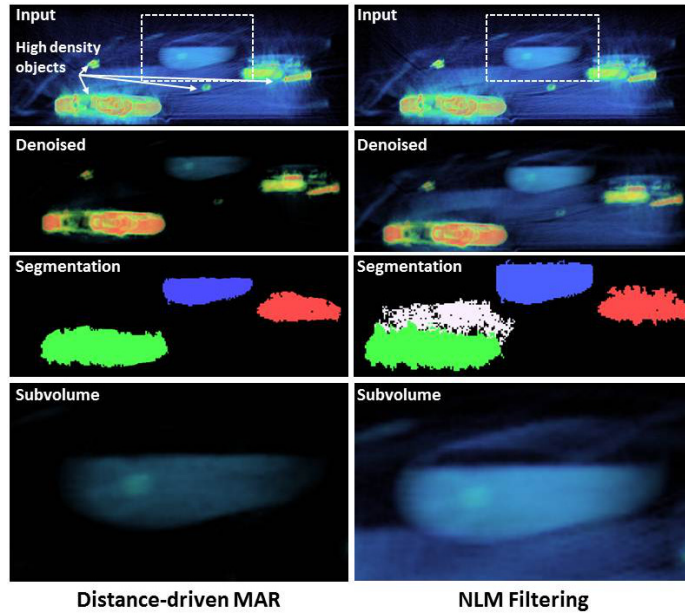


Figure 13: Comparison of DDMar (false negative) and NLM (true positive) bottle segmentation/classification. DDMar removes greater degree of artefacts, eliminating the empty region in the bottle and causing a misclassification.

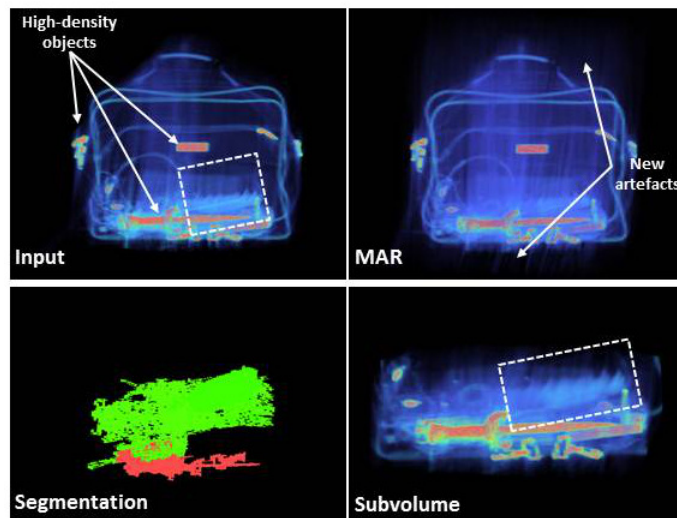


Figure 14: Example of LIMar false-negative classification of a bottle (outlined).