Object Classification in 3D Baggage Security Computed Tomography Imagery using Visual Codebooks

Greg Flitton[a], Andre Mouton[a,*], Toby P. Breckon[b]

[a] *School of Engineering, Cranfield University, U.K.*
[b] *School of Engineering and Computing Sciences, Durham University, U.K.*

Abstract

We investigate the performance of a Bag of (Visual) Words (BoW) object classification model as an approach for automated threat object detection within 3D Computed Tomography (CT) imagery from a baggage security context. This poses a novel and unique challenge for rigid object classification within complex and cluttered volumetric imagery. Within this context it extends the BoW model to 3D transmission imagery (X-ray CT) from its conventional application in 2D reflectance (photographic) imagery. We explore combinations of four 3D feature descriptors (Density Histogram (DH), Density Gradient Histogram (DGH), Scale Invariant Feature Transform (SIFT) and Rotation Invariant Feature Transform (RIFT)), three codebook assignment methodologies (hard, kernel and uncertainty) and seven codebook sizes. Optimal performance is achieved using the DH and DGH descriptors in conjunction with an uncertainty assignment methodology. Successful detection rates in excess of 97% for handguns and 89% for bottles and false-positive rates of approximately 2-3% are achieved. We demonstrate that the underlying imaging modality and the irrelevance of illumination and scale invariance within the transmission imagery context considered here, result in the favourable performance of simpler density histogram descriptors (DH, DGH) over 3D extensions of the well-established SIFT and RIFT feature descriptor approaches.

Keywords: 3D Object classification, Bag of (Visual) Words, 3D descriptors, SIFT, RIFT, baggage-CT

1. Introduction

Baggage screening plays a central role within the aviation security domain [1]. Recent advances in airport-security regulations (European Civil Aviation Conference (ECAC) Standard 3 screening regulations [2]) will see high-speed variants of 3D X-ray Computed Tomog-

*Principal Corresponding Author
Email address: andremouton.email@gmail.com (Andre Mouton)

raphy (CT) scanners, which have enjoyed much success in medical imaging, introduced to the security-screening domain in an attempt to mitigate the limitations of conventional 2D X-ray based scanners (particularly object occlusion, clutter and density confusion) [3]. In summary, high-speed CT imaging gives full 3D voxel representation that overcomes the inherent (and very apparent) problem of inter-object occlusion within conventional X-ray (even with 2 or more views) and furthermore allows for the recovery of 3D materials-based information (e.g. effective atomic number [4, 5]) at each voxel location. In turn, this has led to increased research interest in the potential use of object detection and classification techniques to perform automated-analyses tasks on such 3D-baggage imagery [6–10].

To date, the most prominent application of CT within the security-screening domain has been the materials-based detection of explosives [1]. Dual-Energy Computed Tomography (DECT) [4], whereby objects are scanned at two distinct X-ray energies, provides an effective means for performing such materials-based discrimination (e.g. via the recovery of effective atomic numbers [4, 5]).

As a result of this primary explosives detection objective within the aviation-security domain, DECT baggage scanners have become dominant, offering both CT density as well as materials-based information. Despite overcoming the inherent problem of occlusion within 2D X-ray, demand for high throughput has often meant that 3D baggage-CT imagery typically contains substantial noise, metal-streaking artefacts and voxel resolutions of significantly poorer quality than the modern medical-CT equivalent [1] (Figure 1). Prior work has considered denoising and metal artefact reduction in baggage-CT imagery [11–14], although the impact on object classification within this space remains unproven [6–10].

While there exists a rich resource of literature addressing the topic of automated object recognition in 3D medical imagery, it is important to emphasise that these are typically targeted at specific organs and/or pathologies. This allows for the development of task-specific algorithms and crucially, the incorporation of *a priori* information [15, 16] (which is not viable in the unconstrained security-scanning context considered here).

The simultaneous segmentation of multiple anatomical structures, which is typically addressed as a voxel classification problem, is perhaps more closely related to the classification of multiple objects in baggage-CT imagery. The most significant contributions in this field address the issue of multi-organ segmentation in varied CT imagery [17–21]. In these approaches anatomical context is captured via context-rich features, which describe the relative position of visual patterns in the local anatomy [21]. These features are then used to build,

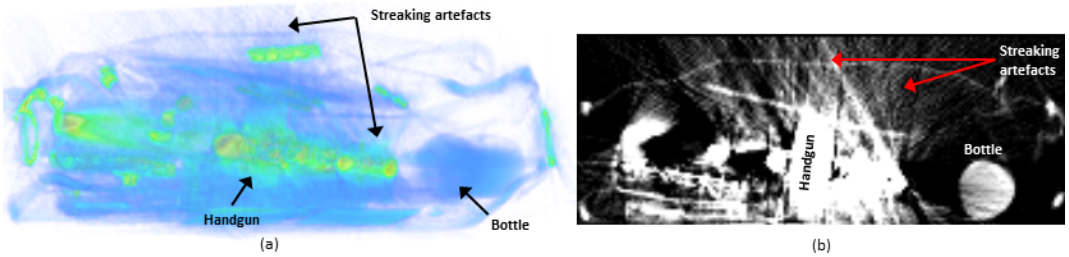


Figure 1: Example baggage-CT scan containing handgun and bottle. (a) Volumetric rendering. (b) Single 2D axial slice. Handgun, bottle and artefacts indicated.

for example, a random-forest-based spatial-context model [20, 21], which is ultimately used in the voxel classification. Therefore, despite targeting the classification of multiple organs, a depedence on *a priori* information is still prominent.

In this study we consider the automatic recognition of two distinct object types, namely handguns and bottles. While this would suggest that the incorporation of *a priori* information (particularly shape characteristics) would be beneficial, it is important to emphasise that the restriction to these two object types has been dictated by limitations in the currently available datasets. In reality, the ultimate objective in baggage screening is more complex. Particularly, it is a requirement to detect the sub-parts of disassembled objects (e.g. components of handguns, weapons and/or explosives). Therefore, while *a priori* information related to the geometric properties, the X-ray attenuation characteristics and the spatial relations of the structures being scanned is readily available in medical imagery, this is generally not the case in the baggage-CT domain, rendering many of the state-of-the-art medical techniques infeasible.

To these ends, approaches that rely on strong shape priors [22–24] break down against this inherent requirement to detect isolated sub-parts of threat objects (e.g. the barrel of a handgun). Such approaches [22–24] extended to 3D, are commonly used to recognise objects in their entirety or under occlusion - not where additional shape boundaries are introduced by potential disassembly in actual 3D real-world space (making any 3D shape signature invalid in the conventional sense - as new non-occlusion bounded edges are introduced). The flexibility of the BoW model [25], which is structureless in the object feature sense, is therefore considered highly correlated with the unique challenges of object classification in this problem space.

There is limited prior work in the area of automatic recognition of items in scanned baggage-CT imagery. A larger resource of literature exists concerning automated object recognition in 2D X-ray imagery. In particular, several techniques have been presented for the detection of handguns. Nercessian et al. [26] use edge detection to characterise handgun features but only considering handguns in fixed orientations. A small dataset was used (40

handgun images, 400 clutter images) with two simple examples of handgun detection shown but no statistical results presented. Oertel and Bock [27] present an approach for that may be categorised as an instance of specific item recognition: one type of handgun was characterised using the distinguishing features of its trigger, hammer and spring. Regions of interest are created for each pixel on an edge contour and a descriptor is constructed from the distribution of white and black pixels in the local neighbourhood. While it is unclear if the approach is invariant to rotation in the horizontal plane, it is not invariant if the weapon is rotated out of this plane. A small dataset was used (30 training images; 10 test images) and again no quantitative results were presented. Gesick et al. [28] search for the handgun trigger in the edge information of the scans in a similar fashion to [27]. The proposed approach is not rotation invariant and no quantitative results have been presented.

Bastan et al. [29] recently applied the BoW model to colour 2D dual-energy X-ray images of baggage items to detect handguns and are the first to report detection results. Investigation of a variety of interest point detectors (DoG, Hessian-Laplace, Harris, FAST, STAR) coupled with three descriptors (SIFT, SURF, BRIEF) was made. Whole baggage items were considered, as opposed to adopting a sliding-window approach, which raises the complexity of the recognition task. In the classification of baggage containing handguns they reported that the method does not work well in isolation but results can be improved using the extra information available from the colour image (indicating material type). In an extension of [29], Turcsany et al. [30] present a BoW model using the Speeded-Up Robust Features (SURF) [31] and a Support Vector Machine (SVM) [32] classifier for automated object recognition within 2D X-ray baggage imagery. Correct classification rates in excess of 99% and a false positive rate of approximately 4% are demonstrated on a diverse dataset.

Comparatively few studies consider 3D volumetric baggage-CT imagery. Bi et al. [33] attempted handgun detection in baggage-CT imagery. The work did not involve processing the 3D data directly as the problem was reduced to searching for the characteristic cross-section of the handguns and appeared preliminary in nature as no explicit quantitative or qualitative detection results are presented. Further work by the same author [34] proposed a methodology for the detection of planar materials within CT-baggage imagery using a 3D extension to the Hough transform [35]. Megherbi et al. [36] investigated the detection of potential threats in CT volumetric data using curvature information captured via a normalised histogram of shape index descriptor [37]. While correct classification rates in excess of 98.0% are presented on a relatively small dataset, curvature properties are shown to be beneficial

for particular exemplar items only (namely, bottles) and are severely limited in the presence of noise.

The observation that curavature information is of limited value in the current context has been substantiated by Mouton [38], who demonstrates the ineffectiveness of curvature-based descriptors in the representation of typical 3D objects segmented from baggage-CT imagery.

Extending upon their earlier work, Flitton et al. [7] present an experimental comparison of four interest point descriptors (the Density Histogram (DH) descriptor [7]; the Density Gradient Histogram descriptor (DGH) [7]; 3D SIFT [6] and 3D RIFT [39, 40]) in the detection of known rigid objects within low resolution, cluttered volumetric CT imagery. It is shown that the simpler density-statistics-based descriptors (DH and DGH) outperform the more complex 3D descriptors (SIFT and RIFT) due to the low-resolution imagery and the high level of noise and artefacts. It is worth emphasising, however, that object detection is achieved using a traditional instance-specific recognition approach, whereby a particular reference object is identified (e.g. a handgun) and pose estimated within a given unknown volume. The study does not consider the task of generalised object classification by type, as we consider in this study.

In the domain of object class recognition, the Bag of (Visual) Words (BoW), or codebook, approach has been met with considerable success [25, 41–43]. In terms of object recognition in 3D baggage-CT imagery, Flitton et al. [8] compare the performance of a 3D visual cortex-based approach to a BoW model using the 3D SIFT descriptor [6]. The cortex-based approach is shown to outperform the BoW approach in the detection of handguns and bottles in subvolumes. Mouton *et al.* [10] demonstrate a further improvement over the 3D visual cortex model in terms of classification accuracy and processing time using a codebook approach based on Extremely Randomised Clustering (ERC) forests [44], a dense feature sampling strategy [45] and an SVM classifier [46]. In particular, correct classification rates in excess of 98% and false positive rates of less than 1%, in conjunction with a reduction of several orders of magnitude in processing time are demonstrated for the 3D object classification in subvolumes. While these studies have not considered the impact on performance of the BoW model parameters (e.g. codebook size, descriptor type, cluster assignment strategy), related works [7, 47] suggest the potential for improved recognition performance provided the optimal BoW model parameters are determined.

We expand on these earlier works by providing a comprehensive performance evaluation of the suitability of the BoW model for the automated recognition of rigid threat-like ob-

jects (handguns and bottles) in low resolution, noisy and complex 3D baggage-CT imagery. The study examines the effectiveness of this object-classification paradigm within complex and cluttered transmission imagery (X-ray CT) and challenges convention with regard to relative 3D descriptor performance within this imaging modality when compared to regular reflectance (photographic) imaging. We determine the optimal combination of model parameters by evaluating and comparing the performance of the four descriptor types described in [7] (DH, DGH, 3D SIFT, 3D RIFT) and three cluster assignment methodologies (hard, kernel, uncertainty) within a supervised machine learning framework based on the Support Vector Machine (SVM) classifier [46].

2. 3D Interest Point Detection

An overview of the basis of interest point detection and description is provided here - for a comprehensive description, the reader is referred to the literature [7].

2.1. Interest Point Detection

We perform interest point detection using a 3D extension to the SIFT algorithm [6, 48].

Scale-space extrema detection: The 3D input volume $I(x, y, z)$ is convolved with a 3D Gaussian filter $G(x, y, z, k\sigma)$ at multiple scales to generate a series of multi-scale Difference-of-Gaussian (DoG) volumes:

$$DoG(x, y, z, k) = I(x, y, z) \star G(x, y, z, k\sigma_s) - I(x, y, z) \star G(x, y, z, (k-1)\sigma_s) \qquad (1)$$

where $\star$ represents the convolution operator; k is an integer in the range $\{1..5\}$ representing the scale index; $\sigma_s = \sqrt[3]{2}$ and (x, y, z) are defined in voxel coordinates. Subsequently a three level pyramid ($L = 0, 1, 2$) is constructed up by subsampling the Gaussian filtered volume for $k = 4$ and repeating the process. Initial candidate keypoints are identified as the local extrema of the DoG images across scales. A given voxel is considered a maximum/minimum if it is a maximum/minimum amongst its 26 neighbours at the same scale and its 27 corresponding neighbours in each of the neighbouring scales in $\mathbb{R}^3$ voxel space.

Keypoint refinement: The candidate keypoints are refined by rejecting poor-contrast keypoints and keypoints poorly localised on edges. The removal of low-contrast keypoints (determined by a density threshold $\tau_c = 0.05$) eliminates points that are likely to produce unstable descriptors and points associated with metal artefacts. Keypoints located along edges are likely to produce unstable descriptors in the presence of noise. Such keypoints are removed via a second threshold $\tau(e)$:

$$\text{Reject if} \quad \frac{Trace^3(H)}{Det(H)} > \frac{(2\tau_e + 1)^3}{(\tau_e)^2} \tag{2}$$

where H is the 3×3 Hessian matrix at a given candidate point; $Det(H)$ and $Trace(H)$ are respectively the determinant and trace of the Hessian. In this work, a value of $\tau_e = 40$ is used and, hence, points where $\frac{Trace^3(H)}{Det(H)} > 332.15$ are rejected.

Keypoint localisation: Finally sub-voxel estimates of the true locations of the extrema are determined by quadratic interpolation of the DoG volumetric data.

2.2. Interest-Point Neighbourhood Weighting

A Gaussian window function, $w(d, \sigma)$, is used to limit the contribution of voxels around the point of interest to those in the local neighbourhood [48]:

$$w(d, \sigma) = exp\left[-\left(\frac{d}{\sigma}\right)^2\right] \tag{3}$$

where d is the voxel distance from the point of interest to the contributing voxel and σ defines the extent of the local contribution. This function is used in conjunction with each of the four descriptors described in Sections 3.1 - 3.4. Given the definition of distance in voxel units, this window will remain consistent with the resolution of the volume being examined.

3. 3D Point-of-Interest Descriptors

Based on the definitions provided in [7], four interest point descriptors are considered (in increasing complexity): Density Histogram (DH); Density Gradient Histogram; 3D SIFT and 3D RIFT [40]. On completion, all descriptors are normalised to unit area. For a more comprehensive description, the reader is again referred to [7].

3.1. Density Histogram (DH) Descriptor

This descriptor defines the local density variation at a given interest point as an N-bin histogram defined over a continuous density range. Given the local area function $w(d_k, \sigma)$ (Equation 3), an addition of $w(d_k, \sigma)$ is made to the appropriate histogram bin where d_k is the voxel distance from the point of interest to the voxel k.

3.2. Density Gradient Magnitude Histogram Descriptor, (DGH)

The Density Gradient Histogram (DGH) for a given interest point quantifies the distribution of the density *gradient* magnitude in the local neighbourhood of that point. The density

gradient magnitude is calculated for all voxels in the volume using a central difference formulation to ensure that the gradient location is aligned with the voxel grid. An addition of $w(d_k, \sigma)$ is again made to the appropriate histogram entry. It is worth noting that, due to the rotational variance of the objects under consideration for detection, the gradient *magnitude* is used as opposed to the gradient *orientation* (frequently used for recognition tasks in 2D [49]).

3.3. Rotation Invariant Feature Transform, (RIFT)

Lazebnik et al. [40] developed the Rotation Invariant Feature Transform (RIFT). The RIFT descriptor examines the local neighbourhood gradients with reference to a radial vector emanating from the point of interest. Histograms are constructed from the gradient orientation and magnitude. Multiple histograms are derived following segmentation of the local neighbourhood into a series of rings centred on the point of interest. RIFT has been shown to operate well in standard 2D imagery and is used in our work as it is more complex than the simple histograms described above but is not as complex as the SIFT descriptor [40, 48]. For a detailed description of the 3D extension of RIFT used here, see [7].

3.4. 3D Scale-Invariant Feature Transform (SIFT)

This descriptor is closely modelled on that used in [6, 50], considering the correct definition of 3D orientation, based on azimuth, elevation and tilt. Volume gradients are examined in a two stage process to locally establish an invariant orientation to be used in the subsequent keypoint description. Firstly, a dominant *direction* for a keypoint is determined by computing a 2D direction histogram that groups the Gaussian filtered volume gradients in bins dividing azimuth and elevation into 45° sections (with $N_a = 8$ azimuth bins and $N_e = 4$ elevation bins). A regional weighting is applied to the gradients according to their voxel distance from the keypoint location. The dominant directions of the keypoint are determined by locating the peaks in this histogram and refining them via interpolation. Histogram peaks within 80% of the largest peak are also retained as possible secondary directions [48]. Secondly, the keypoint *orientation* is determined by calculating the tilt angle for each derived direction. This is accomplished by re-orientating the volume around the keypoint and constructing a 1D tilt histogram that resolves the gradients orthogonal to the dominant direction. This histogram is again built in 45° bins using the same regional weighting method as for the direction histogram. The interpolated peaks in the tilt histogram are used to derive an

estimate of the keypoint tilt. Again, peaks within 80% of the largest peak are retained to give secondary orientations.

Keypoint description is obtained by constructing an $N_g \times N_g \times N_g$ grid of gradient histograms, with each histogram being computed from a $N_v \times N_v \times N_v$ voxel grouping. Each gradient histogram is derived by splitting both azimuth and elevation into 45° bins. Consequently, each descriptor, normalised to unity, contains $N_g^3 \times N_a \times N_e$ elements. For further details of this 3D SIFT extension the reader is referred to [6].

4. The Bag of (Visual) Words (BoW) Model

Given the requirement for detecting the components of disassembled objects and the related limitations of traditional shape-based approaches (see Section 1), we seek a shape-independent representation/model for object recognition. The Bag of (Visual) Words (BoW), or codebook, model [25] is one such approach which has enjoyed success in various object recognition and image classification tasks. Traditionally, the BoW model is composed of the following steps: [51]: 1) feature detection and description (Section 2 and 3); 2) visual codebook generation and 3) classification. A similar approach is adopted here.

A k-means clustering algorithm [52] is employed to calculate a set of cluster centres from a set of descriptors obtained from a series of baggage items. For each baggage item, a single BoW vector is then obtained via vector quantisation of the set of descriptors describing that bag [25]. The assignment methodology used in the vector quantisation is known to have an impact on performance [40]. We consider the performance of three assignment methodologies: hard assignment, kernel assignment and uncertainty assignment.

Hard assignment is the original and the most basic codebook assignment approach whereby every descriptor is assigned to the cluster centre, c_i, where the distance $D\left(d_m, c_i\right)$ is minimal over all c_i. Essentially the closest cluster to the descriptor is taken as the only possible assignment and an allocation to that cluster proceeds:

$$CB_H(i) = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k} \begin{cases} 1 & \text{if } w_i = \arg\min\left(D\left(d_m, c_i\right)\right) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where k is the number of clusters and M is the number of descriptors in the volume. Note the normalisation by M to ensure that the same histogram is constructed, regardless of the number of contributing descriptors.

Kernel assignment: In order to overcome the problems associated with a hard assignment methodology (e.g. quantisation errors [47]) an assignment that considers the possibility

of error is used. A simple Gaussian kernel is used to provide the assignment ambiguity in the codebook - assigning values to the codebook as a function of the distance from the descriptor to the cluster centre:

$$CB_k(i) = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k} \exp\left(-\frac{1}{2} \left(\frac{D\left(d_m, c_i\right)}{\sigma} \right)^2 \right) \tag{5}$$

where k is the number of clusters; M is the number of descriptors in the volume and σ is the smoothing parameter defining the degree of assignment 'fuzziness' (i.e. the degree to which assignments to adjacent clusters are made). A known shortcoming of kernel assignment occurs when the smoothing parameter of the kernel is much smaller than the nearest cluster centres (in terms of Euclidean distance), resulting low weightings being assigned to 'important' clusters.

Uncertainty assignment: The aforementioned limitation of the kernel assignment procedure may be addressed by adopting a normalisation scheme whereby each descriptor contributes the same cumulative (sum) value to the codebook. This normalisation ensures that each descriptor only contributes a sum total of 1.0 to the codebook and eliminates the low values (weak contributions) that can occur using kernel assignment. The uncertainty assignment is then given by:

$$CB_U(i) = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k} \frac{\exp\left(-\frac{1}{2} \left(\frac{D(d_m, c_i)}{\sigma} \right)^2 \right)}{\sum_{j=1}^{k} \exp\left(-\frac{1}{2} \left(\frac{D(d_m, c_j)}{\sigma} \right)^2 \right)} \tag{6}$$

where k is the number of clusters; M is the number of descriptors in the volume and σ the smoothing parameter. Van Gemert et al. [47] have demonstrated that uncertainty assignment approach yields the highest true-positive rates in the task of 2D scene classification.

5. Object Classification Methodology

A set of training descriptors is extracted from a given set of training volumes (Sections 2 and 3). These descriptors are passed to the k-means algorithm to derive a set of cluster centres or visual words (Section 4). The k-means algorithm is initialised using the algorithm proposed by Arthur and Vassilvitskii [53], which has been shown to improve the speed of convergence over a random initialisation of the cluster centres. The algorithm is prone to sub-optimal solutions when the initial cluster centres are randomly assigned. Therefore, the algorithm is executed 10 times and the result with the minimal cluster compactness is chosen. The resulting cluster centres constitute the codebook words over which training descriptors then

undergo vector quantisation using a chosen assignment method (Section 4). Classification is accomplished via a Support Vector Machine (SVM) classifier [46] with a Gaussian Radial Basis Function (RBF) kernel (controlled by two parameters: cost C and the kernel width γ). The optimal values of $(C;\gamma)$ are derived using a grid-search approach over a 10-fold cross-validation. A further 10-fold cross-validation procedure is used to obtain the final classification results. Traditional True-Positive Rates (TPR) and False-Positive Rates (FPR) are used as performance measures.

6. Evaluation Imagery

We consider the classification of two target objects, namely handguns and bottles. The restriction to these two object types has been dictated by limitations in the currently available datasets, which has been derived from a recognised test set of baggage items used for the validation of human-in-the-loop studies such as threat resolution [54], Explosive Detection Systems (EDS) [1] and threat image projection [9, 55].

The data was obtained from a CT80-DR dual-energy baggage-CT scanner manufactured by Reveal Imaging Inc, designed specifically for materials-based explosives detection. A fan-beam geometry was employed with a focus-to-isocentre distance of 550mm, a focus-to-detector distance of 1008.4mm and nominal tube voltages of 160kVp and 80kVp. A volumetric representation of a given bag is obtained by *stacking* its FBP-reconstructed 2D axial slices (512×512). The data is characterised by anisotropic voxel resolutions of $1.56 \times 1.61 \times 5.00$mm.

Anisotropic voxel resolutions are known to negatively impact both human and computerised detection rates [1, 11]. Therefore, the anisotropic volumes have been resampled (using cubic spline interpolation) to create isotropic voxel resolutions of 2.5x2.5x2.5mm. The interpolation results are not hard-limited to the range $\{0.0 \Rightarrow 1.0\}$ - the working voxel value range is thus extended to $\{-1.0 \Rightarrow 2.0\}$. Use of this extended voxel value range in subsequent descriptor formulations (Section 3) needs to be noted.

We construct two distinct datasets for each of the target classes (handguns and bottles). Each dataset is composed of the given target object scanned in random poses (to obtain rotational invariance) and isolated (in a sliding window fashion) prior to feature extraction (Figure 2). The two object classes are considered independently of one another. All non-target objects are considered as clutter (e.g. clothing, books, mobile phones etc.) and are chosen to provide an environment that is representative of that encountered within the transport infrastructure. The handgun dataset consists of 284 target volumes and 971 clutter volumes,

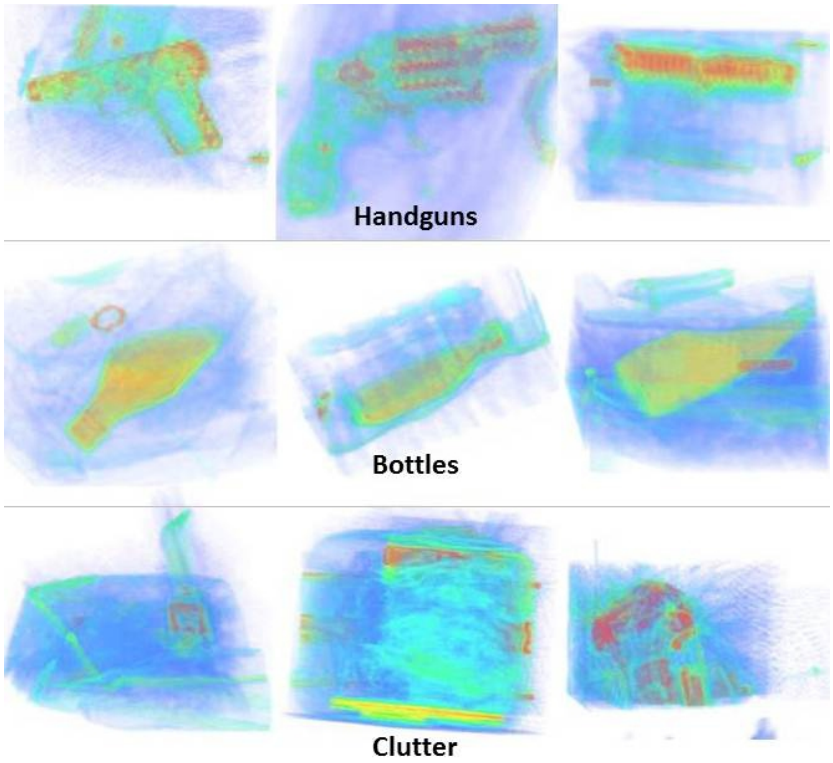


Figure 2: Examples of subvolumes used in study.

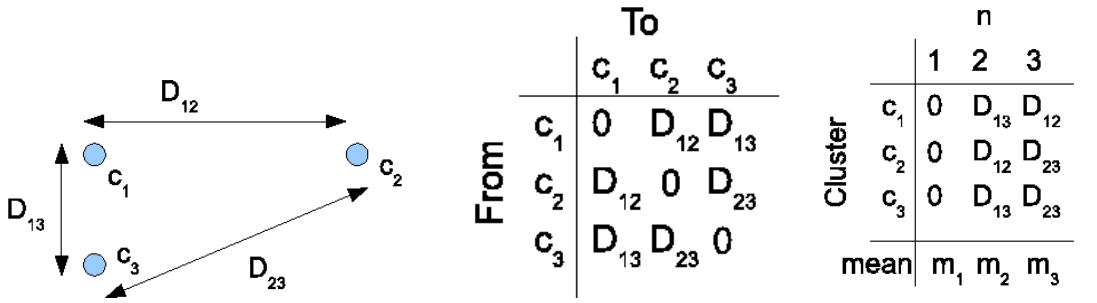
while the bottle dataset consists of 534 target volumes and 1170 clutter volumes.

We emphasise that ideally a larger dataset, incporating an extended set of target objects is desirable. Unfortunately, given the sensitive nature of security-CT imagery (particularly threat-containing data), as well as the tight regulations surrounding access to baggage-CT scanners, the data gathering process is not straightforward. Expanding the current dataset has therefore not been feasible for this work but is highlighted as an important area for future work.

7. Results

Four sets of descriptors (DH, DGH, RIFT and SIFT) were calculated for each of the aforementioned datasets. Codebooks were generated using varying combinations of assignment methods (hard, kernel or uncertainty) and parameter values (k, σ (Section 4)). For each resulting codebook a 10-fold cross validation procedure was performed and the mean and standard deviation for both true-positive and false-positive rates recorded. In the interest of brevity, we have chosen to display only the optimal results for each experiment.

The kernel and uncertainty-assignment methods rely on the smoothing parameter, σ, to determine the influence on neighbouring clusters (Equations 5 - 6). The choice of a suitable



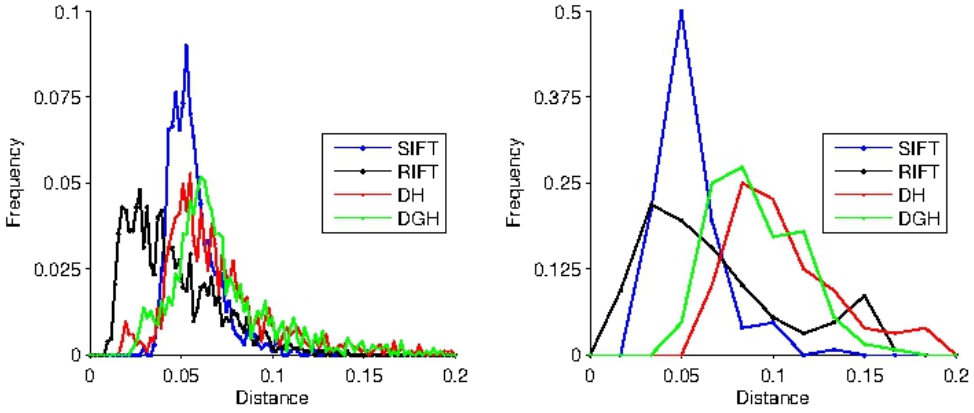
(a) Three example cluster centres (b) Distance matrix (c) Sorted distance matrix

Figure 3: Cluster distance and sorting

value for σ is dependent on the cluster spacings for a given dataset as well as the value of k. Optimal values for the parameters k and σ used in the clustering and assignment procedures were thus determined according to procedure illustrated in Figure 3. In Figure 3 (a) $k = 3$, resulting in three cluster centres (c_1, c_2, c_3) separated by the distances indicated. These distances are stored in a square distance matrix (3 (b)). The rows of this matrix are sorted in ascending order. Figure 3 (c) illustrates these distances as well as the mean of each column, m_n, representing the mean distance to the n^{th} closest cluster. The second column of this sorted matrix contains the mean distance to the nearest adjacent cluster centre for a given k-means clustering operation. A histogram over the nearest adjacent cluster distances is constructed using the information in column 2 of the sorted distance matrix (Figure 3 (c)). For a given value of k, this histogram exhibits peaks (for each descriptor type) for some range of distances. The range which captures the peaks of all four descriptor types then provides an indication of an appropriate range of values for σ using the a codebook of size k.

In this way, we computed the σ ranges $k = 1024$ and $k = 128$. are determined. For $k = 1024$ (Figure 4 (a)), the SIFT, DH and DGH descriptors have peaks in the region 0.05 to 0.06, while the RIFT descriptor peak is nearer to 0.025, though its distribution is quite broad. These values indicate that $0.02 \leq \sigma \leq 0.06$ should be used for the kernel and uncertainty assignment methodologies when using 1024 clusters. For $k = 128$ (Figure 4 (b)) the SIFT descriptor peaks at a distance of 0.05, while the DH and DGH distributions have shifted from ≈ 0.06 (for $k = 1024$) to ≈ 0.09. As these results were not conclusive, we ultimately chose to use a range of σ values (as per prior works [47, 56]), where the ranges were selected based on the peaks in Figures 4 (a) and (b). We further evaluated performance for a range of codebook sizes (for all assignment methods): $k = 32$, 64, 128, 256, 512, 1024, 2048.

Caution associated to airport security-screening domain dictates the acceptability of a



(a) Case: $k = 1024$ (b) Case: $k = 128$

Figure 4: Nearest adjacent-cluster distance histograms

Descriptor	**k**	**TPR (%)**	**FPR (%)**
DH	256	96.1 ± 2.0	1.4 ± 1.3
DGH	2048	97.2 ± 2.2	2.1 ± 1.2
RIFT	512	83.0 ± 3.5	4.5 ± 1.7
SIFT	256	83.0 ± 5.4	3.4 ± 2.3

Table 1: Handgun detection: optimal detection results using hard assignment

low false-positive rate at the expense of even lower false negatives (i.e. missed items) [57]. Therefore, the optimal parameters for each experimental configuration presented hereafter, were chosen as those maximising the True-Positive Rate (TPR) and hence minimising the false-negative rate.

7.1. Handgun Results

The following optimal handgun TPRs were recorded when using hard assignment (Table 1): DH = 96.1% ($k = 256$); DGH = 97.2% ($k = 2048$); RIFT = 83.0% ($k = 512$) and SIFT = 83.0% ($k = 256$). Surprisingly, the DH and DGH descriptors significantly outperformed the SIFT and RIFT descriptors, both in terms of true-positive TPRs and FPRs (incidentally, this was true for all values of k).

In addition to varying k, kernel assignment performance was measured over a range of σ = 0.02, 0.04, 0.08, 0.16. For each descriptor type, the following optimal detection results were recorded (Table 2: DH = 97.3% ($k = 1024$; $\sigma = 0.04$); DGH = 96.8% ($k = 2048$; $\sigma = 0.04$); RIFT = 86.9% ($k = 1024$; $\sigma = 0.02$) and SIFT = 85.8% ($k = 2048$; $\sigma = 0.08$).

The differences in the optimal TPRs recorded for the hard-assignment and the kernel assignment methodologies for each descriptor fall within the margin of error measured across the 10 folds and are therefore unlikely to be statistically significant. Based on the measured

Descriptor	**k**	σ	**TPR (%)**	**FPR (%)**
DH	1024	0.04	97.3 ± 3.4	1.8 ± 1.7
DGH	2048	0.04	96.8 ± 2.6	1.4 ± 1.3
RIFT	1024	0.02	86.9 ± 5.4	4.7 ± 2.0
SIFT	2048	0.08	85.8 ± 4.3	3.3 ± 1.8

Table 2: Handgun detection: optimal detection results using kernel assignment

Descriptor	**k**	σ	**TPR (%)**	**FPR (%)**
DH	2048	0.02	97.2 ± 2.1	1.6 ± 1.4
DGH	512	0.04	97.2 ± 2.8	2.1 ± 1.3
RIFT	2048	0.01	87.3 ± 3.9	5.1 ± 2.3
SIFT	1024	0.02	87.0 ± 5.4	3.8 ± 2.4

Table 3: Handgun detection: optimal detection results using uncertainty assignment

measurement error, however, the DH and DGH descriptors outperform SIFT and RIFT. Although the values of the smoothing parameter, σ, are relatively coarse, the values used for DH and DGH substantiate the observations in Figure 4. Likewise, the setting for RIFT is in line with its lower distance histogram result (Figure 4 (a)). With SIFT a higher setting for σ = 0.08 is observed. This setting is above the peak in the distance histogram (Figure 4 (a)), suggesting that the SIFT clusters are less distinct as visual words and require greater spread in the assignment.

Uncertainty assignment performance was similarly measured over a range of σ = 0.005; 0.01; 0.02; 0.04; 0.08; 0.16 . This extended range was selected based on initial experimentation which showed a noticeable decline in performance for all descriptors at $\sigma = 0.08$ and $\sigma = 0.16$. The optimal handgun detection results are illustrated in Table 3: DH = 97.2% (k = 2048; σ = 0.02); DGH = 97.2% (k = 512; σ = 0.04); RIFT = 87.3% (k = 2048; σ = 0.01) and SIFT = 87.0% (k = 1024; σ = 0.02).

The density histogram (TPR = 97.2%; FPR = 1.6%) and density-gradient histogram (TPR = 97.2%; FPR = 2.1%) descriptors again yield the optimal detection rates, while SIFT (TPR = 87.0%; FPR = 3.8%) and RIFT (TPR = 87.3%; FPR = 5.1%) perform significantly poorer.

7.2. Bottle Results

The optimal bottle detection rates using hard assignment are shown in Table 4: DH = 87.0% (k = 64); DGH = 80.3% (k = 1024); RIFT = 73.9% (k = 128) and SIFT = 78.5% (k = 64). Surprisingly, the DH and DGH descriptors significantly outperformed the SIFT and RIFT descriptors, both in terms of true-positive TPRs and FPRs (incidentally, this was true for all values of k). Although the DH and DGH descriptors outperformed the SIFT and RIFT

Descriptor	k	TPR (%)	FPR (%)
DH	64	87.0 ± 4.9	2.7 ± 1.1
DGH	1024	80.3 ± 4.7	4.7 ± 2.3
RIFT	128	73.9 ± 5.8	5.8 ± 1.2
SIFT	64	78.5 ± 6.6	5.0 ± 1.7

Table 4: Bottle detection: optimal detection results using hard assignment

Descriptor	k	σ	TPR (%)	FPR (%)
DH	512	0.08	89.3 ± 5.5	3.0 ± 1.4
DGH	2048	0.04	84.4 ± 5.4	3.5 ± 2.0
RIFT	1024	0.04	78.1 ± 5.5	4.6 ± 1.6
SIFT	2048	0.16	82.8 ± 5.7	5.6 ± 1.5

Table 5: Bottle detection: optimal detection results using kernel assignment

descriptors (in TPR and FPR) for all values of k, the results were significantly poorer than the handgun results (Table 1) for all four descriptors.

The optimal performance values using kernel assignment (measured over $\sigma = 0.02$, 0.04, 0.08, 0.16) for the botte subvolumes are shown in Table 5: DH = 89.3% ($k = 512$; $\sigma = 0.08$); DGH = 84.4% ($k = 2048$; $\sigma = 0.04$); RIFT = 78.1% ($k = 1024$; $\sigma = 0.04$) and SIFT = 82.8% ($k = 2048$; $\sigma = 0.16$). There was a general decline in performance for all four descriptors relative to the handgun dataset (Table 2. With the exception of the SIFT descriptor, kernel assignment (when optimally tuned) did, however, signficantly outperform hard assignment (Table 4) - substantiating the findings of [47, 56].

The performance of each descriptor using uncertainty assignment was again measured over and extended range of $\sigma = 0.005$; 0.01; 0.02; 0.04; 0.08; 0.16. Optimal results are illustrated in Table 6: DH = 88.2% ($k = 512$; $\sigma = 0.04$); DGH = 87.2% ($k = 512$; $\sigma = 0.04$); RIFT = 78.2% ($k = 2048$; $\sigma = 0.01$) and SIFT = 82.7% ($k = 2048$; $\sigma = 0.02$).

The differences in performance between the kernel assignment (Table 5) and uncertainty assignment methodologies are negligible, with the DH and DGH descriptors again significantly outperforming SIFT and RIFT. Similarly to the hard assignment results, there was a general and significant decline in performance relative to the handgun dataset (Table 3). Although not evident in Table 6, it is worth noting that performance varied significantly across the range of values for σ. In particular, setting σ too high (≥ 0.08) or too small (≤ 0.01) led to large drops in performance - emphasises the necessity for the correct adjustment of σ for each setting of k.

Descriptor	k	σ	TPR (%)	FPR (%)
DH	512	0.04	88.2 ± 4.7	2.2 ± 1.3
DGH	512	0.04	87.2 ± 6.8	4.0 ± 1.8
RIFT	2048	0.01	78.2 ± 6.4	5.6 ± 2.4
SIFT	2048	0.02	82.7 ± 7.0	4.2 ± 1.2

Table 6: Bottle detection: optimal detection results using uncertainty assignment

8. Discussion

We examine more closely the volumes that were misclassified in order to determine the possible causes of the misclassifications in each of the experiments.

Several examples of misclassified images in the handgun subvolume experiments are shown in Figure 5. The missed handguns (Figure 5 (a)) do not indicate any obvious reason for the error (e.g. a particular handgun and/or orientation). Several commonalities exist in the objects found in the false-positive volumes (Figure 5 (b)), which may be triggering the errors. For the DH and DGH descriptors, common objects include: batteries, electrical transformers, in-line roller skates and electronic equipment. Metallic objects also appear frequently in these images. These objects all have similar material characteristics to the materials constituting the handguns and therefore will result in similar CT numbers [58]. Since the DH and DGH descriptors capture information related to the distribution of these densities (which is dependent on material characteristics), it is understandable that materials with similar properties would trigger false positives. The SIFT and RIFT false-positive images, however, differ substantially from the DH and DGH images - with a far greater number of metal-free images triggering false-positives - this is likely due to the fact that the original SIFT and RIFT concepts [40, 48] were designed around use in reflectance (photographic) imagery (as opposed to transmission/attenuation imagery). This observation is discussed in further detail below.

The soft-assignment methodologies (uncertainty and kernel) consistently outperform the more traditional hard assignment approach. While the uncertainty approach performs marginally better than the kernel approach, the high error margins do not allow for a definitive conclusion. A closer examination of those volumes that led to erroneous classifications does not indicate in obvious reasons for the misclassifications. Further investigation of the causes of these errors is left as an area for further work.

A universal decline in performance is observed for the bottle subvolume dataset, suggesting that the proposed methodology is less well-suited to this object class. Despite fairly substantial variations in the optimal true-positive detection rates for the various descriptors (see Table 7 vs. Table 8), an examination of the threat-containing volumes which have been

Descriptor	Assignment Method	k	σ	TPR (%)	FPR (%)
DH	Kernel	1024	0.04	97.3 ±3.4	1.8 ±1.7
DGH	Hard	2048	-	97.2 ±2.2	2.1 ±1.2
	Uncertainty	512	0.04	97.2 ±2.8	2.1±1.3
RIFT	Uncertainty	2048	0.01	87.3 ±3.9	5.1 ±2.3
SIFT	Uncertainty	1024	0.02	87.0 ±5.4	3.8 ±2.4

Table 7: Handgun detection: optimal settings for each descriptor

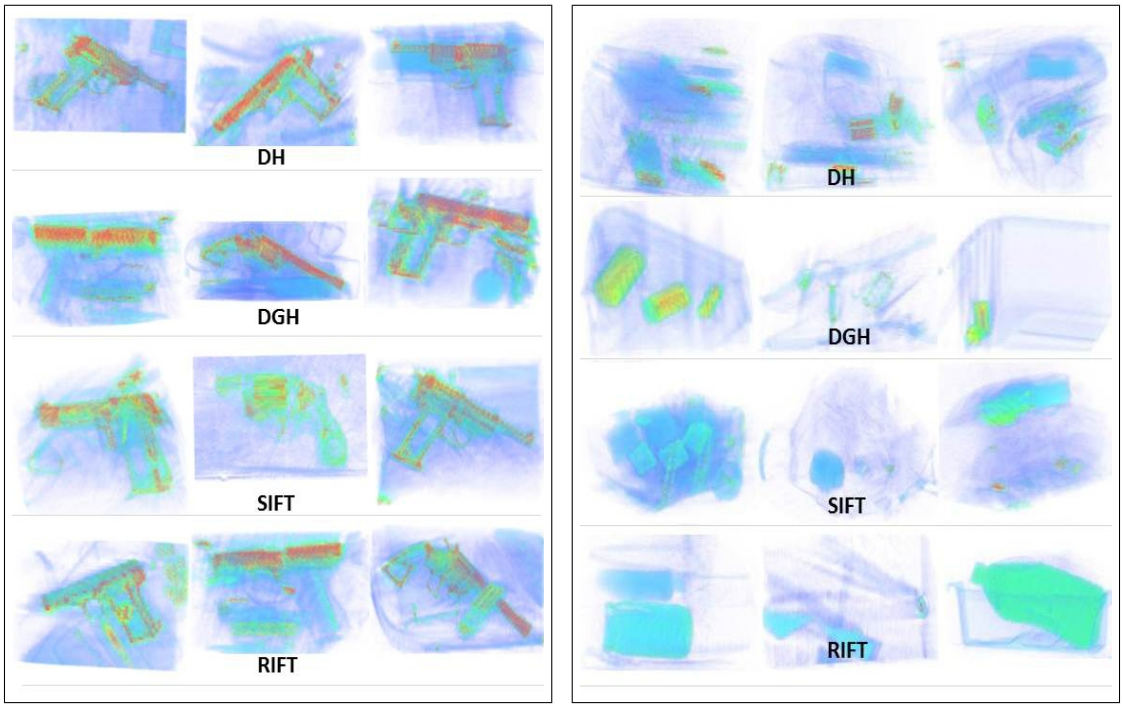
Descriptor	Assignment Method	k	σ	TPR (%)	FPR (%)
DH	Kernel	512	0.08	89.3 ±5.5	3.0 ±1.4
DGH	Uncertainty	512	0.04	87.2 ±6.8	4.0 ±1.8
RIFT	Uncertainty	2048	0.01	78.2 ±6.4	5.6 ±2.4
SIFT	Uncertainty	2048	0.02	82,7 ±7.0	4.2 ±1.2

Table 8: Bottle detection: optimal settings for each descriptor

incorrectly classified as clutter, again does not indicate any obvious characteristic that triggers the misclassification (Figure 6 (a)). Furthermore, the misclassified bottles do not appear to be particularly challenging in nature. The most likely reason for the decline in performance is the large intraclass variation in this dataset (in terms of bottle types and contents). The most obvious solution to this is to increase the size of the dataset, but it may perhaps be worth subdividing the bottle dataset into smaller subsets (according to contents, bottle type etc.). It therefore remains unclear as to why some bottle instances are misclassified while others are not. While we suggest an expansion of the dataset as an area for future work, it is worth acknowledging the difficulties related to gathering security-sensitive baggage-CT data (hence the limitations in the currently available data).

Similarly to the missed detections, the false-positive detections using the RIFT descriptor (Figure 6 (b)) bear little resemblance to bottles. The poor overall performance of the RIFT descriptor suggests a codebook that poorly characterises bottles, making it a poor choice for this particular problem.

In contrast, the DH, DGH and SIFT false-positives (Figure 6 (b)) do exhibit several distinct characteristics that appear to have triggered the errors. While the DH false-positives contain little evidence of any bottle-shaped items, there do appear to be several regions in these images that are similar in density to the liquids used in the training set. This observation highlights an obvious shortcoming of the codebook approach: namely the inability to capture spatial/shape information [59, 60]. The DGH false-positives contain several, virtually empty deodorant bottles as well as items with similar gradients to those from genuine bottle objects (e.g. perspex rods). Similarly, the SIFT false-positives contain several metallic objects whose



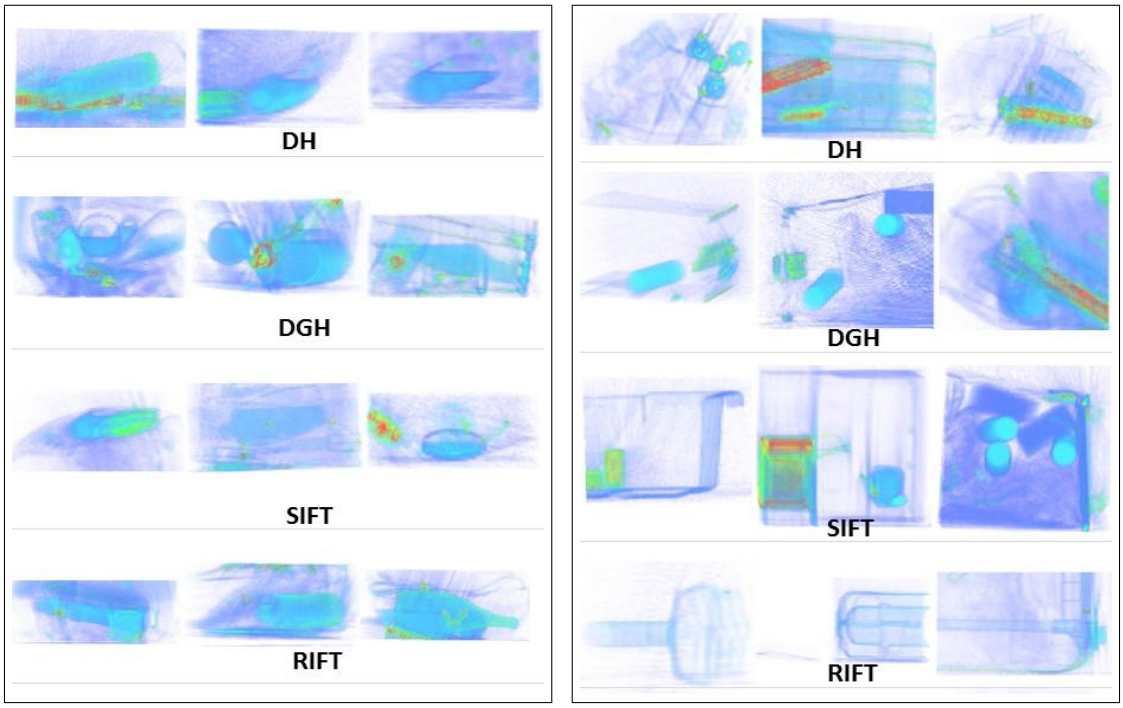
(a) Missed handguns (b) False-positives

Figure 5: Handgun misclassifications

features, when normalised during the SIFT descriptor generation, likely resemble bottles in shape. Particularly, several electrical transformers, batteries and perspex rods are present whose circular cross sections are similar to that of a full bottle. These observations suggest that items with circular cross-sections are particularly challenging. It is worth noting, however, that not all instances of such items were misclassified.

We emphasise that, in practice, whole volumes are scanned as a series of sub-volumes, in the same way one would perform object detection (localisation) in images using scanning-windows. As such, the subvolumes used here represent a real-world scenario in the same way as we consider False-Positive Per Window (FPPW) and alike in 2D imagery [61, 62]. In the current context whole volumes are bags of varying shapes and sizes. Therefore, each is comprised of a different number of 'non-empty' sub-volumes. If whole volumes are used, the system is biased by the very large and the very small bags with a very large or a very small number of empty (i.e. trivially easy) regions (for measures such as FPPW). As such, we propose that subvolumes are both statistically valid and real-world in this context.

The most significant observation in this study has been the superior performance of the DH and DGH descriptors relative to the more complex SIFT and RIFT descriptors (in all experiments). We attribute this to the inherent illumination and scale redundancies in the SIFT



(a) Missed bottles (b) False-positives

Figure 6: Bottle misclassifications

and RIFT descriptors [40, 48] as these were initially developed for illuminated optical imagery containing surface reflectance colour information. Since the densities in CT imagery depict the linear attenuation properties of the materials being imaged and not the reflectance properties, these built-in redundancies in SIFT and RIFT become largely irrelevant and appear to have a negative impact on performance.

Within transmission imagery such as X-ray CT, captured via a parallel (orthographic) projection as opposed to a perspective projection at the sensor level [63], scale invariance at the feature level is a significantly lessor issue as the size of an object does not vary with its distance to the sensor (i.e. measurements can be recovered from CT-scan imagery in real-world values). Furthermore, although variations in the transmission attenuation and surface reflectance are present, these are largely resolved within the in image recovery stages of modern scanners. Attenuation variation that is present is indicative of material characteristics (e.g. physical density) and arguably implicitly aids in object classification, rather than hindering it, as some form of materials information encoded within the density derived feature descriptors (DH and/or DGH). The information captured by the density-based descriptors (DH and DGH), is independent of what the actual densities represent (e.g. surface reflectance or linear attenuation) and hence the imaging formality. We therefore believe that this fundamental

change in the nature of the data being considered affects the relative performance of the descriptor types, such that simplicity, that is directly driven by the core modality (i.e. DH and DGH), outperforms convention in the wider object classification space (i.e. poor performance of SIFT and RIFT).

In light of the above observation, it is likely that all gradient-based relfectance-imaging descriptors (e.g. Harris3D [64]; Histograms of Oriented Gradients (HoG) [49, 65]) will be less effective in the transmission imaging domain consisered here. In the context of this study, an exhaustive comparison of further reflectance-based imaging descriptors thereby becomes redundant.

Finally, it is worth noting that, due to the size of the dataset, the measurement errors are consistently large. The considerable task of gathering annotated datasets in this problem-space, whilst maintaining realistic variation and levels of background clutter, on a scale representative of operational scenarios (e.g. airport scale) is left as an area for future work.

9. Conclusions

This work has investigated the suitability of the Bag-of-Words (BoW) approach in the classification of two object classes (handguns and bottles) in 3D baggage-CT imagery. In particular, the performance of four descriptor types (density histograms, density gradient histograms, SIFT and RIFT) and three cluster assignment methodologies (hard assignment, uncertainty assignment and kernel assignment) have been compared.

Optimal performance in both the handgun and bottle experiments was achieved using the Density Histogram (DH) and Density Gradient Histogram (DGH) descriptor types. Particularly, the DH descriptor yielded the highest true-positive rates (97.3% for handguns; 89.3% for bottles) as well as the lowest false-positive rates (1.8% for handguns; 3.0% for bottles) in both experiments. The DGH descriptor similarly, with correct detection rates of 97.2% for handguns and 87.2% for bottles and false-positive rates of 2.1% for handguns and 4.0% for bottles. Interestingly, the performance of the more complex SIFT and RIFT descriptors was significantly poorer, yielding both lower detection rates as well as higher false-positive rates. We have attributed this change in relative performance of the descriptor types to the change in the underlying imaging modality from reflectance (photographic) imaging to transmission imaging. This finding, within transmission imagery (X-ray CT), challenges established convention on the relative performance of 3D feature descriptors [7] and is attributed

to the redundancy of the illumination (primarily) and scale (partially) components present in established approaches (e.g. SIFT, RIFT, HoG, Harris etc.) within this imaging modality.

Soft kernel-assignment (uncertainty and kernel) has been shown to outperform hard-assignment. A general improvement in performance has been observed for larger codebooks, with the optimal results achieved using $k = 512; 1024; 2048$ - suggesting that too few visual words (small k) produces codebooks with insufficient salient entries to accurately describe the data. The observations regarding the impact of the assignment methodology and codebook size substantiate the findings of van Gemert et al. [47].

Finally, we hihglight several promising areas for future work. Automated segmentation within this problem domain will enable a true end-to-end classification of baggage-CT imagery and eliminate the need for computationally intensive subvolume generation. An investigation into the impact of image noise and artefacts on performance will enable performance optimisation under operational conditions. An extended target object dataset will provide further robustness to the evaluation protocol in use.

Acknowledgments

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

References

[1] S. Singh, Explosives detection systems (EDS) for aviation security, Signal Processing 83 (1) (2003) 31–55.

[2] European Parliament resolution of 6 July 2011 on aviation security, with a special focus on security scanners (2010/2154(INI)), [6 July 2011].

[3] B. R. Abidi, Y. Zheng, A. V. Gribok, M. A. Abidi, Improving weapon detection in single energy X-ray images through pseudocoloring, IEEE Transactions on Systems, Man, and Cybernetics 36 (6) (2006) 784–796.

[4] T. R. Johnson, Medical radiology/diagnostic imaging: dual energy CT in clinical practice, Springer, 2011.

[5] 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.

[6] G. Flitton, T. Breckon, N. Megherbi, Object recognition using 3D SIFT in complex CT volumes, in: Proceedings British Machine Vision Conference, 11.1–11.12, 2010.

[7] G. Flitton, T. P. 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.

[8] 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, 3634–3641, 2012.

[9] N. Megherbi, T. Breckon, G. Flitton, A. Mouton, Fully automatic 3D threat image projection: application to densely cluttered 3D computed tomography baggage images, in: Proceedings of the IEEE International Conference on Image Processing Theory, Tools and Applications, 153–159, 2012.

[10] A. Mouton, T. Breckon, G. Flitton, 3D object classification in complex volumes using randomised clustering forests, in: IEEE International Conference on Image Processing, 2014 - to appear.

[11] 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, 2057–2060, 2012.

[12] 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, 2334–2338, 2013.

[13] 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, 1063–1068, 2013.

[14] 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.

[15] E. M. van Rikxoort, B. van Ginneken, Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review, Physics in medicine and biology 58 (17) (2013) R187.

[16] H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, D. Comaniciu, Hierarchical, learning-based automatic liver segmentation, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1–8, 2008.

[17] A. Criminisi, J. Shotton, S. Bucciarelli, Decision forests with long-range spatial context for organ localization in CT volumes, in: MICCAI Workshop on Probabilistic Models for Medical Image Analysis, 2009.

[18] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, A. Criminisi, Entangled decision forests and their application for semantic segmentation of CT images, in: Information Processing in Medical Imaging, Springer, 184–196, 2011.

[19] B. Glocker, O. Pauly, E. Konukoglu, A. Criminisi, Joint classification-regression forests for spatially structured multi-object segmentation, in: European Conference on Computer Vision, Springer, 870–881, 2012.

[20] A. Criminisi, J. Shotton, D. Robertson, E. Konukoglu, Regression forests for efficient anatomy detection and localization in CT studies, Medical Computer Vision.Recognition Techniques and Applications in Medical Imaging (2011) 106–117.

[21] A. Criminisi, J. Shotton, S. Bucciarelli, Decision forests with long-range spatial context for organ localization in CT volumes, in: MICCAI Workshop on Probabilistic Models for Medical Image Analysis, 2009.

[22] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Matching 3D models with shape distributions, in: IEEE International Conference Shape Modeling and Applications, 154–166, 2001.

[23] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.

[24] J. Gall, V. Lempitsky, Class-specific Hough forests for object detection, in: Decision Forests for Computer Vision and Medical Image Analysis, Springer, 143–157, 2013.

[25] 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, 1470–1477, 2003.

[26] S. Nercessian, K. Panetta, S. Agaian, Automatic detection of potential threat objects in X-ray luggage scan images, in: Proceedings of the IEEE International Conference on Technologies for Homeland Security, 504–509, 2008.

[27] C. Oertel, P. Bock, Identification of objects-of-interest in X-ray images, in: Proceedings of the IEEE Applied Imagery and Pattern Recognition Workshop, 17, 2006.

[28] R. Gesick, C. Saritac, C. C. Hung, Automatic image analysis process for the detection of concealed weapons, in: Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research, 1–4, 2009.

[29] M. Bastan, M. Yousefi, T. Breuel, Visual words on baggage X-ray images, in: Computer Analysis of Images and Patterns, Springer, 360–368, 2011.

[30] D. Turcsany, A. Mouton, T. Breckon, Improving feature-based object recognition for X-ray baggage security screening using primed visual words, in: Proceedings International Conference on Industrial Technology, 1140–1145, 2013.

[31] H. Bay, T. Tuytelaars, L. V. Gool, SURF: Speeded up robust features, in: Proceedings of the European Conference on Computer Vision, 404–417, 2006.

[32] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, S. Yang Bang, Constructing support vector machine ensemble, Pattern Recognition 36 (12) (2003) 2757–2767.

[33] W. Bi, Z. Chen, L. Zhang, Y. Xing, A volumetric object detection framework with dual-energy CT, in: Proceedings of the IEEE Nuclear Science Symposium Conference Record, 1289–1291, 2008.

[34] W. Bi, Z. Chen, L. Zhang, Y. Xing, Fast detection of 3D planes by a single slice detector helical

CT, in: Proceedings of the IEEE Nuclear Science Symposium Conference Record, 954–955, 2009.

[35] D. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 13 (2) (1981) 111–122.

[36] 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, 1833–1836, 2010.

[37] K. J. J., van Doorn A. J., Surface shape and curvature scales, Image and vision computing 10 (8) (1992) 557–564.

[38] A. Mouton, On Artefact Reduction, Segmentation and Classification of 3D Computed Tomography Imagery in Baggage Security Screening [PhD Thesis], Cranfield University UK, 2014.

[39] 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, II–319 – II–324, 2003.

[40] 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.

[41] J. Knopp, M. Prasad, G. Willems, R. Timofte, L. Van Gool, Hough transform and 3D SURF for robust three dimensional classification, in: Proceedings of the European Conference on Computer Vision, vol. 6, 589–602, 2010.

[42] A. Bolovinou, I. Pratikakis, S. Perantonis, Bag of spatio-visual words for context inference in scene classification, Pattern Recognition 46 (3) (2013) 1039–1053.

[43] L. Zhou, Z. Zhou, D. Hu, Scene classification using a multi-resolution bag-of-features model, Pattern Recognition 46 (1) (2013) 424–433.

[44] F. Moosmann, B. Triggs, F. Jurie, Fast discriminative visual codebooks using randomized clustering forests, Advances in Neural Information Processing Systems 19 (2007) 985–992.

[45] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: European Conference on Computer Vision, 490–503, 2006.

[46] V. N. Vapnik, The nature of statistical learning theory, Springer-Verlag New York Inc, 2000.

[47] 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.

[48] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[49] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 886–893, 2005.

[50] S. Allaire, J. Kim, S. Breen, D. Jaffray, V. Pekar, Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis, in: Proceedings of the IEEE International Conference Computer Vision and Pattern Recognition Workshops, 1–8, 2008.

[51] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification, in: Proceedings European Conference on Computer Vision, 490–503, 2006.

[52] J. Z. C. Lai, Y. C. Liaw, Improvement of the k-means clustering filtering algorithm, Pattern Recognition 41 (12) (2008) 3677–3681.

[53] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, 1027–1035, 2007.

[54] G. Flitton, Extending computer vision techniques to recognition problems in 3D volumetric baggage imagery [PhD Thesis], Cranfield University UK, 2012.

[55] N. Megherbi, T. Breckon, G. Flitton, A. Mouton, Radon transform based metal artefacts generation in 3D threat image projection, in: Proceedings of SPIE Optics and Photonics for Counterterrorism, Crime Fighting and Defence, vol. 8901, 1–7, 2013.

[56] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, Lost in quantization: improving particular object retrieval in large scale image databases, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1–8, 2008.

[57] N. E. L. Shanks, A. L. W. Bradley, Handbook of Checked Baggage Screening: Advanced Airport Security Operation, John Wiley and Sons, ISBN 978-1-86058-428-2, 2004.

[58] J. Hsieh, Computed tomography: principles, design, artifacts, and recent advances, SPIE and John Wiley and Sons, 2003.

[59] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2169–2178, 2006.

[60] A. Bosch, A. Zisserman, X. Muoz, Image classification using random forests and ferns, in: Proceedings of the IEEE International Conference on Computer Vision, 1–8, 2007.

[61] Q. Zhu, M.-C. Yeh, K.-T. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 1491–1498, 2006.

[62] X. Wang, T. X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, in: IEEE International Conference on Computer Vision, 32–39, 2009.

[63] C. Solomon, T. Breckon, Fundamentals of digital image processing: a practical approach with examples in Matlab, Wiley-Blackwell, ISBN 0470844736, 2010.

[64] I. Sipiran, B. Bustos, Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes, The Visual Computer 27 (11) (2011) 963–976.

[65] A. Zaharescu, E. Boyer, K. Varanasi, R. Horaud, Surface feature detection and description with applications to mesh matching, in: IEEE International Conference on Computer Vision and Pattern Recognition, 373–380, 2009.