Pattern Recognition 46 (2013) 2420–2436

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery

Greg Flitton *, Toby P. Breckon [1], Najla Megherbi

Applied Mathematics and Computing Group, School of Engineering, Cranfield University, Bedfordshire MK43 0AL, UK

ARTICLE INFO

Article history:
Received 6 July 2011
Received in revised form
27 January 2013
Accepted 7 February 2013
Available online 16 February 2013

Keywords:
CT baggage scan
Threat detection
Object recognition
3D feature descriptors
CT object recognition
3D SIFT

ABSTRACT

We present an experimental comparison of 3D feature descriptors with application to threat detection in Computed Tomography (CT) airport baggage imagery. The detectors range in complexity from a basic local density descriptor, through local region histograms and three-dimensional (3D) extensions to both to the RIFT descriptor and the seminal SIFT feature descriptor. We show that, in the complex CT imagery domain containing a high degree of noise and imaging artefacts, a specific instance object recognition system using simpler descriptors appears to outperform a more complex RIFT/SIFT solution. Recognition rates in excess of 95% are demonstrated with minimal false-positive rates for a set of exemplar 3D objects.

1. Introduction

X-ray type technologies have been used for airport security checks for several decades but the use of computer vision within this domain is limited to techniques that purely aid human baggage screeners [1]. Heightened regard to the detection of complex articles within baggage and parcels for air transit and other forms of transportation has led to an increased interest in the use of automatic recognition strategies. Items of interest can be generally difficult to detect within this environment due to a range of orientation, clutter and density confusion in a traditional two-dimensional (2D) X-ray projection [2]. An example of this is shown in Fig. 1 where we see (a) an example bag (photograph), (b) an overhead 2D X-ray revealing an item of interest within and (c) a different scan of the same bag with the item of interest in an orientation that does not reveal its salient features. This potential problem of object self occlusion (Fig. 1c) is a limitation of 2D X-ray scanners which makes detection (automatically or by human operators) particularly challenging. In this work we specifically look at the use of increasingly popular Computed Tomography (CT) volumetric imagery where a three dimensional voxel image of the baggage/parcel item is obtained in an attempt to overcome some of these issues.

Recent advances in imaging technology now facilitate the use of dual energy CT scanners for the real time scanning of bags in airport baggage/parcel handling operations [3]. It is from these scanners that we obtain a series of image slices through the bag which can be reconstructed as a traditional CT 3D volume akin to those encountered within medical CT imaging [4]. Prior work on the automatic recognition of objects within this complex 3D volumetric imagery is limited [5,6]. Bi et al. [5] took 3D CT volumes and attempted recognition of an item of interest but reduced the problem to two dimensions by looking at the item characteristic cross-section when extracted from the 3D volumetric image (c.f. 2D X-ray views of Fig. 1). By contrast, our prior work [6] explicitly investigated the use of a 3D SIFT descriptor for object recognition with some reasonable results. Here we examine a range of such descriptors and investigate the quality of detection achievable over a quantifiable larger data set.

It is important at this stage to remember a key aspect of the practicalities of the baggage scanning scenario with relation to the rates of detection: in general we require a high true-positive rate (to ensure that true threats are detected) but a low false-positive rate (to maximize scanning throughput and additionally minimize impact on the aviation/transport industry).

1.1. Complex CT volumetric imagery

An example of a 3D scan of an item of baggage is shown in Fig. 2 where we see the presence of an item of interest amongst more general cluttered items. Within Fig. 2 the data is re-scaled to

* Corresponding author. Tel.: +44 1767 690976; fax: +44 1234 754797.
E-mail addresses: g.t.flitton@gmail.com (G. Flitton), toby.breckon@cranfield.ac.uk (T.P. Breckon), n.megherbi@cranfield.ac.uk (N. Megherbi).
[1] Tel.: +44 1234758246.

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.02.008

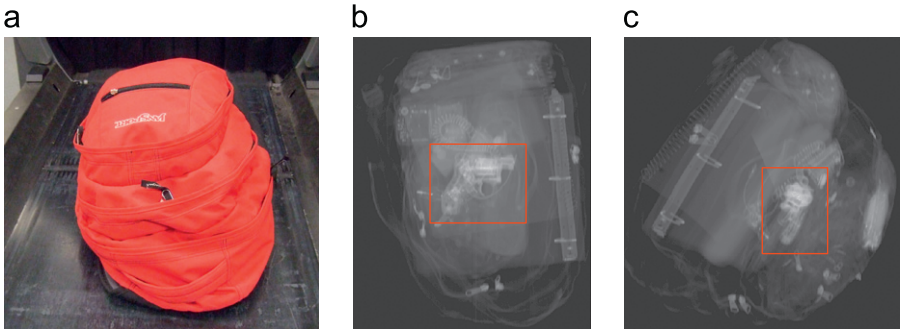

Fig. 1. Bag and X-rays: (a) example bag, (b) X-ray view 1 and (c) X-ray view 2.

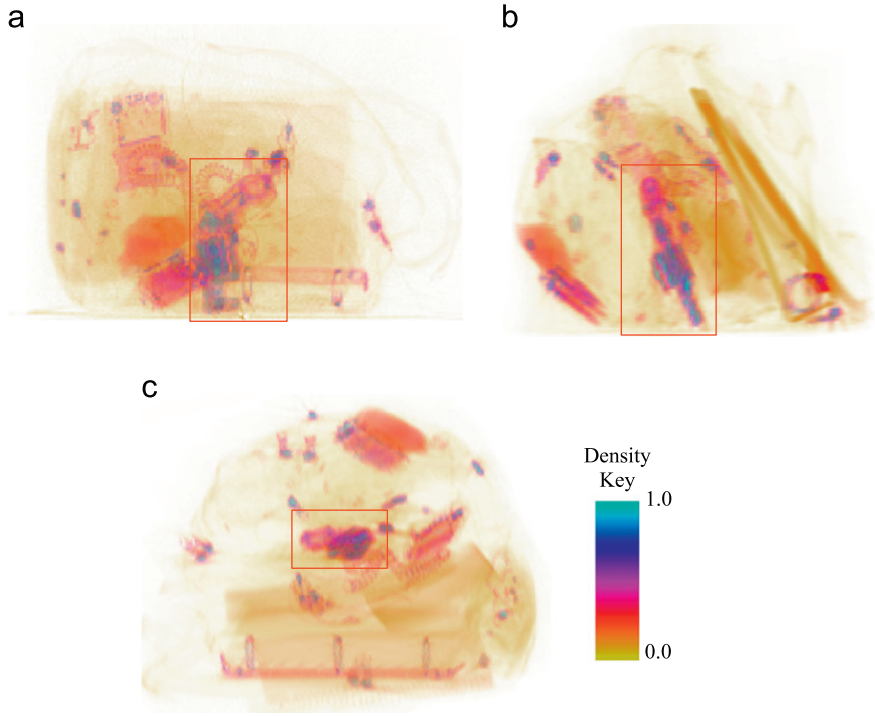

Fig. 2. 3D volume of complex bag containing a revolver: (a) front view, (b) side view and (c) top view.

the continuous range [0.0, 1.0] from the original integer CT scanner output (key as shown).

The type of baggage scanner machine used to capture the CT volumetric imagery for this work is primarily aimed at dual energy explosives detection [3]. As a result of this primary (non-object recognition based) objective two additional consequences are generally observed within the imagery: (1) the presence of metal items causes significant artefacts within the imaging (Fig. 3) and (2) the resolution is anisotropic and limited to $1.6\,\text{mm} \times 1.6\,\text{mm} \times 5\,\text{mm}$. The metal artefacts radiate out in the *xy*-plane and do not remain consistent from one scan to another if the metallic region changes orientation. The 5 mm resolution in the *z* direction will influence the size of target object than can be recognized. Both of these factors are due to the needs of high baggage throughput (speed) and the primary directive of explosives detection such that image quality is sacrificed. It is noted that the artefacts and sampling attributes are significantly different from state-of-the-art medical imaging, where the foremost constraints are not throughput and a need for dual-energy materials detection.

Although prior work has looked at the removal of metal artefacts in medical CT imagery [7–9] this has not been explicitly considered within this work due to constraints on access to the raw CT projection data. Additionally we recognize that the poor resolution gives rise to stair step artefacts [10,11]. Although this poses significant challenges for recognition we consider here the limitation in resolution to be similar to the scale invariance challenge addressed by various interest point feature descriptors in 2D [12–14] and additionally the unpredictable nature of the metal artefacts to be akin to that of recognition in the presence of occlusion – an area in which such interest point detectors [12–14] perform well. Complex imagery of this nature containing dense collections of man made objects scanned at low resolution and in the presence of metal artefacts has not previously been considered within any work on an automated 3D recognition.

2. Object detection in complex CT volumetric imagery

Object detection using interest points and descriptors is a well known approach. Schmid and Mohr [15] proposed the use of Harris features [16] as points of interest in a grayscale image. The interest points were then characterized using a range of rotation invariant descriptors that were then stored in a hash table. Recognition comes by generating the interest points and their descriptors in an image and then looking them up in the hash

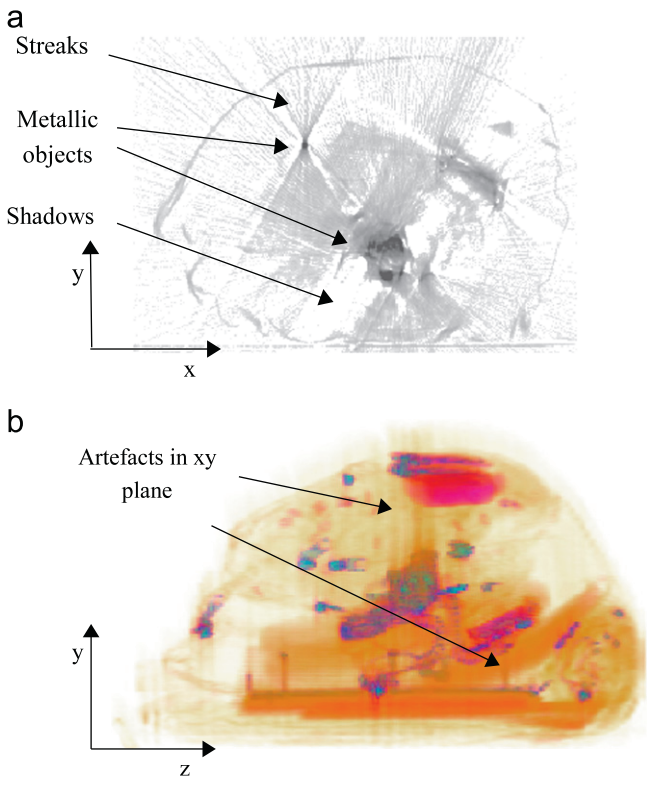

Fig. 3. Metal artefacts in CT baggage imagery: (a) streaks/shadows in CT image slice and (b) artefacts in *xy*-plane only.

table. Lowe [17] introduced the Scale Invariant Feature Transfrom (SIFT) with the aim of object recognition. Refinements by Lowe [12] have led to the SIFT approach being widely used for object recognition in 2D images.

Medical imaging is the primary source of 3D voxel-based data. Computer vision techniques have been used to perform registration in 3D for CT, MRI and ultrasound imagery. Volumetric stitching has also been addressed for ultrasound imagery. The use of computer vision approaches for recognition of objects in medical imagery is of great interest but these tend to be of soft-tissue items. For example, one area of medical analysis is the recognition of colonic polyps – abnormal tissue growth which can be pre-malignant and must be removed [18,19]. Rigid object recognition is seldom of interest in medical image analysis and interest point techniques are restricted to stitching and registration tasks.

There is little published work in the area of automatic recognition of items in scanned baggage (either X-ray or CT). Bi et al. [20] used a CT scanner to detect a handgun in baggage. The work did not involve processing the 3D data directly – the problem was reduced to searching for the characteristic cross-section that the handgun presents and appeared preliminary in nature. No results are presented in respect of detection performance. Further work by the same author [21] presented a methodology for the detection of planar materials within CT-baggage imagery using a 3D extension to the Hough transform [22]. The reported work concentrated on implementing the algorithm on a graphics-processing-unit (GPU) and again no results were presented on the performance of the detection method itself.

Baştan et al. [23] recently applied the bag-of-words machine learning model to color 2D dual energy X-ray images of baggage items to detect handguns and are the first to report detection results. Dual energy X-ray scanners illuminate the baggage item with a high and low power X-ray beam from which an estimate can be made of the material type present at each pixel location. This material type image is colored to aid human operatives recognize objects. 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 rather than cropping threat items 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 color image (indicating material type).

Megherbi et al. [24] investigated detection of bottles in CT volumetric data. Baggage items are segmented from CT volume and analyzed using two approaches. The first approach is to form a descriptor by analysis of the surface of the extracted item. A normalized histogram of shape index [25] is formed as the descriptor from this approach. The second method uses Zernike descriptors [26] which are rotation invariant and produce a finite vector description of the voxel-based object derived from a set of orthogonal basis functions. From a small dataset (79 volumes for training, 126 for testing) classification results in excess of 98.0% accuracy were obtained using the histogram of shape index.

Research into recognition using 2D X-ray imagery has resulted in a number of publications. Nercessian et al. [27] investigated a 2D X-ray imagery of luggage for the detection of handguns. The method uses edges detection to characterize handgun features but only deals with handguns in a fixed orientation. A small dataset was used (40 images with handguns, 400 images clutter) with two simple examples of handgun detection shown but no statistical results presented. Oertel and Bock [28] tackled detection of handguns in 2D gray-scale X-ray baggage imagery. The research is an example of specific item recognition: one type of handgun was characterized 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 neighborhood. It is unclear if is this is rotation invariant in the horizontal plane and the method is certainly not invariant if the weapon is rotated out of this plane. A small dataset was used (40 X-ray images: 30 for training and 10 for testing) and no quantitative results were produced. Gesick et al. [29] were also interested in detecting handguns from 2D X-ray imagery. Edges were extracted from the imagery and the handgun trigger guard was searched for in a similar fashion to [28]. The method was not rotation invariant and no quantitative results were produced.

The popularity of the SIFT algorithm for object recognition methodologies in 2D imagery has led to a number of 3D extensions being presented in the literature [30–33]. Firstly, Scovanner et al. [30] used a form of 3D SIFT to assist in 3D video volume analysis followed by Cheung and Hamarneh [31] who created a 3D SIFT variant to aid in medical image alignment. Ni et al. [32] also extended SIFT to a 3D formulation, derived from Scovanner et al. [30], for use in 3D ultrasound panoramic imagery. It is noted that all of these approaches suffer from a fundamental limitation in their consideration of orientation – the definition of orientation in 3D is incorrectly taken as the direction formed by two angles (azimuth, elevation) in [30–32]. Here, to correctly orientate an object in 3D, we consider three angles – azimuth, elevation and tilt. As shown in Fig. 9a, three angles are required to correctly orientate an object. Fig. 9b shows an example of this with three pistols aiming in the same direction (given by azimuth and elevation) but with differing orientation (given by the addition of tilt). This prior error of [30–32] was previously noted by Allaire et al. [33] and corrected: their subsequent results indicated that the additional tilt angle improves matching as expected. Notably this error originated from the work of [30–32] as a problem of

image registration as opposed to explicit object recognition: a theme also followed by Allaire et al. [33]. Here, by contrast to these earlier works, we fully extend SIFT to 3D for the explicit application of object recognition, taking into consideration the full definition of 3D orientation not considered in earlier works [30–32]. We also compare the system performance using 3D SIFT to that obtained with other descriptors. This extends our previous work of [6].

In this work we explicitly consider the detection of known rigid objects within low resolution, noisy, complex volumetric CT imagery and we examine a range of 3D interest point descriptors for this task. This is facilitated by the use of a traditional specific instance recognition approach whereby a reference volume object is identified and pose estimated within a given unknown volume. The range of descriptors evaluated for this task range from the use of simple density statistics to full 3D extensions of established interest point descriptors from 2D works [12,34]. Parametric settings are derived using an empirical approach and are recorded appropriately within the text. First we detail the detection and localization of these descriptors prior to outlining the descriptor variants which we go on to present in a range of comparative results.

3. 3D interest point detection and localization

We use interest points and local descriptors as the basis for our object recognition algorithm as these methods have been demonstrated in a variety of fields with high degrees of success [17,35–37]. We will now outline our approach for interest point location and local neighborhood definition.

3.1. Interest point detection

The same method of interest point detection is used for each descriptor being tested so that relative system performance is determined by the choice of descriptor rather than interest point detector. We use a 3D extension to the SIFT algorithm [12], as described in [33], to determine the location of interest points. Given a 3D input volume $I(x,y,z)$ and a 3D Gaussian filter $G(x,y,z,\sigma)$ we form multi-scale difference of Gaussian (DoG) volumes as follows:

$$DoG(x,y,z,k) = I(x,y,z) \otimes G(x,y,z,\sigma_s^k) - I(x,y,z) \otimes G(x,y,z,\sigma_s^{k-1}), \quad (1)$$

where k is an integer in the range [0,4] representing the scale index, $\sigma_s = \sqrt[3]{2}$ (following the work of [12]), (x,y,z) are defined in voxel coordinates and $\otimes$ is the convolution operator. Subsequently a three level pyramid ($L=0,1,2$) is built up by subsampling the Gaussian filtered volume by a factor of 2 for $k=3$ (i.e. when $\sigma_s^k = 2.0$) and repeating the process. At the base of the pyramid ($L=0$) each voxel is 2.5 mm^3; at the top ($L=2$) each voxel is 10 mm^3. The base voxel size is determined by the CT scanner available for this work. The number of pyramid levels is chosen to cover the size of features that will occur for objects of interest within the CT imagery.

In a similar vein to the original 2D SIFT methodology [12], DoG local extrema are then located. This requires that a voxel be either a maximum or minimum when compared to its neighboring voxels. Given that each voxel has a $3\times3\times3$ local neighborhood it follows that there are 26 voxels for comparison. It is also a requirement that the voxel is a maxima or minima when compared to the 27 neighborhood voxels in the scale space DoG volumes both above and below ($k+1$, $k-1$). The locations of these extrema form a candidate set of interest point locations.

From this candidate set a number of points are rejected for poor contrast if their density is below a threshold, τ_c ($\tau_c = 0.05$). This removes some erroneous points that are likely to produce unstable descriptors and additionally, in the case of CT volumes, points associated with metal artefacts. Note that the voxel density has been re-scaled into a floating point format (see Section 6). A second stage of candidate point rejection also takes place for points which are poorly localized on an edge. These points are likely to produce unstable descriptors in the presence of noise. A 3×3 Hessian matrix describes the local isosurface curvature at the candidate point,

$$H = \begin{bmatrix} D_{xx} & D_{yx} & D_{zx} \\ D_{xy} & D_{yy} & D_{zy} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}, \quad (2)$$

where D_{ij} are the second derivatives in the DoG volume. Both Allaire et al. [33] and Ni et al. [32] derive a measure to reject points using the Trace and Determinant of H, where

$$\mathrm{Tr}(H) = D_{xx} + D_{yy} + D_{zz}, \quad (3)$$

$$Det(H) = D_{xx}D_{yy}D_{zz} + 2D_{xy}D_{yz}D_{xz} - D_{xx}D_{yz}^2 - D_{yy}D_{xz}^2 - D_{zz}D_{xy}^2. \quad (4)$$

It can be shown [32,33] that the following equation can then be used to reject points:

$$\text{Reject if } \frac{\mathrm{Tr}^3(H)}{Det(H)} > \frac{(2\tau_e+1)^3}{(\tau_e)^2}. \quad (5)$$

We use a value of $\tau_e = 40$ and, hence, points where $\mathrm{Tr}^3(H)/Det(H) > 332.15$ are rejected.

Finally a subvoxel estimate of the extrema true location is achieved using quadratic interpolation on the DoG volumetric data.

3.2. Local point of interest neighborhood function

Following from the identification of interest point locale we now define a localized neighborhood function, extending this from earlier work in 2D [12].

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 neighborhood,

$$w(d,\sigma) = \exp\left[-\left(\frac{d}{\sigma}\right)^2\right], \quad (6)$$

where d is the voxel distance from the point of interest to the contributing voxel and σ is used to determine the extent of the local contribution. The use of this function is given with each of the descriptor formulations described in Section 4. It should be noted that, given the definition of distance in voxel units, this window will remain consistent with the resolution of the volume being examined.

4. 3D point of interest descriptors

Following interest point detection we now wish to characterize the local neighborhood. We detail a range of approaches for this characterization in increasing levels of complexity from a simple local density average, density and gradient histograms, leading on to 3D extensions to RIFT [34] and SIFT [12].

4.1. Simple density descriptor (D)

The density descriptor is a simple Gaussian average around the point of interest as shown in Eq. (7),

$$D_I = \frac{\sum_k \rho_k . w(d_k,\sigma)}{\sum_k w(d_k,\sigma)}, \quad (7)$$

for voxel k, a voxel distance d_k from the interest point location with a density ρ_k. The local neighborhood function, $w(d_k,\sigma)$, is as defined in Eq. (6).

This is a simple detector and is included for comparison to its more complex counterparts.

4.2. Density histogram descriptor (DH)

By contrast this second descriptor, the density histogram (DH), defines the local density variation for a given interest point as an N bin histogram defined over a continuous density range. The density range is taken as [−1.0,2.0] (see Section 6) and is split into N_{DH} bins resulting in each bin having a width of $3.0/N_{DH}$. Each voxel in the local neighborhood contributes to a single histogram bin as follows. The voxel density for point k is ρ_k and this is used to determine which histogram bin is active. Given the local area function $w(d_k,\sigma)$, defined in Section 3.2, 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 voxel k. The descriptor is normalized such that the sum of all histogram entries is 1.0. Fig. 4a shows an example point of interest, I, with one of its neighboring voxels of density ρ_k. Fig. 4 shows an example of a density histogram derived from an interest point that is located near a metallic region. It can be seen from this that the resulting density histogram has a peak due to the high concentration of metal within the neighborhood.

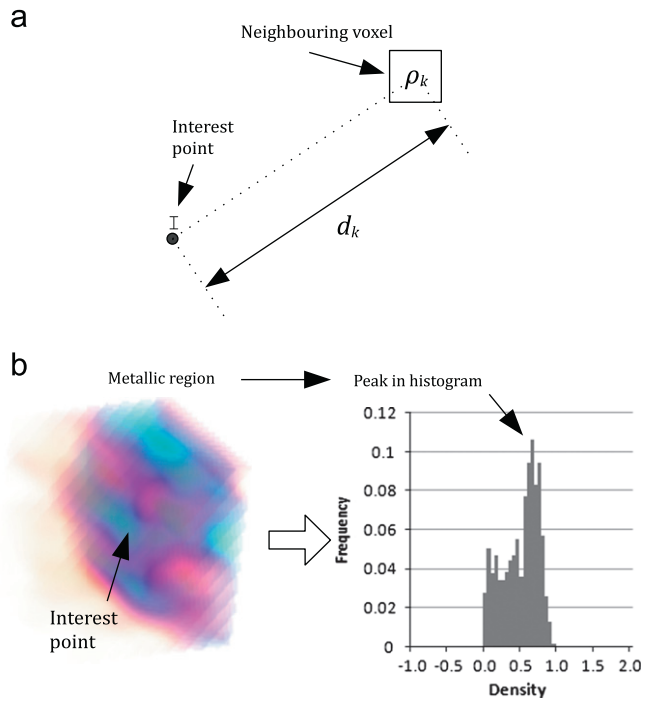

Fig. 4. Density histogram calculation: (a) local neighborhood voxels and (b) example of density histogram descriptor.

4.3. Density gradient magnitude histogram descriptor (DGH)

In a variant of the density histogram descriptor, here we calculate the density *gradient* magnitude in the neighborhood of the interest point and then accumulate these in a histogram. The density gradient magnitude is calculated for all voxels in the volume using a central difference formulation. The density gradient magnitude range ([0.0/cm, 4.0/cm]) (given a voxel dimension of 0.25 cm, and a worst-case change in density between adjacent voxels of 1.0, the vast majority of gradient values lie below a value of 4.0/cm) and is divided into N_{DGH} bins. The voxel gradient magnitude for voxel k is δ_k and this is used to determine which histogram bin is active. Once the active histogram bin is determined, an addition of $w(d_k,\sigma)$ is made to the corresponding histogram entry, with $w(d_k,\sigma)$ again defined by Eq. (6). The descriptor is normalized to unity area on completion as per the previous descriptor (Section 4.2). It is notable here that, due to the rotational variance of the objects under consideration for detection, the gradient *magnitude* is used rather than the gradient *orientation* approach frequently used for recognition tasks in 2D [38].

Fig. 5 shows the same point of interest as for Fig. 4b but now with the density gradient histogram being formed. It is not as obvious how the histogram relates to the volume given the noisy conditions of the imagery.

4.4. Rotation invariant feature transform (RIFT)

Lazebnik et al. [34] developed the rotation invariant feature transform (RIFT). The RIFT descriptor examines the local neighborhood 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 neighborhood 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 [12,34].

Before describing our extension of RIFT to 3D we consider our variant concretely in 2D. Fig. 6a shows a point of interest, I, and

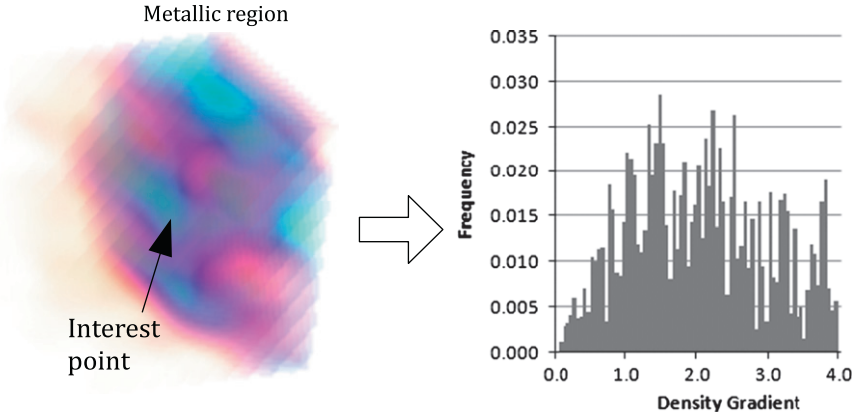

Fig. 5. Density gradient histogram calculation.

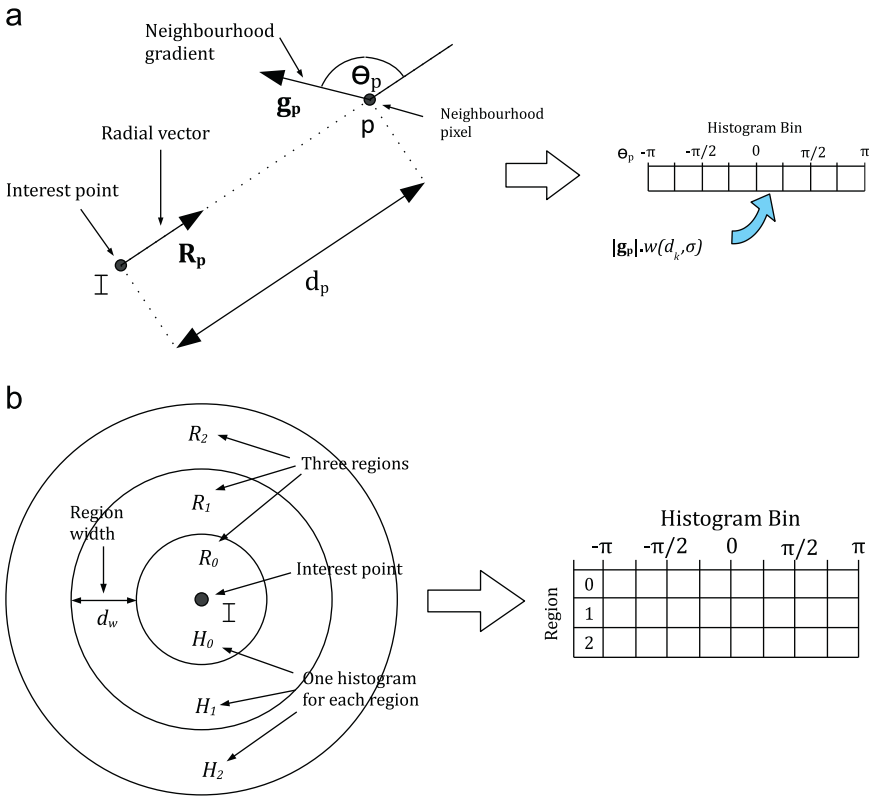

Fig. 6. 2D RIFT descriptor: (a) 2D radial geometry and (b) 2D radial regions.

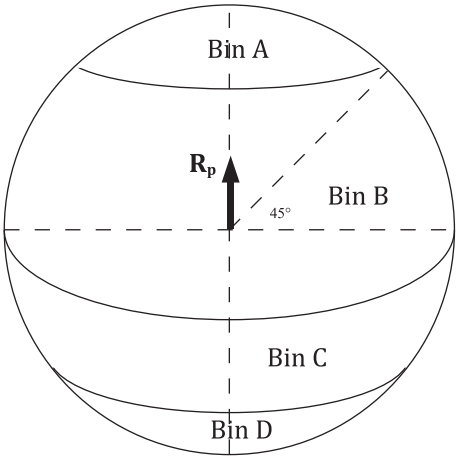

Fig. 7. 3D RIFT bin normalization.

neighboring region. For each neighboring pixel, p, a unit vector $\mathbf{R_p}$ in the direction from I to p is calculated. The gradient at pixel p is $\mathbf{g}_p$. The angle between the gradient ($\mathbf{g}_p$) and radial vector ($\mathbf{R_p}$) is θ_p. A histogram is constructed based on the values of θ_p in the range $[-\pi{:}\pi]$. There are N_b bins in this histogram representing angular regions $2\pi/N_b$ radians in size. For each gradient and angle an addition to the associated histogram bin of $|\mathbf{g}_p|.w(d_k,\sigma)$ is made as shown in Fig. 6a. Note again that the function $w(d_k,\sigma)$ limits the contribution to the local neighborhood. In addition to the histogram, N_r rings, of width d_w pixels, are also defined as shown in Fig. 6b (with $N_r=3$). One histogram is generated for each region and each histogram is normalized by the area of its ring to prevent bias to regions of greater area. The complete descriptor is normalized to unity. The resultant descriptor has $N_r \times N_b$ elements.

The extension of the RIFT descriptor to 3D is straight forward noting that, due to rotation symmetry in 3D, the angular histograms only cover values of θ_p in the range $[0{:}\pi]$ and the normalizations refer to region *volumes* rather than areas. One additional normalization is required in the move to 3D: the histogram summations are normalized by bin surface area to remove bias towards equatorial bins. Fig. 7 shows an example with 4 bins per histogram: bins A, B, C and D. If the volume has unit radius, bins A and D have a surface area of $\pi(2-\sqrt{2})$, whereas bins B and C have a surface area of $\pi\sqrt{2}$. These surface areas are used to normalize the summations for each bin. This step is not required in the 2D case as all histogram bins have the same sector angle. As with the other descriptors, the final step is to normalize the complete descriptor to unity.

Fig. 8 shows the RIFT descriptor generated for the same metallic region as used in the density histogram and density gradient histogram explanations (Figs. 6b and 5). This plot shows that, for this example, the gradients tend to orientate toward to the point of interest (an angle of π) rather than away (an angle of 0).

4.5. 3D SIFT

The 3D SIFT descriptor is closely modelled on that used in [33,6]. We briefly outline our 3D SIFT extension detailing the keypoint orientation and description based upon the initial interest point detection steps and local neighborhood function as outlined in Section 3. Here, as an extension to previous work on 3D SIFT [30–32], we follow the work of [6] and fully consider an object recognition taking into consideration of 3D orientation in terms of azimuth, elevation and tilt as illustrated in Fig. 9.

4.5.1. Keypoint orientation

Once a keypoint location is determined (Section 3) the volume gradients are examined in a two stage process to establish a local

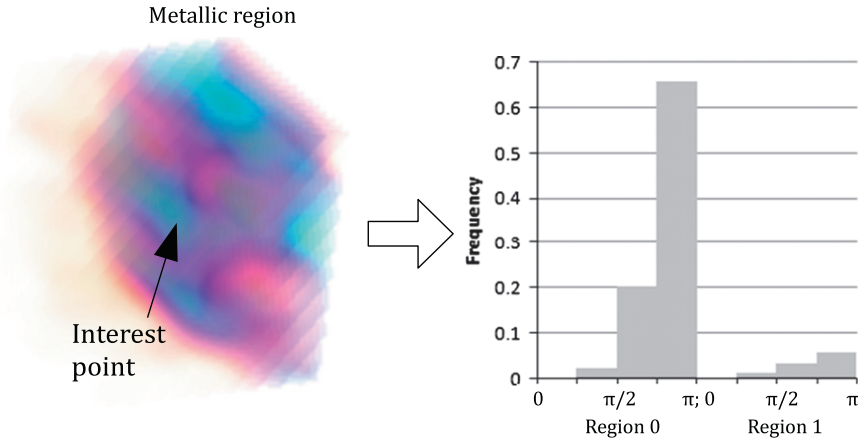

Fig. 8. RIFT descriptor example.

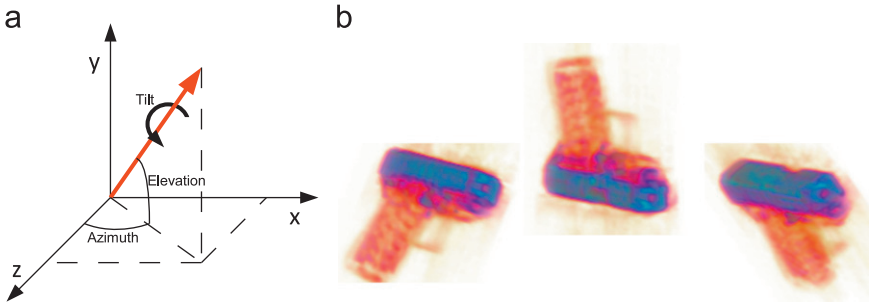

Fig. 9. 3D orientation requires three angles: azimuth, elevation and tilt: (a) our definitions and (b) pistols pointing in the same direction but with differing orientation (i.e. tilt).

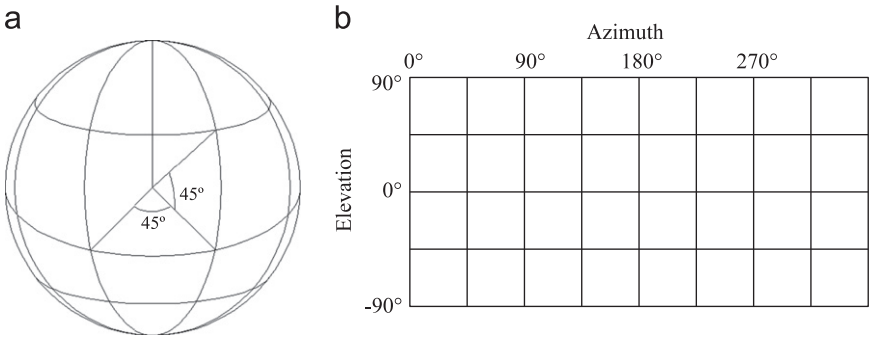

Fig. 10. Direction histogram: (a) splitting Azimuth/elevation into 45° bins and (b) resultant 2D histogram bins.

invariant orientation in the subsequent description. A *direction* in 3D space is defined by the azimuth and elevation angles whereas an *orientation* is defined by the addition of a third angle: tilt (see Fig. 9).

The first step is to determine the dominant *direction* for the keypoint. A 2D histogram is produced by grouping the Gaussian filtered volume gradients in bins which divide the azimuth and elevation into 45° sections, as shown in Fig. 10a (sphere) and Fig. 10b (resulting 2D histogram bins). Consequently there are N_a ($N_a = 8$) azimuth bins and N_e ($N_e = 4$) elevation bins. A regional weighting is applied to the gradients according to their voxel distance from the keypoint location as defined in Eq. (6). Points further than R_{max} voxels from the location are ignored in the current formulation. From a geodetic viewpoint (Fig. 10a) it can be seen that bins near the *equator* in this formulation are larger than those at the poles and this will bias the resulting histogram. This bias is compensated for by normalizing each histogram bin by its solid angle [30]. The output histogram is then smoothed using a Gaussian filter whose coefficients reflect the distance between the bins when located on a sphere. This filtering limits the effects of noise and the dominant directions are determined by searching for peaks and are refined using interpolation. Peaks in this 2D histogram within 80% of the largest peak are also retained as possible secondary directions in line with the formulation of [12].

The second step is to determine the *orientation* by calculating the tilt angle for each dominant direction. This is achieved by re-orientating the volume around the keypoint and calculating a one-dimensional (1D) 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. A simple 1D Gaussian filter is used to limit the effects of noise before peaks in the tilt histogram are used, with interpolation, to derive an estimate of keypoint tilt. Again, peaks within 80% of the largest peak are retained to give secondary orientations. Overall, in this formulation, we see that

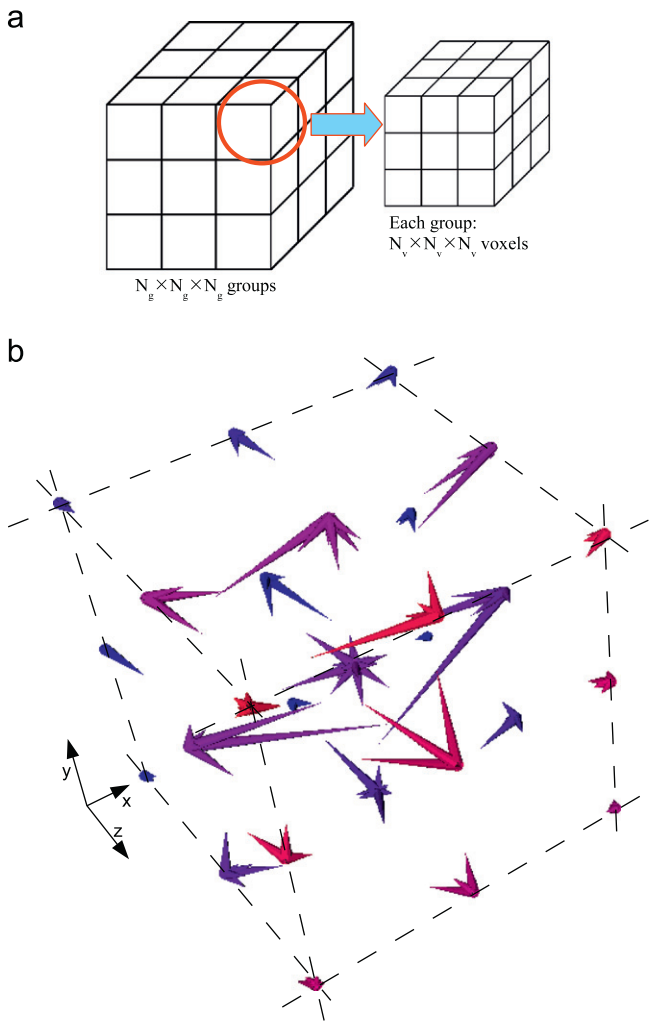

Fig. 11. 3D SIFT descriptor formulation: (a) voxel groupings for descriptor and (b) $3 \times 3 \times 3$ 3D SIFT descriptor shown in 3D space as $3 \times 3 \times 3$ groups of 3D gradient histograms.

keypoints may have more than one possible orientation that will require description.

4.5.2. *Keypoint description*

Once the orientation has been determined the point of interest can be described. In our case we build a $N_g \times N_g \times N_g$ grid of gradient histograms, with each histogram being computed from a $N_v \times N_v \times N_v$ voxel grouping as shown in Fig. 11a. Each gradient histogram is derived by splitting both azimuth and elevation into 45° bins, as described in Section 4.5.1. Consequently, each descriptor, normalized to unity, contains $N_g^3 \times N_a \times N_e$ elements. The final visualization of such a descriptor is shown in Fig. 11b as a 3D grid of gradient histograms.

5. Object detection methodology

An overview of descriptor generation is shown in Fig. 12 where we see the separation of interest point detection from descriptor generation which, in our comparison for this work, can be performed in a number of different ways (as described in Section 4). Interest point locations for an input volume are generated using the SIFT derived methodology described in Section 3. Descriptors for each volume are generated using these locations. The location of the keypoint is stored as part of the descriptor to facilitate a relative position consistency check in the recognition methodology.

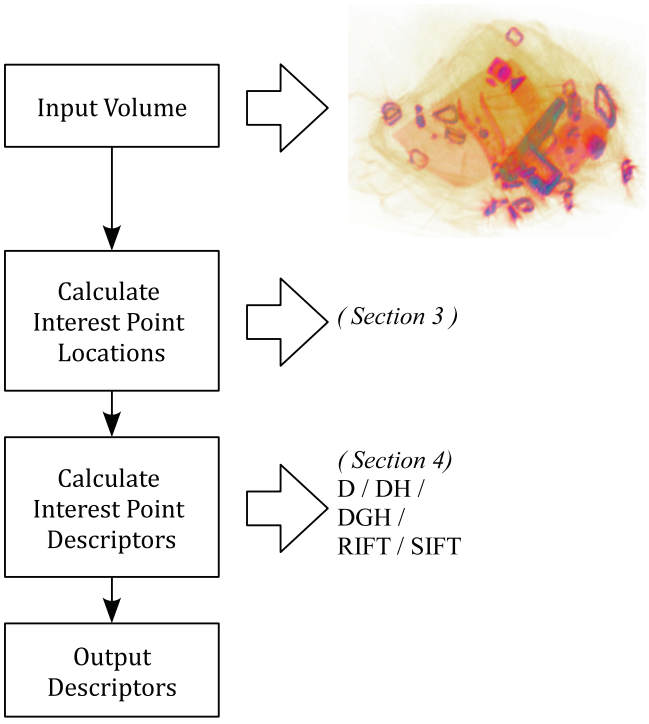

Fig. 12. Descriptor generation.

An object detection system methodology is shown in Fig. 13. Here we start with a known reference item from which descriptors are calculated. A candidate baggage item is received and processed to determine its descriptors. We form a set of matches between the reference and candidate items using Euclidean distance [12,39] as a base case for investigation (performance using alternative distance metrics is left as an area for future work). The matches between the reference and candidate item are filtered in an attempt to retain true matches and remove false matches. The output set of matches from this process are referred to as the correspondence set.

Two methods are examined when forming the correspondence set:

(a) The method of Lowe [12] where a match is accepted to the correspondence set if the ratio of the first and second best match distances is less than 0.8. We refer to this method as the distinction method. We consider this process from the candidate to the reference, i.e. a candidate/reference pair is added to the correspondence set if it is distinct compared to matches between the same candidate and the other reference descriptors.

(b) We reorder the matches from lowest to highest Euclidean distance in the feature space. We then choose a fixed percentage of the best matches as the correspondence set. We refer to this method as the percentile method, with parameter p defining the percentage of matches used.

Given the large number of possible false matches in this formulation we make use of RANSAC [40,41], to find an optimal match using the correspondence set as the input. The RANSAC algorithm [40] has been shown to cope well in the presence of significant outliers (here highly prevalent due to noise). This RANSAC formulation is used to select a set of three possible matches from the correspondence set from which a 3D transformation is derived using a common singular value decomposition (SVD) approach [42].

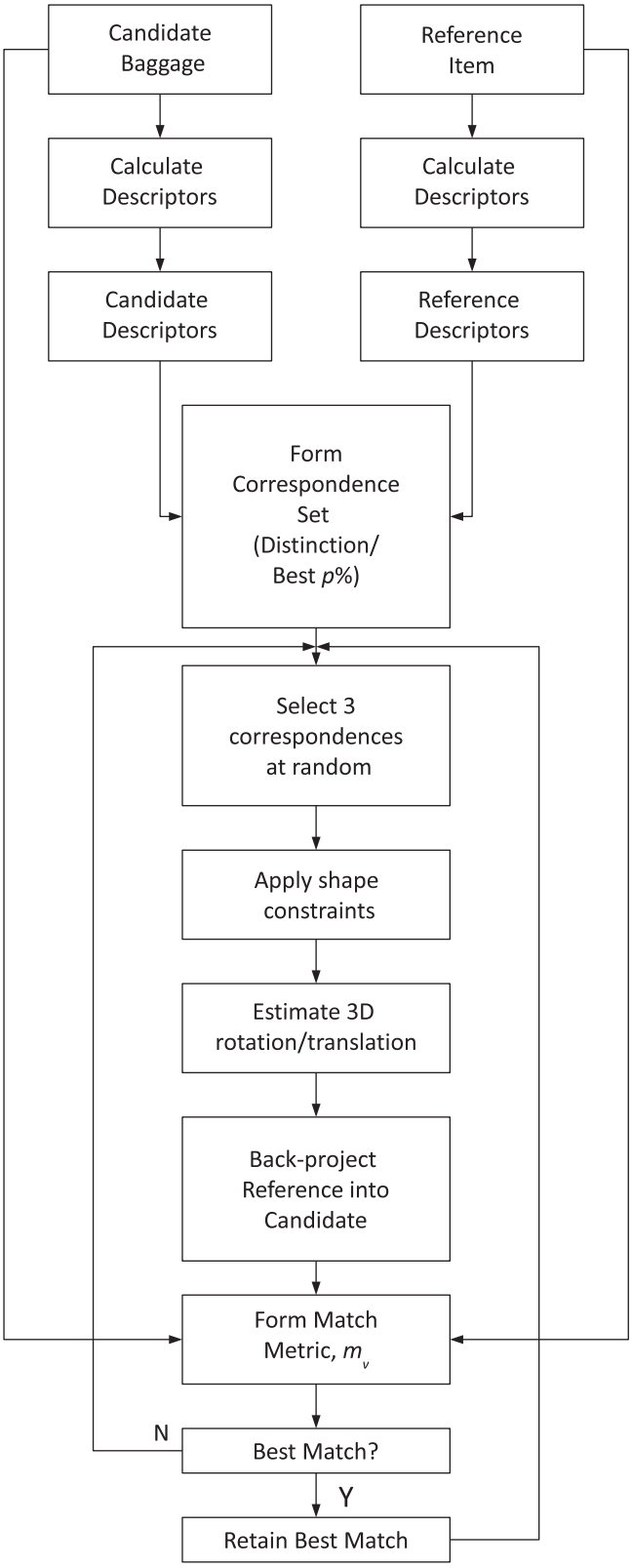

Fig. 13. Object recognition methodology.

Following the estimation of the transformation we further check to see if the three RANSAC selected matches are consistent or likely to provide a reasonable transform:

(a) The reference set and candidate set should be similar shapes: relative distance errors should be less than ϵ_r ($\epsilon_r = 10$ mm).

(b) The physical distance between the points in the reference set should be greater than ϵ_l ($\epsilon_l = 10$ mm). Empirical work has found that poor quality transforms often result if the points are not sufficiently separate as the lack of resolution in the imagery impacts these situations.

It should be noted that the one to one relationship between voxel measurements and real world distances allows the tolerances ϵ_r and ϵ_l to be specified in real world measurements (i.e. mm). These constraints aid the matching process by quickly rejecting poor quality selections prior to the verification stage.

If the relative distance and density criterion is passed a secondary verification is performed. Locations in the reference object with a density above a threshold τ_d ($\tau_d = 0.15$) are recorded to form a set of density verification locations. The threshold is applied in order to reduce the number of low density artefacts in the verification stage. The verification locations are transformed into the candidate baggage item space using the transform estimate provided by the SVD formulation. Given N_{vp} verification points we then form a quality of match metric, m_v, by examining the density differences between the verification locations in the reference item and the candidate baggage item,

$$m_v = \frac{\sum_{k=1}^{N_{vp}} |\rho_k - \psi_k|}{\sum_{k=1}^{N_{vp}} \psi_k}, \tag{8}$$

where ψ_k is the density at the kth verification point in the *reference* item and ρ_k is the density of the voxel closest to the k^{th} transformed verification point in the *candidate* baggage item. The measure is normalized by the sum density of the verification points in the reference item, as shown, to provide a metric that does not vary too greatly between the different reference items. This match metric includes all verification points in its formulation but it may be that a metric which rejects outliers would enhance performance – investigations into alternative match metrics is left as an area for future work. Our initial verification methodology followed the standard RANSAC approach in choosing the transformation that maximized the number of interest point matches in the baggage item, however, we found that higher recognition results were obtained by choosing the transformation that gave the minimum quality of match metric (Eq. (8)).

The optimal configuration for each descriptor was derived through experimentation. In each case the local neighborhood was varied in size as were the underlying configuration parameters and those that produced the best recognition rates were chosen. For instance, with SIFT we used $N_g \times N_g \times N_g$ groups, each containing $N_v \times N_v \times N_v$ voxels, where $N_g = \{1,3,5\}$ and $N_v = \{3,5\}$, yielding a neighborhood range from $3 \times 3 \times 3$ voxels up to $25 \times 25 \times 25$ voxels. We found that the $3 \times 3 \times 3$ group of $3 \times 3 \times 3$ voxels ($9 \times 9 \times 9$ voxel neighborhood) yielded the best detection rates, marginally outperforming $3 \times 3 \times 3$ group of $5 \times 5 \times 5$ voxels. The set of descriptors for comparison, described in Section 3, were computed using the empirically derived parameter settings shown in Table 1. The results of this comparison using the proposed object detection methodology and the parameter settings listed in Table 1 are presented in the next section.

6. Results

The CT scanner used for this work produces volumes with anisotropic voxels (Section 1.1). We choose to re-sample these volumes to create cubic voxels of uniform 2.5 mm dimension using cubic spline interpolation. We do not hard limit the interpolation results to the range [0.0,1.0] with the consequence that the working voxel value range is extended to [−1.0,2.0]. Use

Table 1
Descriptor settings.

Descriptor	Settings	Elements per descriptor
Density	$\sigma = 1.0$	1
Density histogram	$\sigma = 3.0$, $N_{DH} = 60$	60
Density gradient magnitude histogram	$\sigma = 3.0$, $N_{DGH} = 80$	80
RIFT	$\sigma = 3.0$, $N_b = 4$, $N_r = 2$, $d_w = 3.0$	8
SIFT	$\sigma = 4.5$, $N_g = 3$, $N_v = 3$, $N_a = 8$, $N_e = 4$, $R_{max} = 9$	864

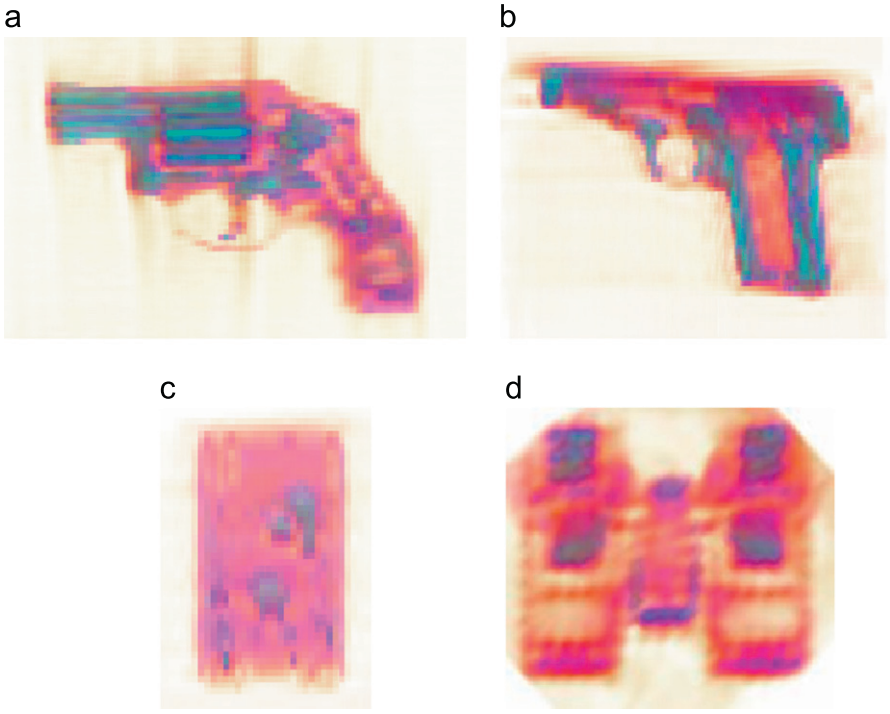

Fig. 14. Reference CT object volumes used for detection: (a) Smith & Wesson revolver, (b) Browning pistol, (c) Apple iPod and (d) binoculars.

of this extended voxel value range in descriptor formulations (Section 4) needs to be noted.

For this comparative study four target items of interest were used (Smith & Wesson revolver, Browning pistol, Apple iPod and compact binoculars) of which scans are shown in Fig. 14. Furthermore a mix of baggage types (e.g. holdalls, suitcases, handbags, etc.) containing a variety of clutter items as would be found in a typical airport scenario, including and excluding these items of interest, were scanned using a Reveal Imaging Technologies 3D CT80 scanner. Table 2 shows the number of baggage items scanned which contained one of these target items or which were left clear of the named targets but still contained regular background clutter.

Each baggage item is "searched" using the object detection methodology outlined in Section 5 for each of the four reference CT object volumes as shown in Fig. 14. From this each baggage item produces a verification match metric result, m_v, as described in Section 5 (a measure of similarity between the reference item and the baggage item). A decision on whether a target item has been detected is made by comparing the verification match metric result, m_v, against a detection threshold, τ_m. Given a ground truth knowledge of which baggage items contain the target items and which do not, we can calculate both a true-positive rate, $TP(\tau_m)$, and a false-positive rate, $FP(\tau_m)$, for a given setting of τ_m. Our analysis uses receiver-operating-characteristic (ROC) plots [43] to investigate the overall system performance as each descriptor type is used. These plots show $TP(\tau_m)$ against $FP(\tau_m)$ and indicate the trade off between the true detection of threat items versus the false detection as the detection threshold, τ_m, is varied. When producing a numerical performance result we choose to quote the true-positive rate for minimal false-positive rate ($< 1\%$) rather than the ROC equal error rate [44] as we feel that this is more applicable to the operating conditions of such a system in an operational security environment (even a moderate false-positive rate is not desirable).

The ROC plot gives one aspect of performance. We also form a plot that shows a measure of tolerance to error in the value of the detection threshold, τ_m, should a fixed value be chosen to decide the presence of the target item. We refer to this as the threshold-quality, $Q(\tau_m)$, where

$$Q(\tau_m) = TP(\tau_m) \times [1 - FP(\tau_m)]. \quad (9)$$

Fig. 15a shows how the true-positive and false-positive rates are combined to form the threshold-quality. The width of the threshold-quality plot indicates the separation between the rise in true-positive rate and the rise in false-positive rate. The height of the threshold-quality peak is also indicative of performance. If the true-positive and false-positive rates are well separated then the threshold-quality will reach a peak value of 1.0, indicating a perfect ROC plot. However, if the true-positive and false-positive

transition regions overlap the threshold-quality peak will be less than 1.0. Fig. 15b and c show threshold-quality plots for two systems, both with perfect ROC plots. It can be seen in Fig. 15b that the threshold-quality peak is narrow indicating that the true-positive transition region is close to the false-positive transition region. A better scenario is shown in Fig. 15c where the threshold-quality peak is broad indicating a large separation between the true-positive and false-positive transition regions. This broad peak indicates that, when allocating a value to the detection threshold (τ_m), a greater tolerance to error in its assignment exists.

Table 2
Items scanned.

Baggage item contents	Scans in collection
Smith & Wesson revolver+clutter	21
Browning pistol+clutter	30
Apple iPod+clutter	15
Compact binoculars+clutter	14
Clutter only	180

First we examine the detection performance using the distinction approach of Lowe [12] to form the correspondence set and then we will look at the performance using the percentile method proposed in our earlier work [6].

ROC plots for detection of the revolver, pistol, iPod and binoculars when using the distinction method are shown in Fig. 16. It can be seen that there is a considerable variation in detection performance between the descriptor types, as well as differing levels of detection of each target item.

For the revolver (Fig. 16a) the best result using the distinction method is obtained using the RIFT descriptor with a detection

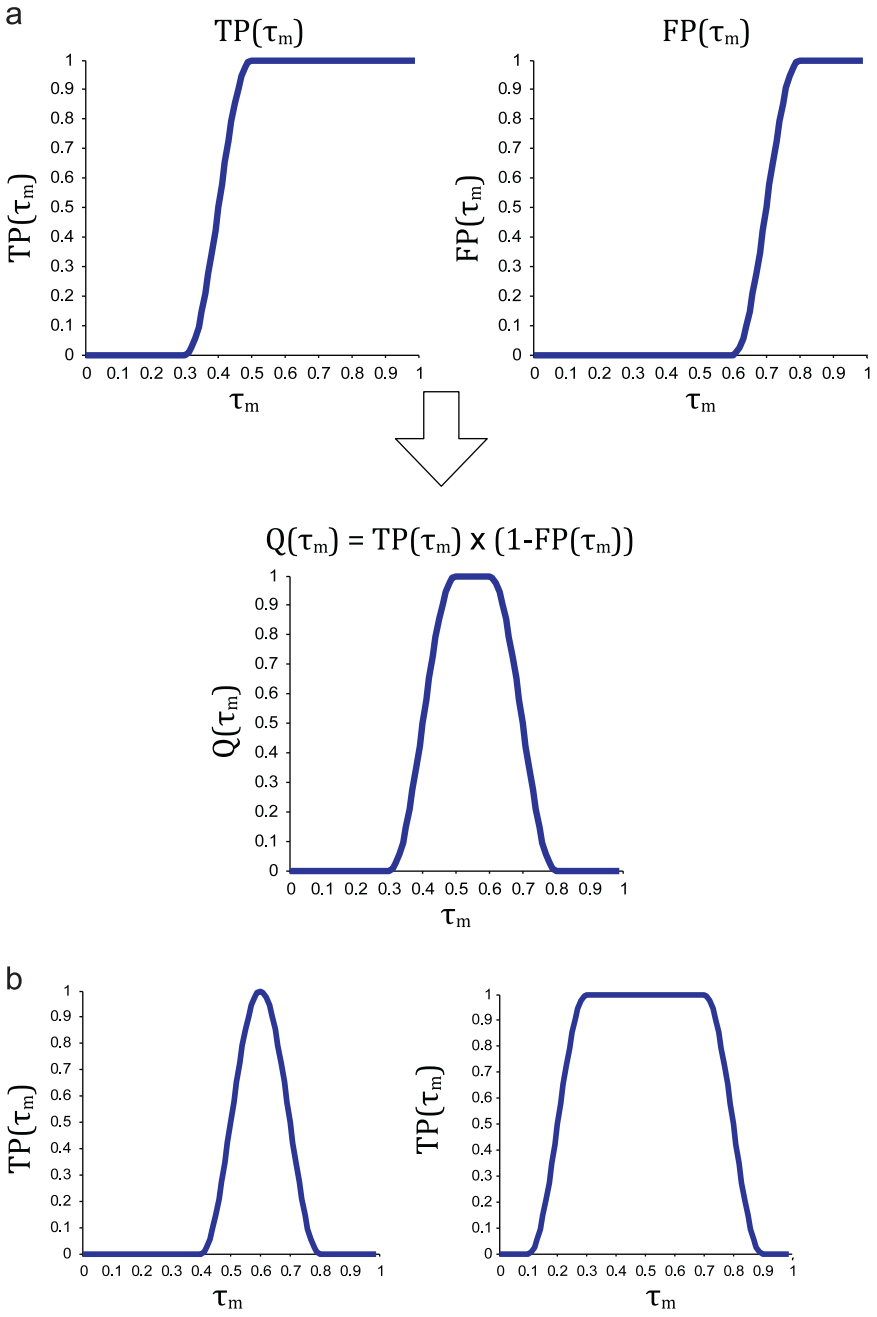

Fig. 15. Threshold-quality: (a) threshold-quality derivation, (b) poor threshold quality and (c) good threshold-quality.

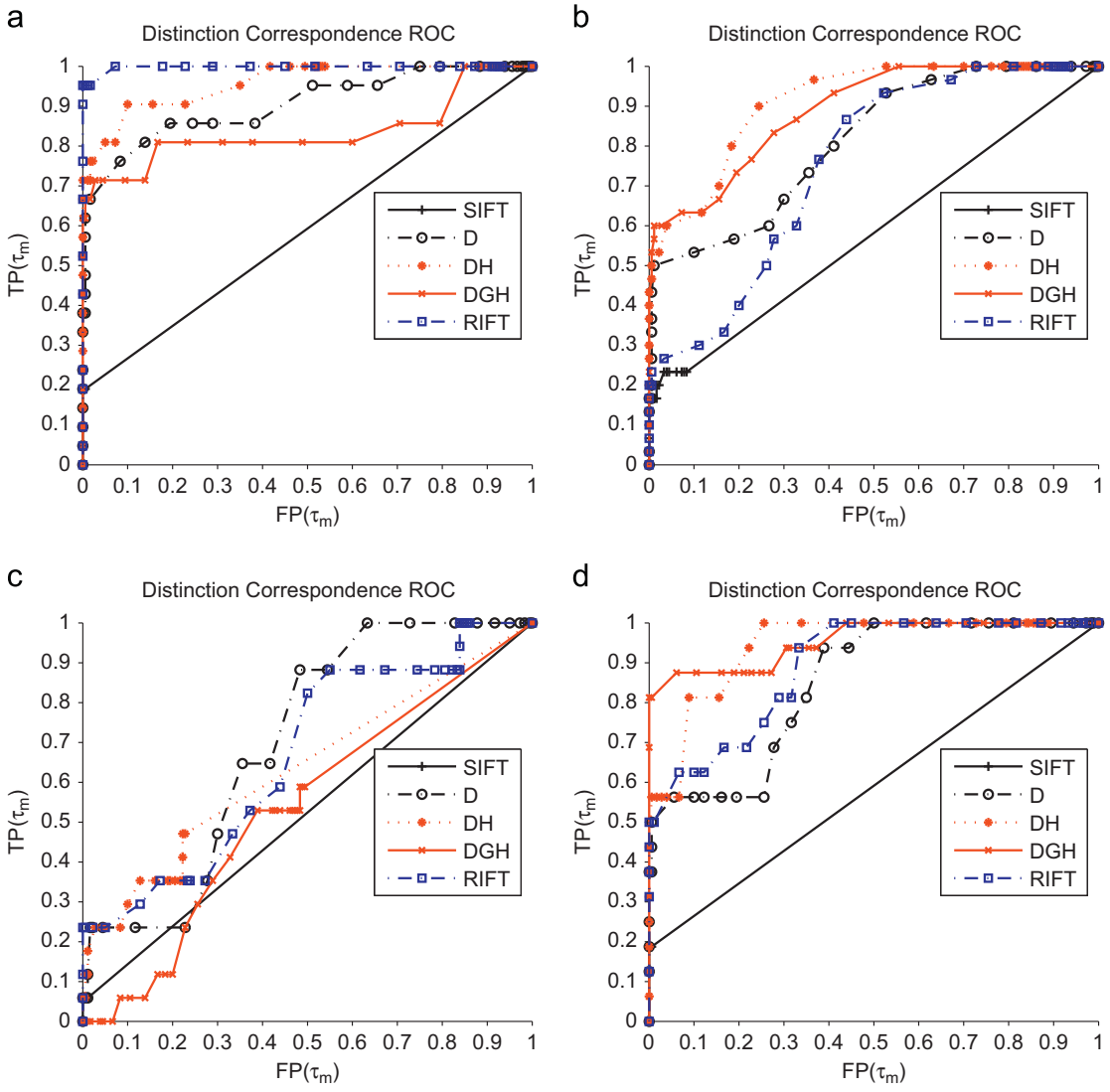

Fig. 16. Target item ROC curves using distinction to form the correspondence set: (a) revolver, (b) pistol, (c) Apple iPod and (d) binoculars.

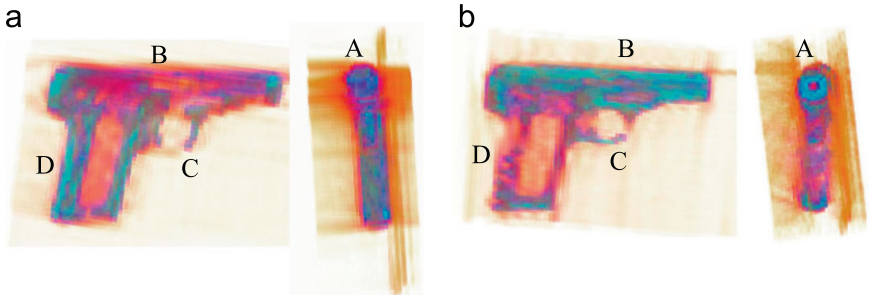

Fig. 17. Browning pistol reference item quality. Note A: muzzle, B: barrel, C: trigger and D: grip. (a) First reference and (b) second reference.

rate of ~95% with detection using D, DH and DGH at ~60/70%. The performance of SIFT is poor with a detection rate of ~20%.

The pistol performance is poorer (Fig. 16b) with a detection rate of ~55% with a negligible false-positive rate using the DGH descriptor. This is closely followed by D and DH descriptors (~50%) with RIFT and SIFT both poor (~20%).

The iPod performance is worst (Fig. 16c) with a detection rate of ~20% using the RIFT descriptor, closely followed by D, DH and DGH (~15%) with SIFT again the worst performing (~5%).

Detection of the binoculars is ~80% (Fig. 16d) with negligible false-positives using the DGH descriptor. Detection using RIFT, D and DH descriptors is ~50% with SIFT again worst with a detection rate of ~20%.

An investigation into the poor quality of the pistol results compared to those of the revolver indicated that the scan quality of the reference item affects performance. Fig. 17a shows the reference used to create the results in Fig. 16b. Fig. 17b shows a different scan of the Browning pistol which is used as a reference.

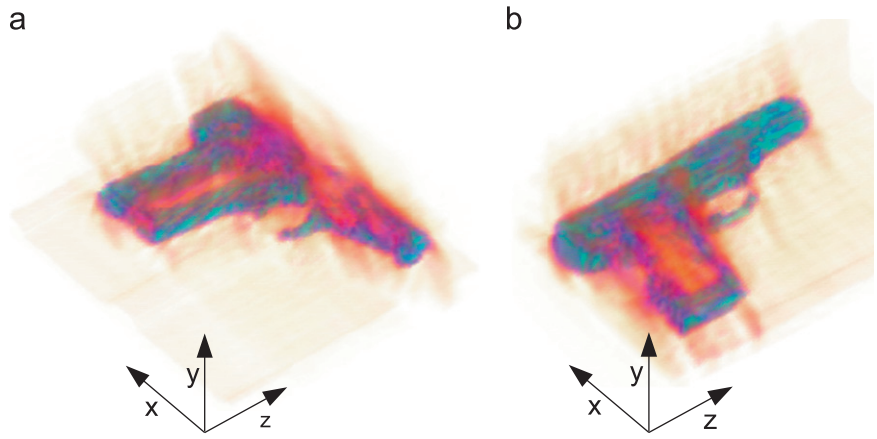

Fig. 18. Browning reference item orientation in CT baggage scanner: (a) first reference and (b) second reference.

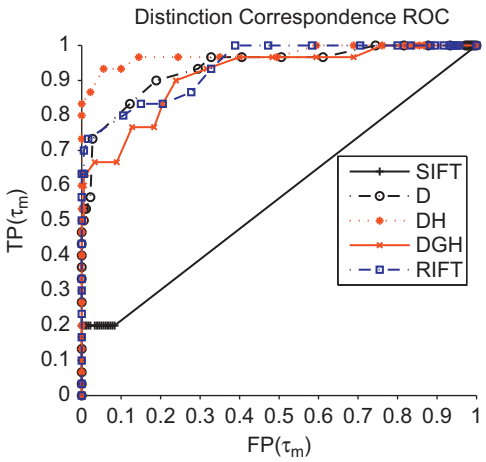

Fig. 19. ROC using second Browning pistol as reference.

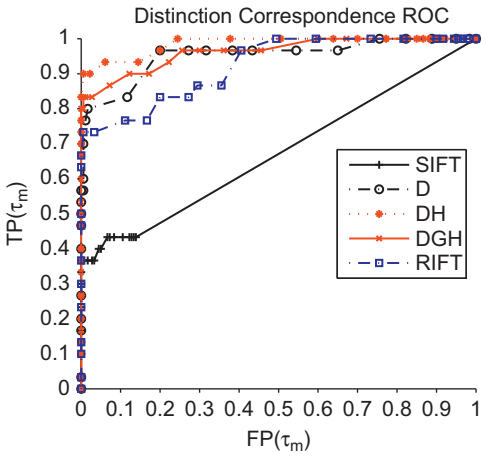

Fig. 20. ROC for combination of pistol results.

Note in this secondary example (Fig. 17b) the clarity of the pistol muzzle (A) compared to Fig. 17a. Also note the apparent density differences in the barrel (B), trigger guard (C) and grip (D) caused by metal artefacts and anisotropic scanning. These differences will affect the resulting descriptors, both in value and location, and this has obvious implications for location of similar points in randomly scanned baggage items. The difference between these scans is the orientation of the pistol relative to the CT scanner *z*-axis, as shown in Fig. 18. The original pistol reference (Fig. 18a) was orientated such that the barrel was parallel to the *xy*-plane resulting in the barrel being scanned with a 5 mm resolution (the CT slice spacing – see Section 1.1). The alternate pistol reference (Fig. 18b) was scanned such that the barrel was orthogonal to the *xy*-plane resulting in a barrel cross-section resolution of $\sim$1.6 mm (the slice pixel resolution – see Section 1.1).

Fig. 19 shows the ROC plot using match distinction to form the correspondence set when using the *alternate* pistol reference. Here we can see a better detection rate of $\sim$85% using the DH descriptor (up from $\sim$50%). The RIFT descriptor has a detection rate of $\sim$70% (up from $\sim$20%) with DGH at $\sim$60% (from $\sim$55%), D at $\sim$50% (unchanged) and SIFT at $\sim$20% (unchanged).

We combined the results for both pistol references by choosing the result with the lowest verification match metric value, m_v, to observe if the combination would provide increased levels of performance. Fig. 20 shows the ROC plot for this situation where we can see that an improvement does occur (compared to the individual reference item results shown in Figs. 16b and 19). The best performance again comes from the DH descriptor with a detection rate of $\sim$90% with negligible false-positives (up from $\sim$85%). The performance using the other descriptors is also improved: D $\sim$75% (up from $\sim$50%); DGH at $\sim$80% (up from $\sim$60%); RIFT up slightly at $\sim$75% (from $\sim$70%); SIFT at $\sim$35% (up from $\sim$20%).

An investigation into why the use of the SIFT descriptor yielded poor detection results was carried out. Analysis of the correspondence set showed that, when using match distinction, very few of the SIFT matches are deemed to be suitable. Table 3 shows the mean correspondence set size (as a % of total matches) for each target item and each descriptor when analyzed over the datasets given in Table 2. For the D, DH, DGH and RIFT descriptors we see correspondence set sizes between 0.80% and 3.08% of the total number of matches. When compared to these descriptors, the SIFT descriptor has very few matches in the correspondence set: between 0.01% and 0.07%. This is indicative of poor quality descriptors (very few pass the distinction criterion) and it would appear that this restricts its performance: true matches are rejected from the correspondence set and not enough are made available to the object detection method for reliable recognition of the target items.

It is notable that the use of the distinction method differs from the selection method used in our prior work [6] where significantly improved SIFT 3D object detection results were obtained.

In light of these results and with the support of the earlier work [6] we vary the method used to form the correspondence set away from the seminal 2D SIFT variation [12] and use our alternative percentile method as previously discussed in Section 5.

Rather than using the match distinction method we instead sort the matches by match distance and then choose a fixed percentage of the best matches as per [6].

Fig. 21 shows the results when the best 2% of matches are chosen to form the correspondence set. For the revolver (Fig. 21a) we can see near 100% detection with minimal false-positives using DH, DGH and RIFT descriptors. Both D and SIFT descriptors have detection rates of $\sim 85\%$. Using the second pistol reference (Fig. 21b) we again see near 100% detection using the RIFT descriptor, closely followed by DH and DGH ($\sim 90\%$) with SIFT at $\sim 65\%$ and D at $\sim 35\%$. Detection of the iPod is still poor (Fig. 21c), though slightly improved, at $\sim 30\%$ (increased from $\sim 20\%$) using DH, followed by DGH, RIFT and SIFT at $\sim 20\%$. The density descriptor has a detection rate of $\sim 0\%$ using our negligible false-positive rate definition. The binoculars show near 100% detection (Fig. 21d) using RIFT, DGH and SIFT, with DH close behind at $\sim 95\%$. The density descriptor is again poor with a detection rate of $\sim 0\%$.

Given a number of ROC plots that appear to show 100% detection rates, mainly due to the limited amount of target items, we can also investigate performance using the threshold-quality, $Q(\tau_m)$, as the detection threshold, τ_m, is varied (Equation 9). Threshold-quality plots relating the ROC plots in Fig. 21 are given in Fig. 22.

Fig. 22a shows the plot in the case of the revolver where we see the superior performance of the DH and RIFT descriptors over the DGH descriptor that it is not possible to see in the ROC plots (Fig. 21a). Both the DH and RIFT descriptors reach a peak when $\tau_m \simeq 0.45$ and then fall off when $\tau_m \simeq 0.6$. The DGH descriptor only reaches a peak for $\tau_m \simeq 0.55$ and then almost immediately

Table 3
Mean correspondence set size (as % of total matches).

Descriptor	Revolver	Pistol	iPod	Binoculars
Density	2.31 ± 0.10	2.47 ± 0.07	3.08 ± 0.10	1.71 ± 0.11
Density histogram	0.80 ± 0.31	1.32 ± 0.30	1.20 ± 0.50	0.96 ± 0.27
Density gradient histogram	1.55 ± 0.23	1.18 ± 0.23	0.93 ± 0.20	0.81 ± 0.18
RIFT	1.39 ± 0.14	1.05 ± 0.15	1.17 ± 0.20	1.15 ± 0.11
SIFT	0.02 ± 0.01	0.07 ± 0.06	0.02 ± 0.01	0.01 ± 0.01

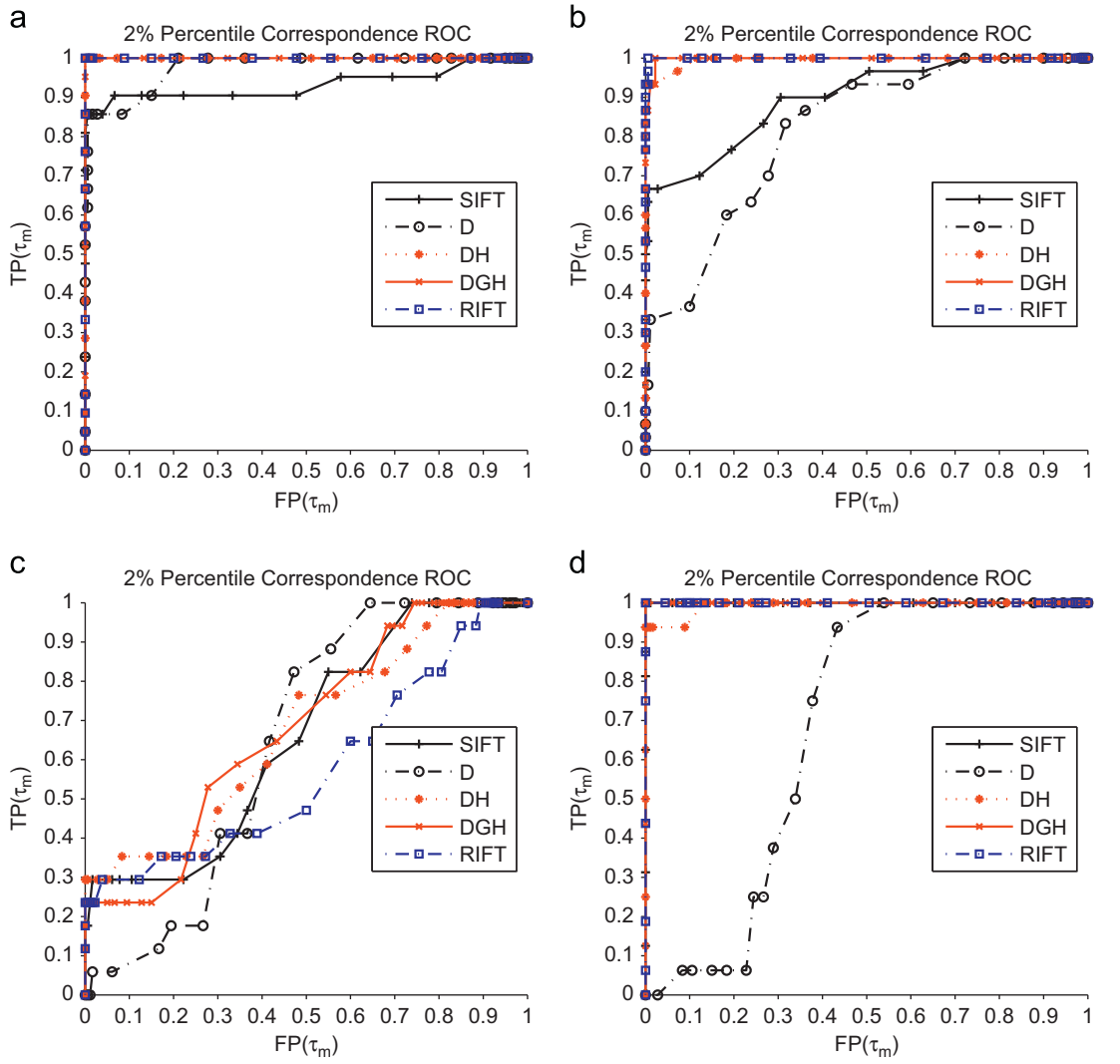

Fig. 21. ROC curves when using percentile matches ($p = 2\%$) for correspondence set: (a) revolver, (b) pistol (second reference), (c) Apple iPod and (d) binoculars.

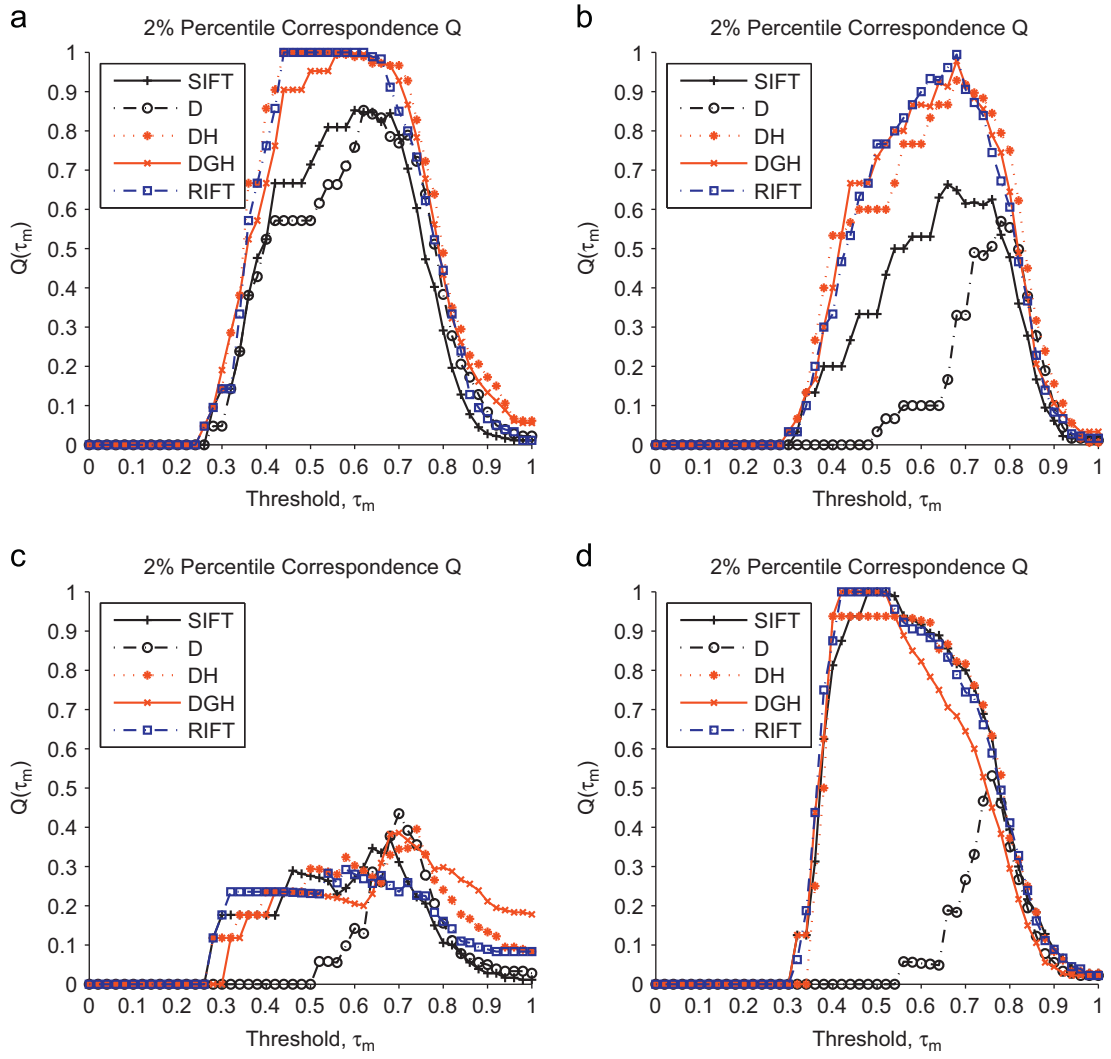

Fig. 22. Threshold-quality for percentile ($p=2\%$) correspondence set: (a) revolver, (b) pistol (second reference), (c) Apple iPod and (d) binoculars.

starts to fall away. The implication for this, in a noisy environment, would be that the DH and RIFT descriptors would be more reliable than the DGH.

Fig. 22b shows the threshold quality for the second pistol reference. Here we see that, although both the RIFT and DGH descriptors reach a peak of 1.0, they quickly fall away. This does not appear to be as good as the revolver.

Fig. 22c shows the results for the Apple iPod. Here we see poor results already indicated by the ROC plot (Fig. 21d).

Fig. 22d shows the results for the binoculars. Here we can see that the RIFT descriptor has the broadest peak, closely followed by DGH. The SIFT descriptor, though apparently with near perfect ROC, only just reaches a peak of 1.0 before falling away. The DH descriptor, though apparently not as good in the ROC plot (Fig. 21d), has the widest peak which would indicate it is more tolerant to detection threshold selection error.

Varying threshold-quality gives us an alternative statistical visualization of the relative performance of the different 3D interest point descriptors within this context.

7. Conclusions

We have shown a comparison of differing 3D point descriptors applied to the problem of object detection in complex 3D CT volumetric imagery. It has been demonstrated that approaches based on simpler density information outperform more complex 3D extensions of common and established point descriptors adapted from 2D image recognition [12,34]. This may be due to the poor quality of imagery available for this work: imaging artefacts and noise hinder a reliable invariant orientation for the SIFT descriptor. Note also that the simpler descriptors also have the advantage that they are not as computationally expensive as the SIFT descriptor: computational efficiency is an aspect of the implementation that will require thought when deployed in a real world situation.

Our results have shown that creation of the correspondence set using the distinction method of Lowe [12] is not the best approach in the case of complex CT imagery containing a large number of artefacts. Better results are obtained if the correspondence set is determined by sorting the matches by Euclidean match distance and then taking a fixed percentage of the best matches [6].

We have shown that an anisotropic scanning system will affect the recognition results. The Browning pistol was scanned in orthogonal orientations and produced very different recognition results. Care thus needs to be taken when choosing a reference item or, as we have demonstrated, multiple reference volumes can be used to improve detection results. The use of multiple reference object scans and methods of determining reference scan quality is also left as an area for future work.

Further work will investigate the use of multiple objects as a derivative for the reference volume and also the evaluation of quality and artefact information within the imagery. At present we have focussed on detection performance without considering the speed of operation. This aspect of the implementation will need to be investigated in the future.

Object class recognition will be investigated as a means of recognising previously unseen objects. This will address one flaw with the specific instance approach: a reference object is required for each potential target item.

Conflict of interest statement

None declared.

Acknowledgments

This project is funded under the Innovative Research Call in Explosives and Weapons Detection (2007), a cross-government programme sponsored by Home Office Scientific Development Branch (HOSDB), Department for Transport (DfT), Centre for the Protection of National Infrastructure (CPNI) and Metropolitan Police Service (MPS). The authors are grateful for the additional support from Reveal Imaging Technologies Inc. (USA).

References

[1] B. Abidi, Y. Zheng, A. Gribok, M. Abidi, Improving weapon detection in single energy X-ray images through pseudocoloring, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 36 (2006) 784–796.
[2] A. Schwaninger, A. Bolfing, T. Halbherr, S. Helman, A. Belyavin, L. Hay, The impact of image based factors and training on threat detection performance in X-ray screening, in: Proceedings of the Third International Conference on Research in Air Transportation, ICRAT, 2008, pp. 317–324.
[3] S. Singh, M. Singh, Explosives detection systems (EDS) for aviation security, Signal Processing 83 (2003) 31–55.
[4] G. Herman, Fundamentals of Computerized Tomography: Image Reconstruction from Projections, 2nd edition, Springer Verlag, 2009.
[5] W. Bi, Z. Chen, L. Zhang, Y. Xing, A volumetric object detection framework with dual-energy CT, in: IEEE Nuclear Science Symposium Conference Record, 2008, pp. 1289–1291.
[6] G. Flitton, T. Breckon, N. Megherbi, Object recognition using 3D SIFT in complex CT volumes, in: F. Labrosse, R. Zwiggelaar, Y. Liu, B. Tiddeman (Eds.), Proceedings of the British Machine Vision Conference, BMVA Press, 2010, pp. 11.1–11.12.
[7] W. Kalender, R. Hebel, J. Ebersberger, Reduction of CT artifacts caused by metallic implants, Radiology 164 (1987) 576.
[8] N. Menvielle, Y. Goussard, D. Orban, G. Soulez, Reduction of beam-hardening artifacts in X-ray CT, in: 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, pp. 1865–1868.
[9] K. Jeong, J. Ra, Reduction of artifacts due to multiple metallic objects in computed tomography, in: E. Samei, J. Hsieh (Eds.), Medical Imaging 2009: Physics of Medical Imaging, vol. 7258, SPIE, 2009, p. 72583E.
[10] G. Wang, M. Vannier, Stair-step artifacts in three-dimensional helical CT: an experimental study, Radiology 191 (1994) 79–83.
[11] J. Barrett, N. Keat, Artifacts in CT: recognition and avoidance, Radiographics 24 (2004) 1679–1691.
[12] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.
[13] T. Kadir, M. Brady, Saliency, scale and image description, International Journal of Computer Vision 45 (2001) 83–105.
[14] H. Bay, T. Tuytelaars, L. Van Gool, Surf: speeded up robust features, Computer Vision-ECCV (2006) 404–417.
[15] C. Schmid, R. Mohr, Local grayvalue invariants for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 530–535.
[16] C. Harris, M. Stephens, A combined corner and edge detector, in: Proceedings of the Fourth Alvey Vision Conference, pp. 147–151.
[17] D. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157.
[18] H. Yoshida, J. Näppi, P. Maceneaney, D. Rubin, A. Dachman, Computer-aided diagnosis scheme for detection of polyps at CT colonography, Radiographics 22 (2002) 963–979.
[19] K. Suzuki, M. Epstein, I. Sheu, R. Kohlbrenner, D. Rockey, A. Dachman, Massive-training artificial neural networks for CAD for detection of polyps in CT colonography: false-negative cases in a large multicenter clinical trial, in: Proceedings of the Fifth IEEE International Symposium on Biomedical Imaging: from Nano to Macro, pp. 684–687.
[20] 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, pp. 1289–1291.
[21] 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, pp. 954–955.
[22] D. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 13 (1981) 111–122.
[23] M. Baştan, M. Yousefi, T. Breuel, Visual words on baggage X-ray images in: P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, W. Kropatsch (Eds.), Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, vol. 6854, Springer, Berlin/Heidelberg, 2011, pp. 360–368.
[24] N. Megherbi, G. Flitton, T. 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, pp. 1833–1836.
[25] J.J. Koenderink, A.J. van Doorn, Surface shape and curvature scales, Image and Vision Computing 10 (1992) 557–564.
[26] M. Novotni, R. Klein, Shape retrieval using 3D zernike descriptors, Computer-aided Design 36 (2004) 1047–1062.
[27] S. Nercessian, K. Panetta, S. Agaian, Automatic detection of potential threat objects in X-ray luggage scan images, in: Proceedings of the IEEE Conference on Technologies for Homeland Security, pp. 504–509.
[28] C. Oertel, P. Bock, Identification of objects-of-interest in X-ray images, in: Proceedings of the 35th IEEE Applied Imagery and Pattern Recognition Workshop, p. 17.
[29] R. Gesick, C. Saritac, C.-C. Hung, Automatic image analysis process for the detection of concealed weapons, in: Proceedings of the Fifth Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1–4.
[30] P. Scovanner, S. Ali, M. Shah, A 3D SIFT descriptor and its application to action recognition, in: Proceedings of the 15th International Conference on Multimedia, ACM Press, New York, NY, USA, 2007, pp. 357–360.
[31] W. Cheung, G. Hamarneh, N-Sift: N-dimensional scale invariant feature transform for matching medical images, in: Proceedings of the Fourth IEEE International Symposium on Biomedical Imaging: from Nano to Macro, 2007, pp. 720–723.
[32] D. Ni, Y. Chui, Y. Qu, X. Yang, J. Qin, T. Wong, S. Ho, P. Heng, Reconstruction of volumetric ultrasound panorama based on improved 3D SIFT, Computerized Medical Imaging and Graphics 33 (2009) 559–566.
[33] 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 Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–8.
[34] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using local affine regions, IEEE Transactions on Pattern Analysis and Machine Intelligence (2005) 1265–1278.
[35] K. Mikolajczyk, B. Leibe, B. Schiele, Local features for object class recognition, in: Proceedings of the Tenth IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1792–1799.
[36] C. Belcher, Y. Du, Region-based SIFT approach to iris recognition, Optics and Lasers in Engineering 47 (2009) 139–147.
[37] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, B. Lu, Person-specific sift features for face recognition, IEEE International Conference on Acoustics, Speech and Signal Processing 2 (2007) 593–596.
[38] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005, vol. 1, 2005, pp. 886–893.
[39] 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 (2010) 1271–1283.
[40] M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.
[41] S. Choi, T. Kim, W. Yu, Performance evaluation of RANSAC family, in: Proceedings of the British Machine Vision Conference, pp. 81.1–81.12.
[42] K. Arun, T. Huang, S. Blostein, Least-squares fitting of two 3D point sets, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (1987) 698–700.
[43] T. Fawcett, An Introduction to ROC analysis, Pattern Recognition Letters 27 (2006) 861–874.
[44] M. Schuckers, Receiver operating characteristic and equal error rate, Computational Methods in Biometric Authentication, Information Science and Statistics, Springer, London, 2010, pp. 155–204.

Greg Flitton received the B.Eng (Hons.) degree in Electronic and Electrical Engineering from the University of Birmingham in 1988 and an M.Sc. in Applied and Modern Optics from the University of Reading in 1989.

A career in mobile telecommunications followed researching DSP algorithms for various telephony standards before a change of tack into machine vision in 2006. He obtained a M.Sc. in Computational and Software Techniques in Engineering (Digital Signal and Image Processing) from Cranfield University in 2007 and a Ph.D. in Computer Vision in 2012 with an interest in transportation security.

Toby P. Breckon received the B.Sc. degree (Hons.) in Artificial Intelligence and Computer Science and the Ph.D. degree in Informatics from the University of Edinburgh, Edinburgh, UK, in 2002 and 2006, respectively.

He has held visiting positions with Northwestern Polytechnical University, Xi'an, China, and Waseda University, Tokyo, Japan. He is currently a Senior Lecturer with the School of Engineering, Cranfield University, Bedfordshire, UK. In addition, he is a visiting member of faculty with Ecole Supérieure des Technologies Industrielles Avancées, France. His key research interests, in the domain of computer vision and robotics, are as follows: 3D sensing and reasoning, 3D visual completion, vision in built environments, sensor fusion, visual surveillance, and robotic deployment in hazardous environments.

Dr. Breckon is a Chartered Engineer and a member of the IET. In addition, he is an Accredited Imaging Scientist and an Associate of the Royal Photographic Society. He led the development of image-based automatic threat detection for the 2008 UK MoD Grand Challenge winning team [R.J. Mitchell Trophy (2008), IET Award for Innovation (2009)]. He has a range of publications and leads several funded research projects in Applied Image Processing and Computer Vision.

Najla Megherbi received the M.Sc. degree in computer science from Évry-Val d'Essonne University, France, in 2003 and the Ph.D. degree in Computer Science from Lille University, France, in 2006. She was a Research Associate in the Digital Imaging Research Centre and Kingston University, UK, in 2006. Since 2008, she has been a Research Fellow in the Digital Image Processing Team within the Applied Mathematics and Computing Group (AMAC) at Cranfield University, UK. Her research interests are mainly focused on video surveillance and 3D Computed Tomography imagery for airport screening systems.