Object Recognition using 3D SIFT in Complex CT Volumes

Greg Flitton
g.t.flitton@cranfield.ac.uk

Toby P. Breckon
toby.breckon@cranfield.ac.uk

Najla Megherbi
n.megherbi@cranfield.ac.uk

Applied Mathematics and Computing Group
School of Engineering
Cranfield University
Cranfield, UK

Abstract

The automatic detection of objects within complex volumetric imagery is becoming of increased interest due to the use of dual energy Computed Tomography (CT) scanners as an aviation security deterrent. These devices produce a volumetric image akin to that encountered in prior medical CT work but in this case we are dealing with a complex multi-object volumetric environment including significant noise artefacts. In this work we look at the application of the recent extension to the seminal SIFT approach to the 3D volumetric recognition of rigid objects within this complex volumetric environment. A detailed overview of the approach and results when applied to a set of exemplar CT volumetric imagery is presented.

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 [9]. 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 within this domain. In this area we specifically look at the use of Computed Tomography (CT) volumetric imagery where a three dimensional voxel image of the baggage/parcel item is obtained. Items of interest can be difficult to detect within this environment due to a range of orientation, clutter and density confusion in a traditional 2D X-ray projection [18]. An example of this is shown in Figure 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. The difference of orientation is a limitation of 2D X-ray scanners which makes detection (automatically or by human operators) particularly challenging.

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 [21]. It is from these scanners that we obtain a series of image slices through the bag which can
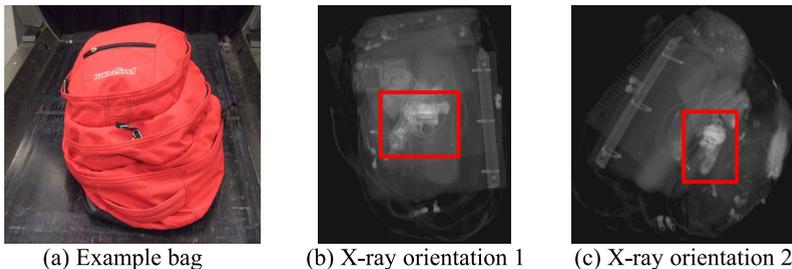
(a) Example bag   (b) X-ray orientation 1   (c) X-ray orientation 2

Figure 1: Bag and X-rays

(a) Front view   (b) Side view   (c) Top view
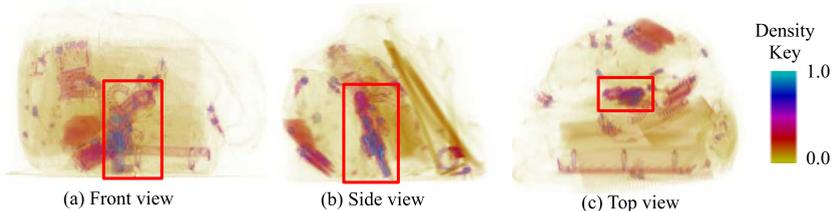
Figure 2: 3D volume of complex bag containing a revolver

be reconstructed as a traditional CT 3D volume akin to those encountered within medical CT imaging. Prior work on the automatic recognition of objects within this complex 3D volumetric imagery is very limited. Only the prior work of 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 Figure 1). By contrast here we consider explicit 3D recognition of items within the 3D CT volume domain.

The Scale Invariant Feature Transform (SIFT) approach [13] is a widely recognized precursor to a substantial body of feature point based object recognition strategies [11, 15]. The extension of the SIFT approach to three dimensional data has been attempted by several researchers [2, 6, 7, 16, 17, 19]. Scovanner *et al.* [19] created a 3D SIFT descriptor for application to action recognition in video volumes and additionally work has been encountered in the application of 3D SIFT to medical registration [2, 6, 17] or panoramic medical image stitching [7, 16]. The use of SIFT for 2D object recognition relies on objects having textures internal to their boundary such that these regions can be reliably described from one image to another. Points of interest that are on an object boundary are more easily corrupted by the presence of other objects. Similarly, in 3D, we anticipate that objects will need to have reliable textures internal to their surface which are not corrupted by the presence of other objects close by. The 3D extension of SIFT for explicit object recognition (its original application [12]) has received little attention in the arena of complex volumetric imagery.

1.1   Complex CT Volumetric Imagery

An example of a 3D scan of an item of baggage is shown in Figure 2 where we see the presence of an item of interest amongst more general cluttered items. Here the voxel density is represented in a continuous range [0,1].

The type of baggage scanner machine used to capture the CT volumetric imagery for this work is primarily aimed at dual energy explosives detection [21]. As a result of this primary (non object recognition based) objective two additional consequences are suffered within the imagery: 1) the presence of metal items causes significant artefacts within the imaging
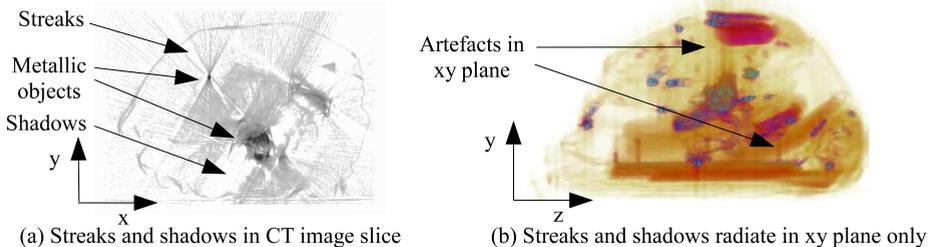
(a) Streaks and shadows in CT image slice (b) Streaks and shadows radiate in xy plane only

Figure 3: An example of metal artefacts in CT baggage imagery

(Figure 3) and 2) the resolution is anisotropic and limited to [1.6mm ×1.6mm × 5mm]. The metal artefacts radiate out in the *x-y* plane and do not remain consistent from one scan to another if the metallic region changes orientation.

Although prior work has looked at the removal of metal artefacts in medical CT imagery [9, 10, 14] this has not been considered at the present time within this work. Additionally we recognize that the poor resolution gives rise to stair step artefacts [4, 22]. Although this poses significant challenges for recognition we consider here the limitation in resolution to be similar to the scale invariant challenges addressed by the SIFT algorithm and additionally the unpredictable nature of the metal artefacts to be akin to that of recognition in the presence of occlusion - again an area in which SIFT [13] has previously excelled. 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 for any work within automated 3D recognition.

2 Extension of SIFT to 3D

A 3D extension of the SIFT algorithm has been recently presented in the literature by a number of authors [2, 6, 16, 19]. Firstly, Scovanner *et al.* [19] used a form of 3D SIFT to assist in 3D video volume analysis followed by Cheung and Hamarneh [6] who created a 3D SIFT variant to aid in medical image alignment. Ni *et al.* [16] also extended SIFT to a 3D formulation, derived from [19], 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 [6, 16, 19]. Here, to correctly orientate an object in 3D, we consider three angles - azimuth, elevation and tilt. As shown in Figure 4a, three angles are required to correctly orientate an object. Figure 4b 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 [6, 16, 19] was previously noted by Allaire *et al.* [2] and corrected: their subsequent results indicated that the additional tilt angle improves matching as expected. Notably this error originated from the work of [6, 16, 19] as a problem of image registration as opposed to explicit object recognition: a theme also followed by [2]. 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 [6, 16, 19].

2.1 3D SIFT approach

Initially we follow the approach of Allaire *et al.* [2] in our 3D SIFT extension with additional parametric differences. Furthermore we extend this work [2] to the explicit recognition of

(a) Definitions used in our work   b) Pistols pointing in the same direction but with differing orientation (i.e. tilt)
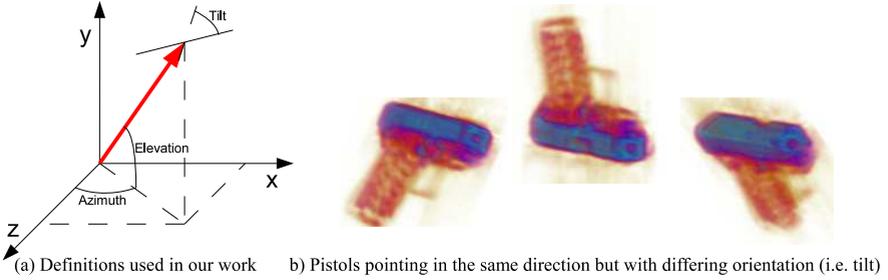
Figure 4: 3D Orientation requires three angles: Azimuth, Elevation and Tilt

objects based on RANSAC driven keypoint match selection, pose estimation and final volumetric object verification. We begin this process with initial keypoint location.

Keypoint Location

The first step in traditional 2D SIFT [13] is the calculation of Difference of Gaussian (DoG) images. Here, given a 3D input volume $I(x,y,z)$ and a 3D Gaussian filter $G(x,y,z,k\sigma)$ we form multi-scale Difference of Gaussian (DoG) volumes as follows:

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

where k is an integer in the range $\{1..5\}$ representing the scale index, $\sigma = \sqrt[3]{2}$ and (x,y,z) are defined in voxel coordinates. Subsequently a three level pyramid ($L = 0,1,2$) is built up by subsampling the Gaussian filtered volume for $k = 4$ and repeating the process.

In a similar vein to the original 2D SIFT methodology [13], DoG local extrema are then located. This requires that a voxel be either a maximum or minimum when compared to its neighbouring voxels. Given that each voxel has a $3 \times 3 \times 3$ local neighbourhood 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 neighbourhood 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, $\tau_c = 0.05$. 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. 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 curvature at the candidate point:

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

where D_{ij} are the second derivatives in the DoG volume. Both [2] and [16] derive a measure to reject points using the Trace and Determinant of H where:

$$Trace(H) = D_{xx} + D_{yy} + D_{zz} \qquad (3)$$

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

It can be shown [2, 16] that the following equation can then be used to reject points:

Figure 5: Keypoint locations for a typical complex baggage item

$$Reject\ Point\ if\ \frac{Trace^3(H)}{Det(H)} < \frac{(2\tau_e + 1)^3}{(\tau_e)^2} \tag{5}$$

We use a value of $\tau_e = 40$ and, hence, points where $\frac{Trace^3(H)}{Det(H)} < 332.15$ are rejected.

Finally a subvoxel estimate of the extrema true location is achieved using quadratic interpolation on the DoG volumetric data. Figure 5 shows some exemplar 3D SIFT keypoints (in black) after all of these stages of rejection have been performed on a typical complex CT volumetric image of a baggage item.
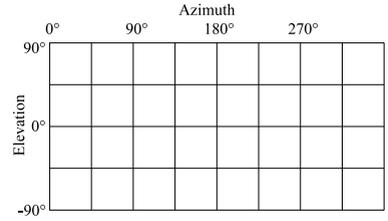
Keypoint Orientation

Once a keypoint location is determined the volume gradients are examined in a two stage process to locally establish an 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 Figure 4).

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 azimuth and elevation into 45º sections, as shown in Figure 6a (sphere) and Figure 6b (resulting 2D histogram bins). A regional weighting is applied to the gradients according to their voxel distance from the keypoint location: we apply a Gaussian weighting of $exp\left[-(2r/R_{max})^2\right]$ for voxels a distance r from the keypoint location. Points further than R_{max} voxels from the location are ignored in the current formulation. From a geodetic viewpoint (Figure 6a) 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 [19]. The output histogram is then smoothed using a Gaussian filter to limit 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 [13].

The second step is to determine the *orientation* by calculating the tilt angle for each derived direction. This is achieved by re-orienting the volume around the keypoint and calculating a 1D histogram that resolves the gradients orthogonal to the dominant direction. This histogram is again built in 45º bins using the same regional weighting method as for the direction histogram. Peaks in the tilt histogram are used, with interpolation, the 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 keypoints may have more than one possible orientation that will require description.
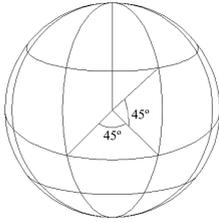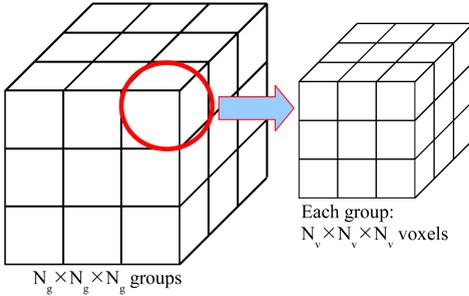
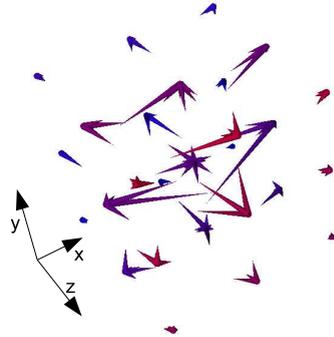(a) splitting the azimuth and elevation into 45º bins

(b) resultant 2D histogram bins

Figure 6: Direction Histogram

(a) Voxel groupings for descriptor

(b) 3×3×3 3D SIFT descriptor

Figure 7: 3D SIFT Descriptor Formulation

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 Figure 7a. Each gradient histogram is derived by splitting both azimuth and elevation into 45º bins, as described in Section 2.1. Consequently, each descriptor, normalized to unity, contains $N_g^3 N_v^3 \times 8 \times 4$ elements. The final visualization of such a descriptor is shown in Figure 7b as a 3D grid of gradient histograms.

3   Object Identification

Following from our extension of SIFT into a 3D voxel formulation we follow a traditional route of object identification following [13] where we search for a reference object in a scene and use a RANSAC based formulation to identify a given set of consistent matches.

A separate scan of the item of interest being considered was taken from which the item is then cropped to provide a reference volume. This reference volume is then subjected to the 3D SIFT generation process creating a reference descriptor set. Figure 8 shows this reference volume with the location of its keypoints at the 3 resolutions in the earlier scale space pyramid. It should be noted that this reference is also subject to the CT artefacts and resolution issues previously discussed (Section 1.1).

Here each example baggage item, when processed as described, will produce a corresponding set of candidate descriptors. The reference descriptors are compared to the candidate descriptors by recording the Euclidean descriptor distance between them [13]. Figure 9a shows a histogram of the Euclidean distances measured in a typical candidate bag. A hard
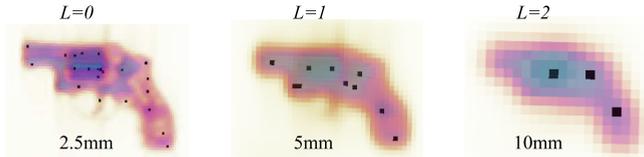
L=0   L=1   L=2

2.5mm   5mm   10mm

Figure 8: Reference item keypoints (in black) at different scale space pyramid resolutions

(a) Histogram of Euclidean distances between reference object descriptors and candidate bag descriptors

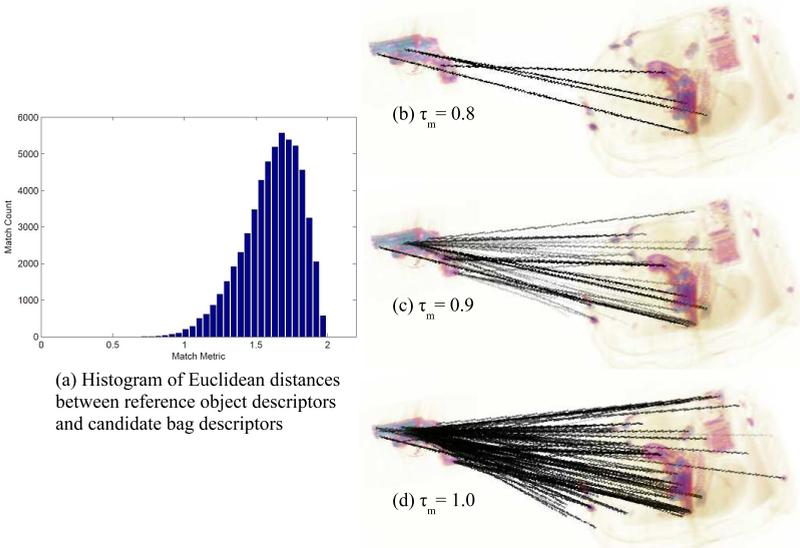(b) $\tau_m = 0.8$

(c) $\tau_m = 0.9$

(d) $\tau_m = 1.0$

Figure 9: Euclidean distance matching between reference object and candidate bag

decision is made on these distance values using a fixed threshold, τ_m, to produce an array of possible 3D SIFT matches. Figure 9b, 9c and 9d show matches from a reference object to a candidate bag as the decision threshold, τ_m, is varied and it can be seen that the number of matches (both true and false) increase as τ_m increases.

Given the large number of possible false matches in this formulation (Figure 9) we make use of RANSAC [8] to find an optimal match between the reference item descriptors and a subset of the candidate descriptors. RANSAC 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 which a 3D transformation is derived using a common place singular value decomposition [4]. An additional constraint is used to enforce consistency between the relative distances of the transformed reference set and the selected candidate match points: any relative distance errors greater than δ_r ($\delta_r = 10mm$) or location errors greater than δ_l ($\delta_l = 10mm$) will result in the transformation being rejected.

If this relative distance criterion is passed a secondary verification is performed using a comparison of CT reference to candidate object density. All locations within the reference object with density above a threshold τ_d ($\tau_d = 0.15$) are compared using L_1 distance on a voxel by voxel basis. This is recorded as the verification match metric. Combined with RANSAC this is used to identify the best candidate match within a complex volume for a given reference item.
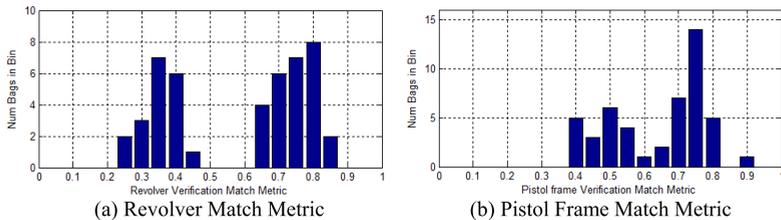
(a) Revolver Match Metric

(b) Pistol Frame Match Metric

Figure 10: Histogram of target verification match metric results

	Number of bags	Correct Identification	Incorrect Identification
Target Present	21	19	2
Clear Bag	25	25	0

(a) Revolver

	Number of bags	Correct Identification	Incorrect Identification
Target Present	27	18	9
Clear Bag	25	25	0

(b) Pistol Frame

Table 1: Object recognition results

4 Results

Results based on our approach are presented using a set of volumes first created from the original [1.6mm ×1.6mm × 5mm] data domain but subsequently resampled (using cubic spline interpolation) to form cubic voxels of uniform 2.5mm dimension. We use: $N_g = 3$ and $N_v = 3$ for descriptor generation (Section 2.1); $R_{max} = 9$ for the Gaussian weighting (Section 2.1); $\tau_m = 1.2$ for the matching decision threshold (Section 3). All data was gathered using a CT-80 model baggage scanner manufactured by Reveal Imaging Technologies.

A number of target items were used to evaluate the target recognition in a variety of cluttered baggage CT images. Firstly a revolver type handgun (.357 Magnum, Figure 2/ Figure 8) was concealed in various baggage items producing a set of 21 3D CT scan images. An additional 25 bag set of negative (target not present) scans were also generated. Over this combined set (46 CT baggage scans) the match metric (Section 3) was evaluated for each bag. In Figure 10a we see a histogram of the match metric result over this set which shows two distinct regions (i.e. peaks) from which a decision threshold on this distribution can be set to determine target identification. Using a match metric threshold τ_i ($\tau_i = 0.55$) over this distribution (Figure 10a) yields the target detection result shown in Table 1a. Here (Table 1a) we see a strong result of positive item detection and a few incorrect identifications. Overall the revolver is correctly located and identified in 90.5% of the examples (19/21) with a low false positive rate of (0.0%, 0/25). Figure 11 shows the keypoints from the revolver *reference* item superimposed into a baggage item indicating correct identification of the target item in this case.

Notably, particular items of interest may be dismantled for concealment in the common place airport baggage screening scenario [20]. Here we consider a dismantled Glock 9mm
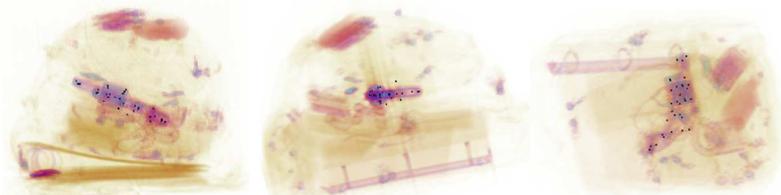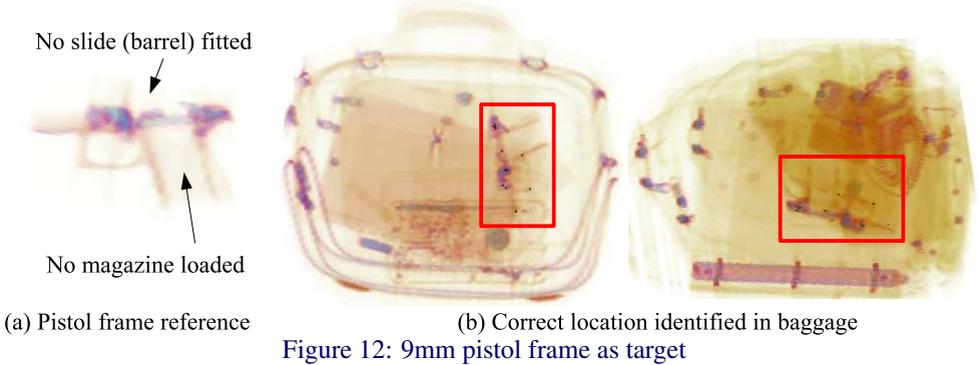

Figure 11: Correct identification of revolver

No slide (barrel) fitted

No magazine loaded

(a) Pistol frame reference

(b) Correct location identified in baggage

Figure 12: 9mm pistol frame as target

pistol with solely its frame (handle and trigger) introduced as the target item (Figure 12a). For this example a number of scans were taken (28 with target; 25 negative). Figure 10b shows a histogram of the match metric results for this target from which we can see that a decision threshold is less obvious (than in Figure 10a). Taking a threshold value $\tau_i (\tau_i = 0.6)$ yields the results presented in Table 1b where we see this more difficult target correctly located 67% of the time with a low false positive rate (0.0%). Two examples of correct identification are shown in Figure 12b.

The lesser performance in this secondary example (pistol frame, Figure 12a) can be attributed to the fact that this item is largely made from plastic with a small amount of metal where the pistol slide (barrel) would be attached. Here metal artefacts that are generated as part of the CT scanning process (Section 1.1) have a similar density to genuine parts of the pistol frame and consequently the 3D image gradients (a key part of the SIFT approach) around points of interest are more easily corrupted by noise. This makes matching in this case more complex and is clearly an area for future work. It had been envisaged that the keypoints derived from the frame of the pistol (target item Figure 12a) would enable location of a fully assembled pistol. Experimentally this has proved invalid as a complete pistol has significantly different keypoints in both location and description (Section 2.1) due to the material changes that occur on reassembly - the pistol frame lacks the internal features that would be unaffected when the rest of the pistol is attached.

Additionally the combined set of data (21 bags containing revolver; 27 bags containing pistol frame; 25 bags clear) were combined into a single data set that was processed to identify any cross related errors of individual item identification. The results of this are represented as a confusion matrix in Table 2 where we can see a clear diagonal correlation between the identification of clear bags and of the two targets (revolver/pistol frame) but we can additionally see a difficulty in the generalized identification of the pistol frame. This is shown as a precursor to future work in more generalized object recognition within complex CT baggage imagery. Within aviation screening in general [20] the identification of disassembled weaponry (such as a pistol) is considered to be a challenging task for human operators and automatic recognition alike.

Overall we can see from these examples the successful recognition of a complex 3D volumetric object over a set of complex volumetric images using a novel application of an extension of 3D SIFT to object recognition.

Identification Result

		Clear	Revolver	Pistol Frame
	Clear	25	0	0
True Status	Revolver	2	19	0
	Pistol Frame	9	0	18

Table 2: Confusion Matrix of {clear bag, revolver, pistol frame}

5   Conclusion

Our results have shown that the use of 3D SIFT to recognize known objects in complex CT volumes that contain significant metal artefacts and relatively poor resolution is possible with a relative degree of success. The detection of a revolver in complex baggage items shows a high true positive rate (90.5%) and a low false positive rate (negligible) which is a requirement for an airport baggage screening scenario. However, the relatively poor resolution coupled with its anisotropic nature leads to issues in the identification of smaller items and generalized item sub parts (Glock 9mm pistol, Figure 12, Table 1b). This is an area for future work.

In general the presence of CT artefacts is thought to be the primary cause behind false matches in the results presented - the image gradients are corrupted thus rendering the SIFT gradient histograms subject to a large degree of noise. Future work will explore the use of alternative descriptors that may offer more robustness to this inherent level of noise and additionally feature preserving volumetric noise removal techniques.

Further testing on a larger data set is required to determine the statistical uncertainty of the true positive and false positive detection results.

References

[1] B. Abidi, Y. Zheng, A. Gribok, and 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(6):784–796, 2006.

[2] S. Allaire, J. Kim, S. Breen, D. Jaffray, and V. Pekar. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008.*, pages 1–8, June 2008.

[3] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, Sept. 1987.

[4] J. Barrett and N. Keat. Artifacts in CT: Recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004.

[5] W. Bi, Z. Chen, L. Zhang, and Y. Xing. A volumetric object detection framework with dual-energy CT. In *IEEE Nuclear Science Symposium Conference Record, 2008.*, pages 1289–1291, October 2008.

[6] W. Cheung and G. Hamarneh. N-Sift: N-Dimensional Scale Invariant Feature Transform For Matching Medical Images. *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007*, pages 720–723, April 2007.

[7] R. Dalvi, I. Hacihaliloglu, and R. Abugharbieh. 3D ultrasound volume stitching using phase symmetry and Harris corner detection for orthopaedic applications. In B. M. Dawant and D. R. Haynor, editors, *Medical Imaging 2010: Image Processing*, volume 7623, page 762330. SPIE, 2010.

[8] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[9] K. Y. Jeong and J. B. Ra. Reduction of artifacts due to multiple metallic objects in computed tomography. In E. Samei and J. Hsieh, editors, *Medical Imaging 2009: Physics of Medical Imaging*, volume 7258, page 72583E. SPIE, 2009.

[10] W. Kalender, R. Hebel, and J. Ebersberger. Reduction of CT artifacts caused by metallic implants. *Radiology*, 164(2):576, 1987.

[11] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,*, volume 2, pages 2169 – 2178, 2006.

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, volume 2, pages 1150–1157, 1999.

[13] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[14] N. Menvielle, Y. Goussard, D. Orban, and G. Soulez. Reduction of Beam-Hardening Artifacts in X-Ray CT. *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005.*, pages 1865–1868, 2005.

[15] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[16] D. Ni, Y. Chui, Y. Qu, X. Yang, J. Qin, T. Wong, S. Ho, and P. Heng. Reconstruction of volumetric ultrasound panorama based on improved 3D SIFT. *Computerized Medical Imaging and Graphics*, 33(7):559–566, 2009.

[17] M. Niemeijer, M. K. Garvin, K. Lee, B. V. Ginneken, M. D. Abràmoff, and M. Sonka. Registration of 3D spectral OCT volumes using 3D SIFT feature point matching. In J. P. W. Pluim and B. M. Dawant, editors, *Medical Imaging 2009: Image Processing*, volume 7259, page 72591I. SPIE, 2009.

[18] A. Schwaninger, A. Bolfing, T. Halbherr, S. Helman, A. Belyavin, and L. Hay. The impact of image based factors and training on threat detection performance in X-ray screening. In *Proceedings of the 3rd International Conference on Research in Air Transportation, ICRAT 2008*, pages 317–324, 2008.

[19] P. Scovanner, S. Ali, and M. Shah. A 3-dimensional SIFT descriptor and its application to action recognition. In *Proceedings of the 15th international conference on Multimedia*, pages 357–360. ACM Press New York, NY, USA, 2007.

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

[21] S. Singh and M. Singh. Explosives detection systems (EDS) for aviation security.

Signal Processing, 83(1):31–55, 2003.

[22] G. Wang and M. Vannier. Stair-step artifacts in three-dimensional helical CT: an experimental study. *Radiology*, 191(1):79–83, 1994.