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

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 complex 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. Prior work on the automatic recognition of objects within this complex 3D volumetric imagery is very limited. The only prior work of Bi *et al.* [3] 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. By contrast here we consider explicit 3D recognition of items within the 3D CT volume domain.

An example of a 3D scan of an item of baggage is shown in Figure 1 where we see the presence of an item of interest amongst more general cluttered items.

(a) Front view (b) Side view (c) Top view

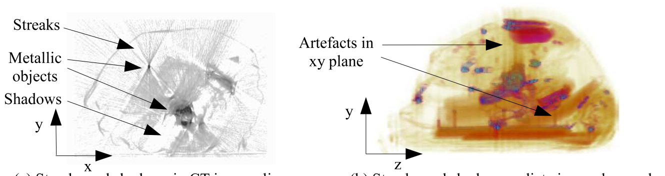
Figure 1: 3D volume of complex bag containing a revolver

The CT imagery suffers from: significant artefacts caused by the presence of metallic objects (Figure 2); 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.

The extension of the SIFT approach [5] to three dimensional data has been attempted by several researchers: we follow the approach of Allaire *et al.* [1] in our 3D SIFT extension with additional parametric differences. Furthermore we extend this work [1] to the explicit recognition of objects based on RANSAC driven keypoint match selection, pose estimation and final volumetric object verification.

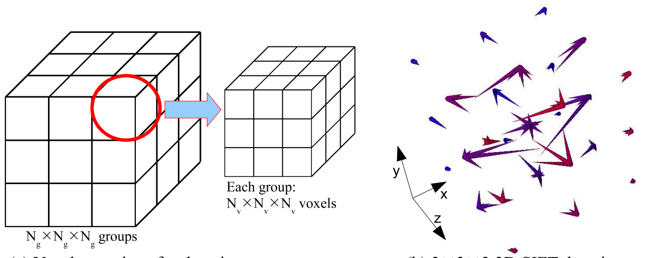
The formulation of the 3D SIFT descriptor takes the form of a $N_g \times N_g \times N_g$ grid of gradient histograms, with each histogram being computed from a $N_v \times N_v \times N_v$ voxel grouping as shown in Figure 3a. Each gradient histogram is derived by splitting both azimuth and elevation into 45º bins. 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 3b as a 3D grid of gradient histograms.

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. Each example baggage item 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 [5]. A hard decision is made on these distance values using a fixed threshold, τ_m, to produce an array of possible 3D SIFT matches. Given the large number of possible false matches in this formulation we make use of RANSAC [4] to find an optimal match between the reference item descriptors and a subset of the candidate descriptors. This RANSAC formulation is used to select a set of three possible matches from which a 3D transformation is derived using a common



(a) Streaks and shadows in CT image slice (b) Streaks and shadows radiate in xy plane only

Figure 2: An example of metal artefacts in CT baggage imagery



(a) Voxel groupings for descriptor (b) 3×3×3 3D SIFT descriptor

Figure 3: 3D SIFT Descriptor Formulation

Identification Result

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

Table 1: Confusion Matrix of {clear bag, revolver, pistol frame}

place singular value decomposition [2]. 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 and is used to identify the best candidate match within a complex volume for a given reference item.

We concentrate on the location of two items of interest: revolver (0.357 Magnum); pistol frame (Glock 9mm). A combined set of data (21 bags containing revolver; 27 bags containing pistol frame; 25 bags clear) were processed to identify any cross related errors of individual item identification. The results of this are represented as a confusion matrix in Table 1 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.

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 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. This is an area for future work.

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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