A 3D Extension to Cortex Like Mechanisms for 3D Object Class Recognition

Greg Flitton, Toby P. Breckon, Najla Megherbi
School of Engineering, Cranfield University, U.K.
{g.t.flitton, toby.breckon, n.megherbi} @cranfield.ac.uk

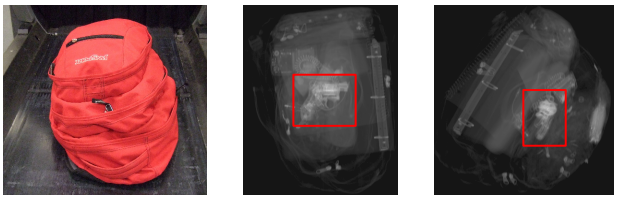
Abstract

We introduce a novel 3D extension to the hierarchical visual cortex model used for prior work in 2D object recognition. Prior work on the use of the visual cortex standard model for the explicit task of object class recognition has solely concentrated on 2D imagery. In this paper we discuss the explicit 3D extension of each layer in this visual cortex model hierarchy for use in object recognition in 3D volumetric imagery. We apply this extended methodology to the automatic detection of a class of threat items in Computed Tomography (CT) security baggage imagery. The CT imagery suffers from poor resolution and a large number of artefacts generated through the presence of metallic objects. In our examination of recognition performance we make a comparison to a codebook approach derived from a 3D SIFT descriptor and demonstrate that the visual cortex method out-performs in this imagery. Recognition rates in excess of 95% with minimal false positive rates are demonstrated in the detection of a range of threat items.

1. Introduction

Object class recognition within 2D imagery can be achieved through modelling of an object as a collection of parts that have known geometric relationships [5]. Alternatively any relationship can be ignored and a 'bag of features' paradigm can be used [3, 23]. An alternative approach is to mimic the functionality of the visual cortex. Investigations into the operation of the visual cortex has a long history [24] and recently software models have been constructed that demonstrate excellent recognition performance in standard 2D photographic imagery [15, 20]. Here we present a novel extension to this work to consider the recognition of 3D objects within complex volumetric imagery applied to an airport security screening context using CT security scan imagery.

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]. 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 2D X-ray projection [17]. 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. This potential problem of object self occlusion (Figure 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. 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 from three views. Upon creation of the CT volume the baggage item can be interrogated from any viewpoint.

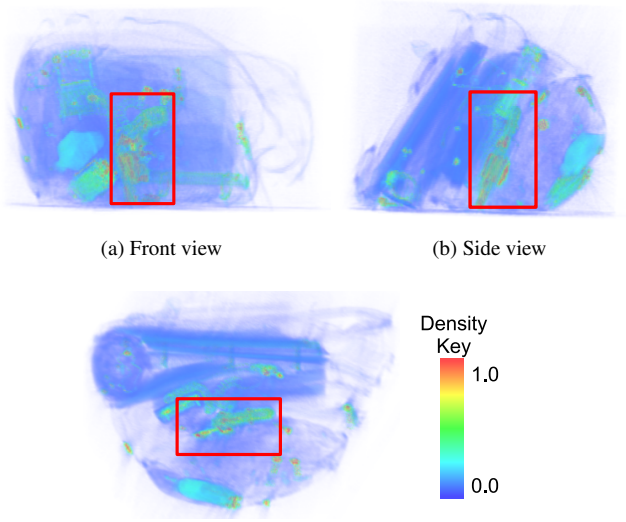


(a) Example bag (b) X-ray view 1 (c) X-ray view 2

Figure 1: Bag and X-rays

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 [22]. 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 (Figure 2) akin to those encountered within medical CT imaging [7]. It is worth noting that, when compared to medical scans, CT baggage imagery is of a inferior quality suffering from both poor resolution and significant artefacts caused by the presence of metallic objects.

Prior work on the automatic recognition of objects within



(a) Front view (b) Side view

(c) Top view

Figure 2: 3D volume of complex bag containing a revolver

this complex 3D volumetric imagery is limited [2, 6, 12]. [2] discussed the recognition of pistols but reduced the problem to an examination of the characteristic cross-section with no published results. [12] examined the recognition of bottles segmented from 3D CT imagery using volumetric shape characteristics. The work produced good results though was only applied to a small dataset. [6] used a 3D extension to the seminal SIFT descriptor [10] for specific instance recognition of a revolver and pistol frame with mixed results. No attempt at object class recognition was performed. Research into action recognition (where video can be viewed as a spatio-temporal volume) has also seen a 3D extension to the SIFT methodology [18] but extensions to the visual cortex approach have not resulted in a full volumetric implementation [9]. Here we extend the visual cortex methodology of [15, 20] to address object class recognition in complex 3D volumetric imagery.

2. Prior Visual Cortex Modelling for Object Recognition

Biological vision has been an area of research interest for many years [24] and computer modelling has recently yielded results that are of interest in the 3D recognition task we are dealing with. The visual cortex appears to be arranged in distinct sub-regions with one sub-region, the primary visual cortex (V1), being the most studied area. It was discovered that V1 is hierarchical in structure with Simple (S) and Complex (C) neurons forming the basis of the hierarchy [16]. Serre *et al*. [20] proposed a hierarchical model comprising alternating Simple and Complex layers (Figure 3).

In this model the V1 region is modelled by layers S1 and

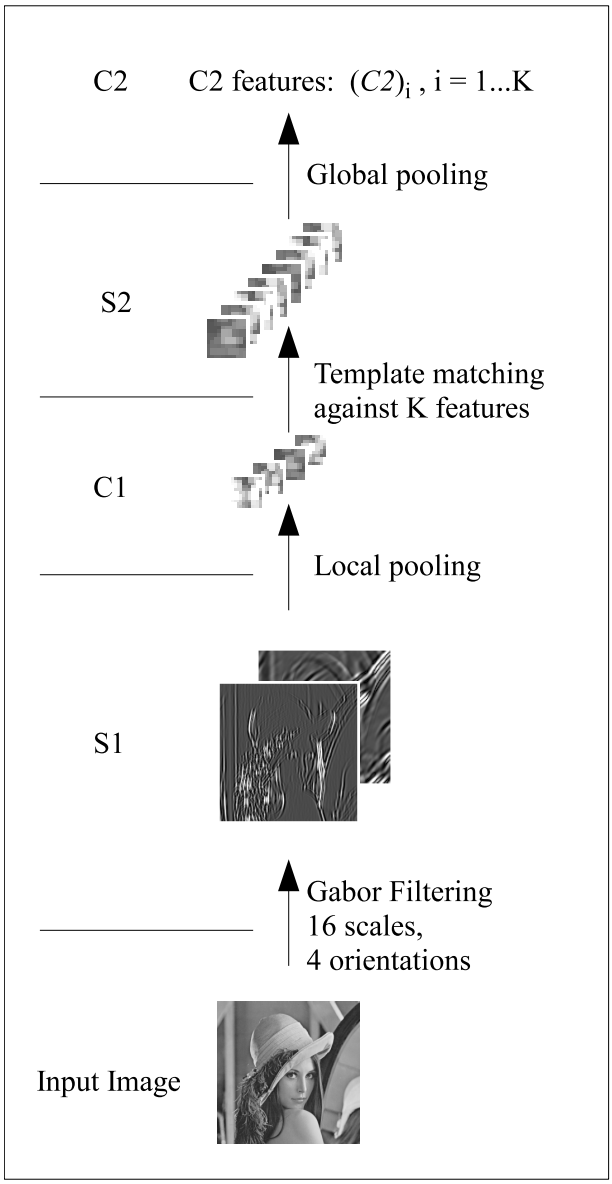


Figure 3: Visual Cortex Model of [20]

C1. The S1 layer takes images and applies a collection of Gabor filters [13] of varying size and orientation. The C1 layer is obtained from S1 though a localized pooling operation that mimics the limited receptive field aspects of the visual cortex. Above S1/C1 the higher regions in the visual cortex such as V4 and the Inferotemporal Cortex are modelled using template matching with learnt salient features (S2), a final pooling operation (C2) and subsequent classification using a SVM (Figure 3).

The work of Serre *et al*. [20] was extended by Mutch and Lowe [15] through the introduction of a scale-space pyramid structure in each layer of the hierarchy (allowing the use of fixed size Gabor and pooling filters) with the addition of some biologically plausible functionality (spar-

sity of input to S2, lateral inhibition from S1/C1, limiting position/scale invariance in C2). Finally the classification features undergo a selection process rather than being randomly selected from the training images (as in [20]) in order to choose features that are significant to the recognition task. Classification on the Caltech 101 dataset [4] was performed and results demonstrated a significant improvement over the model of [20]. Classification rates of 56% with 30 training images were achieved though misclassification rates are unclear (*e.g.* false positive rate).

As noted by [21] the visual cortex hierarchical approach is not inherently invariant to rotation(2D)/orientation(3D). Recognition that shows invariance to such transformations is achieved through training with a set that contains a wide variety of examples in all possible poses [21].

3. Extending Visual Cortex Modelling to 3D Object Recognition

Our extension follows the 2D object recognition work of [15] comprising a hierarchy of layers within which are volumetric pyramids constituting a number of levels (Figure 4). We begin with the image layer: a volumetric scale space pyramid comprising 10 levels. Each level in the pyramid is $2^{1/4}$ smaller in voxel dimension than the former and produced using bi-cubic interpolation from one level to the next. We can view the volumes as a 3D image pyramid akin to that common in 2D image processing with relation to the number of voxels in each dimension. However it is also useful to note that we can also regard each volume as being the same physical size (in *mm*) with the size of voxels increasing as the pyramid is constructed over the same spatial volume (*i.e.* resolution reduction). Interpreting the volumes in this manner is useful during the generation of the C1 layer (Section 3.2) where we must ensure identical absolute location of points between pyramid levels.

In the prior work of [15, 20] input images were rescaled in either width or height to a fixed pixel dimension prior to construction of the scale space pyramid which facilitated scale invariant object recognition. Unlike the prior work on 2D photographic imagery our CT imagery relates directly to physical object dimensions (in *mm*), and does not suffer from the perspective distortion suffered in the 2D imagery equivalent. The CT imagery has voxel dimensions in a similar range to the rescaled imagery of [15, 20], sufficient for the construction of the image layer scale space pyramid, and so we choose not rescale the input volumes that are used to form the image layer (Figure 4).

3.1. S1 Layer

The S1 layer is formed through application of 3D Gabor filters to each volume of the image layer. The derivation of these filters is achieved using a coordinate transform that results in a 3D filter orientation in the specified direction.

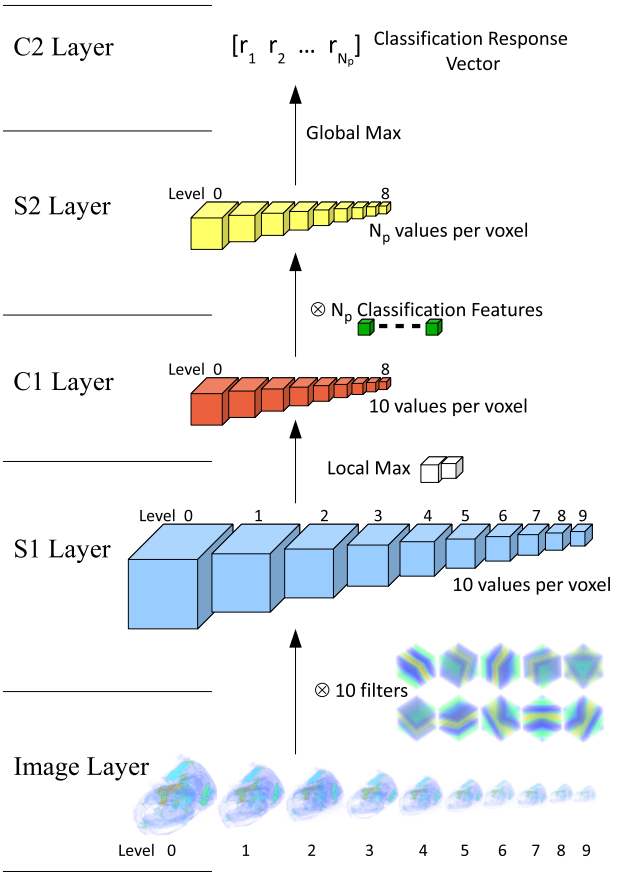


Figure 4: Hierarchical implementation in volumetric scale-space (form inspired by [15])

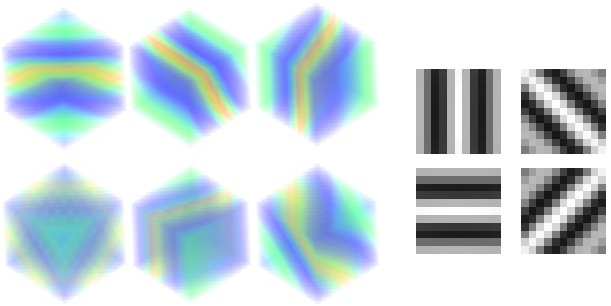
We extend the 2D Gabor definition from [15] into 3D using the directions given from the 20 vertices of a dodecahedron. The vertices are in pairs on opposite sides of the dodecahedron resulting in 10 unique directions and hence 10 Gabor filters. The vertices are defined as coordinates using the golden ratio:

$$\Phi = \frac{1+\sqrt{5}}{2}, \quad \psi = 1/\Phi \tag{1}$$

From which we thus define the 10 direction vectors for our Gabor filtering as follows:

$$\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} = \begin{bmatrix} 0 \\ \psi \\ \pm\Phi \end{bmatrix}, \begin{bmatrix} \pm\Phi \\ 0 \\ \psi \end{bmatrix}, \begin{bmatrix} \psi \\ \pm\Phi \\ 0 \end{bmatrix}, \begin{bmatrix} \pm 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \pm 1 \\ \mp 1 \end{bmatrix} \tag{2}$$

We convert each direction vector to polar coordinates by defining the azimuth, θ $\{-\pi : +\pi\}$, and elevation, ϕ $\{-\pi/2 : +\pi/2\}$, as follows:



(a) 3D (b) 2D as used in [15, 20]

Figure 5: Subset of Gabor filters used in the S1 layer

$$\theta = \arctan(y_v / x_v) \tag{3}$$

$$\phi = \arctan\left(\frac{z_v}{\sqrt{x_v^2 + y_v^2}}\right) \tag{4}$$

From these definitions we create two matrices that specify rotations around the y and z axes: $R_y(\phi)$ and $R_z(\theta)$.

We can now define a coordinate transform in 3D for a given voxel at location $\begin{bmatrix} x & y & z \end{bmatrix}^T$.

Using the rotation matrices, R_y and R_z, we form a new coordinate set:

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = R_y R_z \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{5}$$

From which the Gabor filter is defined:

$$G(\hat{x}, \hat{y}, \hat{z}) = \exp\left[-\frac{1}{2\sigma^2}\left(\hat{x}^2 + \gamma^2\hat{y}^2 + \gamma^2\hat{z}^2\right)\right]\cos\left(\frac{2\pi\hat{x}}{\lambda}\right) \tag{6}$$

where γ is the aspect ratio, σ is the effective width and λ is the wavelength. Following [15] we define the size of each Gabor filter as an $N_G \times N_G \times N_G$ voxel volume with $(N_G = 11)$ where x and y vary between -5 and $+5$. We also set $\gamma = 0.3$, $\sigma = 4.5$ and $\lambda = 5.6$ as defined by [20]. Finally each filter is adjusted to have zero mean and then normalized to give a unity sum of squares.

Figure 5 shows a subset of the 3D volumetric Gabor filters used in the construction of this layer and by contrast the 2D filters used in the prior work. In both cases we can see the varying orientations and truncated extent.

The response to a given volumetric patch of voxels, X, in each volume to a Gabor filter, G, is defined by:

$$R(X,\ G) = \left|\frac{\sum X_i G_i}{\sqrt{X_i^2}}\right| \tag{7}$$

An example of the Gabor filter responses is given in Figure 6 where we see the transformation from an image layer volume containing a pistol and several items of clutter (golf balls, belt buckle, etc.) into a collection of 10 Gabor filtered response volumes. For each pyramid level within the scale space pyramid the Gabor filtered output volumes are aggregated to form a single S1 layer vector volume - each voxel contains a vector of 10 elements representing the response to each of the 10 Gabor filters.

3.2. C1 Layer

The functionality of this layer is to provide local invariance through maximum retention in a localized region (max-pooling) as a means to mimic the functionality of the complex cells in V1 . Extending the work of [15] a max-pooling filter comprising 2 levels (scales) with $10 \times 10 \times 10$ voxels at the lowest scale is scanned through the input S1 aggregate volumes recording the peak S1 response in each Gabor orientation throughout the complete S1 volumetric scale space pyramid. Sub-sampling of the data takes place by adjusting the max-pooling filter location in steps of 5 voxels (at the bottom level). The max-pooling filter has nominally $8.4 \times 8.4 \times 8.4$ voxels in its higher scale level. When considering the pooling operation in volumetric scale space it is useful to consider the voxel positions in real world dimensions (*mm*) rather than in voxel space to ensure correct operation. The result of the C1 layer process is again a pyramid structure comprising 9 scales with smaller volume dimensions which result from the max-pooling volume sub-sampling.

3.3. S2 Layer

The S2 layer is the final filtering stage that performs template matching between the C1 layer and a set of predetermined classification features. This stage represents the beginning of a higher level of recognition within the visual cortex.

Mutch and Lowe [15] found that selection of salient patches improved recognition performance. We wish to use patches that make a strong contribution to the classification of a given volume as either a positive or negative instance of a given object class. The work of Mladenić *et al.* [14] proposed the use of linear Support Vector Machines in the identification of salient features for classification tasks. The SVM derives a hyperplane whose normal can indicate the relative contribution of candidate features to the classification task. By rejecting features with a low contribution (small hyperplane normal coefficient) we can retain only those patches that are salient to the classification task in hand. This approach was used by [15] and we choose to follow that method. We first randomly choose N_r $(N_r = 12,000)$ patches from the C1 layers of the training set of volumes. We split this selection into four sets of 3000

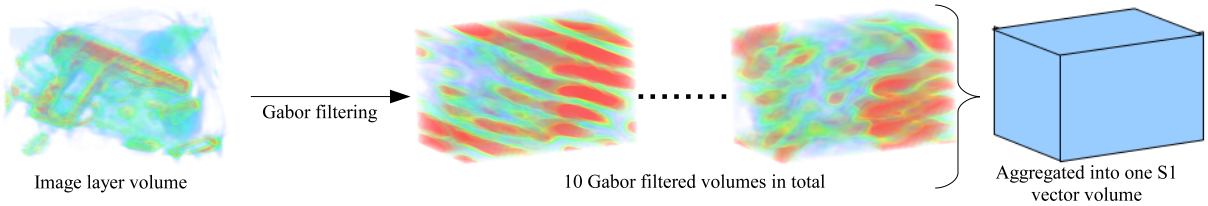


Figure 6: Gabor filtering applied to volume from image layer resulting in S1 vector volume

patches and use the SVM selection method to leave four sets of 1500. These are combined to form 2 sets of 3000 and the process repeated. This continues until we are left with N_p ($N_p = 1500$) patches that are used for the classification process.

Following [15] the filtering response, $R(X,\ P)$, of a patch of C1 units, X, to a predetermined classification feature, P, is given by a radial basis function:

$$R(X,\ P) = \exp\left(-\frac{\|X - P\|^2}{2\sigma^2\alpha}\right) \tag{8}$$

In our work both patches X and P are $n \times n \times n$ voxels in size with each voxel containing a vector of 10 values derived from the 10 Gabor filters used in the creation of the S1 Pyramid Layer (Section 3.1). We follow both [15] and [21] in setting $\sigma = 1.0$. The setting of α is used by [15] to provide a normalisation term for patches of differing size.

For 3D we modify this term to reflect the increased dimension, assuming a lower setting of $n = 4$:

$$\alpha = \left(\frac{n}{4}\right)^3 \tag{9}$$

The response for each salient patch (Equation 8) is calculated at every location in the C1 scale space pyramid with the result that the S2 output is another scale space volumetric pyramid but in this case each voxel contains a vector of N_p response values. This is illustrated in Figure 4 where we see the evaluation of the response between the C1 layer and salient patches resulting in the S2 layer as output. It is worth noting that this process will reduce the size of each volume in the S2 layer when compared to the C1 input due to the size of the salient patch being used.

3.4. C2 Layer

This layer forms the 'bag of features' style vector for presentation to a Support Vector Machine [3] and is achieved through a final pooling stage. We establish a classification feature response vector by taking the largest patch response for each feature in the S2 layer of the baggage item being analyzed. For example, the first element in the feature response vector is obtained by examining the first element in each voxel of the S2 scale space pyramid and retaining the largest value. This is repeated for each element so that, given N_p salient feature patches, we now have a vector of N_p values that describes the volumetric imagery in terms that can be used by a machine learning algorithm for training or classification.

4. 3D Bag of Features Comparison

As an evaluation method we also seek to compare the visual cortex methodology against the conventional bag of features codebook approach [3, 23]. The SIFT descriptor [10] has proven to be adept for this purpose in 2D datasets and so we use the 3D SIFT implementation of [6] to derive a set of descriptors for each baggage item. Interest point locations are obtained using a 3D extension to the difference of Gaussians method [10]. An invariant coordinate system is established at each point (to counter arbitrary orientation) prior to characterization that results in an 864 element descriptor. We use k-means clustering [11] on these descriptors to establish the visual codewords and then use the uncertainty assignment methodology outlined in [25] to establish the codebook for each volumetric image. The uncertainty assignment method of [25] has been shown to improve classification performance over more conventional hard assignment though it does require empirical tuning of an assignment parameter to achieve optimal performance.

5. Experimental Results

The type of baggage scanner machine used to capture the CT volumetric imagery for this work is primarily aimed at dual energy explosives detection [22] and as a result two additional consequences are generally suffered within the imagery- (1) the presence of metal items causes significant artefacts within the imaging and (2) the resolution is anisotropic and limited to [1.6*mm* $\times$1.6*mm* $\times$ 5*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. We choose to resample the anisotropic volumes to create cubic voxels of uniform 2.5*mm* dimension using cubic spline interpolation. The volumetric data is rescaled to the continuous range $\{0.0 \Rightarrow 1.0\}$ from the original integer CT scanner output (Figure 2, key as shown).

We use two classes of object in our experiment: handguns and bottles. The handgun and bottle datasets have 284/971 and 534/1170 for threat/clear scan volumes respectively. Note that both object classes are varied in shape and size with the bottles containing a varied amount of liquid. In

order to achieve orientation invariance for the final classification stage the objects are scanned within baggage items in random poses. In each case the objects are extracted from volumetric imagery with a 30*mm* margin. Formation of the negative dataset in each case is made through dicing of baggage items that do not contain a threat item to produce subvolumes that have a similar size to the threat subvolumes. Figure 7 shows some examples of the data used for this experiment.

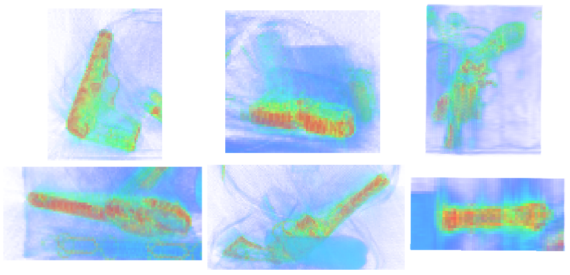
Each volume in the test and training sets is processed to obtain the C1 layer. We then process the training set to obtain the classification features by randomly selecting 12000 volumetric patches $(4 \times 4 \times 4)$ and retaining 1500 that are most salient. The salient patches are then used on the training and test sets to obtain the S2 and C2 layers prior to application to a Support Vector Machine (Section 3).

We use a ten-fold cross validation approach in the evaluation process. A Support Vector Machine using a RBF kernel is used - its settings are obtained through a grid search and a ten-fold cross validation on the C2 layer of the training set [8]. The SVM parameters that achieve the lowest misclassifications are used to retrain the SVM on the complete training set.

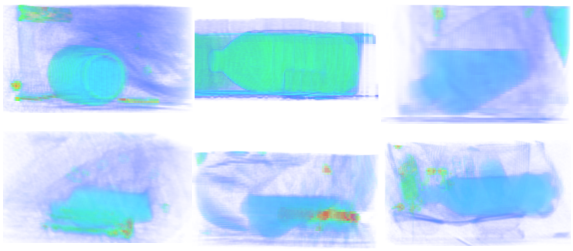
Generation of the visual codebooks using the 3D SIFT descriptor methodology (Section 4) is made using the same data sets. We vary the number of clusters and uncertainty assignment parameters to achieve a setting that minimizes the number of classification errors. A ten-fold cross validation machine learning approach is used on the training set codebooks in the same manner as for the visual cortex approach.

Table 1 shows the classification results for each class where we can see the performance for the visual cortex method with a clear out-performance over the 3D SIFT codebook approach. The visual cortex method produces similar results for both handguns and bottles with a high true positive rate (above 96.0%) and low false positive rate ($\simeq$ 1.0%). The precision and recall are both good (above 0.96 for both) illustrating consistency over multiple objects.

The 3D SIFT codebook approach lags behind with a distinct difference between the handgun and bottle results. For handguns we have a true positive rate of 87.0% and for bottles 82.8% with false positive rates $\simeq$4.0%. The number of codewords used to optimized the recognition rate was 1024 for handguns and 2048 for bottles. The lesser performance of the 3D SIFT codebook method is primarily due to variations in achieving an invariant orientation in the 3D SIFT descriptor caused by metallic artefact disruption in the baggage imagery - this results in a more noisy codebook which hinders the classification process. It has also been noted that for 2D recognition at the descriptor level the visual cortex methodology outperforms the SIFT descriptor [19] which further supports the results obtained within this work.



(a) Handgun data



(b) Bottle data

Figure 7: Example Data

Figure 8 shows some examples of correctly classified handguns and bottles using the visual cortex methodology where we can see the cluttered nature of the dataset.

Figures 9, 10 show a range of misclassifications for the visual cortex methodology. For the majority of these cases a visual examination shows no clear systematic reason for the misclassifications. Notably in Figure 10 a possible attribution may be confusion between a book and a bottle - the book has a similar size and density to a bottle with flat surfaces resembling some bottles.

By way of contrast, Figures 11, 12 show a range of misclassifications for the bag of features codebook methodology that were correctly classified by the visual cortex approach. Again there are is no clear reason for the misclassifications.

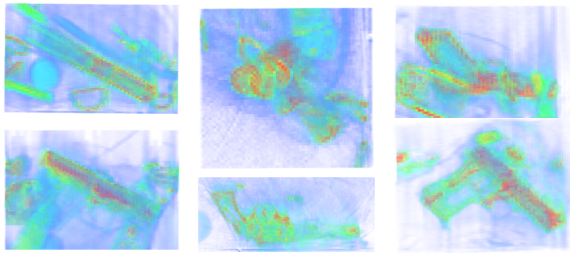
6. Conclusions

In this work we have demonstrated that a 3D extension to the visual cortex models can achieve strong classification results with low false positives for complex 3D volumetric baggage imagery. We have also shown that this method outperforms a codebook methodology that uses a 3D SIFT descriptor as its basis.

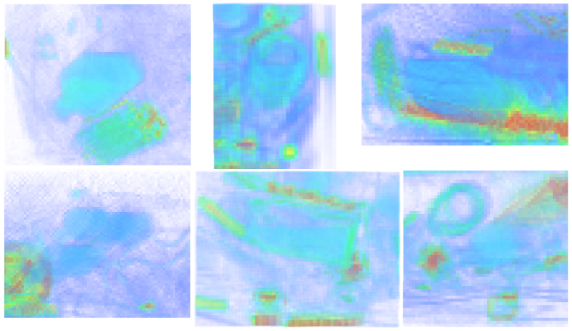
Future work will investigate the use of a volumetric sliding window approach for object localization with large baggage items and additionally consider further objects (*e.g.* knives, electronic circuitry). We will also investigate the application of this volumetric hierarchy to action recognition within spatio-temporal volumes [9, 18].

Method	Class	True Positive Rate (%)	False Positive rate (%)	Precision	Recall
Cortex	Handgun	96.8 ± 2.6	1.1 ± 0.9	0.962 ± 0.029	0.968 ± 0.026
	Bottle	96.6 ± 3.2	1.0 ± 1.6	0.977 ± 0.034	0.966 ± 0.032
Codebook	Handgun	87.0 ± 5.4	3.8 ± 2.4	0.870 ± 0.069	0.870 ± 0.054
	Bottle	82.8 ± 7.0	4.2 ± 1.2	0.900 ± 0.025	0.828 ± 0.070

Table 1: Overall recognition performance for visual cortex and codebook methodologies

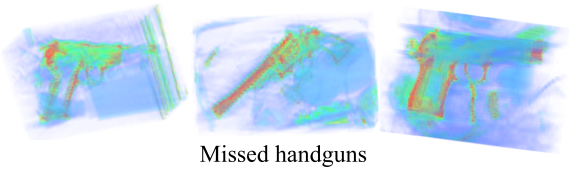


(a) Handguns

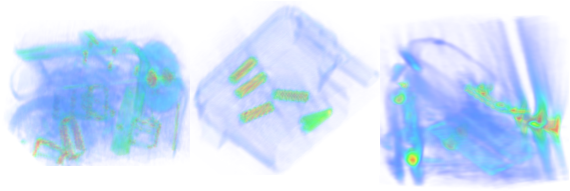


(b) Bottles

Figure 8: Visual Cortex: Examples of correct recognition



Missed handguns

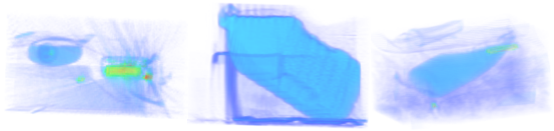


Clutter classified as handgun

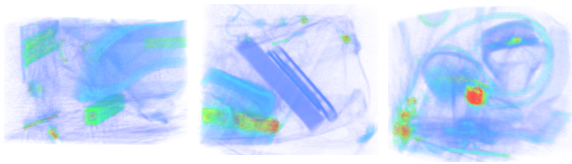
Figure 9: Visual cortex: Example classification errors for handgun data

Acknowledgements

This project is funded under the Innovative Research Call in Explosives and Weapons Detection (2007), a cross-government programme sponsored by a number of govern-

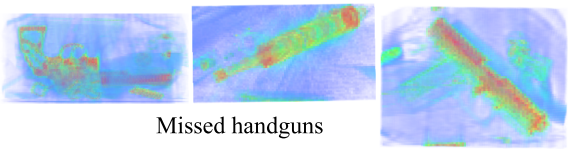


Missed bottles

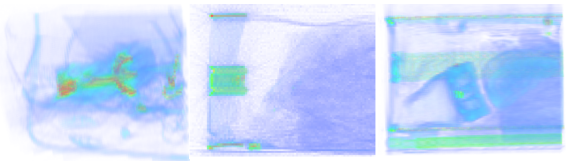


Clutter classified as bottle

Figure 10: Visual cortex: Example classification errors for bottle data

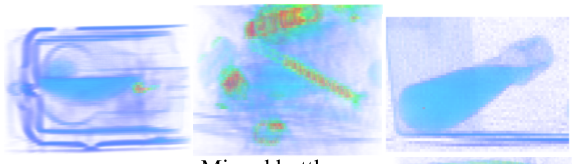


Missed handguns

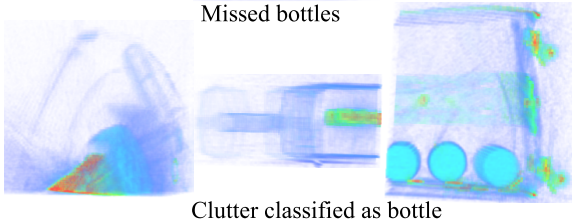


Clutter classified as handgun

Figure 11: Codebook: Example classification errors for handgun data that visual cortex correctly classified



Missed bottles



Clutter classified as bottle

Figure 12: Codebook: Example classification errors for bottle data that visual cortex correctly classified

ment departments and agencies under the CONTEST strategy. We also wish to thank Reveal Imaging Technologies for their assistance.

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. 1

[2] 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*, pages 1289–1291, October 2008. 2

[3] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*, pages 1–22, 2004. 1, 5

[4] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007. 3

[5] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, September 2010. 1

[6] G. Flitton, T. Breckon, and N. Megherbi. Object Recognition using 3D SIFT in Complex CT Volumes. In F. Labrosse, R. Zwiggelaar, Y. Liu, and B. Tiddeman, editors, *Proceedings of the British Machine Vision Conference*, pages 11.1–11.12. BMVA Press, 2010. 2, 5

[7] G. Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Springer Verlag, 2nd edition, 2009. 1

[8] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, 2010. 6

[9] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007. 2, 6

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

[11] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967. 5

[12] N. Megherbi, G. Flitton, and T. Breckon. A Classifier based Approach for the Detection of Potential Threats in CT based Baggage Screening. In *Proc. International Conference on Image Processing*, pages 1833–1836. IEEE, September 2010. 2

[13] R. Mehrotra, K. Namuduri, and N. Ranganathan. Gabor filter-based edge detection. *Pattern Recognition*, 25(12):1479–1494, 1992. 2

[14] D. Mladenić, J. Brank, M. Grobelnik, and N. Milic-Frayling. Feature selection using linear classifier weights: interaction with classification models. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 234–241. ACM, 2004. 4

[15] J. Mutch and D. Lowe. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80(1):45–57, 2008. 1, 2, 3, 4, 5

[16] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025, 1999. 2

[17] 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. 1

[18] 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. 2, 6

[19] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 411–426, 2007. 6

[20] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 994–1000, 2005. 1, 2, 3, 4

[21] T. Serre, L. Wolf, and T. Poggio. A new biologically motivated framework for robust object recognition. Technical report, Massachusetts Institute of Technology, Cambridge, MA., July 2004. CBCL Paper 239 / AI Memo 2004-017. 3, 5

[22] S. Singh and M. Singh. Explosives detection systems (EDS) for aviation security. *Signal Processing*, 83(1):31–55, 2003. 1, 5

[23] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1470–1477, Oct. 2003. 1, 5

[24] N. V. Swindale. The development of topography in the visual cortex: a review of models. *Network: Computation in Neural Systems*, 7(2):161–247, 1996. 1, 2

[25] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010. 5