DOI: 10.1784/insi.2017.59.2.85

Object recognition in X-ray testing using an efficient search algorithm in multiple views

D Mery, V Riffo, I Zuccar and C Pieringer

In order to reduce the security risk of commercial aircraft, passengers are not allowed to take certain items in their carry-on baggage. For this reason, human operators are trained to detect prohibited items using a manually-controlled baggage screening process. In this paper, the use of an automated method based on multiple X-ray views is proposed to recognise certain regular objects with highly-defined shapes and sizes. The method consists of two steps: 'monocular analysis', to obtain possible detections in each view of a sequence, and 'multiple view analysis', to recognise the objects of interest using matching in all views. The search for matching candidates is efficiently performed using a look-up table that is computed offline. In order to illustrate the effectiveness of the proposed method, experimental results on recognising regular objects (clips, springs and razor blades) in pencil cases are shown achieving high precision and recall (P_r = 95.7% , R_e = 92.5%) for 120 objects. We believe that it would be possible to design an automated aid in a target detection task using the proposed algorithm.

Keywords: X-ray testing, computer vision, baggage inspection, image analysis.

1. Introduction

Baggage inspection using X-ray screening is a priority task that reduces the risk of crime, terrorist attacks and the propagation of pests and diseases[42]. Security and safety screening with X-ray scanners has become an important process in public spaces and at border checkpoints[30]. However, inspection is a complex task because threat items are very difficult to detect when placed in closely-packed bags, occluded by other objects or rotated, thus presenting an unrecognisable view[5]. Manual detection of threat items by human inspectors is extremely demanding[34]. It is tedious because very few bags actually contain threat items and it is stressful because the work of identifying a wide range of objects, shapes and substances (metals, organic and inorganic substances) takes a great deal of concentration. In addition, human inspectors receive only minimal technological support. Furthermore, during rush hours, they have only a few seconds to decide whether or not a bag contains a threat item[4]. Since each operator must screen many bags, the likelihood of human error becomes considerable over a long period of time, even with intensive training. The literature suggests that detection performance is only about 80-90%[22]. In baggage inspection, automated X-ray testing remains an open question due to: (i) loss of generality, which means that approaches developed for one task may not transfer well to another; (ii) deficient detection accuracy, which means that there is a fundamental trade-off between false alarms and missed detections; (iii) limited robustness given that requirements for the use of a method are often met for simple structures only; and (iv) low adaptiveness in that it may be very difficult to accommodate an automated system to design modifications or different specimens.

Before 9/11, the X-ray analysis of luggage focused mainly on capturing the images of their contents. The reader can find in[26] an interesting analysis carried out in 1989 of several aircraft attacks around the world and the existing technologies to detect terrorist threats based on thermal-neutron activation (TNA), fast-neutron activation (FNA) and dual-energy X-rays (used in medicine since the early 1970s). In the 1990s, explosive detection systems (EDS) were developed based on X-ray imaging[27] and computed tomography through elastic scatter X-ray (comparing the structure of irradiated material against stored reference spectra for explosives and drugs)[36]. All of these works were concentrated on image acquisition and simple image processing; however, they lacked advanced image analysis to improve detection performance. Nevertheless, the 9/11 attacks increased the security measures taken at airports, which in turn stimulated the interest of the scientific community in the research of areas related to security using advanced computational techniques. Over the last decade, the main contributions were: analysis of human inspection[41], pseudo-colouring of X-ray images[1,6], enhancement and segmentation of X-ray images[35] and the detection of threatening items in X-ray images, based on texture features (detecting a 9 mm Colt Beretta automatic (machine) pistol)[29], neural networks and fuzzy rules (yielding about 80% of performance)[12] and an SVM classifier (detecting guns in real time)[28].

Three-dimensional (3D) recognition from two-dimensional (2D) images is a very complex task due to the infinite number of points of view and different image acquisition conditions[31]. An example is illustrated in Figure 1, where a handgun is very difficult to recognise in the first view. Nevertheless, automated recognition has been possible, in certain cases, through seminal works dedicated to obtaining highly discriminative and local invariant features related to illumination factors and local geometric constraints (see for example[23] for a good review and evaluation). In such cases,

Submitted 25.07.16 / Accepted 16.12.16

Domingo Mery is with the Department of Computer Science (DCC), Pontificia Universidad Católica, Vicuña Mackenna 4860, Santiago, Chile. Email: dmery@ing.puc.cl; URL: http://dmery.ing.puc.cl

Vladimir Riffo is with DIICC, Universidad de Atacama, Copiapó, Chile. Email: vladimir.riffo.b@gmail.com

Irene Zuccar is with DIINF, Universidad de Santiago de Chile, Santiago, Chile.

Christian Pieringer is with DCC, Pontificia Universidad Católica de Chile, Santiago, Chile.

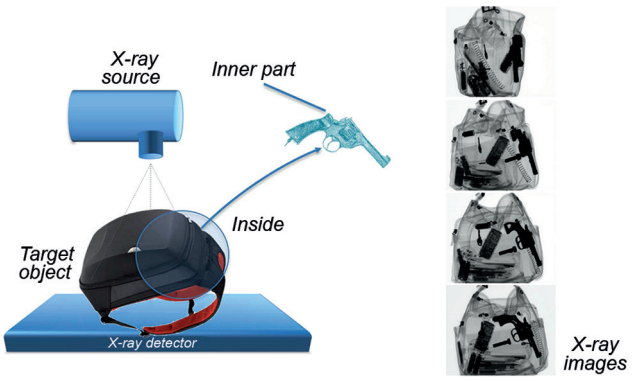

Figure 1. X-ray images of a backpack containing a handgun from different points of view

recognition of a test object can be performed by matching its invariant features with the features of a model.

There are some contributions in computer vision for X-ray testing, such as applications in the inspection of castings, welds, food, cargo and baggage screening[17]. For this work, it is very interesting to review the advances in baggage screening that have taken place over the course of this decade. They can be summarised as follows. Some approaches attempt to recognise objects using a single view of mono-energy X-ray images (for example the adapted implicit shape model based on visual codebooks[33]) and dual-energy X-ray images (for example Gabor texture features[38], bag of words based on SURF features[38] and pseudo-colour, texture, edge and shape features[47]). More complex approaches that deal with multiple X-ray images have been developed as well. In the case of mono-energy imaging, see, for example, the recognition of regular objects using data association in[16,18,19] and active vision[32], where a second-best view is estimated. Other methods perform a synthesis of new X-ray images obtained from kinetic depth effect X-ray (KDEX) images based on SIFT features in order to increase detection performance[2]. In the case of dual-energy imaging (with multiple views), see the use of visual vocabularies and SVM classifiers in[11]. Progress has also been made in the area of computed tomography. For example, in order to improve the quality of CT images, metal artefact reduction and denoising[24] techniques were suggested. Many methods based on 3D features for 3D object recognition have been developed (see for example RIFT and SIFT descriptors[9], 3D visual cortex modelling, 3D Zernike descriptors and the histogram of shape index[14]). There are contributions using known recognition techniques (see for example bag of words[10] and random forest[25]). It can be seen that the progress in automated baggage inspection is modest and still very limited compared to what is needed because X-ray screening systems are still being manipulated by human inspectors.

In baggage screening, the use of multiple view information yields a significant improvement in performance as certain items are difficult to recognise using only one viewpoint. As reported in a study that measures the human performance in baggage screening[40], (human) multiple view X-ray inspection leads to a higher detection performance of prohibited items under difficult conditions; however, there are no significant differences between the detection performance (single *versus* multiple view) for difficult-easy multiple view conditions, *ie* two difficult or two easy views are redundant. We observed that, for intricate conditions, multiple view X-ray inspection is required.

Even though several scientific communities are exploring a range of research directions, adopting very different principles and developing a wide variety of algorithms for very different applications, automated recognition in baggage inspection is far from being perfected, due to: (i) the large variability of the appearance and shape of the test objects, both between and within categories (see Figure 2); (ii) the large variability in terms of object sample depending on its points of view (for example the top view and frontal view of a razor blade are very different, as shown in Figure 3); and (iii) the appearance of a test object can vary due to the conditions of (self-) occlusion, noise and acquisition (see Figure 4).

Figure 2. Large variability in the appearance of guns and knives

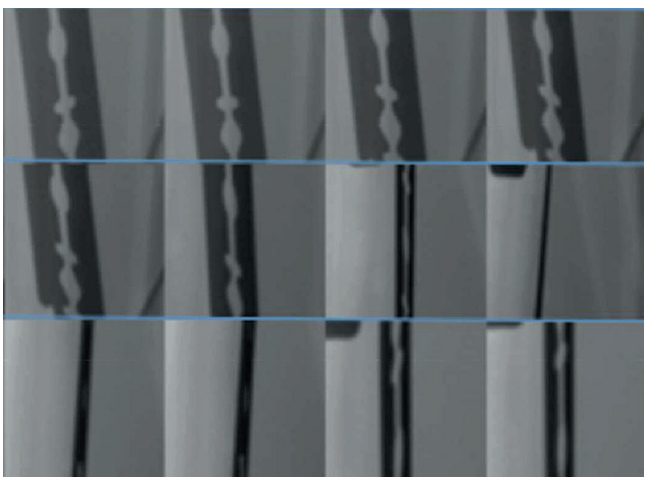

Figure 3. Large variability within a razor blade: some X-ray images of the same blade in different poses

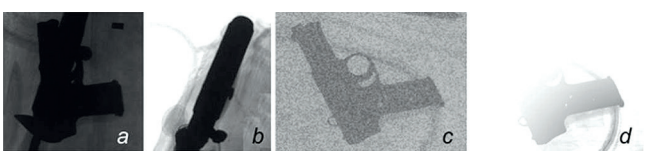

Figure 4. Typical problems in recognition tasks: (a) occlusion; (b) self-occlusion; (c) noise and (d) low contrast

In this paper, a contribution is made to the last two mentioned problems, in which object recognition plays a crucial role. The proposal is based on three potent ideas: (i) detection windows, as they obtain a high performance in recognition and detection problems in computer vision; (ii) multiple views, as they can be an effective option for examining complex objects where uncertainty by analysing only one angle of perspective can lead to misinterpretation; and (iii) efficient visual search, given the speeds involved when searching for objects. We believe that our framework is a useful alternative for recognising objects because it is based on an efficient search in multiple views using corresponding multiple view windows.

In this paper, a framework based on computer vision and machine learning techniques is proposed in order to deal with the problem of 3D recognition. We believe that this solution also allows a general and adaptive methodology for X-ray testing to be proposed that can be tested in several detection problems, such as

the characterisation of materials and airport security. Additionally, we think that it would be possible to design an automated aid in a target detection task using the proposed algorithm.

The rest of the paper is organised as follows: the proposed approach (Section 2), the results obtained in several experiments (Section 3) and some concluding remarks and suggestions for future research (Section 4). A preliminary version of this paper was published in[21].

2. Proposed method

The strategy of the method is illustrated in Figure 5. The target object, for example a backpack, is irradiated from different viewpoints (image acquisition). The key idea of the method is to detect, in each image of the sequence, individual potential objects (monocular detection) and corroborate these detections across the multiple views. Thus, a matching in 2D and a 3D analysis is performed using geometric constraints in order to eliminate false monocular detections. This means that a monocular detection that does not find any correspondence will be filtered out in the next steps. Finally, an appearance analysis is carried out (final detection).

In this section, the proposed method is explained in further detail. The explanation consists of two main stages: offline and online. In the first stage the appearance model and the geometric model are established, whereas in the second stage the above-mentioned strategy is followed.

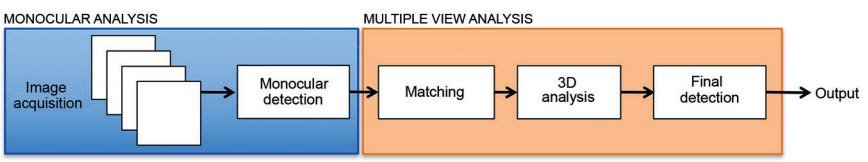

Figure 5. Strategy of the proposed method

2.1 Offline stage

The first stage, performed offline, consists of two main steps: (i) training a model that is used for the recognition; and (ii) estimation of a multiple view geometric model that is used for data association.

2.1.1 Learning

In this step, a classifier h is trained to recognise parts of the objects that we are attempting to detect. It is assumed that there are $C + 1$ classes (labelled as '0' for a non-object class and '1', '2', …, 'C' for C different objects). Images are taken of representative objects of each class from different points of view. In order to model the details of the objects from different poses, several keypoints per image are detected and for each keypoint a descriptor $\mathbf{d}$ is extracted using, for example, LBP, SIFT, HOG and SURF, among others[23]. In this supervised approach, each descriptor $\mathbf{d}$ is manually labelled according to its corresponding class $c \in \{0, 1, \ldots, C\}$. Given the training data $(\mathbf{d}_t, c_t)$, for $t = 1, \ldots, N$, where N is the total number of descriptors extracted in all training images, a classifier h is designed, which maps $\mathbf{d}_t$ to the classification label c_t, thus $h(\mathbf{d}_t)$ should be c_t. This classifier will be used in the online stage by monocular and multiple view analysis.

2.1.2 Geometry

The strategy deals with multiple monocular detections in multiple views. In this problem of data association, the aim is to find the correct correspondence among different views. For this reason, multiple view geometric constraints are used to reduce the number of matching candidates between monocular detections. For an image sequence with n views $\mathbf{I}_1 \ldots \mathbf{I}_n$, the fundamental matrices $\{\mathbf{F}_{ij}\}$ between consecutive frames $\mathbf{I}_i$ and $\mathbf{I}_{j=i+1}$ are computed for $i = 1, \ldots, n-1$. In this approach, the fundamental matrix $\mathbf{F}_{ij}$ of the epipolar geometry (see Figure 6) is calculated from projection matrices $\mathbf{P}_i$ and $\mathbf{P}_j$ that can be estimated using calibration or bundle adjustment algorithms[17].

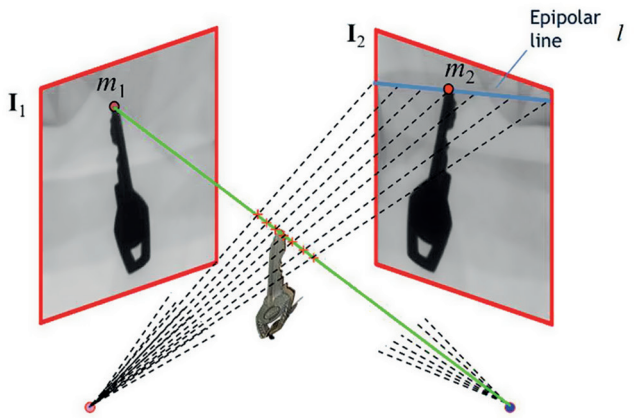

Figure 6. Epipolar geometry: the corresponding point of m_1 in the left image is m_2 in the right image. They are the projections of the same 3D point of the key. The epipolar constraint says that m_2 must lie on the epipolar line of m_1, that is line l

The geometric constraints are expressed in homogeneous coordinates. Therefore, given a point $\mathbf{m}_i = [x_i\ y_i\ 1]^T$ in image $\mathbf{I}_i$, a corresponding point $\mathbf{m}_j = [x_j\ y_j\ 1]^T$ in image $\mathbf{I}_j$ must fulfill: (i) the epipolar constraint: $\mathbf{m}_j$ must lie near the epipolar line $l = \mathbf{F}_{ij}\mathbf{m}_i$; and (ii) the location constraint: for small variations of the points of view between $\mathbf{I}_i$ and $\mathbf{I}_j$, $\mathbf{m}_j$ must lie near $\mathbf{m}_i$. Thus, a candidate $\mathbf{m}_j$ must fulfill:

$$\frac{\left|\mathbf{m}_j^T \mathbf{F}_{ij} \mathbf{m}_i\right|}{\sqrt{l_1^2 + l_2^2}} < e \text{ and } \left\|\mathbf{m}_i - \mathbf{m}_j\right\| < r \quad \ldots\ldots\ldots\ldots (1)$$

In order to accelerate the search of candidates, the use of a look-up table is proposed as follows. Points in images $\mathbf{I}_i$ and $\mathbf{I}_j$ are arranged in a grid format with rows and columns. For each grid point (x, y) of image $\mathbf{I}_i$, we look for the grid points of image $\mathbf{I}_j$ that fulfill the Equation (1), as illustrated in Figure 7. Therefore, the possible

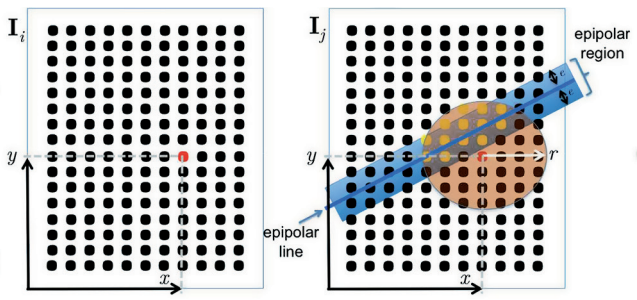

Figure 7. Given the grid point illustrated as the red point at (x, y) in image $\mathbf{I}_i$, the set of possible corresponding points in image $\mathbf{I}_j$ can be those grid points (yellow points) represented by the intersection of the epipolar region (blue rectangle) and neighbourhood around (x, y) (orange circle with radius r centred at the red point). The use of grid points allows us to use a look-up table in order to search the matching candidates in $\mathbf{I}_j$ efficiently

corresponding points of (x, y) will be the set $\mathbf{S}_{xy} = \left\{\left(x_p, y_p\right)\right\}_{p=1}^{q}$, where $x_p = X(x, y, p)$, $y_p = Y(x, y, p)$ and $q = Q(x, y)$ are stored (offline) in a look-up table. In the online stage, given a point $\mathbf{m}_i$ (in image $\mathbf{I}_i$), the matching candidates in image $\mathbf{I}_j$ are those that lie near to $\mathbf{S}_{xy}$, where (x, y) is the nearest grid point to $\mathbf{m}_i$. This search can be efficiently implemented using k-d tree structures[3].

In a controlled and calibrated environment, it can be assumed that the fundamental matrices are stable and it is not necessary to estimate them in each new image sequence, *ie* the look-up tables are constant. Additionally, when the relative motion of the point of view between consecutive frames is the same, the computed fundamental matrices are constant, *ie* $\mathbf{F}_{ij} = \mathbf{F}$, and only one look-up table needs to be stored.

2.2 Online stage

The online stage is performed in order to recognise the objects of interest in a test image sequence of n images $\{\mathbf{I}_i\}$, for $i = 1, \ldots, n$. The images are acquired by rotation of the object being tested at β degrees (in the experiments, $n = 4$ and $\beta = 10^o$ were used). This stage consisted of two main steps: monocular and multiple view analysis, which will be described in further detail as follows.

2.2.1 Monocular analysis

This step is performed in each image $\mathbf{I}_i$ of the test image sequence, as illustrated in Figure 8 in a real case. The whole object contained in image $\mathbf{I}_i$ is segmented from the background using threshold and morphological operations. SIFT keypoints, or other descriptors, are only extracted in the segmented portion. The descriptor $\mathbf{d}$ of each keypoint is classified using classifier $h(\mathbf{d})$, trained in the offline stage and explained in Section 2.1.1. All keypoints classified as class c, where c is the class of interest, with $c \in \{1 \ldots C\}$ are selected. As can be seen in Figure 8 for the classification of a razor blade, there are many keypoints misclassified. For this reason, neighbouring keypoints are clustered in the 2D space using a mean shift algorithm[7]. Only those clusters that have a large enough number of keypoints are selected. They will be called detected monocular keypoints.

2.2.2 Multiple view analysis

Multiple view analysis performs the recognition of objects of interest in three steps (see Figure 9): (i) data association; (ii) 3D analysis; and (iii) final analysis. The input is the detected monocular keypoints obtained by the mentioned monocular analysis of Section 2.2.1. The output is c', the assigned class for each detected object.

Data association

In this step, matchings are found for all detected monocular keypoints in all consecutive images $\mathbf{I}_i$ and $\mathbf{I}_{j=i+1}$, for $i = 1, \ldots, n-1$, as follows:

- For each detected monocular keypoint in image $\mathbf{I}_i$ (located at position (x_i, y_i) with descriptor $\mathbf{d}_i$), the nearest point (x, y) (see the red point in Figure 7 (left)) is sought in a dense grid of points using a k-d tree structure.
- $\mathbf{S}_{xy}$, the set of matching candidates in image $\mathbf{I}_{j=i+1}$ arranged in a grid manner, is determined by reading the look-up table explained in Section 2.1.2 (see the yellow points in Figure 7 (right)).
- The detected monocular keypoints in image $\mathbf{I}_j$ that are located in the neighborhood of $\mathbf{S}_{xy}$ are sought, again using a k-d tree structure. They will be called neighbour keypoints. When no neighbour keypoint is found, no match is established for (x_i, y_i).
- From the neighbour keypoints, the one (located at position (x_j, y_j) with descriptor $\mathbf{d}_j$) with minimum distance $\|\mathbf{d}_i - \mathbf{d}_j\|$ is selected. In order to ensure the similarity between matching points, the distance should be less than a threshold ε. If this constraint is not satisfied, again no match is established for (x_i, y_i).

3D analysis

From each pair of matched keypoints (x_i, y_i) in image $\mathbf{I}_i$ and (x_j, y_j) in image $\mathbf{I}_{j=i+1}$ established in the previous step, a 3D point is reconstructed using the projection matrices $\mathbf{P}_i$ and $\mathbf{P}_j$ of the geometric model mentioned in Section 2.1.2 (see the triangulation algorithm in[17]). Similarly to the monocular detection approach, neighbour 3D points are clustered in the 3D space using the mean shift algorithm[7] and only those clusters that have a large enough number of 3D points are selected.

Final analysis

For each selected 3D cluster, all 3D reconstructed points belonging to the cluster are reprojected onto all images of the sequence using the projection matrices of the geometric model (see Figure 10). The extracted descriptors of the keypoints located near these reprojected points are classified individually using classifier h (defined in Section 2.1.1). The cluster will be classified as class c' if there is a large number of keypoints individually classified as c' and this number represents a majority in the cluster.

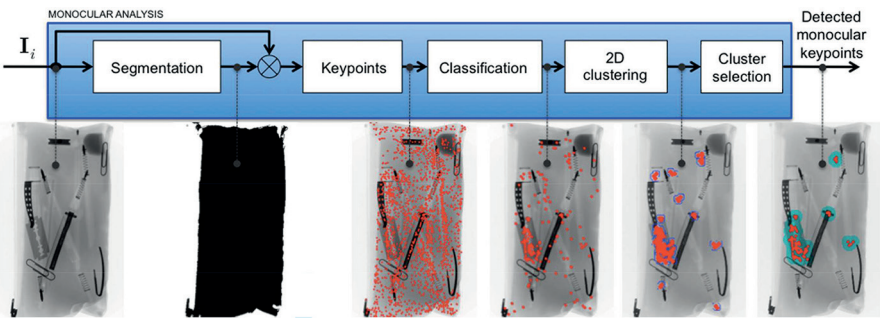

Figure 8. Monocular analysis for each image of the sequence, *ie* for $i = 1, \ldots, n$. In this example, the class of interest is 'razor blade'

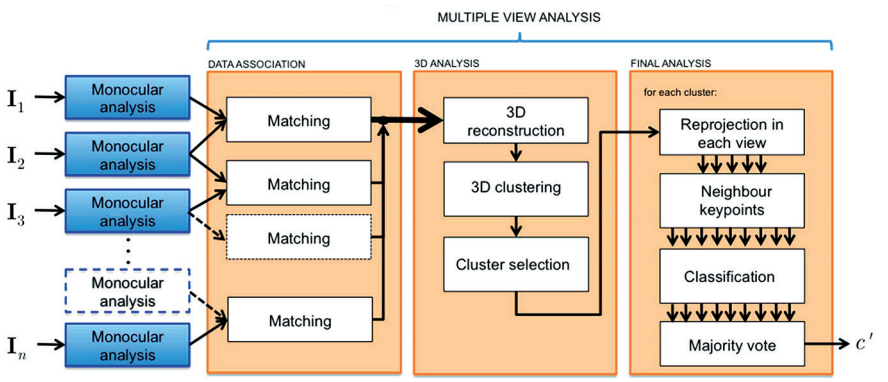

Figure 9. Multiple view analysis. An explanation of the last step (final analysis) is illustrated in Figure 10

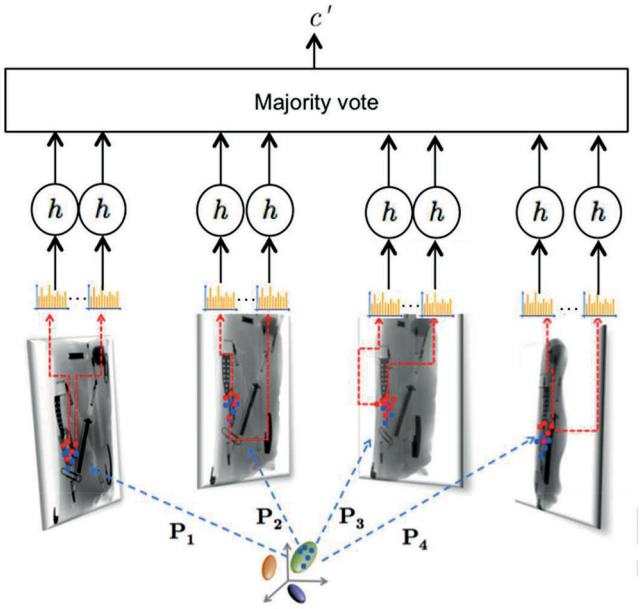

Figure 10. Final analysis: using the geometric model, the reconstructed 3D points in each cluster are reprojected in each view (blue points). The keypoints that are near to the reprojected points are identified (red points). The descriptors of these keypoints (orange histograms) are classified using trained classifier *h*. The class *c*′ of this cluster is determined by majority vote. In this example of *n* = 4 views, only the green cluster is represented

This majority vote strategy can overcome the problem of false monocular detections when the classification of the minority fails. A cluster can be misclassified if the part that we are trying to recognise is occluded by a part of another class. In this case, there will be keypoints in the cluster assigned to both classes; however, we expect that the majority of keypoints will be assigned to the true class if there are a small number of keypoints misclassified.

3. Experiments and results

In the experiments, the task was to recognise three different classes of object present in a pencil case (see, for example, a sequence in Figure 11(a)). These classes are: 'clips', 'springs' and 'razor blades'. The recognition approach explained in Section 2 was followed.

In the offline stage a structure from a motion algorithm was used in order to estimate the projection matrices of each view[1]. Additionally, in the learning phase, only 16 training images of each class were used.

Due to the small intra-class variation of the classes, this number of training images was deemed sufficient. The training objects were posed at different angles. SIFT descriptors were extracted as explained in[13] and a *k*-nearest neighbour (KNN) classifier with $k = 3$ neighbours was ascertained using the SIFT descriptors of the four classes[2]. Other descriptors (such as LBP and HOG) and other classifiers (such as SVM or KNN with other values of *k*) were also tested, although the best performance was achieved with the aforementioned configuration.

In order to illustrate step-by-step the online stage, the

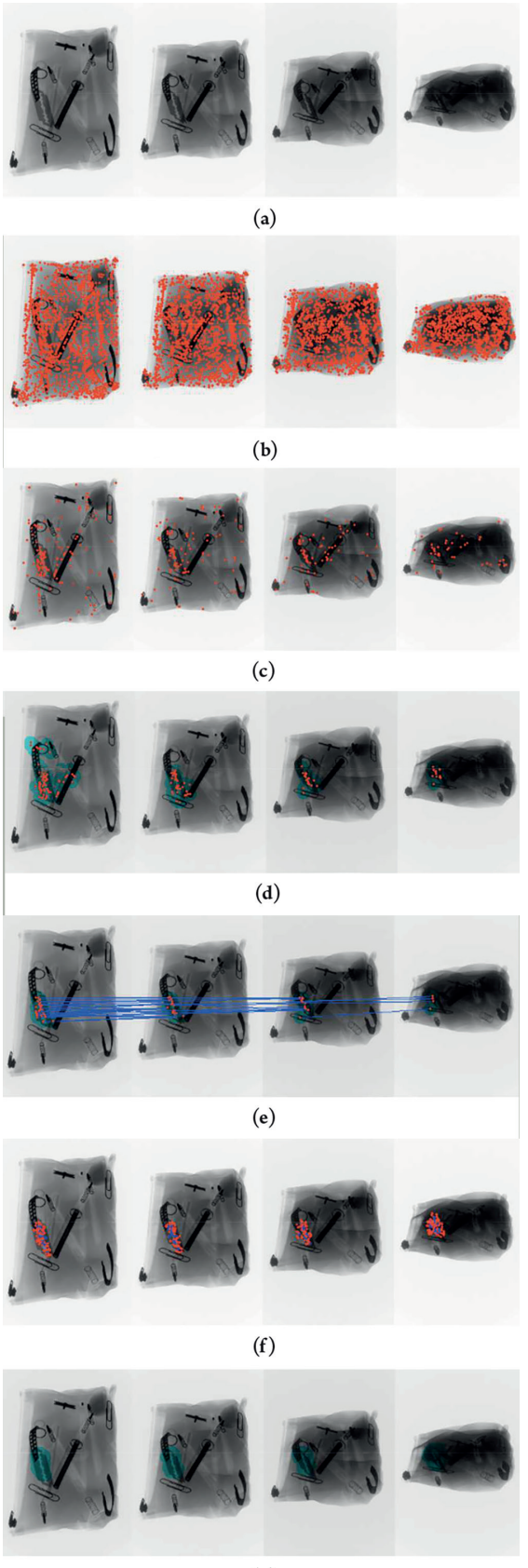

Figure 11. Recognition of a razor blade using the proposed approach: (a) original sequence; (b) keypoints; (c) classified keypoints; (d) detected monocular keypoints; (e) matched keypoints; (f) reprojected 3D points (blue) and neighbour keypoints (red); (g) final detection

[1]A fast implementation of multiple view geometry algorithms from the BALU Toolbox[15] was used in the experiments.

[2]Fast implementations of SIFT and KNN (based on *k*-d tree) from the VLFeat Toolbox[39] were used in the experiments.

recognition of a razor blade is illustrated in Figures 11(a) to 11(d) for monocular analysis and in Figures 11(e) to 11(g) for multiple view analysis[3]. Other examples are illustrated in Figure 12. It is worth mentioning that in monocular detection there are false alarms; however, they can be filtered out after multiple view analysis. The reason is that false alarms cannot be tracked in the sequence or because the tracked points, when validating the corresponding points in other views of the sequence, do not belong to the class of interest. Other results with some degree of overlap, where the task was the recognition of springs and clips, are illustrated in Figure 13.

The performance of the method is given in terms of precision-recall (*Pr*, *Re*), defined as follows[17]:

$$Pr = \frac{TP}{TP + FP}, Re = \frac{TP}{TP + FN} \quad \text{.....(2)}$$

where true positive (*TP*) is the number of targets correctly classified and false positive (*FP*) is the number of non-targets classified as targets. The false positives are known as 'false alarms' and 'Type I errors' and false negatives (*FN*) are the number of targets classified as no-targets. The false negatives are known as 'Type II errors'.

On one hand, precision gives the ratio of the number of true positives to the number of detections ($D = TP + FP$). On the other hand, recall gives the ratio of the number of true positives to the number of existing targets, known as 'ground truth' ($GT = TP + FN$). Ideally, a perfect detection means all existing targets are correctly detected without any false alarms, *ie* $Pr = 1$ and $Re = 1$.

Testing experiments were carried out by recognising the three mentioned classes (clips, springs and razor blades) in 45 different sequences of four views (15 sequences for each class)[4]. The size of an individual image was 1430 × 900 pixels. In these experiments there were 30 clips, 75 springs and 15 razor blades to be recognised. A summary of the results using the proposed algorithm is presented in Table 1 and in Figure 14, where the performance in the recognition of each class is presented in two different parts of the algorithm: after monocular analysis (mono) and after multiple view analysis (multi). These parts are illustrated in Figures 11(d) and 11(g), respectively, for a razor blade. In this Table, ground truth (*GT*) is the number of existing objects to be recognised. In the experiments, precision (*Pr*), computed as $Pr = TP / (TP+FP)$, is 71.4% and 95.7% in each part and recall (*Re*), computed as $Re = TP / GT$, is 90.8% and 92.5% in each step.

Figure 12. Detection of clips, springs and razor blades in the same sequence

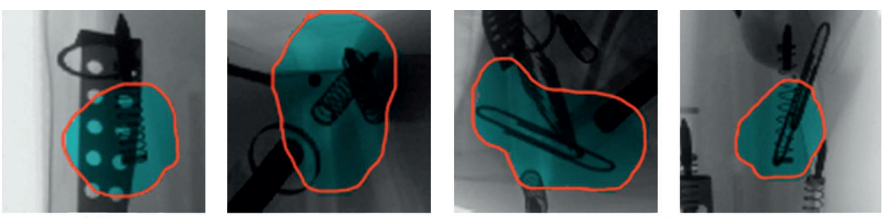

Figure 13. Recognition using the proposed approach in cases with some degree of overlap: (a) one spring; (b) two springs; (c) one clip; (d) one clip and one spring. Each Figure shows a part of one image of the whole sequence

As can be seen in Figure 14, if single *versus* multiple view detection are compared, both precision and recall are incremented. Precision, however, is drastically incremented because the approach achieves good discrimination from false alarms.

Since, in this approach, the geometric model and tracking are estimated by matching SIFT keypoints from different views, it was expected that the methodology would not allow for large viewpoint angles because matching performance with SIFT features decreases with large changes in viewpoint[13]. For this reason, the experiments were performed in cases where the viewpoint angles between consecutive views are small (up to 10º on the rotation axis).

The amount of time required in the experiments was about

[3]A fast implementation of mean shift from the PMT Toolbox[8] is used in the experiments.

[4]The images tested in the experiments came from the public GDXray database[20].

Table 1. Recognition performance

Class	Mono					Multi				
	TP	*FP*	*GT*	*Pr*	*Re*	*TP*	*FP*	*GT*	*Pr*	*Re*
Clip	114	127	120	0.4730	0.9500	26	2	30	0.9286	0.8667
Spring	263	30	300	0.8976	0.8767	71	3	75	0.9595	0.9467
Razor blade	59	18	60	0.7662	0.9833	14	0	15	1.0000	0.9333
	436	175	480	0.7136	0.9083	111	5	120	0.9569	0.9250

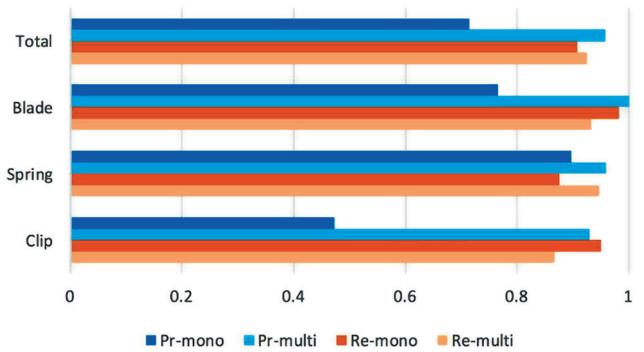

Figure 14. Graphic representation of Table 1

10 min for the offline stage and about 10 s for testing each sequence on a Mac Mini Server OS X 10.10.1 with 2.6 GHz Intel Core i7 processor with 4 cores and memory of 16 GB RAM 1600 MHz DDR3. The code of the program (implemented in Matlab) is available on the authors' website.

4. Conclusions

In this paper, a new method that can be used to recognise certain parts of interest in complex objects using multiple X-ray views was presented. The proposed method filters out false positives resulting from monocular detection performed on single views by matching information across multiple views. This step is performed efficiently using a look-up table that is computed offline. In order to illustrate the effectiveness of the proposed method, experimental results on recognising regular objects (clips, springs and razor blades) in pencil cases are shown achieving high precision and recall (*Pr* = 95.7%, *Re* = 92.5%) for 120 objects. We believe that it would be possible to design an automated aid in a target detection task using the proposed algorithm. In future work, the approach will be tested in more complex scenarios recognising objects with a larger intra-class variations.

Acknowledgements

This work was supported in part by Fondecyt Grant Nos 1130934 and 1161314 from CONICYT, Chile.

References

1. B R Abidi, Yue Zheng, A V Gribok and M A Abidi, 'Improving weapon detection in single-energy X-ray images through pseudo-colouring', IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, Vol 36, No 6, pp 784-796, November 2006.
2. O Abusaeeda, J P O Evans, D Downes and J W Chan, 'View synthesis of KDEX imagery for 3D security X-ray imaging', 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), pp 1-6, 2011.
3. J L Bentley, 'Multidimensional binary search trees used for associative searching', Communications of the ACM, Vol 18, No 9, pp 509-517, 1975.
4. G Blalock, V Kadiyali and D H Simon, 'The impact of post-9/11 airport security measures on the demand for air travel', The Journal of Law and Economics, Vol 50, No 4, pp 731-755, November 2007.
5. A Bolfing, T Halbherr and A Schwaninger, 'How image-based factors and human factors contribute to threat detection performance in X-ray aviation security screening', In: HCI and Usability for Education and Work, Vol 5298, Andreas Holzinger Editor, pp 419-438, Springer, Berlin Heidelberg, 2008.
6. J Chan, P Evans and Xun Wang, 'Enhanced colour-coding scheme for kinetic depth effect X-ray (KDEX) imaging', 2010 IEEE International Carnahan Conference on Security Technology (ICCST), pp 155-160, October 2010.
7. D Comaniciu and P Meer, 'Mean shift: a robust approach towards feature space analysis', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 24, No 5, pp 603-619, 2002.
8. P Dollár, 'Piotr's Image and Video Matlab Toolbox (PMT)', http://vision.ucsd.edu/_pdollar/toolbox/doc/index.html
9. G Flitton, T P Breckon and N Megherbi, 'A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery', Pattern Recognition, Vol 46, No 9, pp 2420-2436, September 2013.
10. G Flitton, A Mouton and T P Breckon, 'Object classification in 3D baggage security computed tomography imagery using visual codebooks', Pattern Recognition, Vol 48, No 8, pp 2489-2499, August 2015.
11. T Franzel, U Schmidt and S Roth, 'Object detection in multiview X-ray images', Pattern Recognition, 2012.
12. Dongmei Liu and Zhaoxia Wang, 'A united classification system of X-ray image based on fuzzy rule and neural networks', 3rd International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2008, Vol 1, pp 717-722, November 2008.
13. D Lowe, 'Distinctive image features from scale-invariant keypoints', International Journal of Computer Vision, Vol 60, No 2, pp 91-110, November 2004.
14. N Megherbi, J Han, T P Breckon and G T Flitton, 'A comparison of classification approaches for threat detection in CT-based baggage screening', 19th IEEE International Conference on Image Processing (ICIP), pp 3109-3112, 2012.
15. D Mery, 'BALU: a Matlab toolbox for computer vision, pattern recognition and image processing', http://dmery.ing.puc.cl/index.php/balu
16. D Mery, 'Automated detection in complex objects using a tracking algorithm in multiple X-ray views', 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 41-48, 2011.
17. D Mery, Computer Vision for X-Ray Testing, Springer, 2015.

18. D Mery, 'Inspection of complex objects using multiple X-ray views', IEEE/ASME Transactions on Mechatronics, Vol 20, No 1, pp 338-347, 2015.
19. D Mery, G Mondragon, V Riffo and I Zuccar, 'Detection of regular objects in baggage using multiple X-ray views', Insight: Non-Destructive Testing and Condition Monitoring, Vol 55, No 1, pp 16-20 and 23, 2013.
20. D Mery, V Riffo, U Zscherpel, G Mondragón, I Lillo, I Zuccar, H Lobel and M Carrasco, 'GDXray: the database of X-ray images for non-destructive testing', Journal of Nondestructive Evaluation, Vol 34, No 4, pp 1-12, 2015.
21. D Mery, V Riffo, I Zuccar and C Pieringer, 'Automated X-ray object recognition using an efficient search algorithm in multiple views', IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2013), pp 368-374, 2013.
22. S Michel, S M Koller, J C de Ruiter, R Moerland, M Hogervorst and A Schwaninger, 'Computer-based training increases efficiency in X-ray image interpretation by aviation security screeners', 41st Annual IEEE International Carnahan Conference on Security Technology, pp 201-206, October 2007.
23. K Mikolajczyk and C Schmid, 'A performance evaluation of local descriptors', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 27, No 10, pp 1615-1630, October 2005.
24. A Mouton, G T Flitton and S Bizot, 'An evaluation of image denoising techniques applied to CT baggage screening imagery', IEEE International Conference on Industrial Technology (ICIT 2013), 2013.
25. A Mouton and T P Breckon, 'Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening', Pattern Recognition, Vol 48, No 6, pp 1961-1978, June 2015.

26. E E Murphy, 'A rising war on terrorists', Spectrum, IEEE, Vol 26, No 11, pp 33-36, November 1989.
27. N C Murray and K Riordan, 'Evaluation of automatic explosive detection systems', Proceedings of the 29th Annual IEEE International Carnahan Conference on Security Technology, pp 175-179, October 1995.
28. S Nercessian, K Panetta and S Agaian, 'Automatic detection of potential threat objects in X-ray luggage scan images', 2008 IEEE Conference on Technologies for Homeland Security, pp 504-509, May 2008.
29. C Oertel and P Bock, 'Identification of objects of interest in X-ray images', 35th IEEE Applied Imagery and Pattern Recognition Workshop 2006, AIPR 2006, p 17, October 2006.
30. European Parliament, 'Aviation security with a special focus on security scanners', European Parliament Resolution (2010/2154(INI)), pp 1-10, October 2012.
31. T Poggio and S Edelman, 'A network that learns to recognise 3D objects', Nature, 1990.
32. V Riffo and D Mery, 'Active X-ray testing of complex objects', Insight: Non-Destructive Testing and Condition Monitoring, Vol 54, No 1, pp 28-35, 2012.
33. V Riffo and D Mery, 'Automated detection of threat objects using adapted implicit shape model', IEEE Transactions on Systems, Man and Cybernetics: Systems, Vol 46, No 4, pp 472-482, 2016.
34. 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', Proceedings of the 3rd International Conference on Research in Air Transportation, ICRAT 2008, pp 317-324, 2008.
35. M Singh and S Singh, 'Optimising image enhancement for screening luggage at airports', Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, CIHSPS 2005, pp 131-136, 2005.
36. H Strecker, 'Automatic detection of explosives in airline baggage using elastic X-ray scatter', Medicamundi, Vol 42, pp 30-33, July 1998.
37. D Turcsany, A Mouton and T P Breckon, 'Improving feature-based object recognition for X-ray baggage security screening using primed visual words', IEEE International Conference on Industrial Technology (ICIT 2013), pp 1140-1145, 2013.
38. I Uroukov and R Speller, 'A preliminary approach to intelligent X-ray imaging for baggage inspection at airports', Signal Processing Research, 2015.
39. A Vedaldi and B Fulkerson, 'VLFeat: an open and portable library of computer vision algorithms', 2008. http://www.vlfeat.org/
40. C C von Bastian, A Schwaninger and S Michel, 'Do multiview X-ray systems improve X-ray image interpretation in airport security screening?', Vol 52, GRIN Verlag, 2010.
41. A Wales, T Halbherr and A Schwaninger, 'Using speed measures to predict performance in X-ray luggage screening tasks', 43rd Annual International Carnahan Conference on Security Technology, pp 212-215, October 2009.
42. G Zentai, 'X-ray imaging for homeland security', IEEE International Workshop on Imaging Systems and Techniques (IST 2008), pp 1-6, September 2008.
43. N Zhang and J Zhu, 'A study of X-ray machine image local semantic features extraction model based on bag-of-words for airport security', International Journal on Smart Sensing and Intelligent Systems, Vol 8, No 1, pp 45-64, 2015.