Detection of Regular Objects in Baggages Using Multiple X-ray Views

Domingo Mery, German Mondragon, Vladimir Riffo and Irene Zuccar

Abstract

In order to reduce the security risk of a commercial aircraft, passengers are not allowed to take certain items in carry-on baggage. For this reason, human operators are trained to detect prohibited items using a manually controlled baggage screening process. In this paper, we propose the use of a method based on multiple X-ray views to detect some regular prohibited items with very defined shapes and sizes. The method consists of two steps: 'structure estimation', to obtain a geometric model of the multiple views from the object to be inspected (a baggage), and 'parts detection', to detect the parts of interest (prohibited items). The geometric model is estimated using a structure from motion algorithm. The detection of the parts of interest is performed by an *ad-hoc* segmentation algorithm (object dependent) followed by a general tracking algorithm based on geometric and appearance constraints. In order to illustrate the effectiveness of the proposed method, experimental results on detecting regular objects –razor blades and guns– are shown yielding promising results.

Keywords: X-ray testing, baggage screening, luggage scan, multiple view imaging, computer vision.

D. Mery is with DCC - Pontificia Universidad Católica de Chile. E-mail: dmery@ing.puc.cl,

G. Mondragon is with DCC - Pontificia Universidad Católica de Chile. E-mail: german.mondragon@gmail.com

V. Riffo is with DCC - Pontificia Universidad Católica de Chile and DIICC - Universidad de Atacama. E-mail: vriffo1@uc.cl

I. Zuccar is with DCC - Pontificia Universidad Católica de Chile. E-mail: irene.zuccar@usach.cl

1. Introduction

Since 9/11 aviation security screening with X-ray scanners has become a very important issue in airports. The inspection process, however, is complex because threat items are very difficult to detect when placed in close packed bags, superimposed by other objects, and/or rotated showing an unrecognizable view[1]. In baggage screening, where human security plays an important role and inspection complexity is very high, human inspectors are still used. Nevertheless, during rush hours in airports, human screeners have only a few seconds to decide whether a bag contains or not a prohibited item, and detection performance is only about 80-90%[2].

For these reasons, digital imaging and computer vision techniques have been developed in order to increase the effectiveness and automation of the inspection task. Before 9/11, however, the X-ray analysis of luggage mainly focused on capturing the images of their content: the reader can find in[3] an interesting analysis done in 1989 of several aircraft attacks in the world, and the existing technologies to detect the terrorists threats based on Thermal-Neutron Activation (TNA), Fast-Neutron Activation (FNA) and dual energy X-rays (used in medicine since early 70). In the 90's, Explosive Detection Systems (EDS) were developed based on X-ray imaging[4], and computed tomography through elastic scatter X-ray (comparing the structure of irradiated material, against stored reference spectra for explosives and drugs)[5].

All these works were concentrated on image acquisition and simple image processing but they lack advanced image analysis to improve the detection performance. Nevertheless, the 9/11 attacks increased the security policies at airports, which also produced the interest of the scientific community for researching topics related to security using advanced computational techniques. In the last decade, the main contributions were: analysis of human inspection[6], pseudo-coloring of X-ray images[7], enhancement and segmentation of X-ray images[8] and detection of threat items in X-ray images based on texture features (detecting a 9mm Colt Beretta machine pistol)[9], neural networks and fuzzy rules (yielding about 80% of performance)[10], and SVM classifier (detecting guns in real time)[11].

Recently, some algorithms based on multiple X-ray views were reported in the literature. For example: synthesis of new X-ray images obtained from Kinetic Depth Effect X-ray (KDEX) images based on SIFT features in order to increase the detection performance[12]; active vision with X-ray, which allows modifying the

viewpoint of the target object in order to obtain better X-ray images to analyze (detecting razor blades in different cases)[13]; and tracking across multiple X-ray views in order to verify the diagnoses performed using a single view[14]. The key idea of this method is: *i*) to segment potential parts (regions) of interest in each view using an application dependent method that analyzes 2D features in each single view ensuring the detection of the object parts of interest (not necessarily in all views) and allowing false detections, *ii*) to match and track the potential regions based on similarity and geometrical multiple views constraints eliminating those that cannot be tracked, and *iii*) to analyze the tracked regions including those views where the segmentation fails (the positions can be predicted by re-projection). This algorithm will be explained in next section in further details because it is the core of this paper.

In baggage screening, the use of multiple view information yields a significant improvement in performance because certain items are difficult to be recognized using only one viewpoint, as we illustrate in Figure 1, where we detected a razor blade in a pencil case using the proposed method. It is clear, that this detection performance could not be achieved with only the first view of the sequence.

In this work, we use the general methodology proposed by us in[14] and implement algorithms to automatically detect regular objects in baggages (like razor blades and guns) with multiple X-ray views. We show the robustness of the approach against poor segmentation or noise because these false detections are not attached to the object and therefore they cannot be tracked.

The rest of the paper is organized as follows: the multiple view approach is summarized in Section 2, the *ad-hoc* single view detectors are explained in Section 3, the results obtained in several experiments are shown in Section 4, and some concluding remarks are given in Section 5.

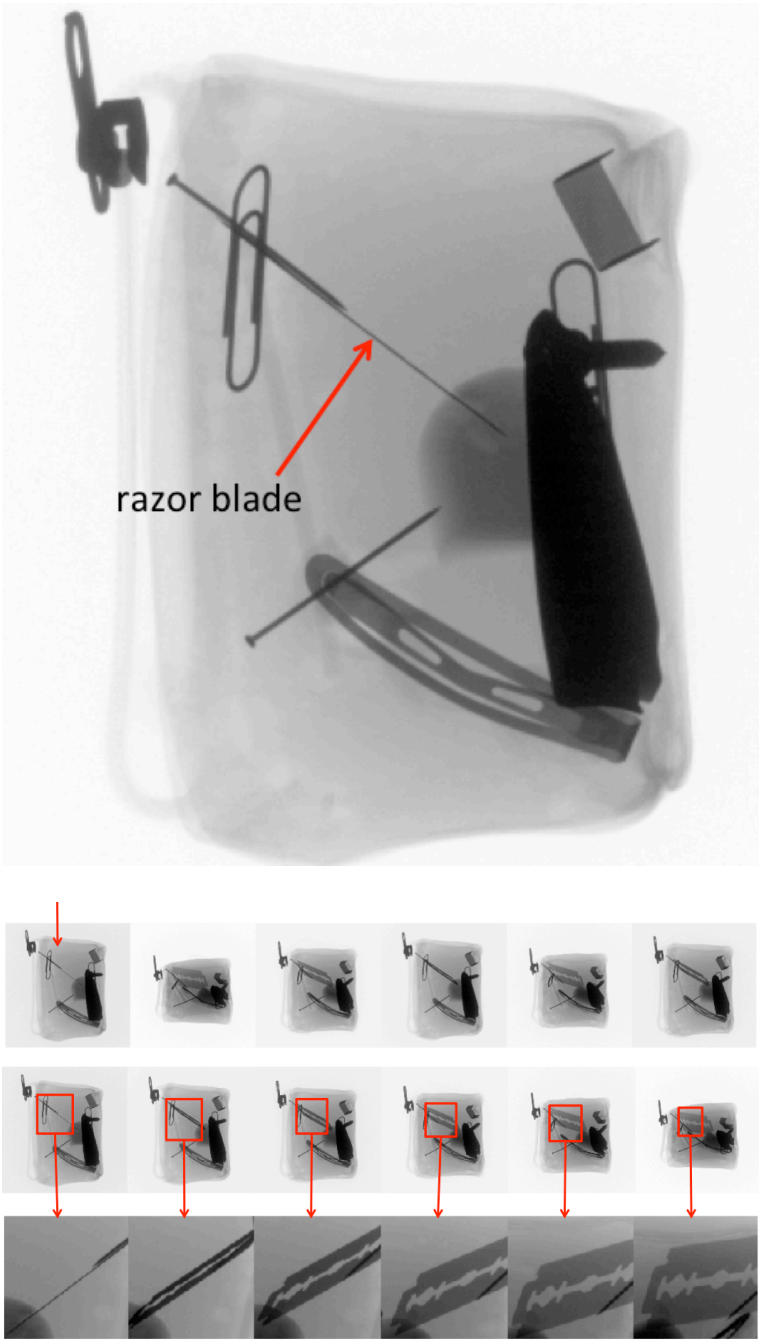

Figure 1. Detection of a razor blade in a pencil case using our approach: First X-ray image of the sequence, 1430 × 900 pixels (top). Unsorted and sorted sequences with six images (middle). Detection in each image (bottom).

2. Multiple view approach

In this Section we summarize the multiple view approach outlined in [14] using *ad-hoc* single view detectors for regular objects. The proposed method follows two main steps: 'structure estimation', to obtain a geometric model of the multiple views from the object itself, and 'parts detection', to detect the object parts of interest.

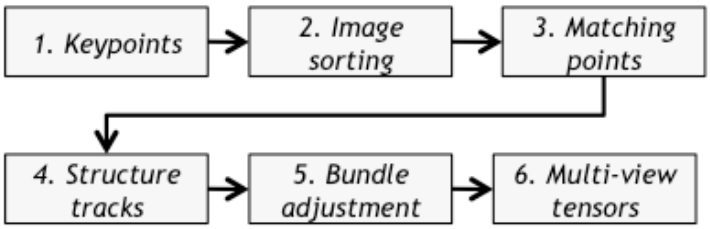

Figure 2. Block diagram of structure estimation.

2.1 Structure estimation

The approach outlined in this section is based on well known structure from motion[15] estimated from a sequence of m images taken from a rigid object at different viewpoints (see Figure 2).

The original image sequence is stored in m images $\mathbf{J}_1$, …, $\mathbf{J}_m$. For each image, SIFT keypoints are extracted[16]. Thus, not only a set of 2D image positions $\mathbf{x}$, but also descriptors $\mathbf{y}$, are obtained. The images of the sequence are sorted using a visual vocabulary tree in order to obtain a sequence with small changes between consecutive frames[17] as shown in Figure 1. For two consecutive and sorted images, $\mathbf{I}_i$ and $\mathbf{I}_{i+1}$, SIFT keypoints are matched using the algorithm suggested by Lowe[16] that rejects too ambiguous matches. Afterwards, the Fundamental Matrix between views i and $i + 1$, $\mathbf{F}_{i,i+1}$, is estimated using RANSAC[15] to remove outliers. We look for all possible structure tracks with one keypoint in each image of sequence.

The determined tracks define n image point correspondences over m views. They are arranged as $\mathbf{x}_{i,j}$ for $i = 1, ..., m$ and $j = 1, ..., n$. Bundle adjustment estimates 3D points $\hat{\mathbf{X}}_j$ and camera matrices $\mathbf{P}_i$ so that $\sum\|\mathbf{x}_{i,j} - \hat{\mathbf{x}}_{i,j}\|$ is minimized, where $\hat{\mathbf{x}}_{i,j}$ is the projection of $\hat{\mathbf{X}}_j$ by $\mathbf{P}_i$. A RANSAC approach is used to remove outliers. Bundle adjustment[15] provides a method for computing bifocal and trifocal tensors from projection matrices $\mathbf{P}_i$, that will be used in the next section.

2.2 Parts detection

In this section we give details of the algorithm that detects the object parts of interest. The algorithm consists of following two main steps: identification and tracking as shown in Figure 3.

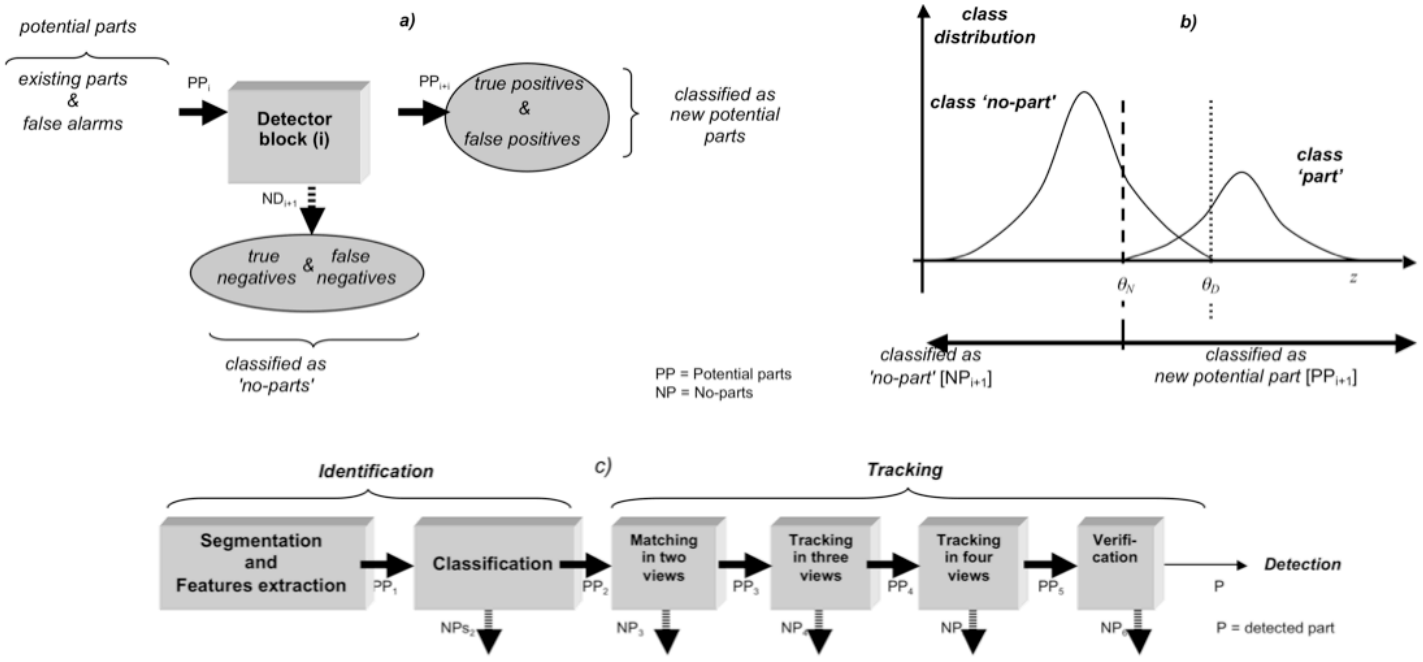

Figure 3. Block diagram of parts detection: each block separates the new potential parts from the no-parts (a) according to the class distribution (b). The whole diagram follows a cascade schema as shown in (c).

The strategy is to ensure the detection of the 'existing parts of interest' in first step, allowing the inclusion of 'false alarms'. The discrimination between both is achieved in second step using *multiple view analysis*, where the attempt is made to track the potential parts of interest along the image sequence.

In the identification, potential parts of interest are segmented and classified in each image $\mathbf{I}_i$ of the sequence. It is an *ad-hoc* single view detector that depends on the application. In Section 3, two algorithms will be explained for the detection of regular objects (razor blades and guns).

An existing part of interest can be successfully tracked in the image sequence because its appearance in the images is similar and their projections are located in the positions dictated by geometric conditions. In contrast, false alarms can be successfully eliminated in this manner, since they do not appear in the predicted places on the following images and, thus, cannot be tracked. The tracking in the image sequence is performed

using algebraic multi-focal constraints: bifocal (epipolar) and trifocal constraints among others[15] obtained from projection matrices estimated in previous step outlined in Section 2.1 where the geometric model is obtained from the target object itself.

Each sub-step i of the automated multiple view analysis can be understood as a detector block as shown in Figure 3. The potential parts (PP_i) consisting of existing parts and false alarms are classified as either new potential parts (PP_{i+1}) or no-parts (NP_{i+1}) (Figure 3a). In a training phase, each detector block is tuned so that the maximal number of false alarms is eliminated from the potential parts without discriminating the existing defects (see θ_S in Figure 3b). The throughput cycle can be considerably incremented if we use an additional decision boundary (see θ_D in Figure 3b) which guarantee the detection of defects in previous stages without computing the next steps.

The reader is referred to [14] for a detailed description of the tracking algorithm.

3. Object dependent single view detector

An object dependent algorithm must be defined to detect automatically potential parts of interest in a single test image. As mentioned above, in order to test our method, we developed two algorithms that are able to detect razor blades and guns. In this section, they will be explained in further details.

3.1 Detection of razor blades

The algorithm is based on matching of SIFT keypoints[16] and was proposed by us in the first part of [13] for active vision. In our approach, we use a SIFT description of the target object in all feasible poses by rotating two axes in nine steps. All extracted descriptors are stored in an arrange $\mathbf{P}$, where $\mathbf{p}_j$ means the j-th descriptor, for $j = 1 \ldots m$. Each descriptor $\mathbf{p}_j$ has a corresponding pose r_j. In our example, $r_j \in [1, 81]$ for 9×9 poses. All SIFT descriptors of the test image of the inspection object are extracted and stored in an arrange $\mathbf{Q}$, where $\mathbf{q}_i$ means the i-th descriptor of the test image for $i = 1 \ldots n$. Now, all duplets $(\mathbf{q}_i, \mathbf{p}_j)$ that fulfill the condition $\|\mathbf{q}_i - \mathbf{p}_j\| < \theta_E$ for $i = 1 \ldots n$ and $j = 1 \ldots m$ are selected, where θ_E is a minimum distance threshold, and $\|\mathbf{q}_i - \mathbf{p}_j\|$ means the Euclidean distance between both vectors. Afterwards, for each selected descriptors the corresponding

pose r_j is obtained. The selected descriptors and their corresponding poses will be stored in $\boldsymbol{Q}$ and $\boldsymbol{R}$ respectively. Thus, we have *i*) $\boldsymbol{Q}$: all keypoints of the test image that have been matched with keypoints of the target object, and *ii*) $\boldsymbol{R}$: the corresponding poses for the selected keypoints $\boldsymbol{Q}$.

The detection is performed in the following two steps: *i*) Clustering: in $\boldsymbol{Q}$, we find all keypoints of the same pose that are close to each other in the test image. Thus, we define subwindows $\boldsymbol{W_B}$ that have at least θ_B keypoints of the same pose. In our experiments, we set the size of $\boldsymbol{W_B}$ equal to 80×80 pixels, and $\theta_B = 3$. *ii*) Merging: all subwindows $\boldsymbol{W_B}$ that are connected or overlapped, will be merged in a new larger subwindow $\boldsymbol{W_G}$. The subwindow that encloses the highest number of keypoints of the same pose will be selected if this number is equal or greater than θ_G, ensuring at least θ_G descriptors of the same pose in the selected window. In our experiments, we set $\theta_G = 2$ in order to ensure the detection (allowing false alarms). The selected subwindow will be called $\boldsymbol{W_S}$ and it corresponds to a potential target object. If no subwindow fulfills this condition, then no potential target object is detected.

3.2 Detection of guns

In computer vision community many object detection and classification problems have been recently solved –without segmentation– using *sliding-windows*. Sliding-window approaches have established themselves as state-of-the-art in computer vision problems where an object must be separated from the background (see for example successful applications in face detection[18]). In sliding-window methodology, a detection window is passed over an input image in both horizontal and vertical directions, and for each localization of the detection window, a classifier decides to which class belongs the corresponding portion of the image according to its features. Multiple detection can be eliminated using non maximum suppression[18].

We used sliding-windows to detect guns, however, since there are many types of guns, our approach to detect a gun is based on the detection of its trigger. We observed, that the shape of triggers has a smaller variability in comparison of the shape of the guns. For these reason, we collected 100 images of guns from Google Images and cropped their triggers. Afterwards, we build a dataset with positive classes (trigger images) and negative classes (no trigger images). We trained a classifier using this dataset. We tested with several

features and classifiers. A simple and fast solution was achieved using a Mahalanobis distance classifier with seven geometric features (a Hu moment, a Fourier descriptor, center of mass, minor and major axes of a fitted ellipse).

4. Experimental results

We experimented on X-ray images from 2 different objects: *i*) detection of razor blades and, *ii*) detection of guns.

4.1 Experiments on razor blades

We tested on four sequences of razor blades with four-six images with very good results. Figure 4 illustrates the detection using the single view detector outlined in Section 3.1. We observe that the razor blade was identified, however, there are two false alarms. They were filtered out after tracking steps.

Figure 5 shows the results obtained in each step. We can see that the razor blade was not identified by the single view detector in the last image, however, after tracking using the geometric model, it was possible to re-project its position in this view (see the last dashed rectangle). Another example was illustrated in Figure 1, again, the razor blade was not identified in the first image, however, it could be re-projected.

Another experiment with the same result is shown in Figure 6. In this experiment there were 35 potential razor blades identified in the six images sequence (only six of them were real existing razor blades), however, after tracking algorithm, the 29 false alarms were eliminated.

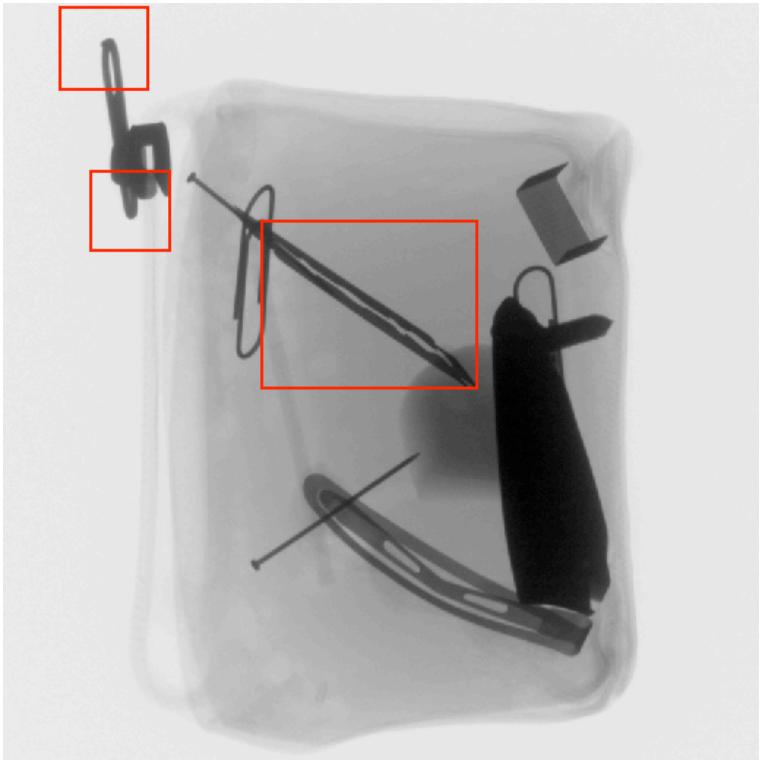

Figure 4. Identification of potential razor blades in a single view: there are two false alarms (upper left) and a true positive (large rectangle in the middle).

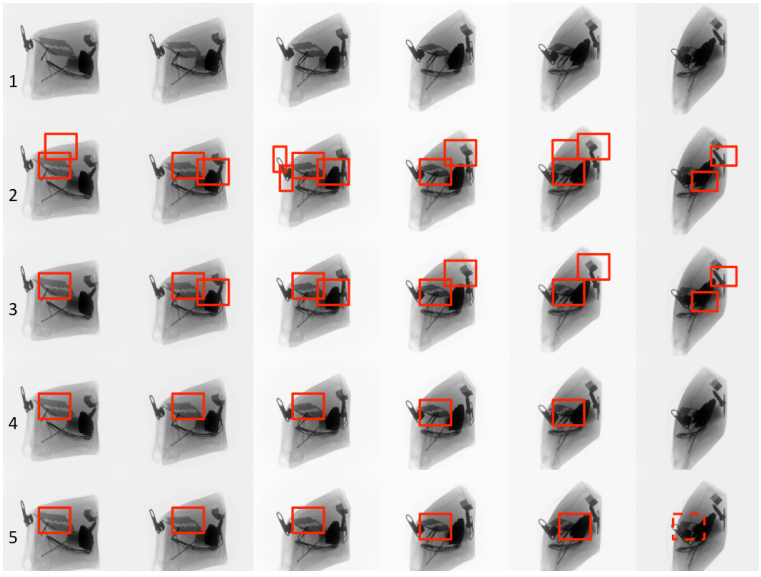

Figure 5. Results in each step: 1) sorted image sequence, 2) detection of potential razor blades using the single view detector, 3) remaining potential razor blades after matching in two views, 4) after three views, 5) four views.

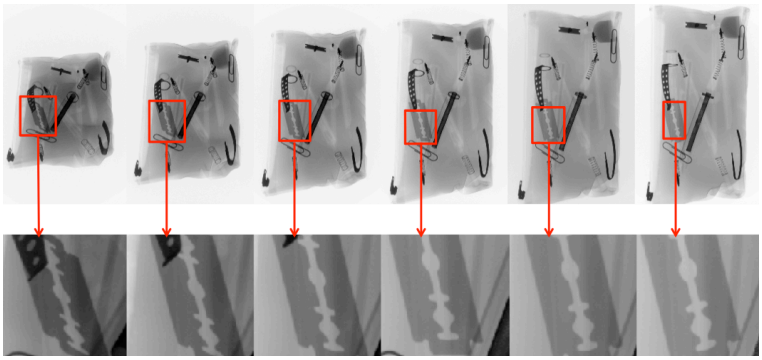

Figure 6. Detection of a razor blade in a pencil case. Top: sequence with 6 X-ray images, 1430 × 900 pixels. Bottom: detection.

4.2 Experiments on guns

We tested on ten sequences of four-five images of bags and backpacks containing a gun. The detection was achieved in sequences where the trigger was distinguishable. An example of the single view detector is shown in Figure 7. In this case the multiple false alarms were filtered out by tracking algorithm as shown in Figure 8. Another experiment that shows a very good detection with some occlusion is illustrated in Figure 10. Nevertheless, in intricate sequences (see for example Figure 9) the gun could not be detected because it was too occluded.

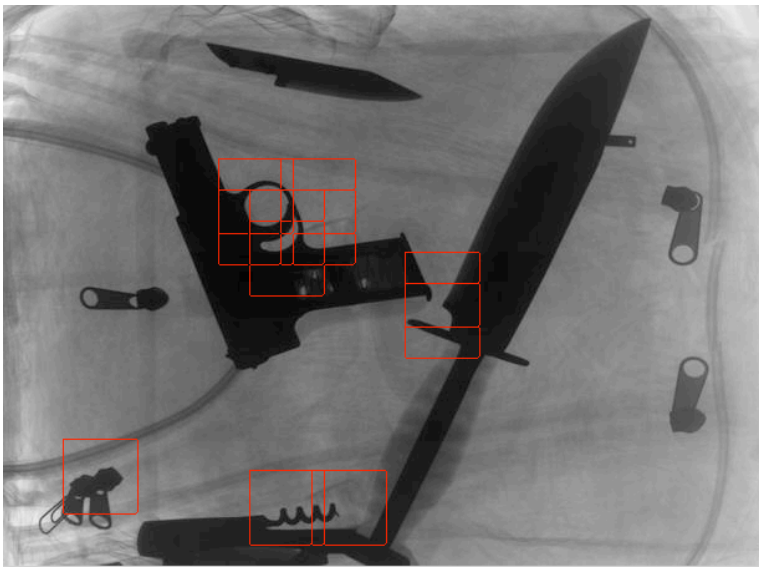

Figure 7. Single view detection of a gun, we observe that there are several false alarms that will be eliminated after tracking as shown in Figure 8.

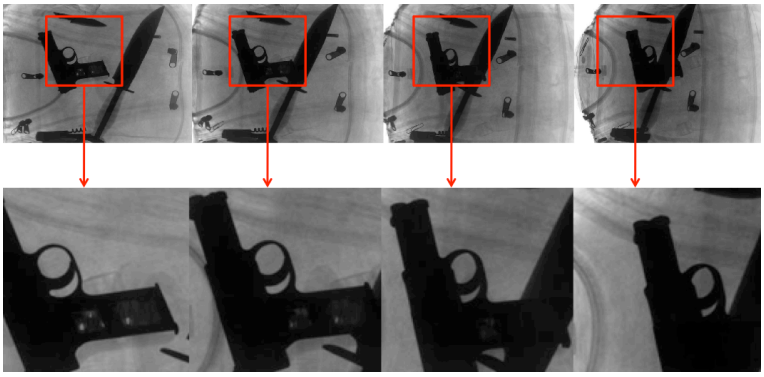

Figure 8. Detection of a gun in a bag. Top: sequence with 4 X-ray images, 452 × 612 pixels. Bottom: detection.

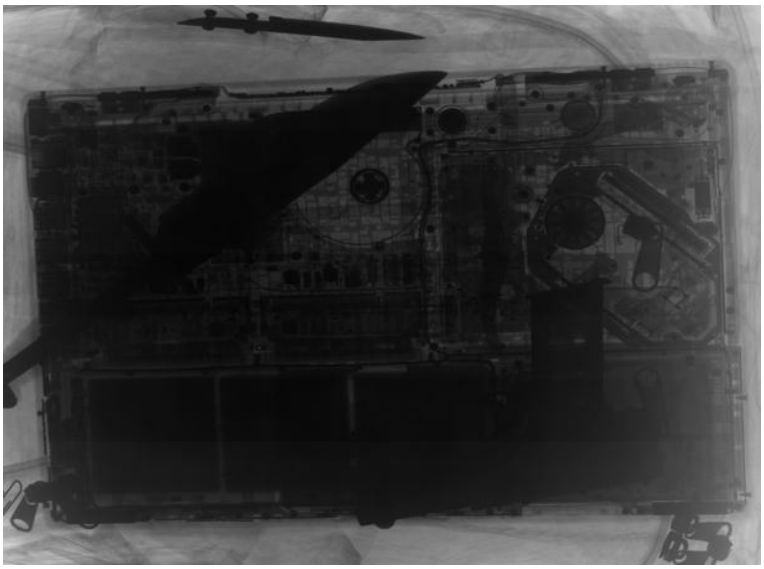

Figure 9. X-ray image of a gun and a knife on a laptop: the detection of the gun was impossible. See performance statistics in Table 2, Seq. 9.

4.3 Performance

Tables 1 and 2 summarizes the results on razor blades and guns with 14 sequences (64 X-ray images). Some of them are illustrated in the mentioned figures. m corresponds to the number of images in the sequence. SIFT/m means the average of the number of SIFT keypoints extracted per image. BA is the number of structure tracks found by bundle adjustment algorithm. n_1 is the number of segmented potential regions in the whole image sequence, and n_1/m is the average of segmented regions per image. n_1, is the number of l-tuplets tracked in the

sequence. n_d is the number of detected parts. GT is the number of existing parts (ground truth). FP and TP are the number of false and true positives after eliminating multiple overlapped detections. Ideally, FP = 0 and TP = GT. If we include all sequences the average performance is given by: *Precision* = TP/(TP+FP) = 70% and *Recall* = TP/GT = 86%. Nevertheless, if we exclude the last two gun sequences (they are not allowed in baggage screening because laptops must be removed from bags) *Precission* = 86% and *Recall* = 100%.

Table 1. Detection of razor blades*

Seq.	size	m	SIFT/m	BA	n_1	n_1 /m	n_2	n_3	n_4	n_q	n_d	GT	FP	TP
1	1430 × 900	6	2372	30	35	6	18	4	1	1	1	1	0	1
2	850 × 850	6	1679	4	14	2	13	8	1	1	1	1	0	1
3	850 × 850	6	1312	2	12	2	12	4	1	1	1	1	0	1
4	1430 × 900	4	5135	58	26	7	15	6	2	2	2	1	1	1
Total	–	22	–	–	–	17	–	–	–	–	5	4	1	4

* Variables used in this Table are explained in Section 4.3

Table 2. Detection of guns**

Seq.	size	m	SIFT/m	BA	n_1	n_1 /m	n_2	n_3	n_4	n_q	n_d	GT	FP	TP
1	459 × 612	4	2226	24	73	18	268	162	66	5	5	1	0	1
2	459 × 612	5	2253	6	114	23	573	347	164	15	15	1	0	1
3	459 × 612	4	2222	39	44	11	171	71	32	5	5	1	0	1
4	459 × 612	4	2242	35	38	10	162	87	8	2	2	1	1	1
5	459 × 612	4	2192	113	33	8	182	192	91	8	8	1	0	1
6	459 × 612	4	2297	39	88	22	596	166	48	7	7	1	0	1
7	459 × 612	4	662	33	103	26	1058	1407	1108	38	38	1	1	1
8	459 × 612	4	662	33	8	2	14	9	3	1	1	1	0	1
9	459 × 612	5	2246	162	180	36	3041	2916	1221	71	71	1	2	0
10	459 × 612	4	1473	62	93	23	600	509	376	17	17	1	1	0
Total	–	42	–	–	–	179	–	–	–	–	169	10	4	8

** Variables used in this Table are explained in Section 4.3

4.4 Implementation

We implemented our approach in a Matlab Graphic User Interface (Figure 10). We used the implementation of SIFT, visual vocabulary, etc. form VLFeat[19]. The rest of algorithms were implemented in MATLAB. For multiple view matching, ε_2 = 30 pixeles, ε_3 = 50 pixels. The computing time depends on the application,

however, in order to present a reference, for Figure 1 the results were obtained after 30 seconds on a iMac OS X 10.6.6, processor 3.06GHz Intel Core 2 Duo, 4GB RAM memory. The code of the MATLAB implementation is available on our webpage[20].

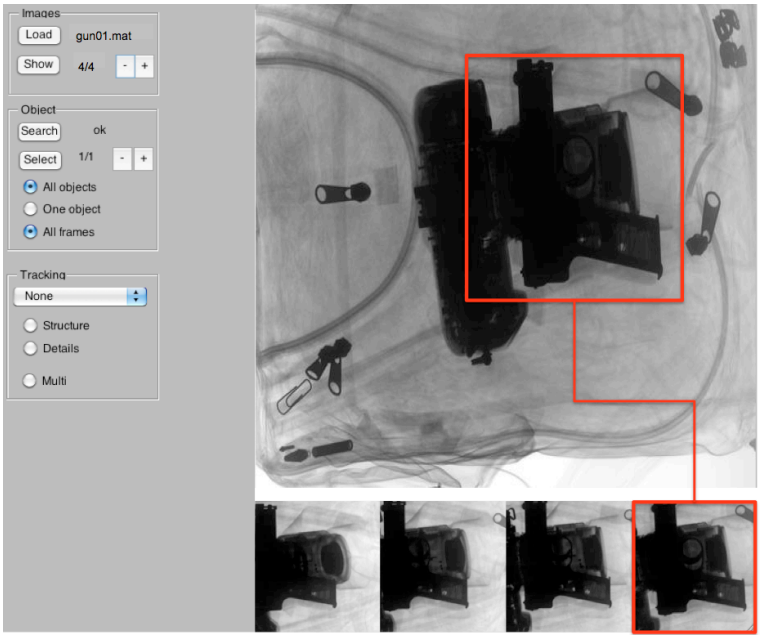

Figure 10. Developed graphic user interface (GUI) showing the detection of an occluded gun.

5. Conclusions

In this paper we presented the use of a generic methodology that can be used to detect regular prohibited items (like razor blades and guns) in baggages automatically yielding promising results. The proposed approach is an application of state-of-art computer vision techniques. It filters out false positives resulting from segmentation steps performed on single views of an object by corroborating information across multiple views.

Using multiple views (instead of one) the matching accuracy and robustness (*i.e.*, tolerance to false-positive detections) of the detection of physical features on an object is increased. The detection method is image-based (2D appearance-based detection). By using multiple views, the method is able to increase the detection rates and robustness of 2D feature detection, in comparison to application of the same method in a single image. We believe that our methodology is a useful alternative for assisting human operators in baggage screening.

Acknowledgements

This work was supported by grant Fondecyt No 1100830 from CONICYT – Chile.

References

1. G. Zentai, "X-ray imaging for homeland security," IEEE International Workshop on Imaging Systems and Techniques (IST 2008), pp. 1–6, Sept. 2008.
2. S. Michel, S. Koller, J. de Ruiter, R. Moerland, M. Hogervorst, and A. Schwaninger, "Computer-based training increases efficiency in XRay image interpretation by aviation security screeners," in Security Technology, 2007 41st Annual IEEE International Carnahan Conference on, Oct. 2007, pp. 201–206.
3. E. Murphy, "A rising war on terrorists," Spectrum, IEEE, vol. 26, no. 11, nov 1989, pp. 33 –36.
4. N. Murray and K. Riordan, "Evaluation of automatic explosive detection systems," in Security Technology, 1995. Proceedings. Institute of Electrical and Electronics Engineers 29th Annual 1995 International Carnahan Conference on, oct 1995, pp. 175 –179.
5. H. Strecker, "Automatic detection of explosives in airline baggage using elastic X-ray scatter," in Medicamundi, vol. 42, jul. 1998, pp. 30–33.
6. A. Wales, T. Halbherr, and A. Schwaninger, "Using speed measures to predict performance in X-ray luggage screening tasks," in Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on, oct. 2009, pp. 212 –215.
7. J. Chan, P. Evans, and X. Wang, "Enhanced color coding scheme for kinetic depth effect X-ray (KDEX) imaging," in Security Technology (ICCST), 2010 IEEE International Carnahan Conference on, oct. 2010, pp. 155 –160.
8. M. Singh and S. Singh, "Optimizing image enhancement for screening luggage at airports," in Computational Intelligence for Homeland Security and Personal Safety, 2005. CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on, 31 2005-april 1 2005, pp. 131 –136.
9. C. Oertel and P. Bock, "Identification of objects-of-interest in X-Ray images," in Applied Imagery and Pattern Recognition Workshop, 2006. AIPR 2006. 35th IEEE, oct. 2006, p. 17.
10. D. Liu and Z. Wang, "A united classification system of X-ray image based on fuzzy rule and neural networks," in Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on, vol. 1, nov. 2008, pp. 717 –722.
11. S. Nercessian, K. Panetta, and S. Agaian, "Automatic detection of potential threat objects in X-ray luggage scan images," in Technologies for Homeland Security, 2008 IEEE Conference on, may 2008, pp. 504–509.

12. O. Abusaeeda, J. Evans, D. D., and J. Chan, “View synthesis of KDEX imagery for 3D security X-ray imaging,” in Proc. 4th International Conference on Imaging for Crime Detection and Prevention (ICDP-2011), 2011.
13. V. Riffo and D. Mery, “Active X-ray testing of complex objects,” Insight, vol. 54, no. 1, pp. 28–35, 2012.
14. D. Mery, “Automated detection in complex objects using a tracking algorithm in multiple X-ray views,” in Proceedings of the 8th IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS 2011), in Conjunction with CVPR 2011, Colorado Springs, 2011, pp. 41–48.
15. R. I. Hartley and A. Zisserman, Multiple view geometry in computer vision, 2nd ed. Cambridge University Press, 2003.
16. D. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
17. J. Sivic and A. Zisserman, “Efficient visual search of videos cast as text retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 591–605, 2009.
18. P. Viola and M. Jones, “Robust real-time object detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
19. A. Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms (http://www.vlfeat.org/),” 2008.
20. D. Mery, “BALU: A toolbox Matlab for computer vision, pattern recognition and image processing (http://dmery.ing.puc.cl/index.php/balu),” 2011.