Journal of Nondestructive Evaluation (2019) 38:66
https://doi.org/10.1007/s10921-019-0602-9

Handgun Detection in Single-Spectrum Multiple X-ray Views Based on 3D Object Recognition

Vladimir Riffo[1] · **Ivan Godoy**[1] · **Domingo Mery**[2]

Received: 27 November 2018 / Accepted: 1 July 2019 / Published online: 6 July 2019

Abstract
In the last years, many computer vision algorithms have been proposed for baggage inspection using X-ray images. In these approaches, the idea is to detect automatically threat objects. In baggage inspection, however, a single view is insufficient because there could be occluded parts or intricate projections that cannot be observed with a single view. In order to avoid a misinterpretation based on a single view, we propose the use of single-spectrum multiple X-ray views. Our approach computes a 3D reconstruction using Space Carving, a method that reconstructs a 3D object from its 2D silhouettes (that have been segmented using Geodesic Active Contours). The detection is performed by analyzing 3D features (obtained from the 3D reconstruction). Instead of dual-energy, that is typically used in baggage inspection to analyze the material of the reconstructed objects, we propose simply to use a single-spectrum X-ray system for the detection of threat objects that can be recognized by analyzing the shape, such as handguns. The approach has been successfully tested on X-ray images of travel-bags that contain handguns. In the evaluation of our method, we have used sequences of X-ray images for the 3D reconstruction of objects inside travel-bags, where each sequence consists of 90 X-ray images, we obtained 0.97 in both recall and precision. We strongly believe that it is possible to design an automated aid for the human inspection task using these computer vision algorithms.

Keywords X-ray testing · Threat objects detection · X-ray baggage security · 3D reconstruction applications

1 Introduction

In recent years, X-ray screening systems have been used to safeguard environments in which access control is of paramount importance [1]. Security checkpoints have been placed at the entrances to many public places such as airports, government buildings, stadiums and large event venues to detect (e.g., to detect handguns and explosives) [2] as shown in Fig. 1. 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, presenting an unrecognizable view [3]. In baggage inspection, a single view of a test object is insufficient because there could be occluded parts or intricate projections that cannot be observed with a single view. It is well known, that multiple view X-ray inspection leads to a higher detection performance of prohibited items in difficult conditions [4]. For this reason, multiple view analysis is a powerful tool that can be used in X-ray testing, especially in the inspection of complex objects in cases where certain items are very difficult to recognize using a single viewpoint.

In our work, we propose to use single-spectrum multiple X-ray views in order to avoid a misinterpretation based on a single view. After multiple views are acquired, our approach computes a 3D reconstruction of the objects that are present inside of the test object. For this end, we use a reconstruction approach based on *Space Carving* [5,6], a method that reconstructs a 3D object from its 2D silhouettes. The 2D segmentation in our approach has been performed using *Geodesic Active Contours* [7]. The detection is performed by analyzing 3D features obtained from the 3D reconstruction. Instead of dual-energy, that is typically used in baggage inspection to

✉ Vladimir Riffo
vladimir.riffo@uda.cl
http://www.diicc.uda.cl/profesores/vriffo/index.html

Domingo Mery
domingo.mery@uc.cl
http://dmery.ing.puc.cl

[1] Departamento de Ingeniería Informática y Ciencias de la Computación, Universidad de Atacama, Copiapó, Chile

[2] Department of Computer Science, Pontificia Universidad Católica de Chile, Santiago, Chile

Fig. 1 Baggage inspection and typical problems in recognition of a *gun*: *a* occlusion, *b* self-occlusion, *c* noise, *d* wrong acquisition, *e* variability in different poses

analyze the material of the reconstructed objects, we propose simply to use a single-spectrum X-ray system for the detection of threat objects that can be recognized by analyzing the shape, such as handguns. The approach has been successfully tested on X-ray images that contain handguns. It is worth mentioning that our method can only distinguish textures and shapes (and not materials). In our experiments, we did not test the proposed method on fake (plastic) handguns. Probably, our algorithm could detect these fake objects as threat object. In this sense, an algorithm based on dual-energy could achieve more accurate results, because it can discriminate plastic materials, however, in order to ensure the security, we believe that a human inspector would like to double-check everything that seems like a threat object independently of its material.

This paper attempts to make a contribution to the field of object recognition in X-ray testing by evaluating an approach based on 3D reconstruction of single-spectrum X-ray images. We strongly believe that it is possible to design an automated aid for the human inspection task using these kind of computer vision algorithms. Thus, human operators could achieve a better performance when taking into account potential threat objects detected by a computer vision algorithm [8].

The rest of the paper is organized as follows: In Sect. 2 a literature review is presented. In Sect. 3, the computer vision methods used in our experiments are briefly explained. In Sect. 4, the experimental results are presented. Section 5 concludes the paper.

2 State of the Art

Baggage inspection using X-ray screening is a priority task that reduces the risk of crime, terrorist attacks and propagation of pests and diseases [1]. Security and safety screening with X-ray scanners has become an important process in public spaces and at border checkpoints [2]. 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 unrecognizable view. Manual detection of threat items by human inspectors is extremely demanding [3]. 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) takes a great deal of concentration. In addition, human inspectors receive only minimal technological support. Furthermore, during rush hour, they only have a few seconds to decide whether a bag contains any threat item or not [9]. 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% [10]. 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 of different specimens [11].

There are some contributions in computer vision for X-ray testing such as applications on inspection of castings, welds, food, cargos and baggage screening [12]. 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 summarized as follows: some approaches attempt to recognize objects using a single view of X-ray images of single-spectrum (e.g., the adapted implicit shape model based on visual codebooks [13], adaptive sparse representations [14], models based on logarithmic X-ray images [15,16]) and dual-energy X-ray images (e.g., Gabor texture features [17], bag of words based [18–20] and pseudo-color, texture, edge and shape features [21,22]). More complex approaches that deal with multiple X-ray images have been developed as well. For the recognition of regular objects from X-ray images of single-spectrum, methods like data association [23,24] and active vision [25,26], where a second-best view is estimated, have been explored. In the case of dual-energy imaging [27, 28], visual vocabularies and SVM classifiers have been used, as shown in Reference [29].

In recent years, models based on *Deep Learning* have been successfully used in image and video recognition [30,31]. The key idea is to replace *handcrafted* features with features that are *learned* efficiently using a hierarchical feature extraction approach. There are several deep architectures such as deep neural networks, convolutional neural networks, energy based models, Boltzmann machines, deep belief networks, deep residual learning among others [31,32]. Convolutional neural networks (CNN) has been used in recognition of threat objects in X-ray images (see, for example, in X-ray images of single-spectrum [11] and in dual-energy images [33–35]).

3D object recognition from 2D images is a very complex task in computer vision in general, not only with X-ray images but also with conventional photographic images, given the infinite number of viewpoints, different acquisition conditions, and objects that are deformable, occluded or embedded in clutter [36]. In certain cases, automated recognition is possible through the use of approaches focused on obtaining highly discriminative and local invariant features related to lighting conditions and local geometric constraints (see, for example [37] for a good review and evaluation of descriptors including the well-known SIFT [38] and SURF [39] features) or texture features (see for example [40]). A test object can be recognized by matching its invariant features to the features of a model. Over the past decade, many approaches have been proposed in order to solve the problem of 3D object recognition. Certain approaches focus on learning new features from a set of representative images (see, for example visual vocabularies [41], implicit shape models [42], mid-level features [43], sparse representations [44], and hierarchical kernel descriptors [45]). For instance, fisher vectors [46] and vector of locally aggregated descriptors (VLAD) [47] on SIFT features has been used successfully in recognition problems. In addition, sparse representation has been widely used in computer vision [48,49]. In many computer vision applications, under the assumption that natural images can be represented using sparse decomposition, state of the art results have been significantly improved. However, these methods may fail when the learned features cannot provide a good representation of viewpoints that have not been considered in the representative images. Additionally, some approaches include multiple view models (see, for example, an interconnection of single-view codebooks across multiple views [50], a learned dense multiple view representation by pose estimation and synthesis [51], a model learned iteratively from an initial set of matches [52], a model learned by collecting viewpoint invariant parts [53], 3D representations using synthetic 3D models [54], and a tracking-by-detection approach [55]). These methods may fail, however, when objects have large intra-class variation.

Progress also has been made in the area of computed tomography (CT). For example, in order to improve the quality of CT images, metal artifact reduction and de-noising [56] techniques were suggested. Many methods based on 3D features for 3D object recognition have been developed [57]. See for example, RIFT and SIFT descriptors [58], 3D visual cortex modeling 3D Zernike descriptors and histogram of shape index [59]. There are contributions using known recognition techniques (see, for example, bag of words [60] and random forest [61]) as well.

In order to address the problem of 3D recognition from 2D images, 3D image acquisition and reconstruction techniques (3D scanners) have been developed along with approaches for reconstructing those images by processing multiple views of a single scene. This family of approaches uses 3D data (points, meshes or CAD models) for 3D object category classification. In such cases, the reconstructed 3D object serves as a query and is matched against the shape of a collection of 3D objects [62]. Full reviews of the different approaches that comprise the state of the art can be consulted in References [63–65]. Following a practical paradigm, the results of benchmarks that evaluate the effectiveness of the existing methods in different contexts also are available (e.g., SHREC, shape retrieval contest [66]). In these benchmarks, several detectors and descriptors were evaluated with promising results (e.g., Harris 3D [67], Mesh-HoG [68], scale invariant spin image [69], center-symmetric local binary pattern (CSLBP) [70], 3D shape context (3DSC) [71], Signature of Histograms of OrienTations (SHOT) [72], unique shape context (USC) [73] and fast point feature histogram (FPFH) [74]).

As we can see, the progress in automated baggage inspection is modest and very limited compared to what is needed because X-ray screening systems are still being manipulated by human inspectors. Automated recognition is far from perfect given that the appearance of the object of interest can become extremely difficult to comprehend due to problems of (self-) occlusion, noise, acquisition and clutter, among others (as illustrated in Fig. 1).

3 Proposed Method

In this Section, we present the proposed method following Fig. 2. The block diagram corresponds to a pattern recognition schema: (i) In training stage, we have training images that are processed in order to build a training database, in this case a set of features that characterize the target object (handguns) using descriptors extracted from the 3D reconstructed object. (ii) In testing stage, we perform the 3D reconstruction of objects from the testing images in order to extract the same 3D descriptors used in training stage. The detection is performed by matching the testing descriptors with the training descriptors.

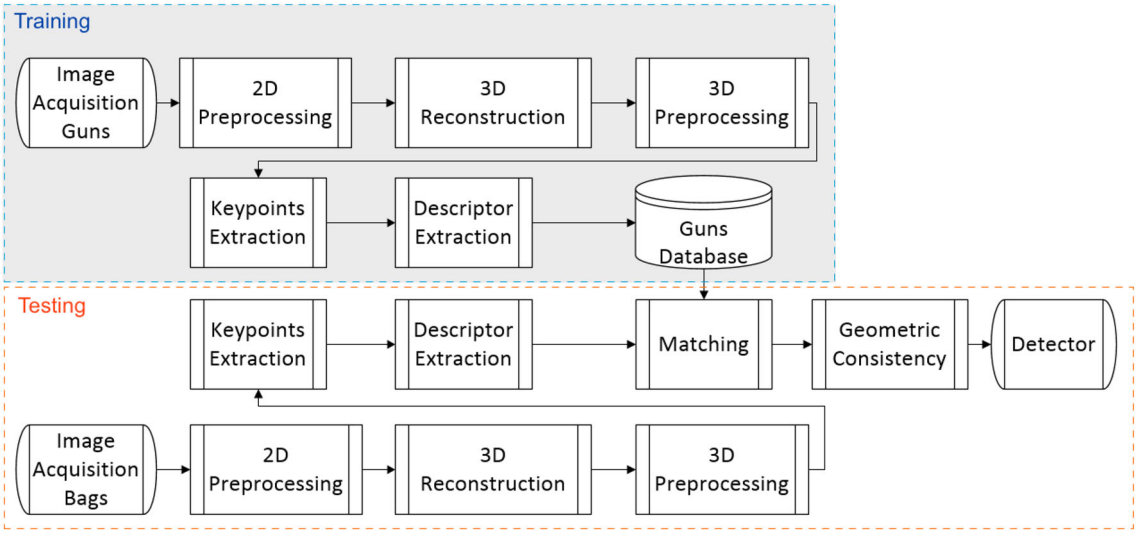

Fig. 2 Block diagram of the proposed method

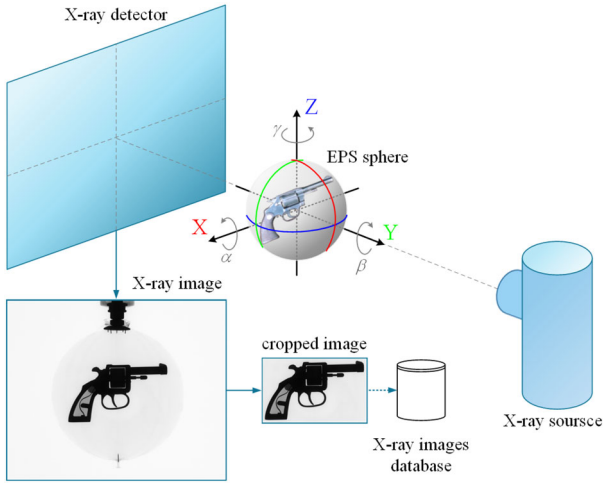

Fig. 3 Acquisition system of X-ray images for characterization of target objects (see acquired X-ray images in Fig. 4)

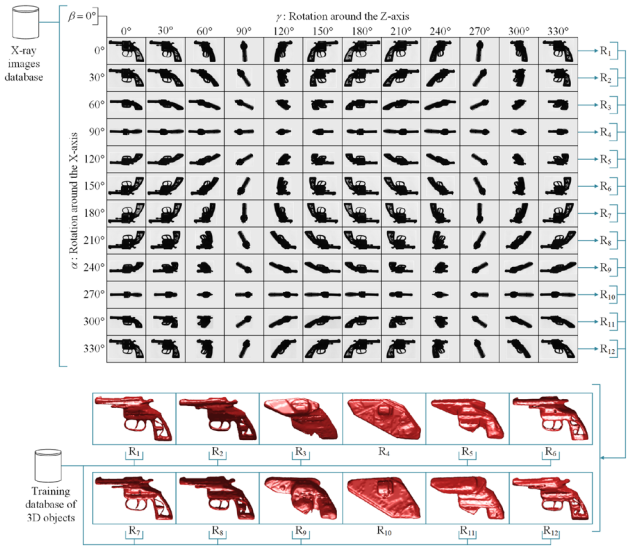

Fig. 4 Set of X-ray images of a target object (handgun) acquired using different angles α, and γ for $\beta = 0°$: (Top) 2D X-ray images using setup of Fig. 3. (Bottom) 3D reconstruction using algorithm of Sect. 3.1.2 (3D reconstructions R4 and R10 are not included in Training database because they are not sufficiently representative)

3.1 Training

In this Section, we describe the Training Stage as illustrated in Fig. 2.

3.1.1 Image Acquisition

In order to acquire representative single-spectrum X-ray images of a target object in different poses, it is necessary to implement an acquisition system that can acquire X-ray images from different points of view, as shown in Fig. 3 for a handgun. The object should be located inside a sphere of expanded polystyrene (EPS). We used a sphere of EPS due to its low X-ray absorption coefficient.

The proposed system allows users to acquire images of an object in many poses by modifying the rotation angles; α, β and γ, associated with each axis, X, Y and Z of the sphere, respectively. All of the images of the handgun are shown in Fig. 4—Top for $\beta = 0°$. The database includes the images acquired in the values of the following angles: $\alpha = 0°, 30°, \ldots, 330°$, $\gamma = 0°, 30°, \ldots, 330°$, and $\beta = 0°, 30°, 60°, 90°$.

3.1.2 3D Reconstruction

In our work, the 3D reconstruction is performed using the well-known approach *Space Carving* [5,6]. For this end, we need a geometric model that relates the 3D coordinates of an object in 3D space (X, Y, Z) to the 2D coordinates of the object viewed in the image of our X-ray imaging system (u, v). The model is obtained using a calibration approach [12]. With this model, we can obtain all points in 3D space that are projected into a unique 2D point in the 2D image. Namely, all 3D points belong to a straight line. In addition, we need a set of binary 2D images, i.e., foreground and background, of the object to be reconstructed acquired form different points of views as illustrated in Fig. 5. In our example, for each image i, we have a binary image i segmented from the X-ray image, and parameters i that are used to define the corresponding model $(X, Y, Z) \rightarrow (u, v)$ for position i.

In Space Carving, the key idea is to model a 'virtual sculpture' of an array of voxels V, in which for each view i we remove from our array V those voxels that produce background pixels in the binary image i. Thus, we start with a voxel array V in a cube that encloses the 3D object to be reconstructed. Space Carving proceeds to remove iteratively (i.e., 'carving') portions of that volume, matching a set of segmented images, representing silhouettes, until it converges in the scene to reconstruct [75]. Using this representation, at each iteration, the coordinates of the world of each voxel (X, Y, Z) are transformed into coordinates of the image (u, v), by means of the 3D $\rightarrow$ 2D model. Finally, it checks if the pixel (u, v) corresponds to a background pixel, resulting in a retention or removal of the voxel from which the pixel was obtained respectively. When all the iterations are completed, only valid voxels will be retained [76]. Finally, in a post-processing step, isolated voxels can be removed using morphological operations.

The geometric model (see Fig. 6) that projects a 3D point (X, Y, Z) into a 2D pixel (u, v) is given in homogeneous coordinates by:

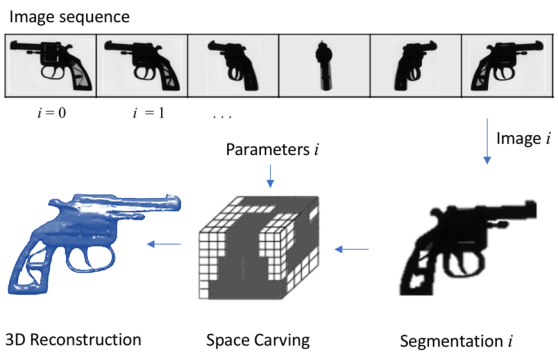

Fig. 5 General overview of the 3D reconstruction based on Space Carving

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P}_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where, λ is a scale factor and $\mathbf{P}_i$ is the projection matrix of view i. It consists of a 3×4 element matrix that depends on scale factors, and rotation and translation variables of position i. They can be estimated using a calibration approach [77].

3.1.3 3D Description

Keypoints of the reconstructed object are extracted using [78,79] at locations where invariant features can be defined. 3D descriptors are local or global characteristics that describe a particular object. These descriptors allow us the detection or recognition of objects in 3D environments. Usually, descriptors must have some invariance, such as rotation, scaling, translation, affine transformation, among others. Local descriptors are used for registration, object recognition and categorization of local surfaces. For these applications, each descriptor is associated with a point, describing the geometry around that point. Global descriptors are highly dimensional representations of object geometry, used for object recognition, geometric categorization, and shape reconstruction [80]. The 3D descriptors that were used in this work to characterize the objects are the following:[1]

- *3D shape context (3DSC)* It is a regional shape descriptor that takes into account 3D shape contexts and harmonic shape contexts [71].
- *Signature of Histograms of OrienTations (SHOT)* It is based on a synergy between the design of a repeatable local reference frame and the embedding of an hybrid signature/histogram [72].
- *Unique shape context (USC)* It is based on 3DSC that uses a unique local reference frame to improve the accuracy [73].
- *Fast point feature histogram (FPFH)* It is a faster version of PFH, which extracts pose-invariant local features at a point based on the combination of geometrical relation with its neighbors [74].

3.2 Testing

In this Section, we describe the Testing Stage as illustrated in Fig. 2.

[1] They were taken from the Point Cloud Library (PCL), an open-source library for 3D computer vision [80].

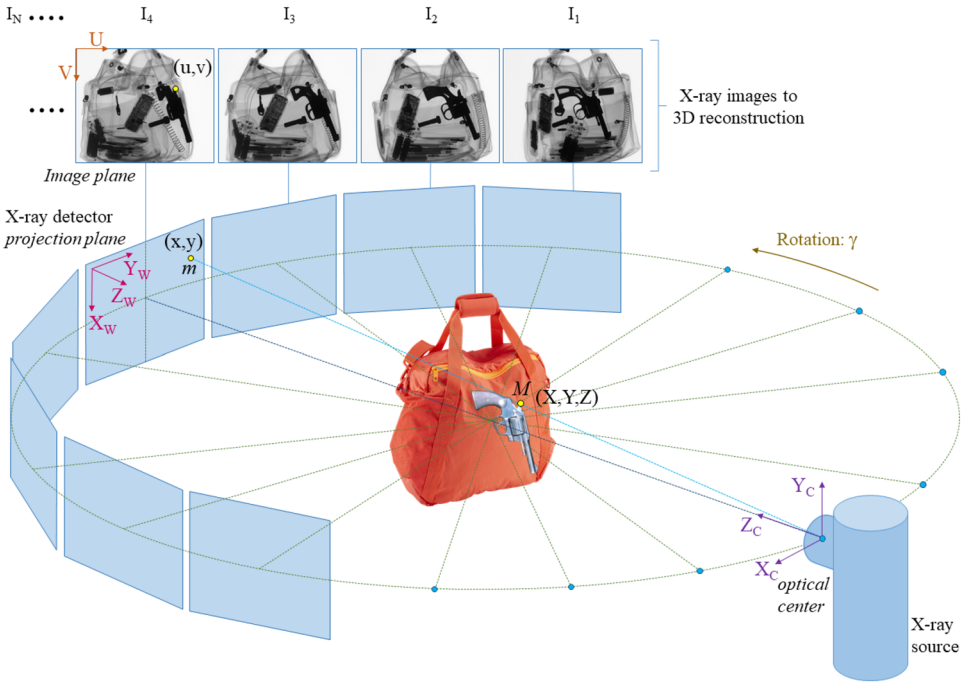

Fig. 6 Image acquisition in the testing stage

Fig. 7 Segmentation: original X-ray image, geodesic contours and binary image

3.2.1 Image Acquisition

The acquisition of single-spectrum X-ray images in the testing stage is illustrated in Fig. 6. It consists of X-ray images taken by rotating the vertical axis. In our case, we use perspectives between 0° and 178° by rotating in 2°, i.e., we acquired 90 views from the testing object.

3.2.2 Image Segmentation

The 2D segmentation in our approach has been performed using *Geodesic Active Contours* [7]. We use for this end the implementation of the function `contour` of Matlab[2] that computes 7 contours. In our work, we select the three darkest contours because they could segment all objects of interest. An example is shown in Fig. 7.

[2] See www.mathworks.com.

3.2.3 3D Reconstruction and Description

The 3D reconstruction is performed following the same method explained in Sect. 3.1.2. In training stage the X-ray image were acquired from isolated objects (as shown in Fig. 3), whereas in the testing stage the X-ray images are taken from whole travel-bags, that means the objects are not isolated at all. An example of the 3D reconstruction is illustrated in Fig. 9. In addition, we extract the same descriptors of the testing stage (see Sect. 3.1.3).

3.2.4 Matching

The matching testing is simply performed using Euclidean distance of the descriptors. We look for the nearest neighbor in the training database. If the minimal Euclidean distance is smaller than a threshold the descriptors are matched.

3.2.5 Geometric Consistency

In order to increase the robustness of our method, it is necessary to verify the spatial consistency of the matching pairs. The idea here is to check the correctness of the reconstructed objects by filtering the potential corresponding pairs based on geometric constraints [81]. A pair C is defined as two 3D points: s, a keypoint in the 3D reconstructed scene (test object), and m, a keypoint in the 3D model of the training database (training object). Pair $c_i = (p_{i,s}, p_{i,m})$ fulfills the matching condition of similar descriptors as explained

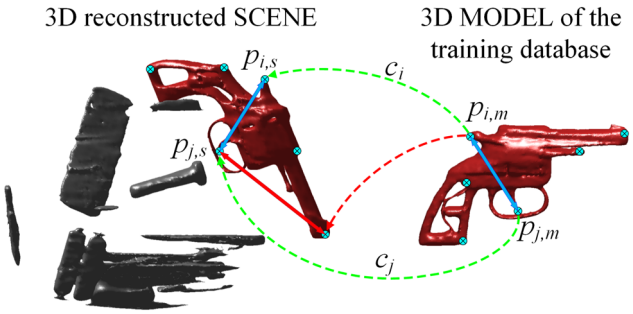

Fig. 8 Geometric consistency between testing object (left) and training object (right). The green and red lines indicate matched pairs that respectively satisfy and do not satisfy the geometric constrain. In this case, the pair of the red line is removed (Color figure online)

in Sect. 3.2.4. In this algorithm, we start with c_i as the best matched pair, and we add a new pair c_j if the distance between the two points in the testing object $\|p_{i,s} - p_{j,s}\|$ is similar to the distance in the training object $\|p_{i,m} - p_{j,m}\|$:

$$|\,\|p_{i,m} - p_{j,m}\| - \|p_{i,s} - p_{j,s}\|\,| < \epsilon. \tag{2}$$

Thus, we guarantee that the distances of added point c_j to previous point c_i in both training and testing object are consistent (in our case $\epsilon = 3$mm). Given a list of corresponding pairs $L = \{c_1, c_2, \ldots, c_n\}$, in our experiments $n = 3$, the grouping procedure for each pair in the list is as follows: (a) use each pair of points as a set of corresponding pairs, (b) for every set, add other pairs if they satisfy (2), (c) repeat for every set, and (d) select the set that has the largest size. An example is illustrated in Fig. 8. See more details in Reference [81].

3.2.6 Detection

A target object, e.g., a handgun, will be detected if a large number of its descriptors are matched with a training object satisfying the geometric constraints. For this end, a threshold is used. Some examples are shown in red in Fig. 9.

4 Experimental Results

In this Section, we show the experiments that we conducted to validate the proposed method.

4.1 Description of the Experiments

In our experiments, we use X-ray images selected from public $\mathbb{GDX}$ray database [82]. $\mathbb{GDX}$ray is a public data-base for X-ray testing with more than 8000 images for baggage inspection.[3]

We use for our experiments two handguns as training objects. For each one, we captured the X-ray images as illustrated in Fig. 3 in order to build the Training Database as shown in Fig. 4. For testing purposes we use 19 travel-bags that contain 0, 1 or 2 handguns. Some examples are shown in Fig. 9—Top. The X-ray images of each travel-bag were captured according to Sect. 3.2.1, i.e., 90 views from 0° and 178° were used for 3D reconstruction. In testing stage, after 3D reconstruction, we follow the method of Sect. 3.2: description, matching, geometric consistency and detection. In our experiments, we tested the performance of the four 3D descriptors: 3DSC, SHOT, USC and FPFH (see Sect. 3.1.3).

The performance is given in terms of Recall–Precision curves, considering all test objects. The variables recall (Re) and precision (Pr) are defined as follows:

$$Re = \frac{\text{TP}}{N_p}, \quad Pr = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{3}$$

where, TP is the number of true positives, FP is the number of false positives and N_p is the total number of target objects to be detected. Ideally, a perfect detection means all existing targets are correctly detected without any false alarms, i.e., $Re = 1$ and $Pr = 1$.

In our experiments, it is worthwhile to mention that the recognition is a challenging task, because the images used in training stage were captured from isolated handguns (in a sphere of EPS) as illustrated in Fig. 3 and explained in Sect. 3.1.1, whereas the images used in testing stage were capture from cluttered travel-bags containing around 18 objects such as cellphones, pencases, knives, keys, screws, bolts, etc. as shown in Fig. 9—Top, and the total number of target objects (handguns) in test database is $N_p = 18$, distributed as follows: (i) five sequences without handguns, (ii) ten sequences with only 1 handgun, and (iii) four sequences with 2 handguns.

We repeat the same test experiments by training with only one of the two handguns following the protocol *leave-one-gun-out*. The best performance is summarized in Table 3 for descriptor 3DSC. As we can see, the performance is very similar.

4.2 Results

Some good results are given in Fig. 9—Bottom, where the effectiveness of the proposed method is evident. However, the method is not perfect because there are some false positives and false negatives (see an Example in Fig. 10). Table 1

[3] The X-ray images included in $\mathbb{GDX}$ray can be used free of charge, for research and educational purposes only. Available at http://dmery.ing.puc.cl/index.php/material/gdxray/.

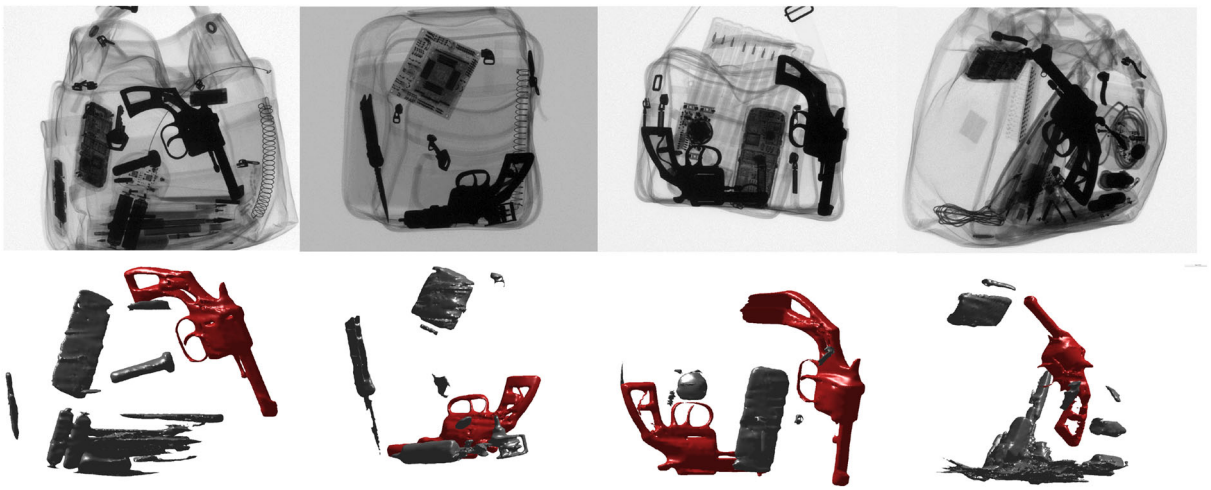

Fig. 9 Example of 3D reconstruction of the segmented objects in different travel-bags. In red we show the detected handguns after the matching and geometric consistency steps (Color figure online)

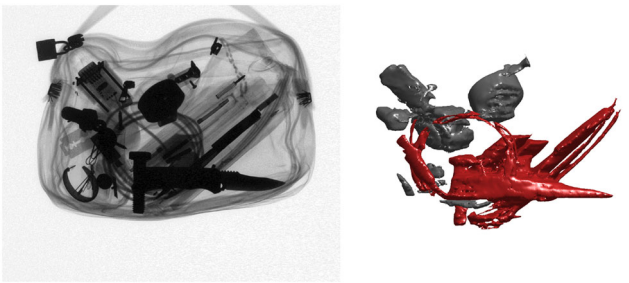

Fig. 10 False positive (Color figure online)

Table 1 Best precision and recall for each descriptor

3D Descriptor	Recall (Re^*)	Precision (Pr^*)
USC	0.89	0.89
FPFH	0.80	0.80
3DSC	0.97	0.97
SHOT	0.91	0.91

and 2 are plotted in Fig. 11, which show the performance achieved by our method, when we modify the threshold value (described by the authors [71–74]) of four different 3D descriptors: USC, FPFH, 3DSC and SHOT. We select the best operation point (Re^*, Pr^*) as the intersection of precision-recall-curve and line $Pr = Re$ as illustrated in Fig. 11. As we can see, in the evaluation of 19 sequences of travel-bags, we obtained 0.97 in both Recall (Re^*) and Precision (Pr^*) for the descriptor 3DSC (Table 3).

4.3 Analysis

In our experiments, we have shown that the proposed method achieved a very good performance obtaining 0.97 in both recall and precision. The method is robust against occlusion in cluttered travel-bags (see for example the fourth image of Fig. 9). The best descriptor was 3DSC, however, it was not the faster (see computational time in Sect. 4.4). Perhaps, in terms of computational time and performance, descriptor SHOT is the most acceptable.

Table 2 Precision and recall for each descriptor

USC		FPFH		3DSC		SHOT	
Re	*Pr*	*Re*	*Pr*	*Re*	*Pr*	*Re*	*Pr*
1.00	0.13	1.00	0.11	1.00	0.09	1.00	0.08
1.00	0.20	1.00	0.17	1.00	0.15	1.00	0.13
1.00	0.26	1.00	0.20	1.00	0.16	1.00	0.16
1.00	0.35	1.00	0.22	1.00	0.17	1.00	0.16
1.00	0.53	1.00	0.25	1.00	0.19	1.00	0.17
0.94	0.71	1.00	0.28	1.00	0.20	1.00	0.17
0.94	0.81	1.00	0.35	1.00	0.21	1.00	0.18
0.89	0.89	1.00	0.38	1.00	0.22	1.00	0.18
0.89	1.00	0.94	0.46	1.00	0.25	1.00	0.18
0.83	1.00	0.94	0.50	1.00	0.35	1.00	0.19
0.83	1.00	0.83	0.52	1.00	0.50	1.00	0.19
0.83	1.00	0.83	0.60	1.00	0.82	1.00	0.19
0.78	1.00	0.83	0.68	1.00	0.95	1.00	0.20
0.78	1.00	0.78	0.88	0.89	1.00	1.00	0.21
0.67	1.00	0.72	0.87	0.72	1.00	1.00	0.25
0.67	1.00	0.67	0.92	0.67	1.00	1.00	0.38
0.61	1.00	0.61	0.92	0.56	1.00	1.00	0.60
0.56	1.00	0.50	0.90	0.33	1.00	0.89	1.00
0.56	1.00	0.28	1.00	0.22	1.00	0.61	1.00
0.56	1.00	0.11	1.00	0.06	1.00	0.50	1.00
0.56	1.00	0.06	1.00	0.06	1.00	0.28	1.00

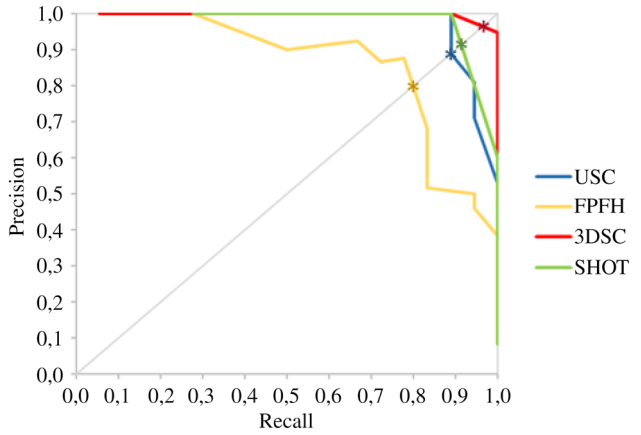

Fig. 11 Precision-recall-curve for each descriptor. The best performance is obtained by descriptor 3DSC

Table 3 Performance using leave-one-gun-out

Handgun	Recall	Precision
A	1.00	1.00
B	0.94	0.94
Average	0.97	0.97

4.4 Implementation

Our method was implemented in Matlab. The computational time for 3D reconstruction for the travel-bags was in average 148 s, and the computational time for future extraction was 2.571, 410, 2.261 and 1.111 s for 3DSC, SHOT, USC and FPFH descriptors respectively. All images and code are available on our webpage.[4]

5 Conclusions

In this work, we focused on recognition of threat objects in baggage inspection using a 3D approach from single-spectrum X-ray images. Our approach computes a 3D reconstruction using Space Carving, a method that reconstructs a 3D object from its 2D silhouettes. In our experiments, the 2D segmentation of the X-ray images was performed using Geodesic Active Contours. The detection is performed by analyzing 3D features obtained from the 3D reconstruction. The approach has been successfully tested on X-ray images of travel-bags with and without handguns. In the evaluation of 19 sequences with 90 X-ray images each, we obtained 0.97 in both precision and recall. We have shown that these preliminary results are promising, because instead of dual-energy, that is typically used in baggage inspection to analyze the material of the reconstructed objects, we simply used X-ray images of single-spectrum for the detection of threat objects that can be recognized by analyzing the shape, such as handguns, achieving a good performance. In addition, we have verified that by means of the method that we propose here, it is possible to detect objects, even in disordered bags, avoiding in this way, the problems of occlusion, typical of 2D images of X-rays. We strongly believe that it is possible to design an automated aid for the human inspection task using these computer vision algorithms. As future work, we would like to test our methodology with other threat objects such as knives and razor blades for example.

[4] See http://dmery.ing.puc.cl/index.php/material/.

Acknowledgements This work was supported in part by DIUDA Grant No. 22277 and No. 22345 from Universidad de Atacama, and in part by Fondecyt Grant No. 1161314 from CONICYT, Chile.

References

1. Zentai, G.: X-ray imaging for homeland security. In: IEEE International Workshop on Imaging Systems and Techniques (IST 2008), pp. 1–6 (2008)
2. Parliament, E.: Aviation security with a special focus on security scanners. European Parliament Resolution (2010/2154(INI)), pp. 1–10 (2012)
3. Schwaninger, A., Bolfing, A., Halbherr, T., Helman, S., Belyavin, A., Hay, L.: 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, pp. 317–324 (2008)
4. von Bastian, C., Schwaninger, A., Michel, S.: Do Multi-View X-ray Systems Improve X-ray Image Interpretation in Airport Security Screening?, vol. 52. GRIN Verlag, Munich (2010)
5. Hemayed, E.E., Farag, A.A.: Object modeling using space carving. In: 2000 International Conference on Image Processing, 2000. Proceedings. Volume 2, IEEE, pp. 760–763 (2000)
6. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. Int. J. Comput. Vis. **38**(3), 199–218 (2000)
7. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. Int. J. Comput. Vis. **22**(1), 61–79 (1997)
8. Hättenschwiler, N., Sterchi, Y., Mendes, M., Schwaninger, A.: Automation in airport security X-ray screening of cabin baggage: examining benefits and possible implementations of automated explosives detection. Appl. Ergonom. **72**, 58–68 (2018)
9. Blalock, G., Kadiyali, V., Simon, D.H.: The impact of post-9/11 airport security measures on the demand for air travel. J. Law Econ. **50**(4), 731–755 (2007)
10. Michel, S., Koller, S., de Ruiter, J., Moerland, R., Hogervorst, M., Schwaninger, A.: Computer-based training increases efficiency in X-ray image interpretation by aviation security screeners. In: 2007 41st Annual IEEE International Carnahan Conference on Security Technology, pp. 201–206 (2007)
11. Mery, D., Svec, E., Arias, M., Riffo, V., Saavedra, J., Banerjee, S.: Modern computer vision techniques for X-ray testing in baggage inspection. IEEE Trans. Syst. Man Cybern. **47**, 682 (2016)
12. Mery, D.: Computer Vision for X-ray Testing. Springer, New York (2015)
13. Riffo, V., Mery, D.: Automated detection of threat objects using adapted implicit shape model. IEEE Trans. Syst. Man Cybern. **46**(4), 472–482 (2016)

14. Mery, D., Svec, E., Arias, M.: Object recognition in X-ray testing using adaptive sparse representations. J. Nondestr. Eval. **35**(3), 1–19 (2016)
15. Heitz, G., Chechik, G.: Object separation in X-ray image sets. 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2093–2100 (2010)
16. Mery, D., Katsaggelos, A.: A logarithmic X-ray imaging model for baggage inspection: simulation and object detection. In: 13th IEEE CVPR Workshop on Perception Beyond the Visible Spectrum (PBVS 2017) (2017)
17. Uroukov, I., Speller, R.: A preliminary approach to intelligent X-ray imaging for baggage inspection at airports. Signal Process. Res. **4**, 1–11 (2015)
18. Turcsany, D., Mouton, A., Breckon, T.P.: Improving feature-based object recognition for X-ray baggage security screening using primed visualwords. In: IEEE International Conference on Industrial Technology (ICIT 2013), pp. 1140–1145 (2013)
19. Baştan, M., Yousefi, M.R., Breuel, T.M.: Visual words on baggage X-ray images. In: Computer Analysis of Images and Patterns, pp. 360–368. Springer (2011)
20. Kundegorski, M.E., Akçay, S., Devereux, M., Mouton, A., Breckon, T.P.: On using feature descriptors as visual words for object detection within X-ray baggage security screening. In: International Conference on Imaging for Crime Detection and Prevention (IET2016) (2016)
21. Zhang, N., Zhu, J.: A study of X-ray machine image local semantic features extraction model based on bag-of-words for airport security. Int. J. Smart Sens. Intell. Syst. **8**(1), 45–64 (2015)
22. Abidi, B.R., Zheng, Y., Gribok, A.V., Abidi, M.A.: Improving weapon detection in single energy X-ray images through pseudocoloring. IEEE Trans. Syst. Man Cybern. C **36**(6), 784–796 (2006)
23. Mery, D.: Inspection of complex objects using multiple-X-ray views. IEEE/ASME Trans. Mechatron. **20**(1), 338–347 (2015)
24. Mery, D., Riffo, V., Zuccar, I., Pieringer, C.: Automated X-ray object recognition using an efficient search algorithm in multiple views. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society, pp. 368–374 (2013)
25. Riffo, V., Mery, D.: Active X-ray testing of complex objects. Insight-Non-Destr. Test. Cond. Monit. **54**(1), 28–35 (2012)
26. Riffo, V., Flores, S., Mery, D.: Threat objects detection in X-ray images using an active vision approach. J. Nondestr. Eval. **36**(3), 44 (2017)
27. Baştan, M., Byeon, W., Breuel, T.M.: Object recognition in multi-view dual X-ray images. In: British Machine Vision Conference BMVC (2013)
28. Baştan, M.: Multi-view object detection in dual-energy X-ray images. Mach. Vis. Appl. **26**(7–8), 1045–1060 (2015)
29. Franzel, T., Schmidt, U., Roth, S.: Object detection in multi-view X-ray images. Pattern Recognition (2012)
30. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
31. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
32. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR arXiv:1512.03385 (2015)
33. Akçay, S., Kundegorski, M.E., Devereux, M., Breckon, T.P.: Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, pp. 1057–1061 (2016)
34. Akcay, S., Kundegorski, M.E., Willcocks, C.G., Breckon, T.P.: Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery. IEEE Trans. Inf. Forensics Secur. **13**(9), 2203–2215 (2018)
35. Akcay, S., Breckon, T.P.: An evaluation of region based object detection strategies within X-ray baggage security imagery. In: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, pp. 1337–1341 (2017)
36. Poggio, T., Edelman, S.: A network that learns to recognize 3D objects. Nature (1990)
37. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
38. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
39. Bay, H.: SURF: speeded up robust features. In: ECCV'06: Proceedings of the 9th European conference on Computer Vision, Berlin, Heidelberg, Springer-Verlag, pp. 404–417 (2006)
40. Shen, L., Jiang, C.J., Liu, G.J.: Satellite objects extraction and classification based on similarity measure. IEEE Trans. Syst. Man Cybern. **46**(8), 1148–1154 (2016)
41. Sivic, J., Zisserman, A.: Efficient visual search of videos cast as text retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 591–606 (2009)
42. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: Workshop on Statistical Learning in Computer Vision, ECCV (2004)
43. Boureau, Y.L., Bach, F., LeCun, Y., Ponce, J.: Learning mid-level features for recognition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2559–2566 (2010)
44. Tošić, I., Frossard, P.: Dictionary learning. Signal Process. Mag. IEEE **28**(2), 27–38 (2011)
45. Bo, L., Lai, K., Ren, X., Fox, D.: Object recognition with hierarchical kernel descriptors. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 1729–1736 (2011)
46. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: BMVC, p. 4 (2013)
47. Arandjelovic, R., Zisserman, A.: All about vlad. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1578–1585 (2013)
48. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., Yan, S.: Sparse representation for computer vision and pattern recognition. Proc. IEEE **98**(6), 1031–1044 (2010)
49. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, pp. 1794–1801 (2009)
50. Thomas, A., Van Gool, L., Tuytelaars, T., Ferrari, V., Leibe, B., Schiele, B.: Towards multi-view object class detection. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society (2006)
51. Su, H., Savarese, S., Sun, M., Fei-Fei, L.: Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, pp. 213–220 (2009)
52. Ferrari, V., Tuytelaars, T., Van Gool, L.: Simultaneous object recognition and segmentation from single or multiple model views. Int. J. Comput. Vis. **67**(2), 159–188 (2006)
53. Savarese, S., Fei-Fei, L.: 3D generic object categorization, localization and pose estimation. In: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, IEEE, pp. 1–8 (2007)
54. Liebelt, J., Schmid, C.: Multi-view object class detection with a 3D geometric model. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1688–1695 (2010)
55. Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E., Van Gool, L.: Robust tracking-by-detection using a detector confidence

particle filter. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, pp. 1515–1522 (2009)
56. Mouton, A., Flitton, G.T., Bizot, S.: An evaluation of image denoising techniques applied to CT baggage screening imagery. In: IEEE International Conference on Industrial Technology (ICIT 2013), IEEE (2013)
57. Mouton, A., Breckon, T.P.: A review of automated image understanding within 3D baggage computed tomography security screening. J. X-ray Sci. Technol. **23**(5), 531–555 (2015)
58. Flitton, G., Breckon, T.P., Megherbi, N.: A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery. Pattern Recogn. **46**(9), 2420–2436 (2013)
59. Megherbi, N., Han, J., Breckon, T.P., Flitton, G.T.: A comparison of classification approaches for threat detection in CT based baggage screening. In: 2012 19th IEEE International Conference on Image Processing (ICIP), IEEE, pp. 3109–3112 (2012)
60. Flitton, G., Mouton, A., Breckon, T.P.: Object classification in 3D baggage security computed tomography imagery using visual codebooks. Pattern Recogn. **48**(8), 2489–2499 (2015)
61. Mouton, A., Breckon, T.P.: Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening. Pattern Recogn. **48**(6), 1961–1978 (2015)
62. Pears, N., Liu, Y., Bunting, P.: 3D Imaging Analysis and Applications. Springer, London (2012)
63. Li, X., Iyengar, S.S.: On computing mapping of 3D objects. ACM Comput. Surv. **47**(2), 1–45 (2015)
64. Yubin, Y., Hui, L., Yao, Z.: Content-based 3D model retrieval: a survey. IEEE Trans. Syst. Man Cybern. C **37**(6), 1081–1098 (2007)
65. Tangelder, T., Veltkamp, R.: A survey of content based 3D shape retrieval methods. Multimed. Tools Appl. **39**, 441–471 (2008)
66. SHREC, H.: Shape Retrieval Contest (SHREC) Home Page. http://www.shrec.net (Visited in year 2018)
67. Sipiran, I., Bustos, B.: A robust 3d interest points detector based on harris operator. In: Proceedings of Eurographics Workshop on 3D Object Retrieval (2010), Eurographics Association, pp. 7–14 (2010)
68. Zaharescu, E. Boyer, K.V., Horaud, R.: Surface feature detection and description with applications to mesh matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009) (2009)
69. Johnson, J.: Spin-images: a representation for 3D surface matching. Ph.D. thesis, Carnegie Mellon University (1997)
70. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with center-symmetric local binary patterns. In: Computer Vision, Graphics and Image Processing, pp. 58–69. Springer (2006)
71. Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: European Conference on Computer Vision, pp. 224–237. Springer (2004)
72. Tombari, F., Salti, S., Stefano, L.D.: Unique signatures of histograms for local surface description. Comput. Vis.—ECCV **2010**(6313), 356–369 (2009)
73. Tombari, F., Salti, S., Di Stefano, L.: Unique shape context for 3d data description. In: Proceedings of the ACM workshop on 3D object retrieval, ACM, pp. 57–62 (2010)
74. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. IEEE International Conference on Robotics and Automation, 2009. ICRA '09 (2009)
75. Kutulakos, K., Seitz, S.: A theory of shape by space carving. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999 (1999)
76. Potmesil, M.: Generating octree models of 3D objects from their silhouettes in a sequence of images. Comput. Vis. Graph. Image Process. **40**(1), 1–29 (1987)
77. Mery, D.: Explicit geometric model of a radioscopic imaging system. NDT E Int. **36**(8), 587–599 (2003)
78. Zhong, Y.: Intrinsic shape signatures: A shape descriptor for 3d object recognition. In: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 689–696 (2009)
79. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. Int. J. Comput. Vis. **89**(2), 348–361 (2010)
80. Aldoma, A., Marton, Z.C., Tombari, F., Wohlkinger, W., Potthast, C., Zeisl, B., Rusu, R.B., Gedikli, S., Vincze, M.: Tutorial: point cloud library: three-dimensional object recognition and 6 dof pose estimation. IEEE Robot. Autom. Mag. **19**(3), 80–91 (2012)
81. Chen, H., Bhanu, B.: 3d free-form object recognition in range images using local surface patches. Pattern Recogn. Lett. **28**(10), 1252–1262 (2007)
82. Mery, D., Riffo, V., Zscherpel, U., Mondragón, G., Lillo, I., Zuccar, I., Lobel, H., Carrasco, M.: GDXray: the database of X-ray images for nondestructive testing. J. Nondestr. Eval. **34**(4), 1–12 (2015)