Automated Detection of Threat Objects Using Adapted Implicit Shape Model

Vladimir Riffo and Domingo Mery, *Member, IEEE*

Abstract—Baggage inspection using X-ray screening is a priority task that reduces the risk of crime and terrorist attacks. Manual detection of threat items is tedious because very few bags actually contain threat items and the process requires a high degree of concentration. An automated solution would be a welcome development in this field. We propose a methodology for automatic detection of threat objects using single X-ray images. Our approach is an adaptation of a methodology originally created for recognizing objects in photographs based on implicit shape models. Our detection method uses a visual vocabulary and an occurrence structure generated from a training dataset that contains representative X-ray images of the threat object to be detected. Our method can be applied to single views of grayscale X-ray images obtained using a single energy acquisition system. We tested the effectiveness of our method for the detection of three different threat objects: 1) razor blades; 2) shuriken (ninja stars); and 3) handguns. The testing dataset for each threat object consisted of 200 X-ray images of bags. The true positive and false positive rates (TPR and FPR) are: (0.99 and 0.02) for razor blades, (0.97 and 0.06) for shuriken, and (0.89 and 0.18) for handguns. If other representative training datasets were utilized, we believe that our methodology could aid in the detection of other kinds of threat objects.

Index Terms—Baggage screening, implicit shape model (ISM), object categorization, object detection, object recognition, threat objects, X-ray testing.

I. Introduction

OVER the past few years, automated inspection systems have been developed to conduct X-ray inspections effectively and efficiently and to perform difficult, tedious, and sometimes dangerous tasks. Reference [1] shows that there are numerous areas in which X-ray testing can be applied. The most important applications are casting inspection, where automated systems are very effective; food quality evaluation, where X-ray imaging is used to describe certain products; welding and cargo, where inspection procedures are semi-automated; and baggage screening, where human inspection is still required. In the latter case, automated X-ray testing remains an open question due to: 1) loss of generality,

which means that approaches developed for one task may not transfer well to another; 2) deficient detection accuracy, which means that there is a fundamental tradeoff between false alarms and missed detections; 3) limited robustness given that requirements for the use of a method are often met for simple structures only; and 4) low adaptiveness in that it may be very difficult to accommodate an automated system to design modifications or different specimens. The objective of this paper is to develop a methodology that automatically recognizes objects in X-ray images. We believe that our method can be a helpful tool in applications such as baggage screening, where recognition of threat objects is a matter of utmost importance.

Baggage inspection using X-ray screening is a priority task that reduces the risk of crime and terrorist attacks. Since the terrorist attack in the U.S. on September 11, 2001, aviation security screening with X-ray scanners has become a very important process at airports. 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 [2]–[5]. Manual detection of threat items by human inspectors is extremely demanding. On the one hand, it is tedious because very few bags actually contain threat items. On the other hand, 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 at airports, they have only a few seconds to decide whether or not a bag contains a threat item. 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% [6].

The detection of objects in X-ray images is a complex task for a human inspector or a computer because the X-ray images are shadow images that correspond to perspective projections of objects. Most X-ray screening systems have been designed to improve the quality of X-ray images through segmentation [7], [8] and pseudo-color algorithms [9]–[11]. Some approaches use automated detection of threat objects with single or multiple views on a single or dual-energy X-ray [1]. Methods based on a bag of visual words [12] have performed well. There is a similar method for detecting firearms [13] using a dual-view approach with pseudo-colored X-ray baggage images (single-energy). Support vector machines and dual-energy multiple views with X-ray images have also been

used [14]. Recently, algorithms based on multiple X-ray views have been reported in the literature. These include a synthesis of kinetic depth effect X-ray images based on scale invariant feature transform (SIFT)-features [15] in order to increase detection performance [16]; X-ray active vision that is able to adjust the viewpoint of the target object in order to obtain better X-ray images (detecting razor blades in different cases) [17]; 3-D-reprojections [18], [19]; and tracking across multiple X-ray views in order to verify the diagnoses performed using a single view [20], [21]. In all of these approaches, detection of threat objects in single views plays a key role.

This research is based on two problems. First, it is very difficult to achieve acceptable recognition rates in the automated detection of threat objects in baggage screening. This is mainly due to the following factors: view difficulty, superposition, and bag complexity. Second, there is a need to support human inspection tasks at screening in order to reduce the probability of human error, mainly due to tiredness and stress. We propose a computer vision strategy that is able to detect threat objects. The system considers them as a collection of independent parts which are logically connected through a star structure. Our approach, the "adapted implicit shape model" (AISM), is based on the well-known "implicit shape model" (ISM) method [22] which was originally developed for recognition of object categories such as cars, people, and animals in photographs. We adapted this methodology in order to detect object categories in single X-ray images that were acquired using a single-energy X-ray system. The proposed method has two main steps. During the training stage, an object category is characterized by estimating a visual vocabulary of the object parts and a measurement of their spatial distribution. In the testing stage, target objects are detected by searching similar visual words and similar spatial distributions. The fundamental difference between the original approach (ISM) and the proposed approach (AISM) is that ISM only uses the occurrence with the highest similarity score as a valid detection. By contrast, AISM merges the occurrences with scores that exceed a certain threshold. In addition, AISM does not require *a priori* knowledge of the number of target objects to be detected. ISM does require such knowledge.

We evaluated the performance of the proposed method in the detection of three threat objects: 1) razors blades; 2) shuriken (ninja stars); and 3) handguns. We show the robustness of the approach for regular shaped objects (razor blade and shuriken) with a high true positive rate (TPR > 0.95) at a low false positive rate (FPR = 0.05). In the case of handgun, we obtained acceptable rates (TPR = 0.89 at FPR = 0.18) due to the irregular shape of the object, which causes scattered detections inside and outside the object. These preliminary results are promising because they establish an approach for the detection of object categories in single X-ray images. We believe that our algorithm could be a helpful tool for human inspectors. In the future, we will improve its performance by incorporating a multiple view analysis and active vision [17], [21].

The rest of this paper is organized as follows. Section II provides a detailed discussion of the proposed method
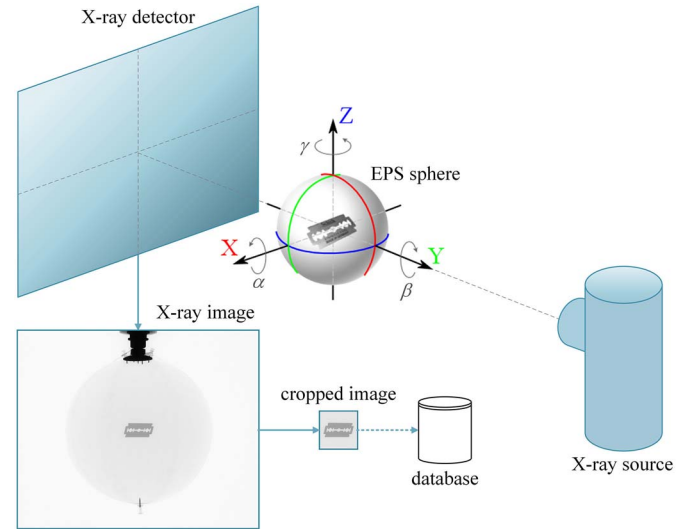

Fig. 1. Acquisition system of X-ray images for characterization of target objects (see acquired X-ray images in Fig. 2).

for the characterization and the detection of threat objects (razor blades, shuriken, and handguns). Section III presents experimental analysis of this paper, which was applied to detection of threat objects in X-ray images. Section IV offers conclusions related to this paper and discusses areas that may be explored in the future.

II. PROPOSED METHOD

In this section, we explain the proposed approach, which can be used to automatically identify objects in X-ray images. We adapted the well-known ISM [22] in order to increase its effectiveness and robustness for the processing of X-ray images. Our proposed method, the AISM, has two main steps: 1) target object characterization and 2) detection. The target object is a threat object to be detected in X-ray images. The steps correspond to training and testing stages, respectively. A more detailed discussion of the steps is provided in the sections that follow.

A. Target Object Characterization

The characterization consists of three steps: 1) training image acquisition: acquisition of representative X-ray images of the threat object; 2) codebook generation: creation of a visual vocabulary using keypoints and local visual descriptors; and 3) occurrence: position estimation of the keypoints related to each visual word of the vocabulary.

1) Training Image Acquisition: A training database of *N* X-ray images is used. In order to acquire representative 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. 1 for a razor blade. 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;
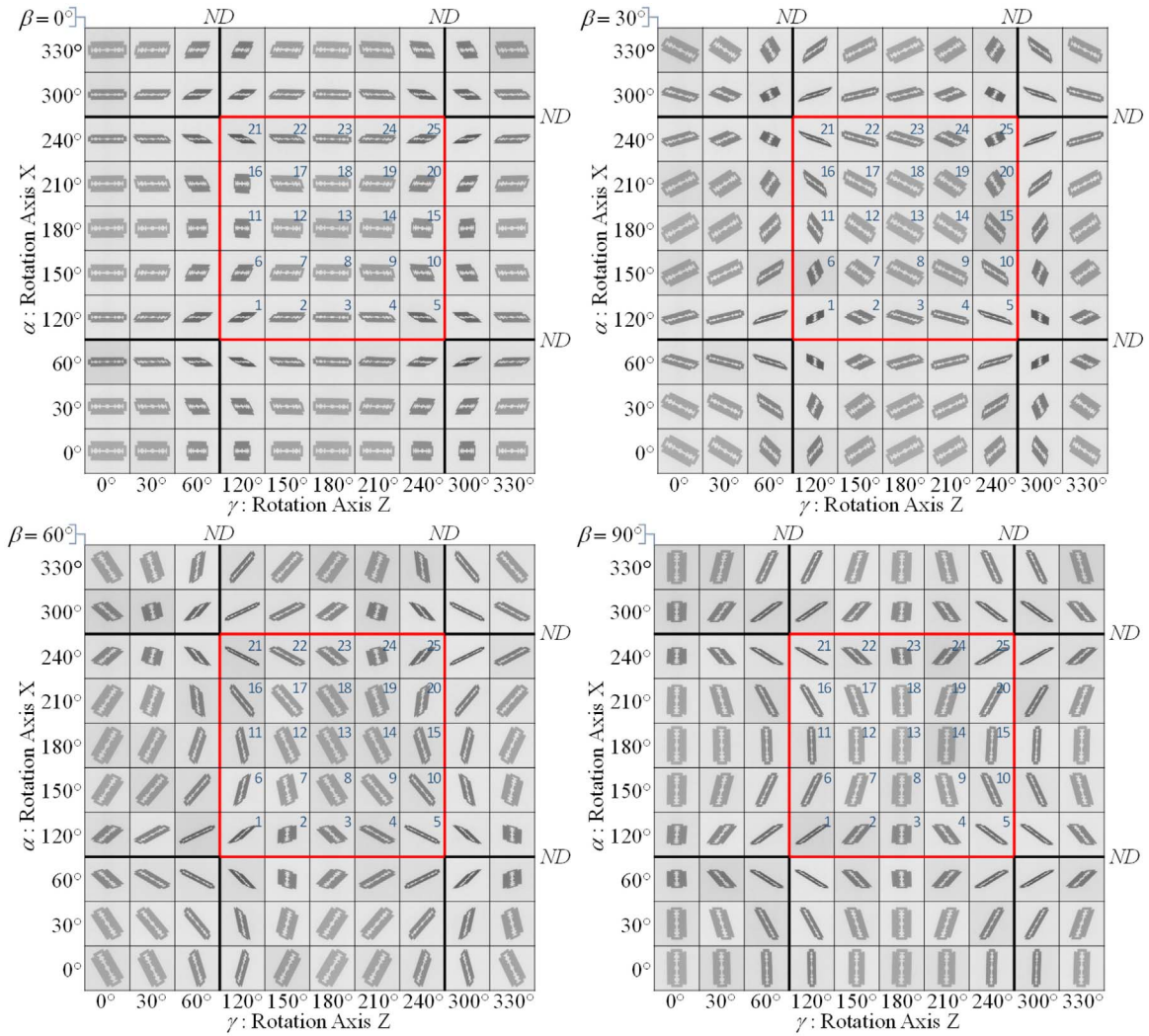
Fig. 2. Set of X-ray images of a target object (razor blade) acquired using different angles α, β, and γ. The useful X-ray images for characterization are enclosed in the red bounding box, i.e., images between α: [120°, 150°, 180°, 210°, 240°] and γ: [120°, 150°, 180°, 210°, 240°]. Each image enclosed in the red bounding box is associated with a pose (1–25).

α, β, and γ, associated with each axis, X, Y, and Z of the sphere, respectively. All of the images of the razor blade are shown in Fig. 2. However, there are quadrants of images that are visually repeated. Therefore, not all images are useful and only a few were stored in the training database. The database includes the images acquired in the values of the following angles (see Fig. 2): $\alpha \in \{120°, 150°, 180°, 210°, 240°\}$, $\gamma \in \{120°, 150°, 180°, 210°, 240°\}$, and $\beta \in \{0°, 30°, 60°, 90°\}$. The self-occluded images or images with positions that do not allow for extraction of discriminating features (e.g., $\alpha = 90°$ or $\gamma = 270°$) have been removed. They are presented in Fig. 2 using the acronym ND (no detection). Note that, the database shown in Fig. 2 was made using a razor blade that has no variability in its object category. For an object with large intraclass variations, the database must include a representative set of images of the object category. This procedure was also performed for the shuriken and handgun.

This process allowed us to obtain a database of N training images. For each training image, the object center (c_x^j, c_y^j) measured in pixels is computed for $j = 1, \ldots, N$.

2) Codebook Generation: In this stage, a target object is represented using a visual vocabulary of parts (category-specific appearance codebook). Keypoints and their local visual descriptors are extracted automatically from all training images of the target object using the well-known SIFT approach [15].[1] A keypoint is a distinguishable point in an image, i.e., it represents a salient image region that can be recognized by changing its viewpoint, orientation, and scale. Thus, SIFT-keypoints are distinguishable image points with high data content in terms of the local change in signal. A keypoint k has a descriptor $\mathbf{f}_k$ and a location $\mathbf{z}_k = (x_k, y_k)$ with respect to (c_x^j, c_y^j), the center of the target object viewed in its corresponding training image j, as shown in Fig. 3.

---

[1]The SIFT descriptors are local features based on the appearance of the object at particular interest points (keypoints). The descriptors are invariant to scale, rotation, lighting, noise, and minor changes in viewpoint. In addition, they are highly distinctive and relatively easy to extract and match against a (large) database of local features. The SIFT descriptor of a keypoint consists of a 128-element vector, which corresponds to a set of 16 gradient oriented histograms of eight bins distributed in a 4×4 grid [15].
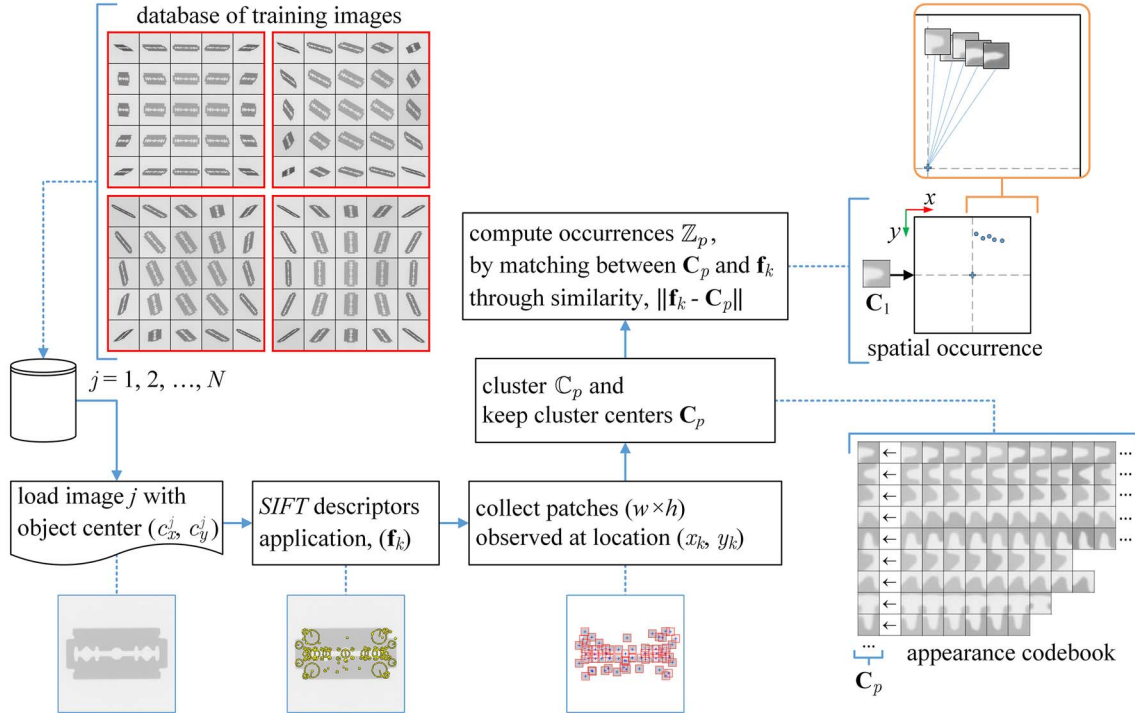
Fig. 3. Target object characterization: training images, codebook generation, and occurrence computation for a razor blade. Here, the keypoints and descriptors are visualized as small patches.

In order to ensure that an effective visual vocabulary will be developed for the various parts of the objects, it is important that the visually similar parts with similar features are clustered to create a codebook of local appearances. In our approach, we use an agglomerative clustering strategy [22]. To initiate the process of agglomerative clustering, each SIFT-feature $\mathbf{f}_k$, i.e., a vector of 128 elements, is assigned to a cluster. The most similar pair of clusters is sequentially merged until a stop criterion is met. The most similar pair of clusters is established by minimizing a distance metric between the clusters. Formally, a cluster p is defined as a set $\mathbb{C}_p$ which contains all SIFT-features $\mathbf{f}_k$ merged into it. As metric distance between two clusters, p and q, we use

$$d\big(\mathbb{C}_p, \mathbb{C}_q\big) = \min\big(d_E\big(\mathbf{f}_p, \mathbf{f}_q\big)\big) \quad \text{for } \mathbf{f}_p \in \mathbb{C}_p, \mathbf{f}_q \in \mathbb{C}_q \quad (1)$$

where $d_E(\mathbf{f}_p, \mathbf{f}_q) = ||\mathbf{f}_p - \mathbf{f}_q||$ is the Euclidean distance between features $\mathbf{f}_p$ and $\mathbf{f}_q$, i.e.,

$$d_E\big(\mathbf{f}_p, \mathbf{f}_q\big) = \sqrt{\sum_{i=1}^{128} \big(f_p(i) - f_q(i)\big)^2} \quad (2)$$

where $\mathbf{f}_p$ and $\mathbf{f}_q$ are SIFT descriptors (with 128 elements each); and $f_p(i)$ and $f_q(i)$ are the ith element of vectors $\mathbf{f}_p$ and $\mathbf{f}_q$, respectively. As suggested in [15] and [23], when comparing histogram-based descriptors—such as SIFT descriptors—the similarity is commonly determined by measuring the Euclidean distance. Thus, the smaller the Euclidean

distance between the descriptors, the more similar the keypoints.[2]

Agglomerative clustering is halted when a certain number of clusters is obtained or when all pair-wise intercluster distances are above a certain threshold. In our implementation, better results were achieved using the former, where the predefined number of clusters (M) for each object category (razor blade, shuriken, and handgun) was set to 400. The center of mass of cluster p is a codeword of our visual vocabulary defined as

$$\mathbf{C}_p = \frac{1}{n_p} \sum_{\mathbf{f}_p \in \mathbb{C}_p} \mathbf{f}_p \quad \text{for } p = 1, \dots, M \quad (3)$$

where n_p is the number of keypoints clustered in set $\mathbb{C}_p$. The codebook is defined as the center of mass of each cluster $\mathbf{C}_p$, and the samples that belong to each cluster $\mathbb{C}_p$, for $p = 1, \dots, M$.

3) Occurrence: In this step, a structure called "occurrence" is computed for each cluster of a target object. The occurrence of cluster p, denoted as set $\mathbb{Z}_p$, for $p = 1, \dots, M$, contains all of the keypoints of the training images whose SIFT descriptors are similar enough to the center of mass of cluster $\mathbf{C}_p$ estimated in (3). Formally, $\mathbb{Z}_p$ is a set of coordinates $\mathbf{z}_k = (x_k, y_k)$ defined as follows:

$$\mathbb{Z}_p = \{\mathbf{z}_k : d_E(\mathbf{f}_k, \mathbf{C}_p) < \theta\} \quad (4)$$

---

[2]Other metrics—such as Mahalanobis distance—can be used as well, but at a higher computational cost. In our preliminary experiments, Euclidean and Mahalanobis metrics achieved similar levels of accuracy. For this reason, in our approach we used Euclidean distance in the similarity measurement.
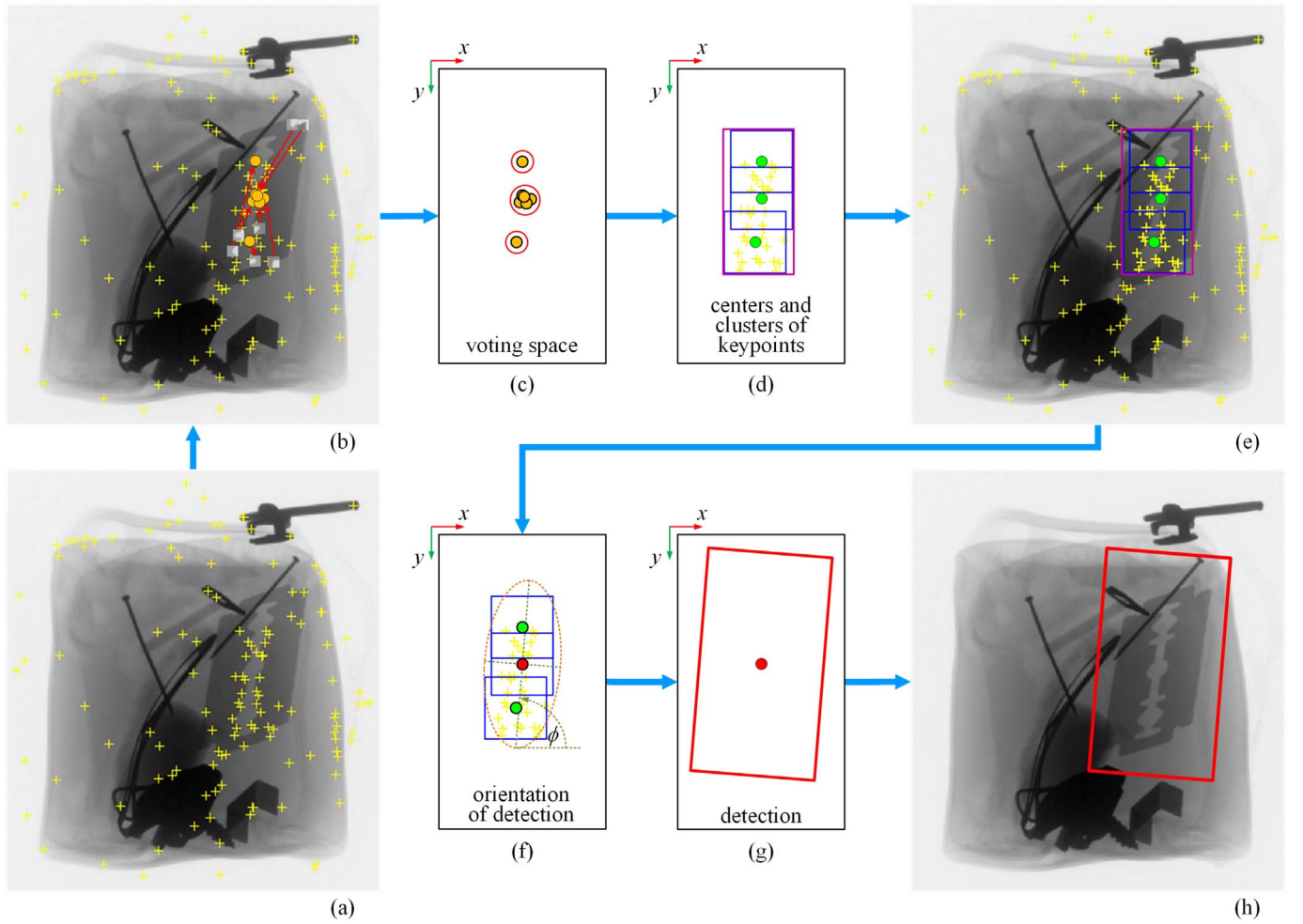
Fig. 4. Detection process for target objects in X-ray images using an ISM approach. (a) Image with useful keypoints only $\hat{\mathbf{f}}$. (b) Image with matched codebook entries and voting space. (c) Detail of voting space, with candidates maxima detected. (d) Detail of voting space, with centers of candidates maxima selected and cluster of keypoints (with blue subwindows W_B). (d) and (e) Merged clusters (with magenta subwindow W_m). (e) Clusters in a previous detection. (f) Ellipse fit to estimate the angle of orientation. (g) and (h) Final detection. (h) Razor blade detected in X-ray image.

where $\mathbf{f}_k$ is the SIFT descriptor of keypoint $\mathbf{z}_k$. As stated in the previous section, the coordinates of the keypoints are defined with respect to the object center.

B. Object Detection

Our method of target object detection in X-ray images has four main stages: 1) feature extraction; 2) matched codebook entries and voting space; 3) merger of candidates detected; and 4) detection. Fig. 4 shows a summary of this process. A detailed explanation of each stage is presented.

1) Features Extraction: We apply an interest point detector to the X-ray test image in order to extract all of the feature keypoints $\mathbf{f}_t$. In general, X-ray images provide a large number of keypoints. We first perform the match between all keypoints $\mathbf{f}_k$ stored in the training database and the keypoints test image $\mathbf{f}_t$ using the expression: $d_E(\mathbf{f}_k, \mathbf{f}_t) < \theta_u$, where θ_u is the minimum distance threshold allowed between $\mathbf{f}_k$ and $\mathbf{f}_t$. Those keypoints that fulfill this expression will be denoted as $\hat{\mathbf{f}}$. Thus, $\hat{\mathbf{f}} \subseteq \mathbf{f}_t$. Therefore, we remove the unnecessary keypoints from the test image and keep only the useful set of keypoints $\hat{\mathbf{f}}$. Fig. 4(a) shows an example of the resulting useful keypoints. We use

the SIFT descriptor as an appearance measure for each interest point.

2) Matched Codebook Entries and Voting Space: This stage of AISM can be described using the probabilistic framework proposed in [22] and [24]. Let $\hat{\mathbf{f}}$ be the descriptor of a useful keypoint observed at location $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ in the X-ray test image. First, each keypoint $\hat{\mathbf{f}}$ is mapped to the closest codeword $\mathbf{C}_p$ of our visual vocabulary learned in "codebook generation" step. The probability that $\hat{\mathbf{f}}$ matches codeword $\mathbf{C}_p$ can be expressed as $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}, \hat{\mathbf{z}})$. Then, a generalized Hough transform algorithm [25] captures the consistent configuration of several visual words. Each matched codeword $\mathbf{C}_p$ generates votes for instances of the object category o_n at different positions $\lambda = (\lambda_x, \lambda_y)$ according to its learned spatial occurrence distribution $\mathcal{P}(o_n, \lambda|\mathbf{C}_p, \hat{\mathbf{z}})$. This probability distribution corresponds to the occurrence locations of keypoints learned during the training process. The previous distribution can be formally expressed by the following marginalization:

$$\mathcal{P}\left(o_n, \lambda|\hat{\mathbf{f}}, \hat{\mathbf{z}}\right) = \sum_p \mathcal{P}\left(o_n, \lambda|\hat{\mathbf{f}}, \mathbf{C}_p, \hat{\mathbf{z}}\right)\mathcal{P}\left(\mathbf{C}_p|\hat{\mathbf{f}}, \hat{\mathbf{z}}\right) \quad (5)$$

for $p = 1, \ldots, M$, where M is the number of codewords in our codebook. Since we have replaced the unknown descriptor $\hat{\mathbf{f}}$ in the X-ray test image with a known interpretation $\mathbf{C}_p$, the first term of (5) can be treated as independent from $\hat{\mathbf{f}}$. In addition, we match descriptors to the codebook independent of their location. Therefore, the equation is reduced to

$$\mathcal{P}\left(o_n, \lambda | \hat{\mathbf{f}}, \hat{\mathbf{z}}\right) = \sum_p \mathcal{P}\left(o_n, \lambda | \mathbf{C}_p, \hat{\mathbf{z}}\right) \mathcal{P}\left(\mathbf{C}_p | \hat{\mathbf{f}}\right) \quad (6)$$

$$= \sum_p \mathcal{P}(\lambda | o_n, \mathbf{C}_p, \hat{\mathbf{z}}) \mathcal{P}(o_n | \mathbf{C}_p, \hat{\mathbf{z}}) \mathcal{P}\left(\mathbf{C}_p | \hat{\mathbf{f}}\right) \quad (7)$$

where $\mathcal{P}(\lambda | o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ is the probabilistic Hough vote for an object position λ given its class label o_n, codeword $\mathbf{C}_p$, and keypoint location $\hat{\mathbf{z}}$. The probability $\mathcal{P}(o_n | \mathbf{C}_p, \hat{\mathbf{z}})$ specifies confidence that the codeword $\mathbf{C}_p$ on keypoint position $\hat{\mathbf{z}}$ is matched on the target category o_n. Finally, $\mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}})$ reflects the probability of the matching between image descriptor $\hat{\mathbf{f}}$ and codeword $\mathbf{C}_p$. If an image feature $\hat{\mathbf{f}}$ found at location $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ matches to a codebook entry that has been observed at position (x_k, y_k) on a training image (stored in the occurrence $\mathbb{Z}_p$), it votes for the following coordinates:

$$x_{\text{vote}} = \hat{x} - x_k \quad \text{and} \quad y_{\text{vote}} = \hat{y} - y_k. \quad (8)$$

Thus, the vote distribution $\mathcal{P}(\lambda | o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ is obtained by casting a vote for each stored observation from the learned occurrence distribution. The ensemble of all such votes together is then used to obtain a nonparametric probability density estimate for the position of the object center.

The score of an object detection hypothesis $h = (o_n, \lambda)$ is obtained by marginalizing all descriptors that contribute to this hypothesis. Considering the hypothesis probability on single-descriptor votes, we arrive at the following equation:

$$\mathcal{P}(o_n, \lambda) = \sum_i \mathcal{P}\left(o_n, \lambda | \hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i\right) \mathcal{P}\left(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i\right) \quad (9)$$

for $i = 1, \ldots, N_f$ where N_f is the number of useful descriptors, $\mathcal{P}(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i)$ is the probability of descriptor $(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i)$ being sampled by the interest point detector for object o_n, and location λ. Nonetheless, we have to tolerate small shape deformations in order to be robust to intraclass variation of the object. We achieve this flexibility by integrating votes over a fixed-size search window $W(\lambda)$ during a mean-shift search [26]. Instead of clustering keypoints with the mean-shift algorithm for any position, we establish that a keypoint can only be joined with another keypoint located near it according to the fixed distance parameter. We set this parameter as 10% of training object size. By basing the decision on single-descriptor votes and assuming a uniform prior for the descriptors, we approximate probability function $\mathcal{P}(o_n, \lambda)$ with the following score:

$$\text{score}(o_n, \lambda) = \sum_i \sum_{\lambda_j \in W(\lambda)} \mathcal{P}\left(o_n, \lambda_j | \hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i\right). \quad (10)$$

We require each sampled descriptor to have the same *a priori* weight in order to avoid error bias. Therefore, we must normalize the vote weights such that the $\mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}})$ and $\mathcal{P}(\lambda | o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ total one. Thus, the weight $\mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}})$ is spread uniformly over all valid descriptor interpretations $\mathbf{C}_p$

TABLE I
TESTING PARAMETERS

Target Object	Razor Blade	Shuriken	Handguns
θ_u	50,000	30,000	30,000
θ_B	5	3	1
θ_m	1	11	4
W_B [pixels]	100×100	150×150	70×70
W_F [pixels]	200×360	820×820	$800 \times 1,300$

by setting $\mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}}) = 1/|C^*|$, with $|C^*|$ the number of matching codebook entries. It would also be possible to let the $\mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}})$ distribution reflect the relative matching scores. In addition to the coordinates of the vote $(x_{\text{vote}}, y_{\text{vote}})$, see (8), we compute its weight w, by setting the occurrence weight $\mathcal{P}(o_n, \lambda | \mathbf{C}_p, \hat{\mathbf{z}}) = 1/|\mathbb{Z}_p|$, with $|\mathbb{Z}_p|$ the number of occurrences for each matched cluster. Thus, a vote is cast in the coordinates $(x_{\text{vote}}, y_{\text{vote}})$ with weight $w = \mathcal{P}(o_n, \lambda | \mathbf{C}_p, \hat{\mathbf{z}}) \mathcal{P}(\mathbf{C}_p | \hat{\mathbf{f}})$. The restricted mean-shift process and normalization of votes is represented in Fig. 4(b) and (c) applied on X-ray test image.

3) Merging of Candidates Detected: The original ISM approach defines the candidate with the highest score as a valid detection. However, this is not necessarily true in X-ray images because they have many dark areas which are similar to one another, and the target object can have multiple possible spatial orientations in baggage (similar descriptors with different occurrences). In our case, we found that the candidates with the highest scores are usually located close to each other. We thus propose keeping the centers (s_x, s_y) of each candidate selected whose score is greater than a threshold value θ_s. This threshold value is adjusted during the analysis of detector performance.

The merging process is similar to the procedure that we designed in [17]. We follow two steps.

a) *Clustering:* The centers of each selected candidate $(> \theta_s)$ are stored in $\mathbf{C}_m$. For each candidate in $\mathbf{C}_m$, we define sub-windows W_B that have at least θ_B keypoints of the same pose. Fig. 4(d)–(f) shows the example of blue squares, W_B, and Table I shows the values of θ_B and size of W_B for all target objects. The pose associated with each useful descriptor was stored in the first stage of cleaning unnecessary descriptors. There are 25 valid poses.

b) *Merging:* All subwindows W_B that are connected or overlap will be merged in a new larger subwindow W_m [see magenta rectangle in Fig. 4(d) and (e)].

4) Detection: The final detection is obtained as follows. First, the subwindow W_m that encloses the highest number of candidates (whose centers $(s_x, s_y) \in \mathbf{C}_m$) will be selected if this number is greater than the threshold θ_m. If no subwindow meets this condition, no potential target object is detected. With this threshold value, we validate the subwindow W_m as the final detection. However, the size and orientation of W_m do not necessarily correspond to the size and orientation of the target object detected. Next, we estimate the orientation of the final detection window W_F, by calculating the centroid of the selected candidates enclosed in W_m, through the use of a well-known K-means clustering algorithm [27], which minimizes the distance between each center (s_x, s_y) using a
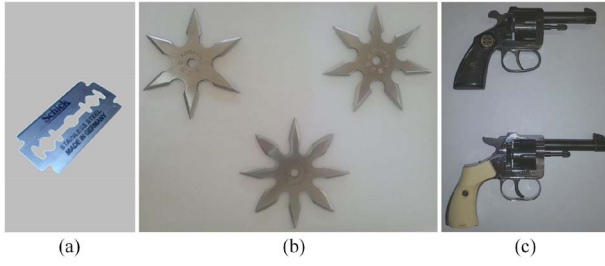
Fig. 5. Threat objects inside of baggage. (a) Razor blade. (b) Shuriken with 6–8 points. (c) Handguns.

Fig. 6. Evaluation criteria for comparing bounding boxes. Interpreting the area of overlap criteria.

single centroid $K = 1$. This is shown in Fig. 4(f) and (g), where the centroid is represented by a red circle. In order to estimate the orientation of the final window W_F, an ellipse is fitted to the keypoints contained in the selected sub-window W_m, as is shown in Fig. 4(f). The angle between the major axis of the ellipse and the horizontal x-axis will be denoted by ϕ. This is precisely the value of the orientation of the final windows W_F (see the size of this window in Table I). This angle estimation is determined for each target object considering the dimensions of the object in the training image with best view, i.e., pose number 13 (see pose 13 in Fig. 2, where $\alpha = 180°$, $\beta = 0°$, and $\gamma = 180°$). An example of oriented final detection of target object is shown in Fig. 4(h).

III. EXPERIMENTAL RESULTS

In this section, we present the experimental evaluation of the proposed method. The X-ray images of our experiments were acquired using a digital X-ray detector (Canon, model CXDI-50G), an X-ray emitter tube (Poskom, model PXM-20BT) and a lead security cabinet to isolate the inspection environment.[3] The size of the X-ray images was 2208×2688 pixels. Our algorithm was tested for the detection of three different threat objects that may be present inside a person's baggage: razor blades, shuriken, and handguns (see examples in Fig. 5). These objects will be the "target objects" to be detected by our algorithm.

A. Training Images

As explained in Section II-A1, X-ray images of each target object must be acquired in representative poses (see Fig. 1 for an example with razor blades). The number of training images used in each experiment is: 100 for razor blades, 100 for shuriken, and 200 for handguns. It is worth noting that the number of training images used in the detection of handguns is higher due to the large interclass variations.

B. Testing Images

We use three sets of testing images (one for each target object). Each testing set consists of 200 X-ray images with the following subsets: 1) 150 X-ray images with one target object each; 2) 30 X-ray images with two target objects each; and 3) 20 X-ray images with no target object. In other words,

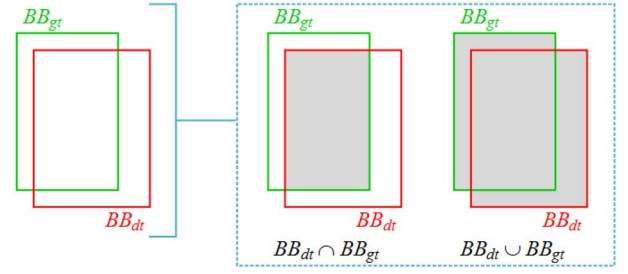[3]All images used in our experiments are available at http://dmery.ing.puc.cl/index.php/material/gdxray.
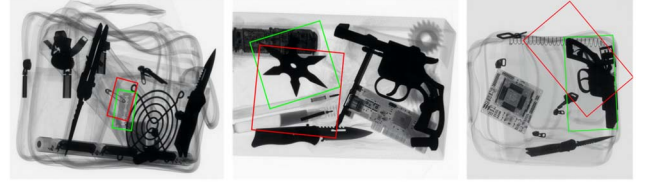

Fig. 7. Detection of target objects (razor blade, shuriken, and handgun) with $\theta_a = 0.4$.

each testing set contains $N_p = 150 + 2 \times 30 = 210$ target objects that correspond to the positive class to be detected. The images were acquired from different suitcases containing several objects in many poses (see examples in Fig. 9). On average, each X-ray image contains 18 objects that are not target objects (pens, CDs, clips, coins, screws, pliers, pins, etc.). As such, each testing set contains $N_n = 18 \times 200 = 3600$ objects that belong to the negative class.

Finally, all target objects are manually annotated by bounding boxes (they correspond to the "ground truth" of our experiments denoted by BB_{gt}). They outline the positions of the target object in the X-ray images (see green bounding boxes in Fig. 9).

C. Evaluation Methodology

The performance of our method is measured using the criteria of quality evaluation of "PASCAL visual object classes challenge" [28], where a detection is considered correct if the normalized area of overlap a_o between the detected bounding box BB_{dt} and the ground truth bounding box BB_{gt} exceeds 0.5, where a_o is defined as follows:

$$a_o = \frac{\text{area}(BB_{\text{dt}} \cap BB_{\text{gt}})}{\text{area}(BB_{\text{dt}} \cup BB_{\text{gt}})} \tag{11}$$

with $BB_{\text{dt}} \cap BB_{\text{gt}}$ the intersection of the detected and ground truth bounding boxes and $BB_{\text{dt}} \cup BB_{\text{gt}}$ their union, as shown in Fig. 6.

Usually, a detection is considered correct if $a_0 \geq \theta_a$ with $\theta_a = 0.5$. However, we included two additional cases: $\theta_a = 0.45$ and $\theta_a = 0.4$. They were taken into account because the orientation of BB_{gt} and BB_{dt} can be very different in certain detections, yielding an intersection area which is considerably smaller than the intersection area that could be obtained for similar orientations (see examples in Fig. 7).

In order to evaluate the performance, we ran our algorithm on each test. The detections obtained were then assessed as follows. We measured the total number of true positives (TPs)

in the entire set, i.e., the number of detections that fulfill $a_0 \geq \theta_a$; and the total number of false positives (FPs), i.e., the number of detections in which $a_0 < \theta_a$. Ideally, TP $= N_p$ (the total number of target objects) and FP $= 0$. As mentioned in Section II-B3, our method includes a threshold (θ_s), which can be tuned to achieve the best tradeoff between correct and false detections. In order to plot a receiver operation characteristic (ROC) curve, we compute the TPR as TPR $=$ TP/N_p, and the FPR as FPR $=$ FP/N_n for different values of θ_s, where N_n is the number of objects that do not correspond to target objects. Ideally, TPR $= 1$ and FPR $= 0$. For each case we calculate the area under the ROC curve in order to measure the performance. It should be 1, and we give TPR$_{0.05}$, which is the value of TPR at FPR $= 0.05$. Additionally, we compute the best operation point (FPR* and TPR*), that is, the point on the ROC curve whose distance to ideal operation point (FPR $= 0$ and TPR $= 1$) is minimal.

D. Parameter Tuning

The proposed method has only six parameters that must be tuned for each object category: thresholds θ_u, θ_B, θ_m, the size of square W_B, and the height and width of rectangle W_F. For tuning purposes we created a "tuning dataset" for each object category, i.e., ten randomly selected X-ray images taken from the 200 X-ray images of the corresponding testing dataset. The parameters were tuned manually using trial and error by maximizing the accuracy on the tuning dataset. A table with values is given in Table I for each object category. Thus, the reported accuracy is obtained using the tuned parameters on the whole testing dataset.

E. Results

TPs, FPs, and ROC curves for the detection of razor blades, shuriken, and handguns are shown in Fig. 8. Table II summarizes the performance achieved in each case. Fig. 9 shows some examples of detection of razor blades, shuriken, and handguns. The ground truth, that is, the target objects to be detected, is presented in green. The detections are shown in red (for TPs) and blue (for FPs).

Based on this evaluation, the detection of razor blades and shuriken is clearly very effective. In both cases, we obtained a high TPR at a very low FPR. The results for the detection of handguns are somewhat lower (it was not possible to obtain a high TPR at a very low FPR). Given that asymmetrical objects have very disjointed occurrences with respect to the real center of the object in our approach, the best results are obtained for symmetrical objects like razor blades and shuriken.

F. Comparison With Other Methods

In this section, we present the results that were obtained by comparing our method to three known methods that can be used in object detection. The baseline methods used here are the following.

1) *SIFT [15]:* We use the single view detection applied in our previous work on active inspection of X-ray images [17].

2) *SURF [29]:* We use the same proposed algorithm AISM with SURF descriptors instead of SIFT descriptors.
3) *ISM [22]:* We use the original ISM approach, which was developed to detect object categories such as cars, people, and animals.

In order to compare the approaches under identical conditions, we used only the first 150 images of the testing set (subset "1") as explained in Section III-B with one razor blade

Razor Blades

θ_s	TP with $a_o \geq$			FP with $a_o \geq$		
	0.5	0.45	0.4	0.5	0.45	0.4
0.108	196	196	196	104	104	104
0.110	190	193	193	103	100	100
0.120	175	177	179	62	60	58
0.130	164	169	170	45	40	39
0.140	140	149	152	39	30	27
0.150	122	127	131	23	18	14
0.160	96	102	104	20	14	12
0.170	80	86	87	16	10	9
0.180	65	69	71	10	6	4
0.190	53	55	57	7	5	3
0.200	39	40	42	5	4	2
0.210	36	36	37	2	2	1
0.220	29	30	31	2	1	0
0.230	27	27	27	0	0	0

Shuriken

θ_s	TP with $a_o \geq$			FP with $a_o \geq$		
	0.5	0.45	0.4	0.5	0.45	0.4
0.0150	189	190	193	106	105	102
0.0165	186	187	190	102	101	98
0.0180	163	165	170	92	90	85
0.0195	134	136	142	79	77	71
0.0210	118	121	124	56	53	50
0.0210	99	101	104	41	39	36
0.0225	67	69	71	32	30	28
0.0240	44	46	46	18	16	16
0.0255	27	27	27	11	11	11
0.0270	19	19	19	6	6	6
0.0285	7	7	7	3	3	3
0.0300	3	3	3	0	0	0

Handguns

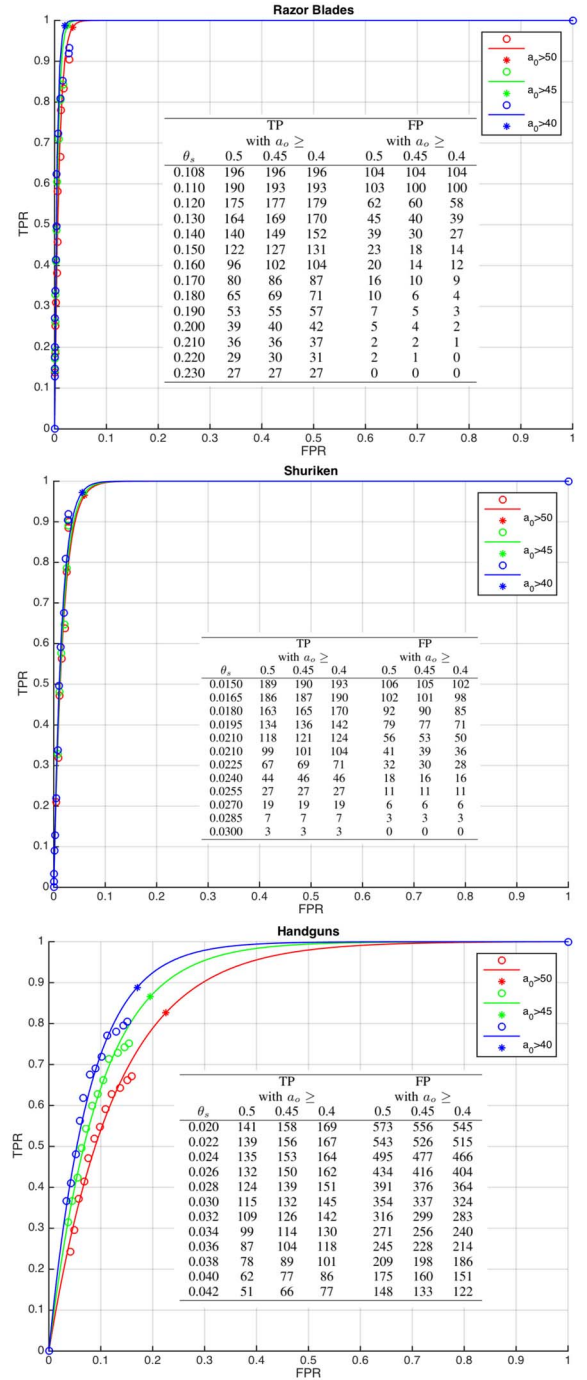| θ_s	TP with $a_o \geq$			FP with $a_o \geq$		
	0.5	0.45	0.4	0.5	0.45	0.4
0.020	141	158	169	573	556	545
0.022	139	156	167	543	526	515
0.024	135	153	164	495	477	466
0.026	132	150	162	434	416	404
0.028	124	139	151	391	376	364
0.030	115	132	145	354	337	324
0.032	109	126	142	316	299	283
0.034	99	114	130	271	256	240
0.036	87	104	118	245	228	214
0.038	78	89	101	209	198	186
0.040	62	77	86	175	160	151
0.042	51	66	77	148	133	122

Fig. 8. ROC curves for the detection of razor blades, shuriken, and handguns for $\theta_a = 0.5, 0.45$, and 0.4. In all cases, the number of positive and negative samples is $N_p = 210$ and $N_n = 3600$, respectively. The measured points are plotted as "o." They are fitted to a curve $y = a(1 - \exp(-\gamma x))$. The best operation point (FPR* and TPR*) is shown as "*."
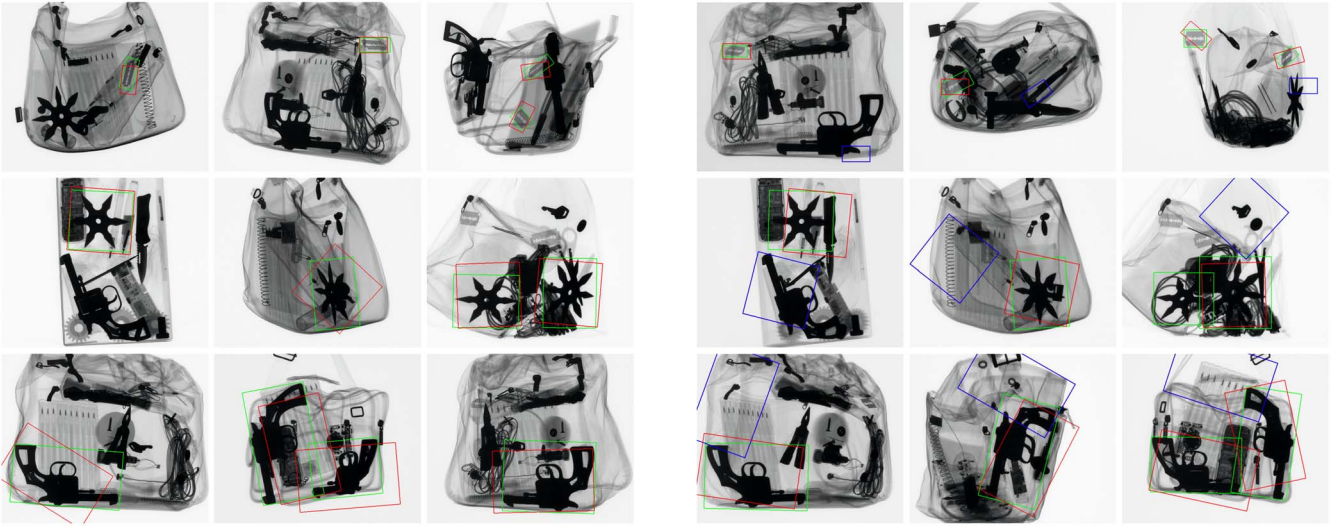
Fig. 9.   Detection of AISM. TPs (left): Examples of test images for which our target object detectors achieved perfect detection results (restricted to $a_o \geq 0.5$). The ground truth annotations BB_{gt} are displayed in green and detections BB_{dt} are displayed in red. X-ray images of the razor blade (first row), shuriken (second row), and handgun (third row) testing set are shown. TPs and FPs (right): Examples of test images for which our target object detectors missed objects or produced false detections. The ground truth annotations are displayed in green. TPs and FPs are displayed in red and blue, respectively. X-ray images of the razor blade (first row), shuriken (second row), and handgun (third row) testing sets are shown.

TABLE II
SUMMARY OF PERFORMANCE OF ASIM

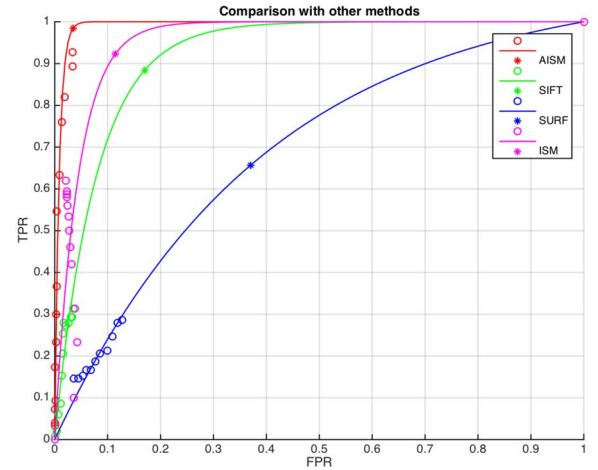| Target Object	Variable	with $a_o \geq$		
		0.5	0.45	0.4
Razor Blade	AUC	0.9915	0.9942	0.9954
	$TPR_{0.05}$	0.9972	0.9998	1.0000
	TPR^*	0.9836	0.9870	0.9876
	FPR^*	0.0350	0.0250	0.0200
Shuriken	AUC	0.9821	0.9832	0.9847
	$TPR_{0.05}$	0.9388	0.9487	0.9621
	TPR^*	0.9650	0.9717	0.9727
	FPR^*	0.0600	0.0600	0.0550
Handgun	AUC	0.8715	0.9029	0.9225
	$TPR_{0.05}$	0.3219	0.4023	0.4754
	TPR^*	0.8261	0.8657	0.8884
	FPR^*	0.2250	0.1950	0.1700

Fig. 10.   ROC curves of our method AISM in comparison with other three known methods. In all cases, the number of positive and negative samples is $N_p = 150$ and $N_n = 2700$, respectively. The measured points are plotted as "o." They are fitted to a curve $y = a(1 - \exp(-\gamma x))$. The best operation point (FPR*,TPR*) is shown as "*."

per image). Thus, the methods were tuned to detect only one target object per image.

The number of TP and FP were computed for each method. The tables are not presented due to space considerations, but the ROC curves are illustrated in Fig. 10 and a summary of the performance of each method is shown in Table III. In order to compare the baseline methods and obtain different values of TP and FP, we had to modify a threshold value for each detector: 1) in our approach (AISM), the activation threshold was the value that keeps candidates detected with higher scores in the image; 2) for our proposed algorithm AISM using SURF descriptors, the activation threshold was the same as in our AISM, i.e., the value that keeps candidates detected with higher scores in the image; 3) in the SIFT-Lowe approach, the activation threshold was the value that keeps only useful SIFT descriptors $\hat{\mathbf{f}}$ in the X-ray image; and 4) in the original ISM approach, the activation threshold was also the value that keeps only useful SIFT descriptors in the X-ray image.

In Fig. 10, we compare the ROC curves of the four methods. Our method has the best performance, achieving TPR = 0.9975 at FPR = 0.05. We show that adapting ISM with SIFT descriptors significantly increases performance, since the recognition accuracies obtained by our method are much higher than those obtained by the other methods.

G. Implementation Details

We used open source libraries like VLFeat [30] for K-means and SIFT to implement the AISM. We use a MATLAB implementation of the Computer Vision System Toolbox for SURF descriptors. The computing time depends on the number of useful descriptors in the image, the spatial distribution and number of occurrences. However, as a reference, the testing

TABLE III
COMPARISON OF ASIM WITH OTHER METHODS

AISM (ours)	AUC	0.9917
	$\text{TPR}_{0.05}$	0.9975
	TPR^*	0.9849
	FPR^*	0.0350
SIFT	AUC	0.9211
	$\text{TPR}_{0.05}$	0.4693
	TPR^*	0.8840
	FPR^*	0.1700
SURF	AUC	0.6162
	$\text{TPR}_{0.05}$	0.1276
	TPR^*	0.6564
	FPR^*	0.3700
ISM	AUC	0.9553
	$\text{TPR}_{0.05}$	0.6734
	TPR^*	0.9237
	FPR^*	0.1150

results for the detection of razor blades were obtained in approximately 35 s per image on an Intel Core i7-3537U CPU @ 2.00 GHz with four cores and memory RAM of 8 GB. The algorithms were implemented in MATLAB R2013b, 64-bit (win64). The code for the MATLAB implementation is available on our webpage.[4]

IV. CONCLUSION

We have presented a new method that can be used to automatically identify certain threat objects in X-ray images. Our approach proposes a training methodology in which a target object is described using several X-ray images from representative points of view. A test X-ray image is then analyzed using that characterization. Our approach, which we have called AISM, is based on ISMs which have been adapted to deal with X-ray images. The main adaptions are as follows: 1) when extracting keypoints from the test image, only useful ones are considered by AISM and 2) in ISM, only the occurrence with the highest similarity score is considered to be a valid detection while AISM merges the occurrences with scores whose values exceed a certain threshold.

Unlike other methods, our approach was tested using very realistic and challenging scenarios in grayscale X-ray images (no pseudo-color). We used single view and single-energy conditions without an image preprocessing algorithm. The bags used in our experiments contained approximately 20 objects, one or two of which were target objects. As our examples show (see Fig. 9), the X-ray images that we used are very complex because the bags were packed full and contained occluded and rotated objects. Experiments were carried out on three different datasets (razor blades, shuriken, and handguns) with 200 X-ray images each.

We show the robustness of the approach for regular shaped objects (razor blade and shuriken) with a high TPR > 0.95 at a low FPR = 0.05 for both objects. In the case of handguns, we obtained acceptable rates (TPR = 0.89 at FPR = 0.18) due to the irregular shape of the object, which caused scattered detections inside and outside of the object. These preliminary results are promising because they establish an approach for the detection of object categories in single X-ray images. In addition,

we compared our algorithm to three other methods that can be used to detect objects. As the ROC curves demonstrate, our proposed method outperforms all of these methods in terms of TPR and FPR.

Our approach was designed to detect only one object category, but we repeated the training and testing strategy in three different object categories in order to show the effectiveness of the proposed method. Thus, several object categories can be detected simultaneously. Using other representative training datasets, our methodology could be used in the detection of other threat objects (e.g., knives). We believe that our algorithm could be a helpful tool for human inspectors.

All training and testing X-ray images used in this research can be downloaded from our webpage.[3] The goal of this database is to share challenging X-ray images with researchers around the world who are working on computer vision and X-ray testing. To facilitate this, we have published the X-ray images for razor blades, shuriken, and handguns as well as the manually annotated bounding boxes for the ground truth and the AISM MATLAB code of AISM.[4]

We suggest that future efforts in this area extend our approach to other threat objects with irregular shapes. We believe that promising results could be achieved in single view detection if we extract SIFT-features only in object regions that are discriminative. Thus, we could avoid for example the description of very dark regions of the objects with high absorption of X-rays and take into account only regions near the contours of the objects (high gradients). Another suggestion is to include this single view detector as the first step in a multiple view strategy in which performance could be improved by considering several single view detectors of the object taken from different points of view.

REFERENCES

[1] D. Mery, "Computer vision technology for X-ray testing," *Insight Non-Destruct. Test. Cond. Monitor.*, vol. 56, no. 3, pp. 147–155, 2014.

[2] G. Zentai, "X-ray imaging for homeland security," in *Proc. IEEE Int. Workshop Imag. Syst. Tech. (IST)*, Crete, Greece, Sep. 2008, pp. 1–6.

[3] 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* (LNCS 5298), A. Holzinger, Ed. Berlin, Germany: Springer, 2008, pp. 419–438.

[4] A. Schwaninger *et al.*, "The impact of image based factors and training on threat detection performance in X-ray screening," in *Proc. 3rd Int. Conf. Res. Air Transp. (ICRAT)*, Fairfax, VA, USA, 2008, pp. 317–324.

[5] S. Michel, S. M. Koller, and A. Schwaninger, "Relationship between level of detection performance and amount of recurrent computer-based training," in *Proc. 42nd Annu. IEEE Int. Carnahan Conf. Security Technol. (ICCST)*, Prague, Czech Republic, 2008, pp. 299–304.

[6] S. Michel *et al.*, "Computer-based training increases efficiency in X-ray image interpretation by aviation security screeners," in *Proc. 41st Annu. IEEE Int. Carnahan Conf. Security Technol.*, Ottawa, ON, Canada, Oct. 2007, pp. 201–206.

[7] N. Megherbi, T. P. Breckon, and G. T. Flitton, "Investigating existing medical CT segmentation techniques within automated baggage and package inspection," *Proc. SPIE Security+ Defence Int. Soc. Opt. Photon.*, vol. 8901, Oct. 2013, Art. ID 89010L.

[8] G. Heitz and G. Chechik, "Object separation in X-ray image sets," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 2093–2100.

[9] B. Abidi, Y. Zheng, A. Gribok, and M. Abidi, "Improving weapon detection in single energy X-ray images through pseudocoloring," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 6, pp. 784–796, Nov. 2006.

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

[10] K. Kase, "Effective use of color in X-ray image enhancement for luggage inspection," M.S. thesis, Univ. Tennessee, Knoxville, TN, USA, 2002.

[11] J. Chan, P. Evans, and X. Wang, "Enhanced color coding scheme for kinetic depth effect X-ray (KDEX) imaging," in *Proc. IEEE Int. Carnahan Conf. Security Technol. (ICCST)*, San Jose, CA, USA, Oct. 2010, pp. 155–160.

[12] M. Baştan, M. Yousefi, and T. Breuel, "Visual words on baggage X-ray images," in *Computer Analysis of Images and Patterns.* Lecture Notes in Computer Science, vol. 6854, P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, and W. Kropatsch, Eds. Berlin, Germany: Springer, 2011, pp. 360–368.

[13] D. Turcsany, A. Mouton, and T. Breckon, "Improving feature-based object recognition for X-ray baggage security screening using primed visualwords," in *Proc. Int. Conf. Ind. Technol. (ICIT)*, Cape Town, South Africa, Feb. 2013, pp. 1140–1145.

[14] M. Baştan, W. Byeon, and T. Breuel, "Object recognition in multi-view dual energy X-ray images," in *Proc. Brit. Mach. Vis. Conf.*, Graz, Austria, 2013, pp. 130.1–130.11.

[15] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[16] O. Abusaeeda, J. P. O. Evans, D. Downes, and J. W. Chan, "View synthesis of KDEX imagery for 3D security X-ray imaging," in *Proc. 4th Int. Conf. Imag. Crime Detect. Prevent. (ICDP)*, London, U.K., 2011, pp. 1–6.

[17] V. Riffo and D. Mery, "Active X-ray testing of complex objects," *Insight Non-Destruct. Test. Cond. Monitor.*, vol. 54, no. 1, pp. 28–35, 2012.

[18] T. Franzel, U. Schmidt, and S. Roth, "Object detection in multi-view X-ray images," in *Pattern Recognition.* Lecture Notes in Computer Science, vol. 7476, A. Pinz, T. Pock, H. Bischof, and F. Leberl, Eds. Berlin, Germany: Springer, 2012, pp. 144–154.

[19] D. Mery, V. Riffo, I. Zuccar, and C. Pieringer, "Automated X-ray object recognition using an efficient search algorithm in multiple views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Portland, OR, USA, Jun. 2013, pp. 368–374.

[20] D. Mery, G. Mondragon, V. Riffo, and I. Zuccar, "Detection of regular objects in baggage using multiple X-ray views," *Insight Non-Destruct. Test. Cond. Monitor.*, vol. 55, no. 1, pp. 16–20, 2013.

[21] D. Mery, "Inspection of complex objects using multiple-X-ray views," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 1, pp. 338–347, Feb. 2015.

[22] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 259–289, 2008.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[24] B. Leibe and B. Schiele, "Interleaved object categorization and segmentation," in *Proc. Brit. Mach. Vis. Conf.*, Norwich, U.K., 2003, pp. 78.1–78.10.

[25] D. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, 1981.

[26] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth mean shift and data-driven scale selection," in *Proc. 8th Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Vancouver, BC, Canada, 2001, pp. 438–445.

[27] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, vol. 1. Berkeley, CA, USA, 1967, pp. 281–297.

[28] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

[29] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[30] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. Int. ACM Conf. Multimedia (ICM)*, Firenze, Italy, 2010, pp. 1469–1472.

Vladimir Riffo was born in Santiago, Chile, in 1971. He received the B.Sc.Eng. degree in electronic engineering from the Universidad de Antofagasta, Antofagasta, Chile, in 1998, and the M.Eng. degree from the Pontificia Universidad Católica de Chile, Santiago, Chile, in 2011, where he is currently pursuing the doctoral degree.

He is an Associate Professor with the Department of Computer Engineering and Computer Science, Universidad de Atacama, Copiapó, Chile. His current research interests include pattern recognition, object detection, and computer vision using multiviews and X-ray testing, as well as ways to combine those approaches.

Prof. Riffo was a recipient of the Ron Halmshaw Award in 2012 and the John Green Award in 2013 from the British Institute for Non-Destructive Testing, which was established for the best papers published in the *Insight Journal on Industrial Radiography*, and the scholarship from the Comisión Nacional de Investigación Científica y Tecnológica. He is a Fellow of the GRIMA, Machine Intelligence Group, Pontificia Universidad Católica de Chile.

Domingo Mery received the M.Sc. degree in electrical engineering from the Technical University of Karlsruhe, Karlsruhe, Germany, in 1992, and the Ph.D. degree with distinction from the Technical University of Berlin, Berlin, Germany, in 2000.

He was a Research Scientist with the Institute for Measurement and Automation Technology, Technical University of Berlin with the collaboration of YXLON X-ray International, Hamburg, Germany. In 2001, he served as an Associate Researcher with the Department of Computer Engineering, Universidad de Santiago, Santiago, Chile. In 2014, he was a Visiting Professor with the University of Notre Dame, Notre Dame, IN, USA. He is currently a Professor with the GRIMA, Machine Intelligence Group, Department of Computer Science, where he served as the Chair from 2005 to 2009, and the Director of Research and Innovation with the School of Engineering, Pontificia Universidad Católica de Chile, Santiago. His current research interests include image processing for fault detection in aluminum castings, X-ray imaging, real-time programming, and computer vision. He has authored 60 technical SCI publications and over 70 conference papers.

Dr. Mery was a recipient of the Ron Halmshaw Award in 2005 and 2012 and the John Green Award in 2013 from the British Institute of Non-Destructive Testing, which was established to recognize the best papers published in the *Insight Journal on Industrial Radiography*, the Best Paper Award from the International Workshop on Biometrics in conjunction with the European Conference on Computer Vision in 2014, and the scholarships from the Konrad Adenauer Foundation and the German Academic Exchange Service. He is the Local Co-Chair of the ICCV2015 (to be held in Santiago, Chile). He served as a General Program Chair of the PSIVT2007, a Program Chair of the PSIVT2009, and a General Co-Chair of the PSIVT2011 (Pacific-Rim Symposium on Image and Video Technology), and the 2007 Ibero-American Congress on Pattern Recognition.