Object Detection in Multi-View X-Ray Images

Thorsten Franzel, Uwe Schmidt, and Stefan Roth

Department of Computer Science, TU Darmstadt

Abstract. Motivated by aiding human operators in the detection of dangerous objects in passenger luggage, such as in airports, we develop an automatic object detection approach for multi-view X-ray image data. We make three main contributions: First, we systematically analyze the appearance variations of objects in X-ray images from inspection systems. We then address these variations by adapting standard appearance-based object detection approaches to the specifics of dual-energy X-ray data and the inspection scenario itself. To that end we reduce projection distortions, extend the feature representation, and address both in-plane and out-of-plane object rotations, which are a key challenge compared to many detection tasks in photographic images. Finally, we propose a novel multi-view (multi-camera) detection approach that combines single-view detections from multiple views and takes advantage of the mutual reinforcement of geometrically consistent hypotheses. While our multi-view approach can be used atop arbitrary single-view detectors, thus also for multi-camera detection in photographic images, we evaluate our method on detecting handguns in carry-on luggage. Our results show significant performance gains from all components.

1 Introduction

Inspection of passenger luggage for dangerous[1] objects is commonplace to increase security in public transportation, especially at airports all around the world (*cf*. [18]). This task is usually carried out by human operators, who search for forbidden objects using X-ray images of the luggage. However, their detection performance can vary substantially as they get tired or distracted. Therefore, computer-based automatic object detection is a promising approach, since it does not suffer from these human limitations. Even though detection of explosives based on material properties is already used in practice, appearance-based detection of objects in X-ray images is not yet common.

In this work, we employ appearance-based object detection approaches that were developed for photographic images, and adapt them to the specific properties of *dual-energy* image data (*cf*. [18]), as acquired by modern X-ray luggage inspection systems. We analyze different sources of object appearance variations that make detecting objects difficult and show how to address them. Specifically, we reduce distortions from the scanner geometry and adapt the feature representation to the specifics of dual-energy X-ray data. We show that single-view detection is significantly influenced by the applied methods. The key challenge, however, stems from the fact that – unlike many object categories in photographic images – dangerous objects may occur in the luggage at any orientation. We first address in-plane rotations using an efficient non-maximum

---

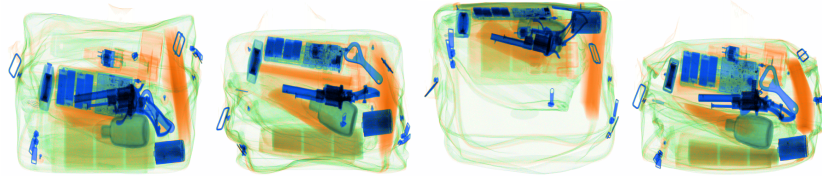[1] http://ec.europa.eu/transport/air/security/doc/info_travellers_hand_luggage.pdf

Fig. 1: **Multi-view X-ray data example.** 4 false color images from dual-energy input.

suppression scheme. Moreover, we specifically focus on the challenge of unfavorable viewpoints due to out-of-plane rotations by exploiting multiple views of the passenger luggage, as provided by modern X-ray systems (Fig. 1). Our main contribution is a multi-view (multi-camera) detection approach that combines independent detections from all available views, which we demonstrate to significantly improve detection performance in comparison to single-view detection.

We evaluate our approach on the task of detecting handguns in carry-on luggage using a challenging large-scale dataset of different bags, some of which contain handguns of various types. Although particularly motivated by the task of screening luggage for dangerous objects in the context of airport security, our multi-view detector is not limited to this scenario, and can be used atop arbitrary single-view detectors as long as the imaging geometry of the multi-camera setup is known.

Related Work. It is important to note that our notion of multi-view detection differs from that in the majority of previous work, which aims to detect object classes from arbitrary viewpoints in a given *single input image* [12,14,15]. Our multi-view detector also addresses this challenge, but additionally aims to do so given *multiple input images* of the 3D scene (*e.g.*, Fig. 1). To that end we take as input the detections from the individual input images (object location and classifier confidence) and fuse them in a voting-based scheme. There is surprisingly little work that considers object detection from multiple input images – also called multi-camera object detection. Recent exceptions [10,11] consider multi-camera detection and tracking in traffic scenes; they exploit geometric properties of the 3D scene, such as ground plane assumptions, temporal background subtraction *etc.*, which are not appropriate for X-ray inspection.

Most closely related to our application scenario is the work of Mery [6], which uses object detection in multiple X-ray views for luggage screening. Detection (*e.g.*, of razor blades) is conducted by comparing a single SIFT [5] descriptor of a reference object to SIFT descriptors of pre-segmented proposal regions in the image. Detections from different views are combined by tracking sparse SIFT features across images. However, [6] only shows results for simple scenes/bags with little clutter to occlude objects of interest and to interfere with reliable feature extraction (in contrast to our work, *cf*. Fig. 7c); furthermore, it is based on ad-hoc, domain-specific segmentation as a preprocessing step. Unlike Riffo and Mery [9], we are constrained by the geometry of existing X-ray inspection systems and thus assume a fixed number of views beyond our control.

Our multi-view fusion/verification approach is related to Ommer and Malik [8], who address object scale as an unknown variable and as a result obtain a line in voting space. They cluster lines to find votes that agree on scales. In the same spirit, we represent
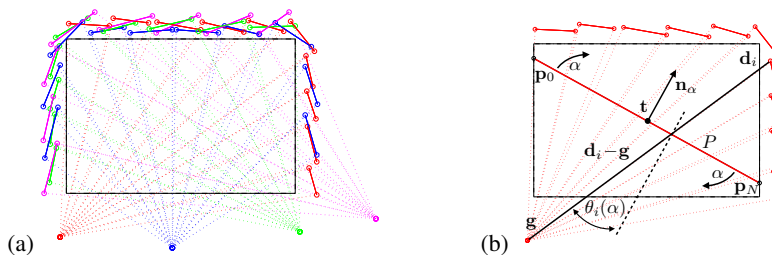
Fig. 2: **X-ray imaging and preprocessing.** *(a)* Arrangement of the four X-ray generators and detectors in the inspection system (color-matched): The detectors are grouped in multiple banks of 64 pixels each. The dotted lines depict the rays between the generators and the first/last pixels in each detector bank. The tunnel is denoted by the black rectangle. *(b)* Illustration of the reprojection approach from individual detector banks to virtual image plane/line (see text); the red line shows the optimized reprojection plane. *Actual geometry differs slightly from this illustration.*

detections from all 2D views with unknown depth to the camera as lines in 3D space; intersecting lines therefore denote detections that agree on 3D object locations.

Although our multi-view approach is not limited to specific single-view detectors to generate detections in each of the input views, we use a standard sliding-window approach with HOG features [1] as the basis. This detector has shown to lead to robust detection results across a wide range of object classes and scenarios, and serves as the backbone of many state-of-the-art systems [4,16]. We here adapt this basic detector to dual-energy X-ray data.

2   Data Acquisition and Preprocessing

The focus of this paper is on detecting dangerous objects in passenger carry-on luggage in the context of airport security check points. Since there is no suitable public dataset available for this task, we recorded one in collaboration with the *German Federal Police* to ensure realism of the data. The recorded X-ray image data contains about 25 different types of real *handguns* placed (randomly and also deliberately concealed) in realistic carry-on luggage. The X-ray inspection system records 4 images of the bag from different viewing angles while it passes through the tunnel (Fig. 2a). Therefore, each *scan* consists of 4 images representing the views of the tunnel (3D scene). The inspection system records with a *dual energy* method: This yields two grayscale X-ray image channels corresponding to a low and a high X-ray energy level, and additional three RGB false-color channels, primarily for visualization, where the colors denote material properties (Fig. 3). In total, we recorded 770 scans of which 652 contain a weapon. All guns in the dataset were annotated with rectangular bounding boxes, which are aligned to the barrel of the gun.

2.1   Image preprocessing for distortion reduction

As can be seen in Fig. 3, the recorded X-ray images appear somewhat distorted (*cf*. the handgun or the bottle opener). This stems from the non-standard imaging process of the
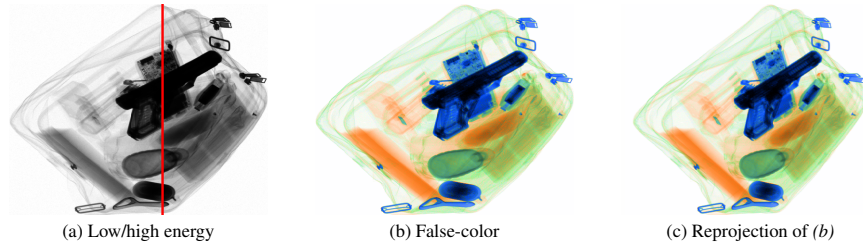
(a) Low/high energy   (b) False-color   (c) Reprojection of *(b)*

Fig. 3: **X-ray image example.** *(a)* Low and high-energy image pair (high energy left of red line) with associated false-color image *(b)*. The effect of reprojection (view 1) is shown in *(c)*.

X-ray inspection system. While this does not pose a major issue for manual inspection, it may significantly affect appearance-based automatic detection approaches as used here. We thus aim to reduce these effects.

As illustrated in Fig. 2a, each of the four 2D X-ray images that comprise each scan are obtained with a 1D fan-shaped line scanner/camera that captures one image row at a time with a perspective projection, while the belt moves the bag along the inspection tunnel. Furthermore, the pixels of the line scanner are not arranged on a straight line, but around the tunnel in an "L"- or "U"-shaped fashion. The combination of these effects leads to the perceived distortion. To alleviate this, we reproject each image row onto a plane in 3D space, which is chosen to approximate the properties of an orthographic projection as best as possible (*i.e.* an orthogonal intersection of rays with image plane), thus leading to a small amount of distortion no matter where the object is located in the bag. The X-ray inspection system is calibrated ahead of time; we thus assume the belt speed and 3D positions of all X-ray generators and detectors (pixels) to be known.

We define the reprojection plane P as the set of all points $\mathbf{x}$ such that $(\mathbf{x} - \mathbf{t})^{\mathrm{T}} \mathbf{n}_\alpha = 0$, where the normal vector $\mathbf{n}_\alpha = [\sin(\alpha), \cos(\alpha), 0]^{\mathrm{T}}$ is parametrized by the rotation angle α around the center of the scanner tunnel $\mathbf{t}$ (Fig. 2b). The normal vector $\mathbf{n}_\alpha$ is defined as perpendicular to the Z axis, since the individual detector banks for a given view all approximately have the same Z coordinates (up to manufacturing tolerances).

We find the normal vector $\mathbf{n}_{\hat{\alpha}}$ that minimizes the square of the angular deviations

$$\hat{\alpha} = \arg\min_\alpha \sum_i \theta_i(\alpha)^2 \quad \text{with} \quad \theta_i(\alpha) = \arccos\left(\frac{\mathbf{n}_\alpha^{\mathrm{T}}(\mathbf{d}_i - \mathbf{g})}{\|\mathbf{n}_\alpha\|\|\mathbf{d}_i - \mathbf{g}\|}\right) \qquad (1)$$

between all X-rays $\mathbf{d}_i - \mathbf{g}$ and the plane normal (Fig. 2b); $\mathbf{d}_i$ is the position of detector pixel i and $\mathbf{g}$ the position of the X-ray generator. The objective in Eq. (1) is periodic $(360°)$ and has a unique minimum in every period, hence a global optimum can be obtained. Subsequently, the image is reprojected onto equidistantly-spaced pixels on the reprojection plane; linear interpolation is used. Note that the four views all have different geometry, hence we require a different reprojection plane P for each of them.

The visual effect of the reprojection can be seen in Fig. 3c; note how the reprojected image looks less distorted with more natural proportions. As we will show in Sec. 3, this reduction in object appearance variation (from distortions) also leads to benefits in terms of detection performance.

3   Detection Approach and Experimental Evaluation

Before explaining our detection approach, it is useful to analyze and separate the object appearance variations that have to be accounted for in our recorded X-ray dataset. As just explained, image distortions inherent to the X-ray imaging process are present, but can be reduced. Other variations are object *intra-class* variations (Intra), *e.g.*, from different types of handguns. Object appearance variations also stem from 3D object rotations, specifically from *in-plane* (In) and *out-of-plane* (Out) rotations. In-plane rotations are object rotations that coincide with the image plane, hence can be undone or accounted for by rotating the 2D image. Out-of-plane rotations correspond to viewpoint changes, which might lead to drastic changes in object appearance. In contrast to in-plane rotations, these cannot be undone by means of a simple image transformation.

To study different kinds of appearance variations individually, and to gain insight to where the challenges are, we create different data subsets. First, we create a hand-selected subset of 300 images that contains handguns viewed from the side, denoted D_{Intra}. Out-of-plane rotations are thus excluded; in-plane rotations are additionally eliminated by *pre-rotating* the images such that the handguns are in a canonical orientation (here possible, since every bag only contains at most one handgun). Hence, the only object variations left are due to intra-class (handgun) variations and from "occlusions" of other unrelated objects (or clutter), which are always present. Next, we use the same subset as described, but do not remove in-plane rotations to gain $D_{Intra+In}$. Then the entire dataset is used, but in-plane rotations are eliminated to yield $D_{Intra+Out}$. Finally, the full dataset without any simplifying alterations or exclusions is referred to as $D_{Intra+In+Out}$.

3.1   Single-view detection

We now explain our detection approach for a single input image (view). It is important to note that this does not mean that the detector need only detect objects from a single, canonical viewpoint. Due to both its popularity as well as robustness for detecting objects in photographic images, our foundation is a standard sliding-window object detector with a linear SVM classifier and Histogram of Oriented Gradients (HOG) features [1]. We train the SVM with the annotated handgun images as positive training examples, which are brought into a canonical orientation and scaled to the SVM window-size of 128×80 pixels; negative training data is randomly cropped from images that do not contain any weapons. We always use three rounds of bootstrapping for SVM training to make results more stable [16]. The different data subsets are divided into approximately $\frac{2}{3}$ training set and $\frac{1}{3}$ test set; positive and negative training examples are divided in the same manner. We evaluate detection performance on each single image (*i.e.* all 4 views of all scans, but considered independently) with the widely used PASCAL criterion [3]. Thus a detection is regarded as correct if the area of overlap with a ground truth bounding box exceeds 50% (and has not been detected before). Based on this, we calculate precision/recall curves and the average precision (AP).

Out-of-plane rotations. Without object out-of-plane rotations in D_{Intra}, we can achieve an average precision (AP) of 88.1% (Fig. 4a). However, with object out-of-plane rotations present in the data ($D_{Intra+Out}$), the performance drops drastically to 26.2% AP.

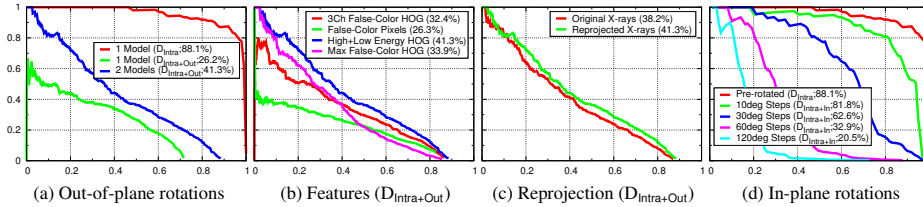|			
(a) Out-of-plane rotations	(b) Features ($D_{Intra+Out}$)	(c) Reprojection ($D_{Intra+Out}$)	(d) In-plane rotations

Fig. 4: **Single-view detection.** Effect of different detector configurations and dataset properties on detection performance. Plots show recall (X axis) vs. precision (Y axis). *Best viewed in color.*

Hence, out-of-plane rotations are a major difficulty for detection in the inspection scenario. To alleviate this, we split the positive training data into two sets (aspect ratios) and train an additional detector ("2 models") with a more narrow search window of 128×40 pixels for handguns roughly viewed from the top. Both detectors are applied independently at test time; their detections are simply concatenated. This boosts performance to 41.3% AP on $D_{Intra+Out}$, hence we always use this approach if out-of-plane rotations are present in the data.

Features. As mentioned in Sec. 2, we obtain low and high-energy image channels with a corresponding false-color image from the X-ray inspection system; the false-color image encodes different material properties with colors, *e.g.* blue for metals and orange for organic materials (*cf*. Fig. 3b). Since we have multiple source images, we evaluate (on $D_{Intra+Out}$) which one yields the best results (Fig. 4b). As baseline we use the false-color pixels directly, which achieves 26.3% AP. With HOG computed on the color images, we achieve 33.9% AP (max. gradient response over color channels), and a modest decline to 32.4% AP when computing HOG separately on all color channels and concatenating the descriptor vectors. The best results of 41.3% AP can be obtained when using HOG separately on the low and high-energy channels, again with concatenation of the descriptors. Consequently, we use this feature in all other experiments.

Reprojection. In Sec. 2.1 we showed that the X-ray images are distorted and explained how we reduce these distortions. We evaluate the effect of this on $D_{Intra+Out}$ and find that the influence of our reprojection method is clearly visible (Fig. 4c): Precision is better for almost all recall levels and the maximum recall is slightly increased. Overall, performance increases from 38.2% AP on the original images to 41.3% AP on the reprojected ones. Hence, we use reprojected images in all other experiments.

In-plane rotations. So far, we conducted every experiment on pre-rotated images, which serves as an upper bound to the performance on real X-ray data. In-plane rotations can often be ignored when trying to detect object classes in regular photographic images. In pedestrian detection [2], for example, it is safe to assume that people mostly appear in an upright position. However, no such assumptions can be made for the X-ray images, since objects in the passenger luggage may appear at arbitrary 3D rotations.

In a realistic setting, we clearly cannot assume pre-rotated images. At test time we thus search at various orientations, because HOG features are not invariant to rotation (besides searching over multiple scales of the image to account for object size variations). We illustrate the effect of this by comparing the performance on $D_{Intra+In}$ (with

several search orientation steps) to the pre-rotated set D_{Intra} (Fig. 4d). Coarse angular steps of $60°$ and $120°$ yield poor performance, but further decreasing step sizes successively improves performance. At $10°$, we achieve an AP of 81.8%, which is quite close to (the upper bound of) 88.1% AP on pre-rotated images. Smaller angular steps will likely increase performance at the expense of computational effort; we always use $10°$ angular steps in all subsequent experiments when searching over object orientations. We present single-view detection results for the full dataset $D_{Intra+In+Out}$ (using rotation-search) in Sec. 4, when comparing to our multi-view integration approach (Fig. 6b).

Non-maximum suppression. After classifying each sliding window (in all scales and rotations) with the linear SVM, we perform non-maximum suppression (NMS) to retain only the local maxima, *i.e.* the most promising object hypotheses with their attached classifier confidence (SVM decision value). One dominant approach for NMS, adopted here, is *pairwise max suppression* (PM) ([2]; *e.g.*, used by [4]), which performs pairwise comparison of all search windows / bounding boxes (BB): If the area of overlap of a pair exceeds a certain threshold, the BB with lower classifier confidence is suppressed. Computing the area of overlap is trivial and computationally efficient if

Fig. 5: Weiler-Atherton.

bounding boxes are axis-aligned, *i.e.* all line segments are either parallel or orthogonal to each other; however, this no longer applies to arbitrarily rotated BBs, as occur here.
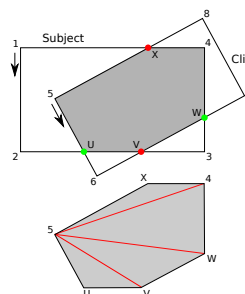
Mittal *et al.* [7] approximate this by simply placing axis-aligned BBs around each rotated BB and then use usual PM, noting (personal communication) that results are very similar as compared to doing accurate calculations. However, this only applies to approximately square BBs, and not to narrow ones as we use to deal with object viewpoint variations. Hence, we are interested in an exact and computationally efficient solution. NMS is performed many times with tens of thousands of BBs, hence naive solutions such as bitmask-based overlap calculations are too inefficient. To perform NMS with rotated BBs efficiently, we propose to use the Weiler-Atherton (WA) algorithm [17], which was originally introduced in computer graphics for polygon clipping. The algorithm efficiently finds the intersection polygon of two input polygons (Fig. 5, top) by alternating between traversing the *subject* and *clip* polygon at every intersection (at *U*, *V*, *W*, and *X* in Fig. 5); the intersection area is then computed through geometric triangulation (Fig. 5, bottom). The runtime is only dependent on the amount of nodes in both input polygons, hence constant when used with rectangles only. We use the Weiler-Atherton algorithm to perform NMS in all our experiments. Its benefits are that it enables accurate overlap calculations with arbitrarily rotated bounding boxes for PM, yet it is fast enough in practice. As far as we are aware, this is the first time that this algorithm has been used in this context.

4   Multi-view Integration and Experimental Results

The preceding experiments on single input images show that out-of-plane rotations are a major challenge for object detection (Fig. 4a). To improve on this, we exploit the multiple input images available to us from modern X-ray inspection systems (*cf*. Fig. 2a).

After running single-view detection separately on each of the (in our case 4) input images (views), we fuse their respective detections. The motivation is to suppress false detections, since they are not likely to coincide in different views; similarly, detections that agree on object locations can reinforce each other. Before going into the details of our integration approach, it is important to hightlight its generality; any single-view detection method can be used. We are not even constrained to X-ray image data, as used here. The input to our integration algorithm is a list of object locations (bounding boxes) with corresponding classifier confidences, and the 3D geometry of all cameras.

The main idea of the proposed multi-view integration is that each detection represents a 2D object location with unknown depth/distance w.r.t. the X-ray generator source (*cf*. Fig. 2). Technically, each 2D bounding box (BB) defines a polyhedron of possible object locations in 3D space (a wedge intersected with the cube of the scanner tunnel). Since it is cumbersome to work with this polyhedron, we use the central pixel of each 2D bounding as its proxy and thus obtain a line/ray in 3D space from the X-ray generator source to the detector (pixel) location (*cf*. Fig. 7a and $\mathbf{d}_i - \mathbf{g}$ in Fig. 2b). In essence, each detection in each view image casts a vote (weighted by the classifier confidence) for the *center* of a particular object location in 3D space with unknown depth. If votes from two or more different views agree, *i.e.* the respective rays intersect in 3D space, we can *(i)* increase our confidence in each of the individual detections, and *(ii)* recover the proposed 3D location of the object's center.

Since detected BBs (and thus their central locations) might be slightly misplaced, we cannot expect perfect intersections from their respective rays and should rather find rays in 3D that come very close to each other. We address this by placing a dense grid of sampling points over the whole 3D volume of the scanner tunnel. Each sampling point is associated with rays in close proximity (*i.e.* below a certain threshold; here less than the distance between sampling points). Confidences of nearby rays are accumulated and finally averaged. To prevent bias from any particular view, we only allow one ray from each view (the one with highest confidence) to contribute to a sampling point.

The result of our integration approach is a 3D confidence map of central object locations (Fig. 7a). We are interested in the most promising locations (local maxima), therefore apply (pairwise) non-maximum suppression where each sampling point suppresses all others with lower confidence in a local neighborhood (neighborhood radius based on object class). The obtained local maxima represent our fused detections, which we project back into the original view images to enable comparison with results from single-view detection; however, they also admit 3D object localization.

Evaluation criterion. The concept of rectangular bounding boxes is lost in our multi-view integration approach. Hence, we evaluate detections with a custom criterion based on relative distance, and not with the usual PASCAL criterion. To that end, we define the relative distance $\text{dist}_{\text{rel}} = \|\text{center}_{\text{gt}} - \text{center}_{\text{dt}}\|/\text{scale}_{\text{gt}}$, between the ground truth (gt) object location $\text{center}_{\text{gt}}$ and the detection (dt) location $\text{center}_{\text{dt}}$, where scale_{gt} is the width of the ground truth bounding box. A detection is considered correct if dist_{rel} is below a threshold (here $\frac{1}{3}$). We do not allow multiple detections of the same object.

Our distance-based evaluation criterion is less restrictive than PASCAL [3], which can be observed in Fig. 6 by comparing single-view detection performance evaluated with both PASCAL (denoted by area) and our criterion (denoted by dist). However,
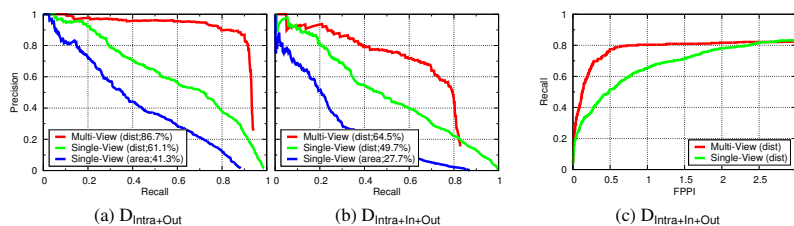
Fig. 6: **Single-view vs. multi-view detection.** Results in *(a)* are obtained with pre-rotated images, *i.e.* denote an upper bound on detection performance on real data *(b,c)* without any simplifying assumptions. Results in *(c)* show false positives per image (FPPI) vs. recall. Please see text for further discussion. *Best viewed in color.*

we believe that it is still restrictive enough to provide good object localization in the targeted scenario, as illustrated in Fig. 7.

Experimental results. We evaluate our multi-view integration approach on the full dataset $D_{Intra+In+Out}$ with our distance-based criterion (Fig. 6b) and find that the average precision (AP) increases substantially from 49.7% in single-view detection to 64.5% with multi-view integration; the difference is also visualized in Fig. 7c. Albeit greatly increased AP, our integration procedure yields lower recall than single-view detection at low precision levels. However, our goal is the detection of dangerous objects in the context of airport security, specifically with the intention of supporting human operators, not replacing them. In this context, we are interested in producing relatively few false alarms/detections per image (FPPI) to only notify operators in case of actual threat items in passenger luggage. We study this effect in Fig. 6c, which shows that on average our multi-view approach is able to detect around 80% of all handguns while producing a single false alarm at every second image; in contrast, single-view detection is able to only find 50% of all handguns at the same false-alarm rate.

5   Summary

In this paper we proposed an approach for object detection in multi-view dual-energy X-ray image data, as it arises for example in luggage inspection systems. We adapted appearance-based object class detection for the specifics of the image data and setting. To that end, we analyzed and addressed several object appearance variations: We reduced distortions in X-ray images, introduce an efficient non-maximum suppression scheme in the context of in-plane rotations, and most importantly introduced a novel multi-view integration approach to deal with out-of-plane object rotations. An experimental evaluation of handgun detection on a challenging dataset of carry-on luggage showed substantial performance improvements from all proposed components of our approach, particularly of our multi-view integration. In the future we plan to further improve the recall at precision levels tolerable for a human operator, in part by closing the gap between performance on pre-rotated (Fig. 6a) and realistic data (Fig. 6b); this may be accomplished by using a rotation-invariant image descriptor (*e.g.* [13]), which has the additional benefit of reducing the computational overhead incurred by searching
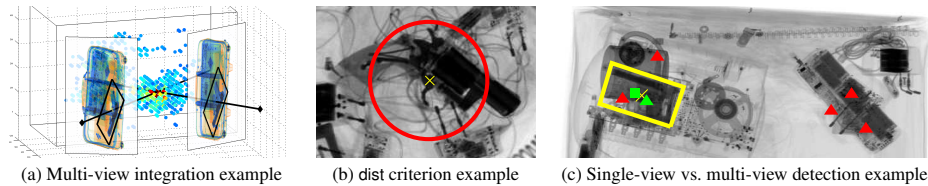
(a) Multi-view integration example   (b) dist criterion example   (c) Single-view vs. multi-view detection example

Fig. 7: *(a)* Partial 3D confidence map (red–high, blue–low) of object locations within scanner tunnel with 2 of 4 views shown to illustrate ray intersections. *(b)* Distance-based evaluation criterion: Only detections within the red circle are considered correct. *(c)* At same recall rate (70%), single-view detection (triangles) produces several false detections (red; correct in green), and multi-view detection none (squares). *Best viewed in color on screen.*

over many image rotations for single-view detection. Future work may also consider applications of our multi-view integration scheme outside of X-ray imagery.

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR 2005
2. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. PAMI 34(4) (2012)
3. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2008 (VOC2008) Results
4. Felzenszwalb, P.F., Girshick, R.B., McAllester, D.A., Ramanan, D.: Object detection with discriminatively trained part-based models. PAMI 32(9) (2010)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2) (2004)
6. Mery, D.: Automated detection in complex objects using a tracking algorithm in multiple X-ray views. In: OTCBVS 2011 workshop (in conj. with CVPR 2011)
7. Mittal, A., Zisserman, A., Torr, P.: Hand detection using multiple proposals. In: BMVC 2011
8. Ommer, B., Malik, J.: Multi-scale object detection by clustering lines. In: ICCV 2009
9. Riffo, V., Mery, D.: Active X-ray testing of complex objects. Insight (2011)
10. Roig, G., Boix, X., Shitrit, H.B., Fua, P.: Conditional random fields for multi-camera object detection. In: ICCV 2011
11. Sankaranarayanan, A.C., Veeraraghavan, A., Chellappa, R.: Object detection, tracking and recognition for multiple smart cameras. Proc. IEEE 96(10) (2008)
12. Savarese, S., Fei-Fei, L.: 3D generic object categorization, localization and pose estimation. In: ICCV 2007
13. Schmidt, U., Roth, S.: Learning rotation-aware features: From invariant priors to equivariant descriptors. In: CVPR 2012
14. Thomas, A., Ferrari, V., Leibe, B., Tuytelaars, T., Schiele, B., Gool, L.V.: Towards multi-view object class detection. In: CVPR 2006
15. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multi-view object detection. PAMI 29(5) (2007)
16. Walk, S., Majer, N., Schindler, K., Schiele, B.: New features and insights for pedestrian detection. In: CVPR 2010
17. Weiler, K., Atherton, P.: Hidden surface removal using polygon area sorting. In: SIGGRAPH 1977
18. Zentai, G.: X-ray imaging for homeland security. In: IST 2008 workshop