A Logarithmic X-ray Imaging Model for Baggage Inspection: Simulation and Object Detection

Domingo Mery
Department of Computer Science
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860(143) – Santiago de Chile
domingo.mery@uc.cl

Aggelos K. Katsaggelos
Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208, USA
aggk@eecs.northwestern.edu

Abstract

In the last years, many computer vision algorithms have been developed for X-ray testing tasks. Some of them deal with baggage inspection, in which the aim is to detect automatically target objects. The progress in automated baggage inspection, however, is modest and very limited compared to what is needed because X-ray screening systems are still being manipulated by human inspectors. In this work, we present an X-ray imaging model that can separate foreground from background in baggage screening. The model can be used in two main tasks: i) Simulation of new X-ray images, where simulated images can be used in training programs for human inspectors, or can be used to enhance datasets for computer vision algorithms. ii) Detection of (threat) objects, where new algorithms can be employed to perform automated baggage inspection or to aid an user in the inspection task showing potential threats. In our model, rather than a multiplication of foreground and background, that is typically used in X-ray imaging, we propose the addition of logarithmic images. This allows the use of linear strategies to superimpose images of threat objects onto X-ray images and the use of sparse representations in order to segment target objects. In our experiments, we simulate new X-ray images of handguns, shuriken and razor blades, in which it is impossible to distinguish simulated and real X-ray images. In addition, we show in our experiments the effective detection of shuriken, razor blades and handguns using the proposed algorithm outperforming some alternative state-of-the-art techniques.

1. Introduction

Baggage inspection using X-ray screening is a priority task that reduces the risk of crime, terrorist attacks and propagation of pests and diseases [29]. Security and safety screening with X-ray scanners has become an important process in public spaces and at border checkpoints [13]. However, as shown in Fig. 1, 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]. Manual detection of threat items by human inspectors is extremely demanding [24]. It is tedious because very few bags actually contain threat items, and it is stressful because the work of identifying a wide range of objects, shapes and substances (metals, organic and inorganic substances) takes a great deal of concentration. In addition, human inspectors receive only minimal technological support. Furthermore, during rush hours, they have only a few seconds to decide whether or not a bag contains a threat item [1]. 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% [15]. In baggage inspection, automated X-ray testing remains an open

Figure 1. Setup of an X-ray imaging system, the X-ray source irradiates the object (a bag containing a handgun) and produces an X-ray image.
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 tradeoff between false alarms and missed detections; iii) limited robustness given that requirements for the use of a method are often met for simple structures only; and iv) low adaptiveness in that it may be very difficult to accommodate an automated system to design modifications or different specimens. There are some contributions in computer vision for X-ray testing such as applications on inspection of castings, welds, food, cargos and baggage screening [10] and very interesting advances in automated recognition of objects [14] that evaluates ten different modern computer vision algorithms. Nevertheless, as we can see in the literature review [21, 27, 30, 11, 20, 6, 17, 4, 9, 5, 16, 13], the progress in automated baggage inspection is far from being perfected given that the appearance of the object of interest can become extremely difficult due to problems of (self)occlusion, noise, acquisition, clutter, etc. We believe, however, that computer vision algorithms can be improved if we could model the X-ray images as a contribution of background and foreground. Thus, the detection can be achieved by analyzing the foreground image only.

In this work, we present an imaging model that can separate foreground from background of an X-ray image. The model can be used in the detection of (threat) objects, where new algorithms can be employed to perform automated baggage inspection or to aid an user in the inspection task showing potential threats. In our model, rather than a multiplication of foreground and background, that is typically used in X-ray imaging, we propose the addition of logarithmic images. This allows the use of linear strategies such as sparse representations in order to segment target objects.

In addition, the model can be used in the simulation of new X-ray images, where simulated images can be used in training programs for human inspectors, or can be used to enhance datasets for computer vision algorithms. Similarly, this allows the use of linear strategies to superimpose images of threat objects onto X-ray images. The simulated X-ray images should be as similar as possible to real X-ray images. This allows the use of linear strategies such as sparse representations in order to segment target objects.

In our experiments, we simulate new X-ray images of handguns, shuriken and razor blades, in which it is impossible to distinguish simulated and real X-ray images. In addition, we show in our experiments the effective detection of shuriken, razor blades and handguns using the proposed algorithm.

The rest of the paper is organized as follows: Section 2 describes the proposed model and the simulation approach, Section 3 outlines the proposed method for threat object detection, Section 4 shows the experimental results, and finally Section 5 gives concluding remarks.

2. X-ray Imaging Model

Typically, X-ray imaging can be modeled using the absorption law which characterizes the intensity distribution of X-rays through matter [8]:

$$\varphi(d) = \varphi_0 e^{-\mu d}$$  \hspace{1cm} (1)

with $\mu$ absorption coefficient, $d$ thickness of the irradiated matter, $\varphi_0$ incident energy flux density, and $\varphi$ energy flux density after passage through matter with the thickness of $d$. In Fig. 2, we can see an example with $n = 3$ materials that can be modeled by:
From (8) and (7), it yields

\[ I_t = (I_f - B) / C. \] (10)

Thus, using the normalized images for total, foreground and background images, we obtain:

\[ J_t = J_f \cdot J_b. \] (10)

Easily, we can compute the total image by

\[ I_t = C \cdot J_f \cdot J_b + B. \] (11)

Indeed, image \( I_t \) in Fig. 3c was simulated from \( I_f \) and \( I_b \) in Fig. 3a and 3b respectively using (11).

3. Application: Object detection

In this section, we explain how to detect threat objects using a sparse representation based on the model of Section 2. The key idea of our strategy is to use a dictionary for the foreground and another dictionary for the background. In the foreground images, only the threat objects are detected as present, whereas in the background, the rest of the objects are present. In this classic strategy, a testing image can be modeled by the sum of the contributions of the foreground and the background, so the detection of the threat objects can be easily achieved by analyzing the contribution of foreground only. However, in our model (10) the testing image is obtained by a multiplication—not by an addition—of two images. We can solve this problem, if we take the logarithm of both sides of the equation (see Fig. 4). Thus,

\[ Z_t = \log(J_t). \] (12)

where the three variables are \( \log(J_t) \), \( \log(J_f) \), \( \log(J_b) \) respectively. We call these images the logarithmic images. In general, the logarithmic image \( Z \) of an X-ray image \( I \) is computed by

\[ Z(J) = \log \left( \frac{I - B}{C} \right). \] (13)
Figure 4. The total image $I_t$ can be modeled by multiplication or addition, in first case the original images are used, whereas in the second case we consider the logarithmic images.

We believe that algorithms based on sparse representations can be used for this general task because in recognition applications, under the assumption that natural images can be represented using sparse decomposition, state-of-the-art results have been significantly improved [26]. In the sparse representation approach, a dictionary is built from the training X-ray images, and matching is done by reconstructing the test image using a sparse linear combination of the dictionary. Usually, the test image is assigned to the class with the minimal reconstruction error.

In our approach, we build two dictionaries: $D_f$ for the foreground and $D_b$ for the background. The first one is obtained from patches of the foreground (see Fig. 5), the second one from patches of the background (see Fig. 6). In both cases, representative X-ray images are used. In our approach, we extract a patch $z$ of size $w \times w$ pixels of the logarithmic images subsampled by $a$ pixels in both directions. For the dictionaries, we use vector $y$ of $(w/a)^2$ elements that corresponds to the intensity values of $z$ given by stacking its columns normalized to unit length in order to deal with different contrast conditions. This operation is represented by function $f$:

$$y = f(z).$$  

(14)

Dictionaries $D_f$ and $D_b$ have $n_f$ and $n_b$ columns respectively. Both dictionaries are concatenated as shown in Fig. 7 yielding a new dictionary of $n = n_f + n_b$ columns and $(w/a)^2$ rows:

$$D = [D_f \ D_b].$$  

(15)

In testing stage, we use the well-known methodology of sliding windows, i.e., a small window $z$ of $w \times w$ pixels (the same size of the patches used in the dictionaries) is sledded over the testing image in both horizontal and vertical directions, and for each localization of $z$, first we compute $y$ using transformation (14), and second we calculate $x$ the sparse representation of $y$ using dictionary $D$:

$$||y - Dx|| \to \min \text{ subject to } ||x||_0 \leq T,$$  

(16)

where $||x||_0$ means the number of non-zero elements of $x$. Thus, in the sparse representation, no more than $T$ elements
of $x$ are allowed to be non-zero. According to (15) and (16), sparse vector $x$ has $n = n_f + n_b$ elements: the first $n_f$ and the last $n_b$ elements correspond to the contribution to the foreground and background respectively. Thus, vector $x$ can be decomposed as:

$$x = \begin{bmatrix} x_f \\ x_b \end{bmatrix},$$  \hspace{1cm} (17)$$

the first one with $n_f$ elements, and the last one with $n_b$ elements. We can investigate, how representations $x_f$ and $x_b$ can reconstruct the original patch $y$ using dictionaries $D_f$ and $D_b$ respectively.

$$y_f = D_f x_f, \hspace{0.5cm} y_b = D_b x_b.$$  \hspace{1cm} (18)$$

The idea here, is the following: for example if the reconstruction $y_f$ is good enough, that means that the reconstruction error $||y - y_f||$ is minimal, then the patch can be classified as foreground, otherwise as background. The reconstruction error for both cases are defined as:

$$e_f = ||y - y_f||, \hspace{0.5cm} e_b = ||y - y_b||.$$  \hspace{1cm} (19)$$

The proposed algorithm is simple and it is shown in Algorithm 1. It follows the sliding-windows strategy. Thus, in each location of the sliding window, we extract a patch $z$ that is transformed into a vector $y$ which is represented as a sparse vector $x$. The idea is to investigate the contribution of $x$ using both foreground and background dictionaries ($D_f$ and $D_e$ respectively). If the reconstruction error of the foreground information is low enough and less than the reconstruction error of the background information, then our algorithm concludes that a part of the threat object is detected. Finally, we select those regions that are large enough and have enough parts that were detected.

4. Experimental Results

In this work, the main contribution is a new model that can be used in X-ray imaging. In this section, we show two experiments: the first one corresponds to the new methodology that can be used to simulate X-ray images following equation (11), the second one refers to the new approach to detect threat objects in X-ray images using Algorithm 1.
In the following two sections, we show some experiments that validate the proposed model and the methods. All X-ray images used in our experiments belong to the GDXray\textsuperscript{1} database \cite{GDXray}.

\subsection*{4.1. Simulation}

In this Section we show how to simulate X-ray images. Simulated images can be used in training programs for human inspectors, or can be used to enhance datasets for computer vision algorithms. The idea is simple, we have to acquire X-ray images of objects that are completely isolated and then we can superimposed them onto X-ray images of cluttered bags. In order to acquire isolated X-ray images, the threat object can be located inside a sphere of expanded polystyrene (EPS)\textsuperscript{2} as suggested in \cite{EPS}. In GDXray we have those kind of images, where a threat object is irradiated from different points of views. Thus, the threat object can be superimposed in many different poses.

In order to illustrate the similarity between original and simulated X-ray images, we show experiments where the original X-ray image has only one threat object and the simulated image has the original threat object and the superimposed threat object, so in the same image we can compare both of them. We tested with the following threat objects: handguns, razor blades and shuriken (ninja stars) in nine different poses. The results are given in Figures 9, 10 and 11 respectively. In our results, the reader can see both threat objects –simulated and original–, and can conclude that both objects are so similar that it is imposible to say which one is the simulated and which one is the original.

\subsection*{4.2. Object detection}

In this Section, we show some experiments using the method explained in Section 3 to detect threat objects in baggage screening. In our experiments, we detect handguns, razor blades and shuriken using sparse representations. In order to illustrate step by step, in Fig. 12 steps of Algorithm 1 are shown. We can see the effectiveness of the proposed method using the sliding-window approach and the classification based on sparse representations and reconstruction error. In addition, we show the output of of Algorithm 1 for a shuriken and a handgun in Figure 13 and 14 respectively. Again, the effectiveness of the proposed method is validated.
Figure 12. Steps of the detection of a razor blade using proposed algorithm. In this example $E_f$ and $E_b$ are the reconstruction error of foreground and background respectively. We observe that for the object to be detected the error is low for the foreground and high for the background.

Figure 13. Detection of a shuriken using proposed algorithm.

Figure 14. Detection of a razor blade using proposed algorithm.

In order to compare our method with other methods that can be used for this task, we followed the evaluation protocol proposed in [14] for the detection of handguns in cropped X-ray images. In this experiment, there are two classes: target and no-target, i.e., handguns and no-handguns. In no-handguns class there are razor blades, shuriken and other objects like pens, clips, etc. The number of target/no-target X-ray images of the sets of training (used to design the detector), validation (used to tune the detectors’ parameters) and testing (used to measure the performance) are: 200/700, 50/300, 100/600 respectively (see more details in [14]). The obtained results (precision, recall and accuracy) are summarized in Table 1. The first 10 rows of Table 1 show the reported results by [14] in this experiment. The last row shows the results obtained by our method (see ‘ours’). As a result, the proposed approach reaches a very good recognition performance, outperforming some alternative state-of-the-art techniques. In order to see the positive effect of the use of the logarithmic images in our model, we repeated this experiment using our method with no-logarithmic images. The results are shown in the row called ‘no-log’. We observe that the use of logarithmic images increases the recall from 0.65 to 0.99.
Table 1. Detection of guns

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW</td>
<td>0.65</td>
<td>0.84</td>
<td>91.3</td>
</tr>
<tr>
<td>Sparse KNN</td>
<td>0.99</td>
<td>0.97</td>
<td>99.4</td>
</tr>
<tr>
<td>Sparse KNN*</td>
<td>0.92</td>
<td>1.00</td>
<td>98.8</td>
</tr>
<tr>
<td>AISM</td>
<td>0.97</td>
<td>0.97</td>
<td>99.1</td>
</tr>
<tr>
<td>XASR+</td>
<td>0.92</td>
<td>0.88</td>
<td>97.2</td>
</tr>
<tr>
<td>GoogleNet</td>
<td>0.83</td>
<td>1.00</td>
<td>97.1</td>
</tr>
<tr>
<td>AlexNet</td>
<td>0.85</td>
<td>0.99</td>
<td>97.4</td>
</tr>
<tr>
<td>SVM</td>
<td>0.90</td>
<td>0.99</td>
<td>98.3</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>1.00</td>
<td>0.87</td>
<td>98.1</td>
</tr>
<tr>
<td>SRC</td>
<td>0.75</td>
<td>1.00</td>
<td>95.2</td>
</tr>
<tr>
<td>no-log</td>
<td>0.93</td>
<td>0.65</td>
<td>94.3</td>
</tr>
<tr>
<td>ours</td>
<td>0.93</td>
<td>0.99</td>
<td>98.7</td>
</tr>
</tbody>
</table>

4.3. Implementation

The experiments were carried out on a iMac OS X 10.12.4 with a 3.7 GHz Quad-Core Intel Xeon E5 processor and 12GB memory (12GB RAM 1866 MHz DDR3). The X-ray images were obtained from GDXRay [12]. The algorithms were implemented in MATLAB [4]. For the sparse representation we used SPAMS library from INRIA [7]. The code of the MATLAB implementation is available on our webpage [3]. For the simulation: Parameters $B$ and $C$ of model [7] were obtained using a calibration approach [10] in which two images of a material with different thickness are acquired. In our case, $B = 0$ and $C = 230$. The simulation of an X-ray image of 6M pixels is performed in 0.09s. For the detection, we give here details of the last experiment (detection of handguns). For the implementation of the dictionaries, we used patches of 180 × 180 pixels subsampled to 18 × 18, i.e., $w = 180$ and $a = 10$. That means that vector $y$ has 100 elements. Around 650,000 patches were extracted for the dictionaries. The sparsity (parameter $T$) was set to 10. The training was performed in 7.45 min, and the detection of threat objects in 65 sec per image (for images of 1M pixel).

5. Conclusions

The main contribution of our paper is a new X-ray imaging model that can separate foreground from background. The model can be used for example in: i) simulation of new X-ray images, and ii) recognition of objects when foreground and background are appropriately defined. On the one hand, simulated images can be used in training programs for human inspectors, or can be used to enhance datasets for computer vision algorithms. On the other hand, detection algorithms can be employed in automated baggage inspection, because it can be used to aid an user in an inspection task.

Rather than a multiplication of foreground and background, that is typically used in X-ray imaging, we propose the addition of logarithmic images. This allows the use of linear strategies such as sparse representations that can be used effectively in the recognition of threat objects. We show in our experiments the detection of shuriken, razor blades and handguns. As a result, the proposed approach reaches a very good recognition performance, outperforming some alternative state-of-the-art techniques. In our experiments, the increase of the performance by including logarithmic images was considerable.

The preliminary results are promising. Nevertheless, since the effectiveness of the proposed model has been verified on a few X-ray images, it is necessary an evaluation on a broader dataset.

Acknowledgments

This work was supported by Fondecyt Grant No. 1161314 from CONICYT, Chile.

References


