Engineering Applications of Artificial Intelligence 67 (2018) 81–90

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Using machine learning to detect and localize concealed objects in passive millimeter-wave images

Santiago López-Tapia *, Rafael Molina, Nicolás Pérez de la Blanca

Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

ARTICLE INFO

Keywords:
Threat detection
Machine learning
Passive millimeter-wave imaging

ABSTRACT

The detection and location of objects concealed under clothing is a very challenging task that has crucial applications in security. In this domain, passive millimeter-wave images (PMMWIs) can be used. However, the quality of the acquired images, and the unknown position, shape, and size of hidden objects render this task difficult. In this paper, we propose a machine learning-based solution to this detection/localization problem. Our method outperforms currently used approaches. The effect of non-stationary noise on different classification algorithms is analyzed and discussed, and a detailed experimental comparative study of classification techniques is presented using a new and comprehensive PMMWI database. The low computational testing cost of this solution allows for its use in real-time applications.

1. Introduction

Millimeter and submillimeter waves are very high-frequency electromagnetic radiations with wavelengths in the ranges 30–300 GHz and 0.3–3 THz, respectively. Images in both ranges can be obtained by employing wave scanners. Depending on wavelength, two types of scanners can be distinguished: active, which direct waves from 350 to 800 GHz to a subject, and collect and interpret the reflected energy; and passive scanners, which create images using background radiation and that emitted by human bodies or objects in the range 30–350 GHz.

Active scanners provide images with higher signal-to-noise ratios (SNR), but problems related to privacy intrusion have prevented their use in many applications. PMMWIs (see Fig. 1) acquired using passive scanners are currently being used in anti-theft and threat detection systems (Alexander et al., 2008) in places like airports and warehouses. Unfortunately, passive millimeter sensors (and consequently, their images) suffer from the following problems: low SNR, low resolution, and space-variant signal intensity. These factors degrade the performance of detection systems as they tend to produce a very large number of false positive detections and miss real threats. Detection systems based on passive scanners must deal with the unknown position, shape, size, and transmission properties of the hidden objects.

In this paper, we use artificial intelligence techniques to propose a new detection approach capable of dealing with the poor quality of PMMWIs. Our approach is based on feature extraction and classification algorithms, in contrast to current methods that only aim to enhance the SNR of images and apply simple detection algorithms.

Despite their enormous potential for security applications, the artificial intelligence community has so far shown little interest in PMMWIs. This is likely because of the absence of large databases of PMMWIs to work with. For this work, we took pictures of people of different complexions wearing 12 objects on 10 parts of the body: forearm, chest, stomach, thigh, ankle (front), waist (side), armpit (side), arm, ankle (lateral) thigh (lateral), and two images without any object. Images of people wearing two objects in different locations were also taken. Real hidden threats were simulated using objects of different sizes with millimeter wave responses similar to real threats. A cutter, gel, a clay bar, a simulated gun, sugar, frozen peas, cologne, a bag with metal pieces, flour, a bottle of water, and a hydrogen peroxide bottle were used. The dataset consisted of 463 pictures of people with no objects, 2144 pictures containing one object, and 702 pictures containing two objects. More details are provided in Appendix. The passive scanner used in this work provided 125 × 195-pixel images. The sizes of the hidden objects of interest ranged from 35×39 to 10×10 pixels, corresponding to roughly 2752.39 cm^2 to 201.64 cm^2 object areas, respectively.[1] In the section detailing the experiments, a discussion of how the size, location, and composition of the threat influence the detection is provided.

[1] The image database used in this paper can be downloaded at http://decsai.ugr.es/pi/pmmwi/testdata.html.

* Corresponding author.
E-mail addresses: sltapia@decsai.ugr.es (S. López-Tapia), rms@decsai.ugr.es (R. Molina), nicolas@ugr.es (N. Pérez de la Blanca).

https://doi.org/10.1016/j.engappai.2017.09.005
Received 16 October 2016; Received in revised form 1 August 2017; Accepted 4 September 2017
0952-1976/

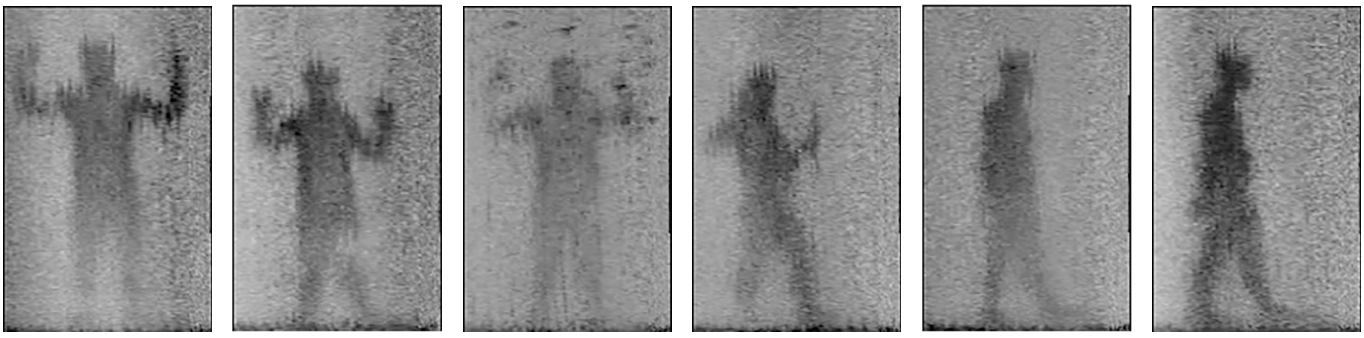


Fig. 1. Examples of PMMWI. Hidden objects correspond to whiter areas within the body. Unfortunately, not all whiter areas correspond to hidden objects.

This paper contains two major contributions. First, it provides a new real time solution to the detection/location of hidden objects in PMMWIs using a machine learning approach. Second, it introduces a new and comprehensive database of PMMWIs that encourages research on this challenging problem. In the remainder of the paper, the terms hidden object and threat are used interchangeably.

The rest of the paper is organized as follows: In Section 2, the current literature on PMMWI acquisition and threat detection is reviewed. Section 3 describes the image enhancement technique we propose to improve the quality of acquired images. Section 4 describes the image patches extracted from the images in the dataset. Section 5 discusses meaningful features extracted from each patch to solve the threat detection problem, and these features are compared in the experimental section. Once features have been extracted from each patch, a formal definition of our solution using a patch detection function and an image classification function is presented in Section 6. These two functions are described in detail in Section 7. An exhaustive experimental validation of the proposed classifiers is presented in Section 8. Conclusions are drawn in Section 9. Appendix contains a complete description of the created dataset.

2. Related work

The seminar work (Yujiri et al., 2003) described the phenomenology that defines the performance of PMMW imaging systems, explained the technological advances that have made these systems a reality, and presented some of the applications for which these sensors can be used (see also Yujiri, 2006). Oka et al. (2008) examined trends in millimeter wave imaging technologies, focusing mainly on applications and technical parameter variations for security surveillance and non-destructive inspections.

As we have indicated, the use of very high frequencies in active sensors compromises the privacy of the people being scanned. Furthermore, due to their very narrow frequency range detection, passive terahertz sensors (0.3–0.35 THz) must be calibrated to detect specific materials making them blind to others. This favors the interest in and the wide range application of PMMW sensors. However, these sensors require the use of robust and efficient algorithms to detect concealed objects.

Image processing techniques have been utilized on PMMWIs. Denoising, deconvolution, and enhancement techniques have been applied to these images, (see Han et al., 2010; Mateos et al., 2016; Liu et al., 2016; Fang et al., 2017; Yang et al., 2010; Yu et al., 2011). In this paper, we develop an image denoising technique tailored to the sensors used during the acquisition process. Note that the use of compressive sensing and super-resolution techniques on these images has also been explored (see Babacan et al., 2011; Saafin et al., 2016).

Few studies have been devoted to the development of robust algorithms for the automatic detection of hidden objects when this type of image is used. In Haworth et al. (2004), an MMW imager which employed a 1D scanned focal-plane array operating at 0.35 THz and produced a real-time head-to-toe video output is utilized. K-means was used to segment MMWIs into three regions: background, body, and threats. However, the method may not detect a threat when its associated region is not connected to the body. To solve this problem, the authors used Active Shape Models (ASM) inside the body. However, this approach does not guarantee adequate body segmentation. In Haworth et al. (2006), Gaussian mixture models were used to characterize background, body, and threat regions and segment the images. Although the reported results were better than those in Haworth et al. (2004), this method also produces an unconnected body segmentation. In Haworth et al. (2007), the method was extended to detect and track metallic objects in a sequence of MMWIs.

In Shen et al. (2008), for a passive terahertz imaging system, noise was first removed from the image using anisotropic diffusion. Following this, the boundaries of the concealed objects were detected. To model the distribution of the temperature inside the image a mixture of Gaussian densities was used. Curves were then evolved along the isocontours of the image to identify the concealed objects. In Martínez et al. (2010), the authors applied noise elimination and then image segmentation using local binary fitting (LBF). Two noise removal algorithms were used: non-local means (NLM) and iterative steering kernel regression (ISKR). Although this method's detection rate was around 90%, its computing time made it impractical for real-time applications. Furthermore, its performance significantly decreased when used on noisy or low quality images. In Sarkis and Mani (2012), a graph-cut algorithm was proposed to segment threats, but its evaluation was inaccurate. In Yeom et al. (2012), a method to detect and recognize threats in outdoor PMMWIs is proposed. The threat detection was carried out through global and local segmentation: at each level (global and local), a Gaussian mixture model whose parameters were initialized using vector quantization (for a different initialization approach see Yu et al., 2015) and optimized through expectation maximization was used. This method was able to detect threats, but was only tested on a small set of images and 2 types of threats. In Yeom et al. (2015), the same segmentation process as in Yeom et al. (2012) was used, with the difference being the initialization of the Gaussian mixture model using k-means clustering. Shape features from the detected object were extracted and compared with the true features with reasonable accuracy. In Agarwal et al. (2014), a mean-standard deviation-based segmentation technique was used and to detect and classify simple shape objects in MMWI images, and a probability density function for this classification was proposed. Furthermore, to improve the quality of the images the authors proposed the use of a neuronal network. In Gómez et al. (2015), a highly time efficient two-step algorithm, based on denoising and mathematical morphology, was proposed. On noisy or low contrast images, it achieved an acceptable detection rate but at the cost of a high false positive detection rate. In Kumar et al. (2016), singular value decomposition and discrete wavelet transform was used to ease the detection in MMWIs

when thresholding was utilized. After this, a neuronal network was used for target identification. In Mohammadzade et al. (2017), the authors used principal component analysis and a two-layer classification algorithm to distinguish between threats and normal objects.

Most of the reviewed techniques were oriented to obtain good quality images which were then used in tandem with simple methods to detect threats. In some cases, the methods were oriented to detect only particular threat types evaluated on small datasets. The approach we follow here is completely different, as it is entirely based on machine learning from the spatial statistical information of the gray level of the image and does not requires of geometric segmentation techniques. We base our approach on image patch classification aiming for the best accuracy/false positive trade-off. The approach is highly efficient; it can achieve a 100% true detection rate with a relatively small false positive detection rate. It can also be used to dissuade, for instance, robbery either by customers or workers. In this case a 100% true detection rate may not be achieved if *good* customers or workers are not to be disturbed. However, a trade-off between accuracy and false positive detection can always be achieved by tuning the model parameters. The proposed approach achieved very good detection scores for medium–low signal-to-noise ratio images. Preliminary results were presented at López-Tapia et al. (2016), here we provide a formal definition of the problem, new material on image pre-processing techniques, additional experiments, and a deeper discussion.

3. Image enhancement

A natural question to consider is whether image filtering (smoothing, contrast enhancement, etc.) helps the detection process. Fig. 2(c)–(d) show the result of applying mean, median, Fig. 2(e)–(f), and bilateral, Fig. 2(g)–(h), filters to observed passive images, Fig. 2(a)–(b).

We observed that to assign to each pixel an estimation of the most frequent value in its neighborhood was a better smoothing criterion. That is, for each pixel i, let $\mathbf{y}_i^{\mathrm{B}}$ denote a block around y_i in the observed image $\mathbf{y}$ and $z_i(1), \ldots, z_i(K)$ denote K independent samples with replacement from $\mathbf{y}_i^{\mathrm{B}}$. We then define:

$$q_i = \frac{\sum_{k=1}^{K} z_i(k)}{K}, \tag{1}$$

assign $\mathbf{y} = \mathbf{q}$ and repeat the process L times. The final processed image is:

$$x_i = \mathrm{median}(\mathbf{y}_i^B) \tag{2}$$

We found that $B = 5 \times 5$ and $L = 5$ produced satisfactorily processed images (see Fig. 2(i)–(j)). Note that additional contrast improvement can be observed in these images. In the experimental section, we analyze how image processing improved the performance of the classifier.

4. Patch extraction

As mentioned in the Introduction, passive scanners provide very low-resolution images (in our case, 125 × 195 pixel images). Because of the small size of the threat regions and the possibility of having several threats in the same image, threat detection was carried out using image patches. A patch is a rectangular image piece centered on a pixel.

Due to the variability of the threat sizes described above, we used patches at three scales on each pixel. These corresponded to patch sizes of 39 × 39, 19 × 19, and 9 ×9 pixels, respectively. Only pixels whose three patches were fully contained in the image (active pixels) were considered. We used $\mathbf{x}_j^P$ to denote the 39 × 39 patch centered on pixel j.

Let P_I be a mapping that selects from an image $\mathcal{I}_i$ in the image dataset $\mathcal{D} = \{\mathcal{I}_i, i = 1, \ldots, N_I\}$ a subset of 39 × 39 patches $\mathcal{P}_I^i$, specifically, $P_I(\mathcal{I}_i) = \mathcal{P}_I^i$. From each image, we extracted one 39 × 39 patch every 2 × 2 pixels, obtaining a total of 3476 patches per image. From the full dataset, we obtained 11,502,084 patches (see Appendix). Patches that fully covered a threat were labeled "positive" (+1); all others were labeled "negative". (−1). Patches that partially overlapped a threat were excluded from the training dataset. We acquired 392,494 positive instances and 9,026,123 negative instances. As the number of negatives was much higher (approximately 23 times), and considering that most of the negative patches were very similar, we used a subset of these for training. We kept one negative sample from every 2 × 2 image block, meaning that we retained 1/4 of the negative patches we acquired. The final number of negative samples (2,256,530) was approximately five times the number of positive ones. These patches constituted the patch dataset $\mathcal{T} = \cup_{i=1}^{N_I} \mathcal{P}_I^i$, associated with the image dataset $\mathcal{D}$. Feature vectors were then extracted from each patch.

5. Feature extraction

All classifiers to be used in this work needed to be fed with feature vectors. The pixel values in our 39×39 patches can be seen as a high-dimensional feature vector where local spatial information is hidden in the correlation between subsets of elements. In order to obtain an adequate vector of more useful characteristics for each patch, where: (a) spatial information is explicitly represented; and (b) individual characteristics are as uncorrelated as possible, two multi-resolution feature extraction techniques were utilized. Although the literature on feature extraction is extensive, (for example, Nixon and Aguado, 2012), we focused on features expected to have good threat detection capabilities.

We used Haar filters (Papageorgiou et al., 1998) and local binary patterns (LBP) codes (Ojala et al., 1996) to create the feature vectors. The corresponding feature vectors are denoted by $\mathbf{x}_j^F$, with $F = LBP$ and $F = Haar$, respectively. It is well-known that both filter banks extract good neighboring features from an image; both features can be computed very efficiently.

5.1. Haar filters

Haar filters (Papageorgiou et al., 1998) compute correlations of different binary patterns with an image region. The pixel image values in a white pattern are added and the difference between these values and the sum of pixel image values in the corresponding black pattern is calculated (see Fig. 3(a)–(j)). These patterns are expected to obtain very positive values at patches containing a threat as body regions have lower pixel values than threat regions. However, their responses will be close to zero on pure background or body regions. Notice, for instance, that the filter shown in Fig. 3(a) shares the pattern of the hidden object in the row 1, column 2 image in Fig. A.9. Similarly, the filters in Fig. 3(b)–(c) are similar to the hidden object areas in the row 1, column 1 and row 1, column 3 images in Fig. A.9, respectively. Although the gray levels of hidden objects and background are similar, hidden objects can be recognized as they are attached to the body, a darker region. We have used 115 filters on each patch resulting in a 3 × 115 feature vector per active pixel, which will be denoted by $\mathbf{x}_j^{Haar}$.

5.2. Local binary patterns

Local binary patterns (LBP) (Ojala et al., 1996) capture image local structures by detecting gray level changes around each pixel. For every pixel in a patch, a binary vector is computed by checking if its value is greater than the value of a fixed number n of pixels located at a given radius r around it. This binary vector is coded as an integer number. The LBP feature vector of a patch is then calculated as the histogram of these numbers obtained for all pixels in the patch. Different radii can be used and the resulting feature vectors can be concatenated resulting in a larger number of features. Regions with an inside strong contrast due to the presence of boundaries between a hidden object and body or background are highlighted. In this experiment, we used the invariant-to-rotation LBP extension proposed in Ahonen et al. (2009)

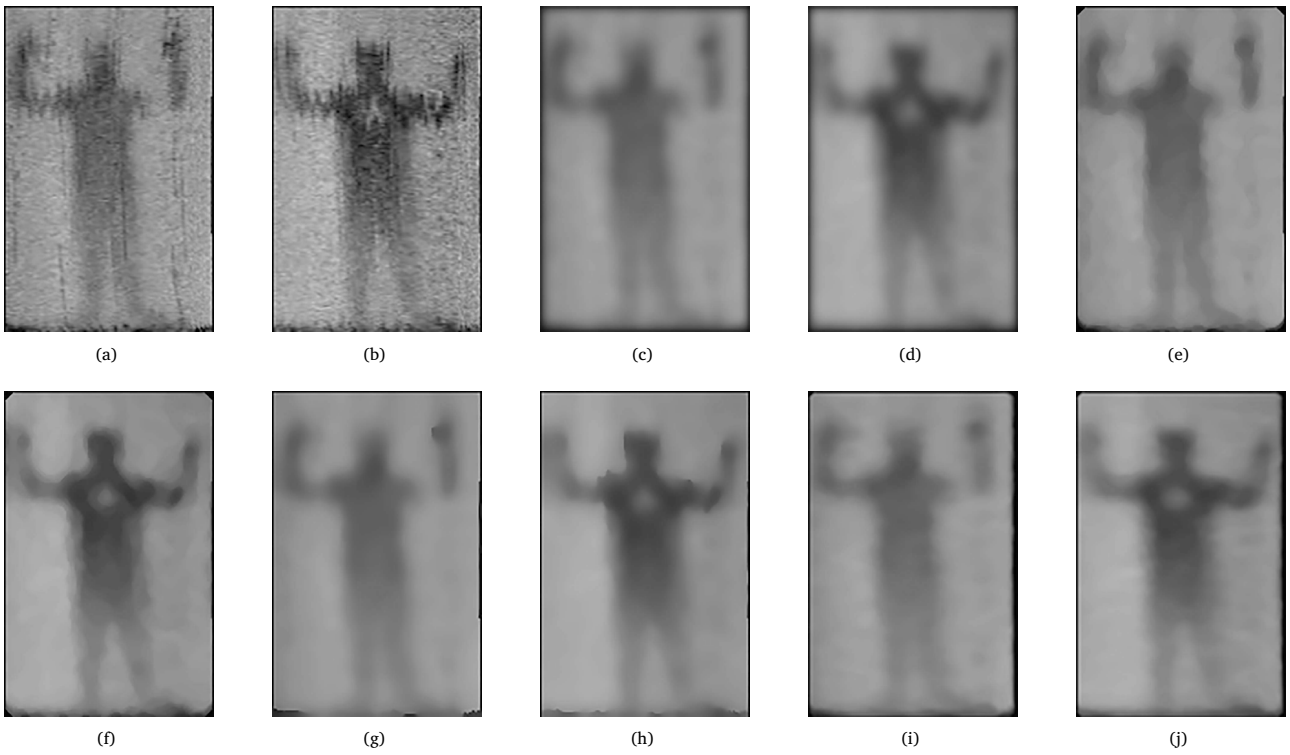


Fig. 2. Results of applying different image filtering techniques to (a) and (b). See text for details.

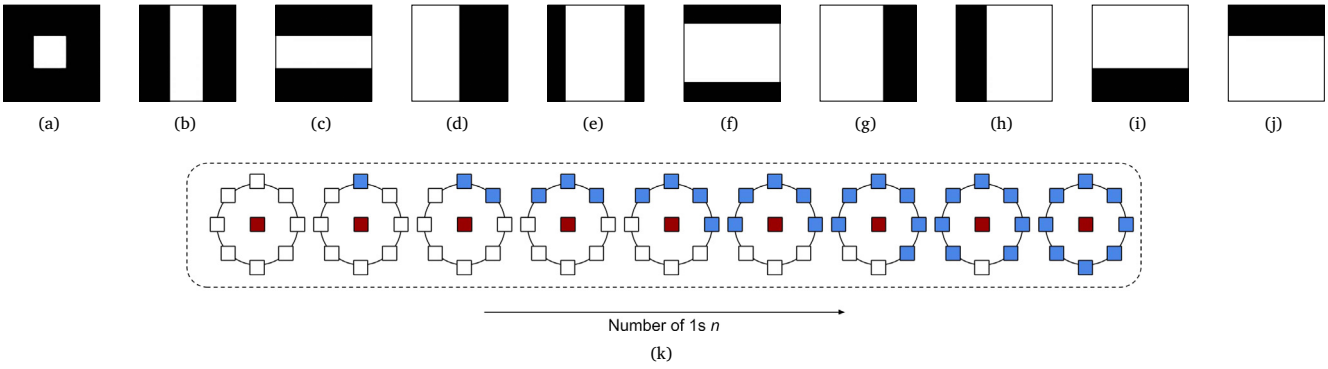


Fig. 3. Examples of Haar filters (a)–(j) and invariant to rotation LBP patterns for an 8 point neighborhood (k).

(see Fig. 3(k)). For each of the three patches described in the previous section and centered on an active pixel, the histogram of all LBP configurations was obtained using a radius $r \in \{1, \ldots, 4\}$, and several points $n = r * 8$ were built. The feature vector on each patch was obtained by concatenating the histograms of its three associated regions. In this case, there are 261 components and the feature vector will be denoted by $\mathbf{x}_j^{LBP}$.

In the following section we formally define the statistical learning framework we used to detect threats in PMMWIs. Its basic building blocks will be the image patches and the features extracted from them.

6. Model

Given our image dataset $\mathcal{D}$ (see Section 4), our goal was to learn a labeling function $f_{\mathcal{D}} : \mathcal{D} \rightarrow \{0,1\}^s$, where s is the image size, with the lowest generalization error. This function assigns to each image pixel a binary value $\{0,1\}$ indicating whether the pixel is part of a threat (1) or not (0). Thus, our problem was defined as an object localization problem using machine learning. We defined the function $f_{\mathcal{D}}$ in two steps. First, we detected potential threat regions (patches). Second, we determined which pixels were considered to belong to real threats. The full image dataset $\mathcal{D}$ was used to learn the function $f_{\mathcal{D}}$, and its generalization error was estimated using cross validation.

We used the patch dataset $\mathcal{T}$ defined in Section 4 to learn a patch detection function $f_{\mathcal{P}} : \mathcal{T} \rightarrow [0,1]$. We assumed, without loss of generality, that the range of this function was the interval $[0,1]$ (note that normalization can always be used). A binary classifier was built by splitting the range of $f_{\mathcal{P}}$ using a threshold to be learned.

Since the patch dataset $\mathcal{T}$ (see Section 4) has five times more negative than positive patches, the set of negative patches used to learn an ensemble of classifiers was partitioned. Let $\mathcal{T}_P, \mathcal{T}_N$ define the subsets of positive and negative labeled patches respectively $\mathcal{T} = \mathcal{T}_P \cup \mathcal{T}_N$. Let us denote by n_P and n_N their cardinals, where $n_P \ll n_N$ and $n_C = n_N/n_P$. Let $\mathcal{T}_N = \cup_{i=1}^{n_C} \mathcal{T}_N^i$ be a random decomposition of the set $\mathcal{T}_N$ in n_C disjointed subsets. We solved n_C learning problems $f_{\mathcal{P}}^k$, $k = 1, \ldots, n_C$, associated to the training sets defined by $\{\mathcal{T}_N^k \cup \mathcal{T}_P\}$ respectively. We repeated the same procedure t times, obtaining an ensemble of $t \times n_C$ functions $f_{\mathcal{P}}^k$, which use the positive and subsets of the negative patches.

Table 1
Hyperparameter grid for the classifiers; see text for details.

Classifier	Parameter	Range	Classifier	Parameter	Range
LR	C_{LR}	[0, 5]	SVM	C_{SVM}	$[10^{-1}, 10^{5}]$
QLR	C_{QLR}	[0, 5]		γ_{RBF}	$[10^{-5}, 10]$
	NT	[100, 300]		NT	[100, 300]
RF	MNF	[8, 30]	ERF	MNF	[8, 30]
	MEL	[1, 50]		MEL	[1, 50]

Following this, the patch detection function was defined as $f_{\mathcal{P}}(\mathbf{x}^P) = \frac{1}{t \times n_C} \sum_k f_{\mathcal{P}}^k(\mathbf{x}^P)$.

We associated to the patch detection function $f_{\mathcal{P}}$ a detection threshold, $thr \in [0, 1]$. We set $f_{\mathcal{P}}(\mathbf{x}^P) = 0$ if $f_{\mathcal{P}}(\mathbf{x}^P) < thr$ and initially did not change $f_{\mathcal{P}}(\mathbf{x}^P)$ if $f_{\mathcal{P}}(\mathbf{x}^P) \geq thr$. To calibrate the threshold value, we used the ROC curve obtained from $f_{\mathcal{P}}$ when used as a binary classifier. Although $f_{\mathcal{P}}(\mathbf{x}^P) \geq thr$ declares $\mathbf{x}^P$ a potential threat patch (not all with the same detection function value), some of these potential threat patches are false positives that appear in the neighborhood of a true threat due to the contamination of the highest patch detection value to surrounding patches. We eliminated these contaminated patches using non-maximum suppression on overlapping patches. On each image, we rejected a detected patch $\mathbf{x}_j^P$ ($f_{\mathcal{P}}(\mathbf{x}_j^P) \geq thr$) if it had an intersection-over-union overlap larger than a learned threshold with a higher scoring patch $\mathbf{x}_l^P$. That is, if $f_{\mathcal{P}}(\mathbf{x}_l^P) > f_{\mathcal{P}}(\mathbf{x}_j^P) \geq thr$, we set $f_{\mathcal{P}}(\mathbf{x}_j^P) = 0$. This process is depicted in Fig. 4.

Due to the strong no-stationary noise presence in the image, this approach can produce a high rate of false positive patches even when the thr value is properly tuned. Our solution to this problem is given in Section 8, where a calibration curve, to estimate the best rate between accuracy and false detection rates, is proposed.

After running the non-maximum suppression algorithm, we assigned to each image pixel, k, the maximum of all $f_{\mathcal{P}}(\mathbf{x}^P)$, for $k \in \mathbf{x}^P$. After this final processing, we acquired an image size vector of values in the interval [0,1]. This vector represents our estimation of the labeling function $f_{\mathcal{D}}$ on each image.

7. Learning the patch detection function $f_{\mathcal{P}}$ and the image classification function $f_{\mathcal{D}}$

We considered six patch detection functions $f_{\mathcal{P}}$. These functions represent different strategies to measure the presence of a threat in a patch: logistic regression (LR), quadratic logistic regression (QLR), support vector machine (SVM), random forest (RF), extreme random trees (ERT), and adaboost (ADA). LR and QLR were the linear approaches used as our baseline in the comparative study, and SVM was the kernel-based approach. These three methods search for the best function from a fixed set of possible functions. ADA, RF, and ERT build a classifier from an ensemble of simpler classifiers. RF and ERT use different subsets of the training dataset for each member of the ensemble, and ADA weights the training dataset with different sets of values. In all cases, the patch detection function range is normalized to $[0, 1]$.

For each classifier M∈ {LR, QLR, SVM, RF, ERT}, the corresponding image classification function f_D^M was estimated and the corresponding training and error estimations were calculated using five-fold cross-validation. This cross–validation partition was performed on the images and on every fold. Each fold contained approximately 600 images, all of which had the same proportion of images with no threat, one threat, and two threats. For adaboost, we used asymmetric boosting (Viola and Jones, 2001) instead of a committee of classifiers.

Hyperparameter estimation was carried out using five-fold cross-validation on every training fold. To reduce the time required to estimate each the hyperparameters of each model, a smaller subset of patches was selected. A uniform sample of patches, using an additional factor reduction of 3, was used. This required selecting one location for every 6 × 6 image block. Finally, for each threat in an image, we added at least one patch that contained it to the training dataset. This guaranteed the inclusion of all the threats in the database. The hyperparameter ranges for each patch detection function are shown in Table 1. They are the regularization strength parameters C_{LR}, C_{QLR}, and C_{SVM} for LR, QLR, and SVM, with RBF kernel with gamma parameter γ_{RBF}, respectively. The estimated hyperparameters of the tree-based models RF y ERT are the number of trees (NT), the maximum number of features (MNF) to consider for splitting, and minimum number of examples per leaf (MEL).

The LR the regularization parameter (C_{LR})for quadratic penalty (weight-decay) was determined using an adaptive search. QLR uses linear and quadratic functions on the features, hence, the LASSO penalty function was utilized to select the relevant features which were then used to learn an LR model with a quadratic penalty. For SVM, RF, and ERT, a grid search was used to estimate the hyperparameters, and feature vectors were normalized by mean and variance before training.

To remove the possible bias introduced by the reduced number of patches used to learn the hyperparameters, we selected the three best sets of hyperparameters using the area value under the ROC curve for each classifier. These three sets were then compared to the complete set of training patches and the one with the largest AUC value was selected.

8. Experimental validation

For a given image, let us denote by $\mathcal{H}_S(\mathbf{h})$ and $\mathcal{P}_S(\mathbf{x}^P)$ the support regions (sets of pixels) associated with a hidden object $\mathbf{h}$ and patch $\mathbf{x}^P$, respectively. Given an existing real hidden threat $\mathbf{h}$ in an image, we considered in the experiments that a patch $\mathbf{x}^P$ was correctly classified as containing a hidden threat when $f_{\mathcal{P}}(\mathbf{x}^P) > thr$ and $\mathcal{H}_S(\mathbf{h}) \cap \mathcal{P}_S(\mathbf{x}^P) > 0.5 * size(\mathcal{H}_S(\mathbf{h}))$. A hidden object in the image was said to be correctly detected when there was at least one patch satisfying both conditions. Finally, an image was correctly classified as positive (containing a hidden threat) when, at least, a hidden object was detected.

We analyze two scenarios: preprocessed images and raw images. Our preprocessing technique was described in Section 3, and Tables 2 and 3 contain a summary of the results for both scenarios. The AUC column represents the area under the ROC curve for the test images when classified as with or without a hidden object. The third column (TP) shows the true positive percentage of detected hidden objects computed by five-fold cross-validation; fold mean and standard deviation are included. The fourth column shows the average number of false positive (FP) patches per image and their standard deviations. Note the high FP deviation values, which are due to the non-uniform quality of the images. The fifth and sixth columns contain the threshold and overlap parameters used for non-maximum suppression.

As Table 2 indicates, the best results are obtained by RF with Haar features extracted from the preprocessed images using the algorithm proposed in Section 3. Note that we also ran experiments with mean, median, and bilateral smoothing algorithms but obtained lower scores. Almost all classifiers obtained similar TP values, RF and ERT were the best ones when used with Haar features.

Table 3 shows the scores for the raw image scenario. In this case, RF and ERT again obtained the best results for Haar and LBP features. The AUC values were lower, and the best value was obtained by the ERT classifier. However, in terms of TP and FP values, the average results were similar for both scenarios. That is, the proposed preprocessing method only slightly helped the detection process, which shows that our machine learning-based classification approach can deal with poor-quality images.

In Table 2 it can also be observed that TP values for Haar and LBP features are very similar. However, when considering FP values, Haar features clearly outperform LBP features. As we have indicated in the Introduction, the average FP per image is a key figure to minimize. When analyzed in percentage terms, all models performed reasonably well. This means that FP values were always below 10% when the threshold was fixed to obtain a 100% TP detection rate (0% FN).

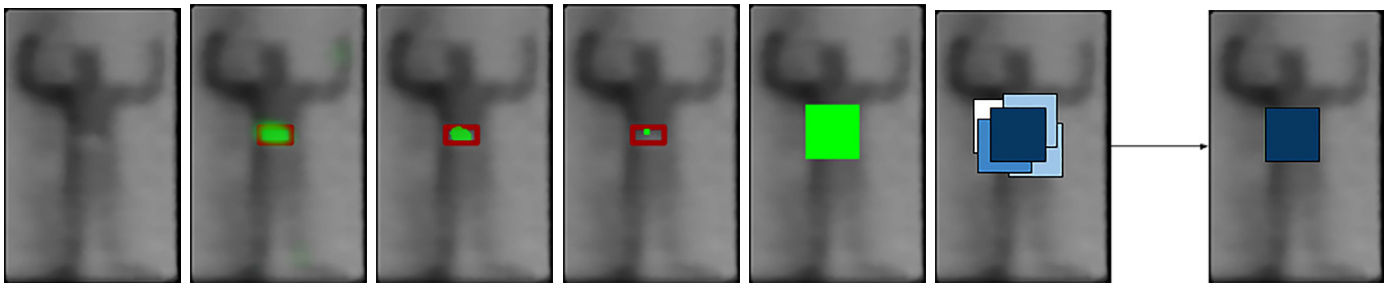


Fig. 4. Detection process: (a) observed image; (b) shows, using green intensities, the initial pixel probabilities (only at patch central pixels) assigned by the patch detection function together with the red bounding box surrounding the threat; (c) shows the active pixels after thresholding by thr; (d) shows the final estimation of the threat center; (e) shows the corresponding patch after non-maximum suppression on (c); (f) shows the details of the non-maximum suppression process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
For the preprocessed images using the preprocessing method presented in 3, this table shows a summary of the performance (AUC) of the classifiers when Haar and LBP features were used. Threshold (Thr) and Overlapping values (Ovl) are also included. See text for details.

Class	Haar features (preprocessed)					LBP features (preprocessed)				
	AUC	TP × 10^2	FP	Thr.	Ovl.	AUC	TP × 10^2	FP	Thr.	Ovl.
LR	0.51	84 ± 1.8	10.5 ± 10.7	0.6	0.5	0.57	92 ± 0.8	10.2 ± 10.5	0.70	0.5
QLR	0.57	91 ± 1.1	8.7 ± 9.1	0.7	0.5	0.57	92 ± 1.0	10.2 ± 10.5	0.70	0.5
SVM	0.52	92 ± 3.0	6.5 ± 6.6	0.75	0.3	0.55	93 ± 1.4	8.8 ± 8.9	0.60	0.4
RF	**0.75**	**94 ± 0.9**	**4.0 ± 3.8**	**0.7**	**0.3**	0.61	92 ± 0.5	6.6 ± 6.5	0.55	0.3
ERT	0.74	94 ± 0.6	5.0 ± 5.0	0.7	0.5	0.63	90 ± 1.2	6.1 ± 6.0	0.55	0.3
ADA	0.57	93 ± 1.1	6.4 ± 6.2	0.5	0.3	0.59	92 ± 1.2	10.0 ± 10.2	0.50	0.5

Table 3
For raw images, this table shows a summary of the performance (AUC) of the classifiers when Haar and LBP features were used. Threshold (Thr) and Overlapping values (Ovl) are also included. See text for details.

Class.	Haar features (raw)					LBP features (raw)				
	AUC	TP × 10^2	FP	Thr.	Ovl.	AUC	TP × 10^2	FP	Thr.	Ovl.
LR	0.51	82 ± 1.5	10.2 ± 10.4	0.6	0.5	0.54	88 ± 1.0	10.4 ± 10.8	0.68	0.5
QLR	0.55	90 ± 1.0	8.2 ± 8.7	0.73	0.5	0.56	89 ± 1.0	10.6 ± 10.9	0.7	0.5
SVM	0.50	94 ± 1.0	6.7 ± 6.6	0.75	0.3	0.52	93 ± 6.0	11.5 ± 12.1	0.7	0.4
RF	0.72	94 ± 1.0	4.0 ± 3.7	0.7	0.3	0.60	86 ± 2.0	6.1 ± 6.1	0.58	0.3
ERT	0.73	94 ± 1.0	4.0 ± 3.8	0.68	0.3	0.59	84 ± 1.0	6.0 ± 6.0	0.57	0.3
ADA	0.63	94 ± 0.6	5.8 ± 5.7	0.50	0.3	0.55	90 ± 1.0	10.3 ± 10.5	0.50	0.5

However, a 10% FP value means that a very large number of patches were incorrectly considered to contain a threat, which means that the model became useless. Consequently, a better compromise between TP and FP scores must be reached by selecting a higher thr_M value which, however, will reduce the TP percentage. Fig. 5(a) shows the ROC curve for the best model combination, RF + Haar features on the preprocessed images. Fig. 5(b) shows the true positive and true negative rate curves for image classification. Their intersection point is slightly above 68% and defines the accuracy of the system to classify new images when FP = 1 − TP, i.e., both false errors are equal. Although this score might be considered low, it is important to note that the high slope of both curves (positive rates, negatives rates) at the intersection point indicates that it is possible to improve the true positive rate with a small threshold increment, but only if the increase in FPs is affordable. Fig. 6 shows the lack of contrast in some FN images and, therefore, the difficulty in detecting some hidden objects without increasing the FP rate.

The success of tree-based methods can be explained by the fact that both RF and ERT minimized mean square error using ensemble voting, which is an effective approach to reduce noise in images. Regarding LBP, the figures in both tables show a high influence of noise on the quality of this feature type. Although LBP features yielded better detection results than Haar features when used with parametric classifiers (LR, QLR, SVM), their FP scores were too high to make them competitive. These results clearly demonstrate the influence of noise on the behavior of classifiers and the importance of preprocessing when LBP features are used.

Fig. 7 shows the analysis of the behavior of our winning combination, Haar + RF. In Fig. 7(a), the histogram of the average number of FPs per image is shown. We observed that the mode of the average of FPs per image was 4. Fig. 7(b) shows the histogram of the first TP position on each image in the list of detected locations, ordered by decreasing probability. This histogram indicates that the clear majority of TPs were among the first two detected positive patches on each image, which validates the use of a high threshold for object detection. Fig. 7(c) shows the histogram of the mean value of the four highest probability values per image, and the correct behavior of the classifier assigning the highest probabilities to regions that overlapped with hidden objects.

Table 4(a)–(b) show how Haar + RF performed depending on threat location and type. Variations in TP and FP were highly dependent on the position of the object. When objects were in the body area (chest, stomach, waist, armpit) detection was simpler. However, when they were mainly surrounded by the background (arm, thigh, ankle), they were more difficult to detect. Table 4(b) shows a clear ranking of the objects depending on their PMMW responses. While the FP values were similar, the TP scores showed a large variation. Objects with higher densities (such as metallic objects) were easier to detect.

We also investigated whether models tailored to body regions improve the quality of the image classification function. For this, we used our best f_p function (RF + Haar) to fit four new models, one for each of the following body regions: arms, legs, chest, and ankles. Thus, only patches from one of these regions were used by each model. For test

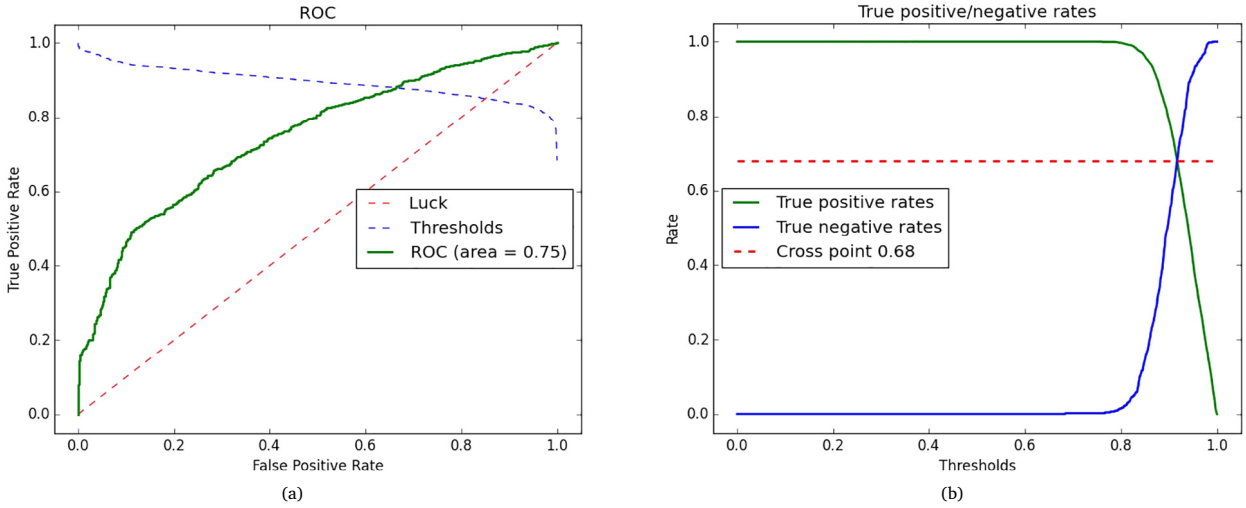


Fig. 5. (a) Shows the ROC when classifying new images. (b) The curves show the accuracy of the model on new images for a range of detection probability thresholds. The cross point is at the 68% of accuracy for both classes.

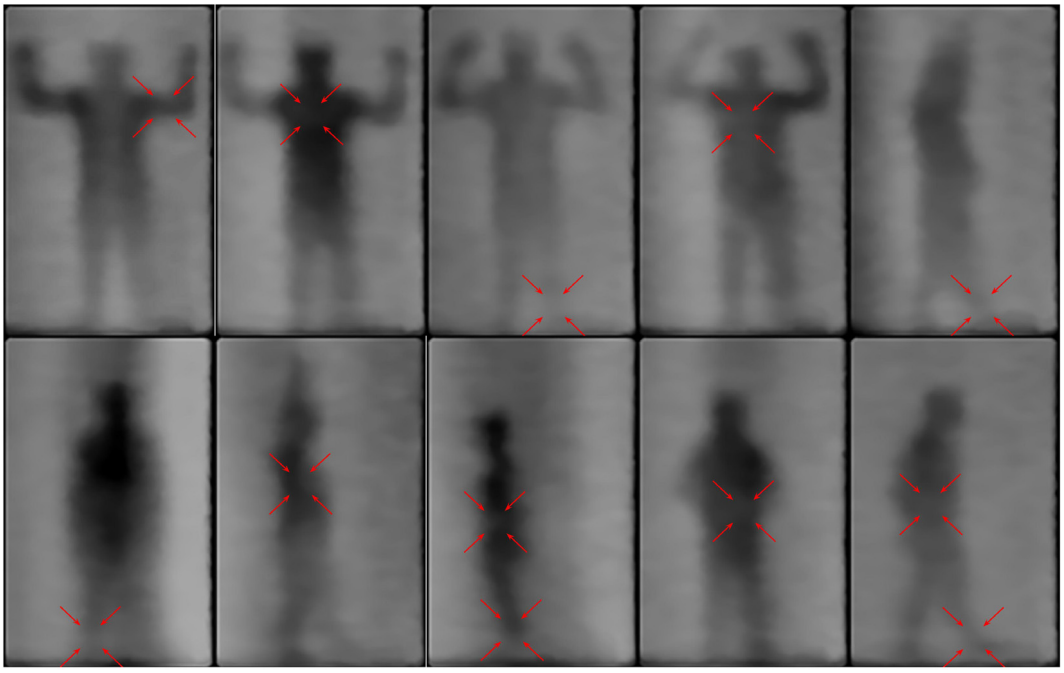
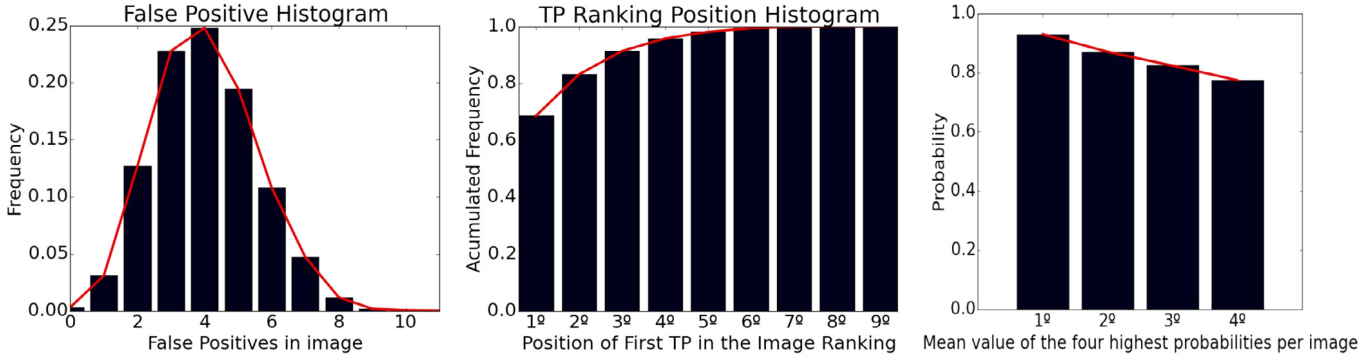


Fig. 6. Examples of hidden objects that RF could not detect when the classifier was calibrated for the same FP and FN rates. Red arrows indicate object locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
The results of the best method (RF with Haar features). (a) Based on the threat location (front (F) and lateral (L)). (b) Based on object type (see Appendix). See text for details.

(a)				(b)			
Loc.	AUC	TP × 10^2	FP	Threat	AUC	TP × 10^2	FP
Forearm	0.73	97 ± 11.2	4.4 ± 4.1	1	0.72	94 ± 7.8	4.5 ± 4.3
Chest	0.82	95 ± 11.6	4.4 ± 4.1	2	0.83	95 ± 10.6	4.5 ± 4.3
Stomach	0.86	96 ± 11.6	4.3 ± 4.1	3	0.68	90 ± 9.3	4.5 ± 4.3
Thigh	0.53	85 ± 14.0	4.7 ± 4.5	4	0.72	94 ± 10.7	4.5 ± 4.2
Ankle-F	0.65	88 ± 8.3	4.4 ± 4.2	5	0.72	93 ± 7.6	4.5 ± 4.3
Thigh-L	0.75	97 ± 9.2	3.1 ± 2.9	6	0.81	96 ± 8.7	4.5 ± 4.3
Waist-L	0.85	95 ± 8.3	3.0 ± 2.8	7	0.81	96 ± 2.9	4.5 ± 4.3
Armpit-L	0.86	99 ± 12.0	3.0 ± 2.7	8	0.82	95 ± 8.0	4.5 ± 4.3
Arm	0.55	85 ± 14.4	5.3 ± 4.9	9	0.79	95 ± 16.4	4.5 ± 4.3
Ankle-L	0.65	92 ± 16.5	3.8 ± 3.5	10	0.64	90 ± 10.8	4.5 ± 4.3
				11	0.79	94 ± 8.1	4.5 ± 4.3
				12	0.83	96 ± 12.7	4.5 ± 4.3



(a) (b) (c)

Fig. 7. Performance histograms of the best model, Haar + RF. See text for details.

Table 5
Testing Total (all the images) and per image (P.I.) times in seconds for each method.

Classifier	Haar		LBP		Classifier	Haar		LBP	
	Total	P.I.	Total	P.I.		Total	P.I.	Total	P.I.
LR	302	0.09	4939	1.49	RF	**740**	**0.22**	**5497**	**1.66**
QLR	341	0.1	4939	1.49	ERT	1202	0.36	5507	1.66
SVM	1277	0.38	6907	1.87	ADA	1965	0.59	7006	2.11

samples, the image localization of the patch determined the model to be used, and a partition of the image in the four body parts was prefixed. In terms of AUC, values there was no significant improvement. However, a 5% improvement on average FP was observed.

In the experiments, we used an Intel Xeon E5-2630 v3 at 2.40 GHz with 8 cores and 128 GB of RAM. Computation times during testing are shown in Table 5. We observed that for both feature types, the methods performed approximately in real time. However, when using Haar features the classifiers performed faster. Finally, for RF and ERT, we used parallel implementations running on 16 threads.

The approach presented in this paper cannot be fairly compared with the methods described in Section 2. The methods reported in Section 2 were tested on small sets of higher quality images, with a reduced number of threat locations, and with reduced variability on the threat characteristics. Nonetheless, we compared our best method, RF + Haar, against the methods described in Haworth et al. (2006) and Gómez et al. (2015) using datasets of different sizes. The best performing method was (Haworth et al., 2006), which, for a dataset of 200 images, obtained a 13% TP and 1.92 mean FP. Our worst result was 47% TP and 1.5 mean FP, and was obtained with 30 raw images. We believe that our method performed better because it is more generally applicable and less adapted to the quality of the acquired images.

Fig. A.8. Simulated threats in the dataset: a cutter (1), 325 g of gel (2), a 200 g clay bar (3), a simulated gun (4), 200 g of sugar (5), 200 g of frozen peas (6), 150 ml of cologne (7), 160 g of gel (8) , a bag with metal pieces (9), 200 g of flour (10), a 50cl water bottle (11), and a 250 ml hydrogen peroxide bottle (12).

9. Conclusions

This paper was devoted to the study of hidden object detection in PMMWIs. The main difficulty in this task arises from the low SNR and non-stationary noise that populates an image. Simple thresholding methods can be used but are most effective with high-quality images. In this study, a machine learning approach to the detection task was developed. This approach deals with the poor quality of passive images and outperforms state-of-the-art threat detection methods for PMMWIs.

Given the lack of publicly available PMMWI datasets, we created one that, to the best of our knowledge, is the largest, and possesses the greatest variety of object types and sizes ever used for this purpose see footnote 1.

Our proposed method is based on a committee of classifiers defined on two highly unbalanced classes of image patches, and performed well on all experiments. We compared different approaches to estimate image classification functions, and found using tree sets to be the most effective, reaching an average 94% TP score with a distribution of the number of false positives in the range of one to seven. The influence of the image quality and the extracted features were also analyzed. Our filtering method slightly helps the detection process; Haar filter banks, very well adapted to our task, performed very well for all classifiers.

The results indicate that large objects with reduced or zero emissions are simpler to detect. The easiest threat locations to detect were those where objects were exposed to the camera in larger areas. Threats in ankles, arms, and thighs were more difficult to detect.

Finally, a comparison between our detection model and other approaches in the literature indicated that our method is less reliant on the quality of the observed images. Furthermore, when a large image training set is available, our method performs very well, which makes a prediction of excellent performance for a wide range of millimeter-based detection systems realistic.

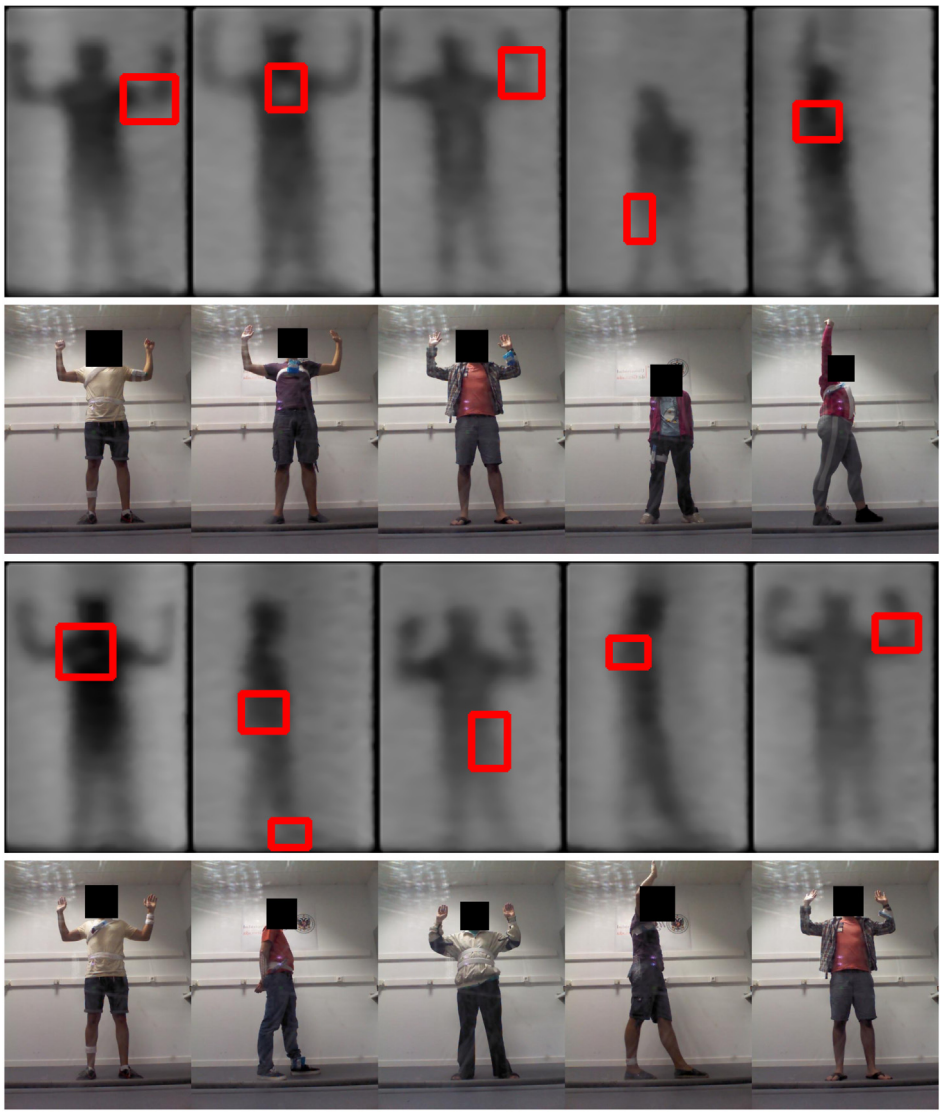


Fig. A.9. First row: Examples of PMMWIs. Red boxes indicate the locations of hidden objects. Second row: Corresponding visual images of the PMMWIs' examples. (Best viewed on a high-resolution color screen). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

Wavecamm (http://wavecamm.com) provided the PMMWIs used in the paper This work was supported by the Spanish MINECO under project TIN2013-43880-R, FEDER funds, and by Project TIC-1692 (Junta de Andalucía).

Appendix. Dataset description

A comprehensive dataset of PMMWIs was created. It consisted of 3309 125 × 195 images of 33 people of different complexions. The hidden objects were in the range 35 × 39 to 10 × 10 pixels which, in our images, corresponded roughly to 2752.39 cm^2 and 201.64 cm^2, respectively. Smaller hidden objects were not considered relevant.

Threats were simulated by bags containing substances with different millimeter wave responses (see Fig. A.8). Note that objects of different sizes were used. We took pictures of each person wearing 12 objects on 10 parts of the body: forearm, chest, stomach, thigh, ankle (front), waist (side), armpit (side), arm, ankle (lateral) thigh (lateral), and 20 images without any object. Images of people wearing two objects in different locations were also taken. Some images were eliminated because of poor quality. The final dataset consisted of 463 pictures of people with no objects, 2144 containing people wearing one object, and 702 containing two objects.

Fig. A.9 shows PMMWIs of subjects with simulated threats on different parts of the body and the corresponding color images taken by a camera located on the scanner. Although objects are visible and not hidden under the clothing, this was irrelevant for the PMMW sensors. Color images were taken at the same time. Threats are marked on the color images by the smallest bounding boxes containing them. To transfer object-bounding boxes from colored to PMMW images, a homography was estimated using a planar calibration pattern. Colored and PMMW images were then registered. The bounding boxes were used to assess the performance of the image classification functions.

References

Agarwal, S., Bisht, A.S., Singh, D., Pathak, N.P., 2014. A novel neural network based image reconstruction model with scale and rotation invariance for target identification and classification for Active millimetre wave imaging. J. Infrared Millim. Terahertz Waves 35 (12), 1045–1067.

Ahonen, T., Matas, J., He, C., Pietikäinen, M., 2009. Rotation invariant image description with local binary pattern histogram fourier features. In: 16th Scandinavian Conference on Image Analysis, pp. 61–70.

Alexander, N.E., Callejero Andrés, C., Gonzalo, R., 2008. Multispectral mm-wave imaging: materials and images. In: SPIE, Vol. 6948, pp. 694803–694812.

Babacan, S.D., Luessi, M., Spinoulas, L., Katsaggelos, A.K., Gopalsami, N., Elmer, T., Ahern, R., Liao, S., Raptis, A., 2011. Compressive passive millimeter-wave imaging. In: 8th IEEE International Conference on Image Processing, pp. 2705–2708.

Fang, H., Shi, Y., Pan, D., Zhou, G., 2017. Iteratively reweighted blind deconvolution for passive millimeter-wave images. Signal Process. 138, 182–194.

Gómez, I., Pérez de la Blanca, N., Molina, R., Katsaggelos, A., 2015. Fast millimetre wave threat detection algorithm. In: 23rd European Signal Processing Conference, pp. 599–603.

Han, B., Xiong, J., Li, L., Yang, J., Wang, Z., 2010. Research on millimeter-wave image denoising method based on contourlet and compressed sensing. In: 2nd International Conference on Signal Processing Systems, Vol. 2, pp. 471–475.

Haworth, C.D., De Saint-Pern, Y., Clark, D., Trucco, E., Petillot, Y.R., 2007. Detection and tracking of multiple metallic objects in millimetre-wave images. Int. J. Comput. Vis. 71 (2), 183–196.

Haworth, C., Gonzalez, B., Tomsin, M., Appleby, R., Coward, P., Harvey, A., Lebart, K., Petillot, Y., Trucco, E., 2004. Image analysis for object detection in millimetre-wave images. In: Passive Millimetre-Wave and Terahertz Imaging and Technology, Vol. 5619. pp. 117–128.

Haworth, C., Petillot, Y., Trucco, E., 2006. Image processing techniques for metallic object detection with millimetre-wave images. Pattern Recognit. Lett. 27 (15), 1843–1851.

Kumar, B., Sharma, P., Upadhyay, R., Singh, D., Singh, K.P., 2016. Optimization of image processing techniques to detect and reconstruct the image of concealed blade for MMW imaging system. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 76–79.

Liu, T., Chen, Z., Liu, S., Zhang, Z., Shu, J., 2016. Blind image restoration with sparse priori regularization for passive millimeter-wave images. J. Vis. Commun. Image Represent. 40, 58–66.

López-Tapia, S., Molina, R., Pérez de la Blanca, N., 2016. Detection and localization of objects in Passive Millimeter Wave Images. In: 24th European Signal Processing Conference, pp. 2101–2105.

Martínez, O., Ferraz, L., Binefa, X., Gómez, I., Dorronsoro, C., 2010. Concealed object detection and segmentation over millimetric waves images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 31–37.

Mateos, J., López, A., Vega, M., Molina, R., Katsaggelos, A., 2016. Multiframe blind deconvolution of passive millimeter wave images using variational Dirichlet blur kernel estimation. In: IEEE International Conference on Image Processing, pp. 2678–2682.

Mohammadzade, H., Ghojogh, B., Faezi, S., Shabany, M., 2017. Critical object recognition in millimeter-wave images with robustness to rotation and scale. J. Opt. Soc. Am. A 34 (6), 846–855.

Nixon, M., Aguado, A.S., 2012. Feature Extraction & Image Processing, second ed. Academic Press.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 29, 51–59.

Oka, S., Togo, H., Kukutsu, N., Nagatsuma, T., 2008. Latest trends in millimeter-wave imaging technology. Progr. Electromag. Res. Lett. 1, 197–204.

Papageorgiou, C., Oren, M., Poggio, T., 1998. A general framework for object detection. In: 6th International Conference on Computer Vision, pp. 555–562.

Saafin, W., Villena, S., Vega, M., Molina, R., Katsaggelos, A., 2016. Compressive sensing super resolution from multiple observations with application to passive millimeter wave images. Digit. Signal Process. 50, 180–190.

Sarkis, M., Mani, L., 2012. Reconstruction of passive millimeter-wave images with Graph Cuts. In: 19th IEEE International Conference on Image Processing, pp. 2053–2056.

Shen, X., Dietlein, C.R., Grossman, E., Popovic, Z., Meyer, F.G., 2008. Detection and segmentation of concealed objects in terahertz images. IEEE Trans. Image Process. 17 (12), 2465–2475.

Viola, P., Jones, M., 2001. Fast and robust classification using asymmetric adaboost and a detector cascade. In: Advances in Neural Information Processing Systems, Vol. 14. pp. 1311–1318.

Yang, J., Wang, J., Li, L., 2010. A new algorithm for passive millimeter-wave image enhancement. In: 2nd International Conference on Signal Processing Systems, Vol. 3, pp. 507–511.

Yeom, S., Lee, D.-S., Jang, Y., Lee, M.-K., Jung, S.-W., 2012. Real-time concealed-object detection and recognition with passive millimeter wave imaging. Opt. Express 20 (9), 9371–9381.

Yeom, S., Lee, D.-S., Son, J.-Y., 2015. Shape feature analysis of concealed objects with passive millimeter wave imaging. Progr. Electromag. Res. Lett. 57, 131–137.

Yu, W., Chen, X., Dong, S., Shao, W., 2011. Study on image enhancement algorithm applied to passive millimeter-wave imaging based on wavelet transformation. In: International Conference on Electrical and Control Engineering, pp. 856–859.

Yu, W., Chen, X., Wu, L., 2015. Segmentation of concealed objects in passive millimeter-wave images based on the Gaussian mixture model. J. Infrared Millim. Terahertz Waves 36 (4), 400–421.

Yujiri, L., 2006. Passive millimeter wave imaging. In: IEEE MTT-S International Microwave Symposium Digest, pp. 98–101.

Yujiri, L., Shoucri, M., Moffa, P., 2003. Passive millimeter wave imaging. IEEE Microw. Mag. 4 (3), 39–50.