Examining Threat Image Projection Artifacts and Related Issues: A Rating Study

Conference Paper - October 2018
DOI: 10.1109/CCST.2018.8585428

CITATIONS
0

READS
45

3 authors:

Robin Riz À Porta
University of Applied Sciences and Arts Northwestern Switzerland
1 PUBLICATION 0 CITATIONS

Yanik Sterchi
University of Applied Sciences and Arts Northwestern Switzerland
8 PUBLICATIONS 10 CITATIONS

Adrian Schwaninger
University of Applied Sciences and Arts Northwestern Switzerland
145 PUBLICATIONS 2,043 CITATIONS

All content following this page was uploaded by Robin Riz À Porta on 05 November 2018.
The user has requested enhancement of the downloaded file.
Examining Threat Image Projection Artifacts and Related Issues: A Rating Study

Robin Riz à Porta, Yanik Sterchi, Adrian Schwaninger
School of Applied Psychology
University of Applied Sciences and Arts Northwestern Switzerland (FHNW)
Olten, Switzerland
and
Center for Adaptive Security Research and Applications (CASRA)
Zurich, Switzerland
Email: robin.rizaporta@fhnw.ch

Abstract— Threat image projection (TIP) is a widely used software function of X-ray machines at airport security checkpoints. TIP projects fictional threat images (FTIs) of actual pre-recorded threat items (mainly guns, knives and improvised explosive devices) into the X-ray images of passenger baggage before they are displayed to security officers (screeners) for screening. TIP increases attention and motivation of screeners and is often used to measure their detection performance. In order to be effective, TIP has to project FTIs in a realistic way. In other words, it should not be possible to detect FTIs by simply detecting visual artifacts resulting from TIP projection. This study was conducted to evaluate TIP quality regarding potential artifacts. First, we interviewed screeners to explore which TIP artifacts they encounter in their day-to-day work. In a second step, we conducted a rating study to quantify the identified artifacts and the quality of TIP images in general. The majority of images (80%) produced by TIP were judged by screeners to appear realistic. However, in some images FTIs were positioned inadequately: the alignment (compared to the surrounding baggage items) appeared artificial (15%) or the placement appeared physically implausible (17%; e.g. an improvised explosive device going through a heel). These two issues also significantly affected the image to appear unrealistic in general. We conclude that in most cases, TIP succeeds in projecting FTIs without creating discernable artifacts. In some cases however, the FTI is positioned inadequately, which could be improved in the future.

Keywords—airport security, X-ray screening, cabin baggage screening, operator performance, threat image projection (key words)

I. INTRODUCTION

Threat image projection (TIP) is a widely used software function of X-ray machines at airport security checkpoints. TIP projects fictional threat images (FTIs) of actual pre-recorded threat items (mainly guns, knives and improvised explosive devices, IEDs) into the X-ray images of passenger baggage before they are displayed to security officers (screeners) for screening. TIP is thought to improve the detection performance of airport security officers (screeners) by artificially increasing the occurrence of threat items. Otherwise, certain threat items would be very rare and rare targets are known to be frequently missed [1], [2]. By exposing screeners to prohibited items that they need to find in X-ray images, the screeners’ task also gets more interesting and more motivating [3]. TIP is also used to measure screeners’ detection performance by calculating the percentage of detected threat items (hit rate) and the percentage of false positive responses (non-TIP alarm rate). However, TIP only can improve detection and be a reliable measure of detection performance if TIP projections are realistic. In other words, the TIP image should look like there was a real threat item in the X-ray image of a passenger bag. Otherwise, screeners might – knowingly or unknowingly – concentrate on finding TIP projection artifacts, instead of finding real threat items. They might also question the usefulness of TIP, which would have a demotivating effect. Hence, if TIP were of bad quality, it might even lower the probability of detecting real threats. Furthermore, TIP performance would not only measure the ability to detect threats but also the ability to detect TIP artifacts.

There are two major types of TIP artifacts. The first type of artifacts refers to physically implausible placement of the FTI or lack of alignment of the FTI with other items in the bag. For example, if an FTI (e.g. IED) is placed in way that it appears to go through a heel. Such a placement is physically implausible and likely to appear unrealistic. Also, if the FTI has another alignment than the rest of the items in its surroundings (for example, in an X-ray image of an orderly packed bag, all items might appear with a horizontal orientation while the FTI is tilted by 45° and therefore appears unrealistic).

The second type of artifacts relates to image merging. TIP has to combine pre-recorded FTIs with X-ray images of real passenger bags during the routine X-ray screening operation. To this end, TIP has to calculate how the X-ray beam would penetrate the FTI and the other objects in front of or behind the FTI and estimate the appropriate color and luminosity information. If this calculation is not accurate, the color and luminosity of the FTI might appear unrealistic. Also, if the image is filtered, for example when edges are enhanced, there might be the risk of the FTI having different looking edges than the other items in the X-ray image.

The objective of this study was to determine which artifacts occur, to quantify them, and to evaluate how strongly different artifacts affect TIP images to appear unrealistic: We first explored which artifacts occur by interviewing screeners about
of the main study had already participated in the interview or consistent with European regulation. None of the participants certified in accordance with the appropriate national authority as for the interviews. All screeners were selected, trained, and close-ended questions. After explaining what artifacts are, we first openly asked them what artifacts they know from work. We then went through a list of potential artifacts and asked for each one specifically whether the screeners had encountered them when working at the airport security checkpoint.

Physically implausible placement or alignment of the FTI within the X-ray image—the first type of artifacts previously described—was mentioned most often (reported by 5 of 9 screeners). Image merging artifacts—the second type of artifacts—were not reported, even when we specifically asked about artifacts regarding color, resolution, and edges. When openly asked about issues with TIP, the screeners reported that they sometimes perceive TIP images as unrealistic because the threat scenarios are unlikely, basically meaning that a terrorist would not hide a threat item in such a way. An unlikely threat scenario, however, is not an artifact according to our definition. We defined artifacts as differences in the appearance of the FTIs from real objects in the X-ray image. If a threat item is not well hidden by a terrorist, it will still be and therefore look like a real threat item. We noticed that the screeners struggled to keep the concepts of artifact and threat scenario apart. For the present rating study, we therefore included a rating about the threat scenario. Our intention was that the screeners would have to make an explicit distinction between threat scenario and artifact.

Based on the results of the interviews, we designed a rating study with the objective to have screeners rate for a representative sample of TIP images whether the images show artifacts. We tested and adapted the selection and wording of the different rating scales with two pilot studies. For the rating studies (two pilot studies and main study), we invited cabin baggage screeners from the same international European airport as for the interviews. All screeners were selected, trained, and certified in accordance with the appropriate national authority consistent with European regulation. None of the participants of the main study had already participated in the interview or pilot studies. The main study was conducted with 51 screeners (29 women and 22 men) aged 25 to 63 (M = 44.7, SD = 11.4) and with 2 to 31 years of work experience (M = 8.7, SD = 4.9). The screeners rated 600 TIP images recorded on X-ray machines (HI-SCAN 6046si) from the airport security checkpoint where the participating screeners used to work. The images were randomly sampled from a pool of images recorded at two different points of time (November 2016 and June 2017) because baggage content typically differs between seasons. Because we wanted to predict image difficulty in a follow-up study and the large majority of TIP images was correctly detected at the airport, we decided to use more TIP images that were missed (and hence likely more difficult). Therefore, we employed stratified sampling with disproportionate allocation.

The screeners rated the images regarding several aspects (from now on called image features) on a scale from 1 to 7 (response levels 1 and 7 were labelled): First, they rated how unrealistic the image appeared in general. Response level 1 was labelled with very unrealistic and response level 7 with very realistic. In this paper, we will refer to the respective image feature as generally unrealistic and response level 7 with very realistic. In this paper, we will refer to the respective image feature as generally unrealistic. Afterwards, they rated from 1 (totally disagree) to 7 (totally agree) whether an unrealistic threat scenario (threat scenario) and whether artifacts (generally artificial) caused the image to appear unrealistic. Next, they rated to what extent the artifacts related to physically implausible placement or alignment of the FTI were present: Alignment (the alignment of the FTI appears unrealistic), placement (the position of the FTI is physically improbable), and fall over (the FTI seems to float and should fall over). While for each of the above-mentioned image features a rating had to be provided, artifacts of the second type (i.e., artifacts related to image merging) were optional to rate, because they were very rarely or not at all mentioned during the interviews and pilot studies. These artifacts were distortion, color, size, resolution, and edges. To estimate the difficulty of detecting the FTI in the TIP image apart from artifacts and threat scenario, we also had screeners rate (1: very low, 7: very high) image based factors (IBF; properties of X-ray images known to affect detection). Screeners provided ratings for view difficulty (the difficulty of detecting the FTI in the displayed rotation), opacity (the proportion of the image which appears black, because the X-ray image cannot penetrate this area), superposition (the extent to which the FTI is covered by other items), and clutter (the amount of clutter in the X-ray image) [5], [6], plus some additional image properties, of which the effect on TIP difficulty is not yet known.

III. RESULTS

We first report the consistency of the ratings between different raters, the so-called inter-rater reliability, which was measured with the intra-class correlation (ICC) [8]. Because each image was rated by at least 12 screeners, the ICC was estimated with 12 ratings per image (the ICC can only be estimated for a fixed number of raters). The ICC(1,12) values ranged from 0.6 to 1, which means that the agreement for each image feature was good to excellent [9]. Therefore, for each image and image feature the ratings were averaged across the 12 raters and rounded to the nearest integer.

Fig. 1 shows the proportion of TIP images that received a certain average rating (over all raters) for each image feature. The majority of the images were rated to be realistic in general (i.e., were rated low on generally unrealistic). However, 20.1% of the images were rated with 5 and more, 5.8% with 6 and more, and 0.1% with 7. Threat scenario was the image feature most often regarded as poor (≥5:26.5%; ≥6:10.7%; 7:0.1%). About one in eight FTIs (≥5:13.5%; ≥6:1.4%; 7:0.0%) were rated to generally appear artificial (generally artificial). As discussed in the introduction, there are two types of artifacts. The first type is physically implausible placement or alignment of the FTI. In

---

1 We oversampled TIP images that were missed but weighed them down to their original proportion when reporting descriptive statistics

2 X-ray images tend to display objects slightly distorted and that distortion depends on the location. It is therefore not unambiguously clear whether distortion should be seen as an artifact of the first or second type.
With the TIP image to appear unrealistic, we analyzed which image features were associated with this category, poor placement was most often (≥5:17.3%; ≥6:5.7%; 7:0.1%) rated to be a reason for the TIP appearing artificial; closely followed by alignment (≥5:15.1%; ≥6:2.2%; 7:0%). A much smaller share of images was rated to appear artificial because the FTI seemed to float (fall over; ≥5:4.6%; ≥6:1.4%; 7:0.0%).

Only very rarely were TIP images rated to be affected by one of the artifacts of the second type, which were related to the image merging and were optional to rate. Therefore, we report here for each of these artifacts how often at least three of the twelve raters reported their occurrence in a TIP image. Color was most often reported (3.8%) as appearing artificial; followed by resolution (1.1%), size (1.1%), and edges (0.2%), while distortion was never reported.

After investigating how frequently and strongly the artifacts occurred, we analyzed which image features were associated with the TIP image to appear generally unrealistic. For this, we used a predictive modeling approach that is suited for the high number of image features. First, we split the dataset into a training (300 images), validation (150 images), and test dataset (150 images) [10]. To determine the image features with the biggest effects, a random forest model was fit to the training data and the conditional variable importances were calculated [11]. In addition to the image feature ratings, the model also included a variable specifying the category of the FTI (category; gun, knife, etc.). In cabin baggage screening, X-ray images do not only display baggage items like bags, suitcases, backpacks or purses, but also loose items like jackets, belts, shoes, etc. Therefore, sometimes FTIs are not placed inside a bag. To test whether this affects generally artificial, we also included a variable to specify whether the FTI was placed inside a bag (FTI in bag). Generally artificial was not included in the model because it was intended to summarize the specific artifacts. Modeling was done with the R package party. The best random forest model was chosen from models varying in the number of trees and number of candidate variables for splitting the nodes (ntree = 1500; mtry = 6). As can be seen in Fig. 2, threat scenario, alignment, placement, view difficulty, fall over and FTI in bag showed the biggest effects.

While random forest offers a flexible modelling approach that allows predictor variables to interact, the resulting model is difficult to interpret beyond the estimation of how important the predictor variables are. We therefore also estimated a multiple linear regression model with the predictor variables identified to be the most important ones by the random forest. As shown in Table I, the linear model could predict generally unrealistic about as well as the more complex random forest, as indicated by the root-mean-square error (RMSE) and the share of explained variance ($R^2$, also known as determination coefficient) for the validation data. We therefore selected the linear model as the preferred model to explain generally unrealistic and the model also showed a high predictive power with the test dataset. Table II shows the estimated coefficients of the linear regression. All the included predictors except for fall over significantly affected generally unrealistic in this model.

### IV. SUMMARY, DISCUSSION & CONCLUSION

TIP is an important tool to increase attention and motivation of screeners and is often used to measure their detection performance. However, TIP is only effective if images produced by TIP appear realistic [3]. There are two types of artifacts to distinguish: The first type is physically implausible placement or alignment of the FTI within the bag image. The second type is artifacts that result from the merging of the FTI with the X-ray image of a passenger bag. We evaluated the occurrence of both types of artifacts by first interviewing screeners about artifacts they encounter in their day-to-day work. Afterwards, another sample of screeners rated a representative sample of TIP images for whether they showed any of the artifacts identified in the interviews.
In general, TIP artifacts do not seem to be a major issue, since four in five TIP images appeared realistic to the screeners. However, in about one in six images, the FTI was positioned inadequately so that the placement of the FTI seemed physically implausible (for example, an IED going partly through a heel). In about the same amount of images, the alignment of the FTI appeared artificial because it differed from the other items in the bag. Not only did these two artifacts occur quite often but they were also significantly associated with the image appearing unrealistic. That the FTI appeared to float in the bag and unrealistic distortion of the FTI did not seem to be an issue. FTIs were also not rated to appear artificial regarding their color, unrealistic distortion of the FTI did not seem to be an issue. FTIs were also significantly associated with the image appearing unrealistic. That the FTI appeared to float in the bag and unrealistic distortion of the FTI did not seem to be an issue. FTIs were also not rated to appear artificial regarding their color, unrealistic distortion of the FTI did not seem to be an issue. FTIs were also not rated to appear artificial regarding their color. The merging of the FTI into the bag image works well in most cases and that the quality of TIP could mainly be improved in the future.

There is some limitation to the finding that unrealistic color was rarely reported to be an issue. Screeners at an airport are only rarely exposed to real IEDs. Therefore, they might not have a very good understanding about the color of real IEDs in X-ray images. Because they are exposed to many IEDs with TIP, they perhaps rated whether the color was realistic based on their experience with TIP and not based on how IEDs would look in reality.

Apart from the assessment of artifacts, our study provided an additional notable finding: About one in four TIP images was perceived to show an unrealistic threat scenario, i.e. the screeners perceived that these TIP images did not represent how a terrorist would hide a threat item. While this is not an artifact in the sense of the FTI looking different from real items, TIP should ideally prepare screeners for realistic threat scenarios. To find a better way of positioning FTIs in X-ray images of passenger bags, ideally taking into account the three-dimensional layout of other items, would not only reduce artifacts, but could also address the problem of unrealistic threat scenarios.

It should also be mentioned that avoiding artifacts and achieving realistic threat scenarios are not the only pre-requisites for effective TIP. Other prerequisites have to be met as well: For instance, frequent and well performed covert-tests, effective computer-based training and large FTI libraries are also important [3].

In conclusion, our study showed that the majority of the evaluated TIP images were of good quality: FTIs were most often projected without creating discernable artifacts. In some images however, the positioning of the FTI was inadequate and could be improved in the future.

### REFERENCES


