Screening Enhancements: Why don't they enhance performance?

Ghylin, K.M.[1], Schwaninger, A.[2,3], Drury, C.G.[1], Redford, J.[1], Lin, L.[1] & Batta, R.[1]
[1] University at Buffalo, State University of New York, Buffalo, New York
[2] University of Zurich, Zurich, Switzerland
[3] Max Planck Institute for Biological Cybernetics, Tubingen, Germany

Data obtained utilizing image enhancements in a carry-on bag x-ray screening task were analyzed to determine whether and how image enhancements affect performance. To complement earlier studies of experienced screeners, sixty-six novices to the screening task used six different overall enhancements. Results indicated a significant worsening of performance, A', between Original images and Negative images, but no performance differences for the other enhancements, similar to effects found for experienced screeners. There was little overall performance learning taking place on the task. More detailed analysis by splitting response times into search and non-search components revealed little more enhancement effect, but a reduction in False Alarm response time as the task progressed. It appears that the locus of lack of positive effects of enhancements is not just a function of familiarity with the current screen view.

INTRODUCTION

The importance of security screening in airports has increased dramatically in the past few years with ongoing research and new technological innovations. Within baggage screening at airports, security screeners are given multiple tools to assist them in efficiently and effectively screening x-ray images of bags. The availability of tools ranges from the use of effective training protocols to improve search strategies, to job rotations, to ensure vigilant screeners, to the use of technology and software to aid the search task.

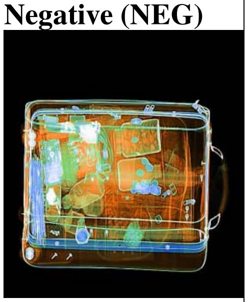
One of the better known software features is image enhancements which enhance various features within x-ray images. Each enhancement feature has the goal of highlighting a different aspect within the image presented, hopefully helping to identify any threat objects. Image enhancements work by changing the search characteristics of the items within the picture. Different types of features tend to emerge from images, such as colors, textures, etc. (Treisman, 1998) Image enhancements work by exploiting this pop-out effect of certain features in order to gain the attention of the observer. The image enhancements use various features of the bag contents (e.g. density) to produce an image where a threat will effectively "pop-out". Instead of relying on the searcher to find the target, the technology will make the target more salient, thus helping gain the searcher's attention. Software engineers have produced many enhancements ostensibly to provide images that are more suitable for specific applications than the original image (Gonzalez & Woods, 2002). This appears to allow the enhancement features to aid the screener's performance. Most screeners use these enhancement features, although different ones use them in different ways. If these enhancements truly help create images that are more suitable for detecting threats, then they are an asset to the screener.

However, few public studies have investigated the performance enhancing aspects of these image enhancements. (Note: Table 1 provides information on the different type of image enhancements utilized in this study.) One study with no details of its 35 participants and no statistical tests (Kase, 2002) found preferences for color over monochrome displays as may be expected. Schwaninger (2005) investigated the effects of image enhancements on the detection of improvised explosive devices (IEDs) in carry-on baggage using experienced screeners. Similarly, Klock (2005) examined if image enhancements improved detection performance for carry-on bags on a range of threat items. Both studies found that organic stripping and inorganic stripping functions resulted in decreased performance. Klock (2005) further found that high penetration decreased performance, while crystal clear, black and white, and low penetration resulted in the best performance. It should be noted however that the original image was not utilized in this study. However Schwaninger (2005) found that the original image produced the best performance while organic stripping, organic only, and luminance negative functions substantially impaired detection of the IEDs. Michel, Koller, Ruh, & Schwainger (2007) confirmed these results finding that original images produced the best performance, while other enhancements substantially impaired detection performance.

Studies of experienced screeners have also found significant effects due to threat type (typically guns, knives and IEDs) as well as the interaction between image enhancement and threat type (Michel, et al, 2007). These findings are unsurprising as previous work has found significant differences in performance of a security inspection task for experienced screeners based on threat type (Drury, Ghylin, & Holness, 2006).

The picture from this scant literature is not encouraging. Colored displays (universally used for many years) are generally better than monochrome displays, but other “enhancements” appear to worsen performance. This superiority of the current standard could be either because it is inherently superior, or could merely reflect its greater familiarity. By utilizing inexperienced subjects, the basic nature of the image enhancements can be explored. In addition, we now have ways of using response time data to explore whether the locus of changes is in that part of the response time spent in active search, or in other activities such as decision or response execution (Ghylin, Drury and Schwaninger, 2006). Fitting cumulative search time distributions has produced excellent fits (r^2 typically >0.9) and interpreting their coefficients has given some further insights into what aspects are changing. The objective of the current study was to determine whether image enhancements affect novices the same as screeners, and to examine whether the analysis of search and non-search components added any insights of use to the security community.

Table 1: Image Enhancement Feature Descriptions

Original (ORI) 	Original image is an unaltered image as produced by the x-ray screening machine. This is the image seen without applying any type of image enhancement filter. This helps to differentiate and/or outline object shapes, helps to differentiate and/or outline type of material (organic vs. inorganic)	**Black & White (BW)** 	Black and white is often referred to as a gray-scale filter. It removes material information from the image, and only shows the luminance value. This feature provides a better resolution, highlights fine details, and helps identify details found in electronic devices.
Negative (NEG) 	Negative filter is often referred to as a Luminance Negative filter. In this image the luminance of the image is inverted. The material value and hue of each pixel remains the same. Changes gray scale from black on white to white on black. It can be used to identify fine wires and other light details.	**Super Enhancement (SEN)** 	The super enhancement filter adaptively adjusts the contrast of each image. The luminance of each pixel is adjusted to the luminance of its surrounding pixels. This enhances the contrast in denser and darker areas.
Organic Only (OO) 	Organic only filters show the organic parts of an image in color, while the metallic pixels are grey. Although similar to a Metallic Stripping enhancement, this filter results in less of the image remaining visible and mixed organic/metal pixels appearing green. Useful in searching for explosives and narcotics as it removes high density objects and metallic clutter images.	**Organic Stripping (OS)** 	Opposite of the organic only filter, the metallic parts of the image remained colored, and the organic parts are shown in a light gray color with low contrast. Resulting image is similar to a Metallic Only enhancement; however in the OS filter the mixed organic/metallic elements are still green. Useful in searching for weapons, components of explosives, and higher density items as it removes images of lower density such s food and textiles (clothing).

METHOD

A within-subject design was utilized to evaluate the effects of six unique image enhancements on screening performance for novice university students (n=65) from the university participant pool. No demographic information was collected.

Images were simulated x-ray carry-on bags with embedded threats which were presented utilizing the X-Ray Tutor system 2.0 (for details see Schwnainger, A. 2004). Six different image enhancements were tested encompassing black and white, negative, organic only, organic stripping, super enhancements, and original images, as described in Table 1.

Eight practice trials of original images were initially given to each participant to give them experience on the task and to give examples of each threat type. Feedback was given to the subjects on the completion of these training images. Each test image set included three unique images of each of three different threat types (guns, knives, IEDs). In the experiment proper, two views of each image were utilized, resulting in 18 unique threat images shown in each of the enhancements. Each image enhancement set contained the same 18 unique threat items, along with a set of blank bags shown in the corresponding enhancement mode. A threat to non-threat ratio of approximately 1:4 was used as a reasonable compromise between face validity and amount of useful data collected per participant. Image sets were presented to participants based on a Latin Square Design. This design allowed us to use only the first trial of each participant as a between-participants design if needed, although this was not found necessary. After the practice trials, no feedback was given to participants. Participants had 10 seconds to respond to each image.

RESULTS

Overall results are presented first, followed by the search and non-search analysis.

Overall Performance

Hit rate (pdh) was defined as the proportion of images which both contain a threat and have been correctly identified as such. False alarm rate (pdfa) is the proportion of images which *do not* contain a threat that participants identify as containing a threat. Detection performance, in terms of A', was calculated and analyzed. To examine the effects of threat type, data was categorized by threat type and image enhancement. Since participants were not asked to identify the threat type, it is not possible to assign specific false alarms rates to each threat type, Therefore average false alarm rates for each participant and each image enhancement were found and utilized to calculate A' for each matching threat type. In this paper the non-parametric form of sensitivity (A') was used rather than the form that makes assumptions about normality and equality of variance (d'). This assured that any lack of agreement with such assumptions would not be an issue.

An ANOVA was performed with the performance measure of A' as a dependent variable, and image enhancement, threat type, and block as dependent variables with participant as a random factor. There were significant effects of image enhancement ($F(5, 1039) = 2.49$, $p < 0.05$), threat type ($F(2, 1039) = 176.03$, $p < 0.001$), and participant ($F(64, 1039) = 2.71$, $p < 0.001$). A tukey's test investigating these effects only found a significant difference between the original enhancement and the negative enhancement ($p = 0.011$) with performance decreasing. All three threat types were significantly different from each other ($p < 0.001$). No other effects were significant. An overall graph can be seen in Figure 1.

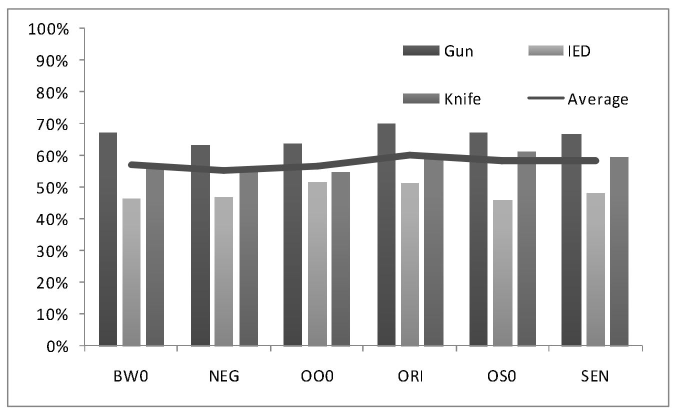


Figure 1: Average A' by image enhancement and threat type

Search and Non-search

To examine the source of changes in performance, the raw data was fitted to search and non-search equations of cumulative time distributions as outlined by Ghylin, Drury, & Schwaninger (2006) and briefly introduced in Appendix 1. Raw data was fitted to the equations, and values for the average time spent in both the search and non-search processes were obtained. For data fitting purposes, only those responses that included at least four hit detections could be utilized for this process. Since each threat item only appeared six times, this greatly reduced the amount of fitted data. Of the 1116 possible data points (65 subjects * 6 enhancements * 3 threat images), only 430 hits and 1092 false alarms points were obtained. Mean r^2 values of 0.98 and 0.85 were obtained for hits and false alarms, respectively, indicating strong fits to the data.

ANOVAs were performed with the dependent variables of fitted search and non-search time for both hits and false alarms. Image enhancement, threat type, and block were again the independent variables, with participant as a random factor. For hits, participant was significant for both search and non-search time [$F(60, 357) = 9.94$, $p < 0.001$ and $F(60, 357) = 12.22$, $p < 0.001$] respectively. For false alarms, a significant effect of non-search time was found for block ($F(5, 1016) = 23.12$, $p < 0.001$) and participant ($F(63, 1016) = 41.66$, $p < 0.001$). Image enhancement approached significance ($F(5, 10) = 2.17$, $p <= 0.055$). A significant effect on false alarm search time was also found for image enhancement ($F(5, 1016) = 2.51$, $p < 0.05$), block ($F(5, 1016) = 22.75$, $p < 0.001$) and participant ($F(63, 1016) = 41.82$, $p < 0.001$). No other effects were significant.

A Tukey's test for false alarm non-search time reveled significant differences between the Negative image enhancement and the Organic Stripping enhancement ($p< 0.05$). Differences between the original image and Organic Stripping were almost significant ($p = 0.10$). An analysis by Blocks revealed significant differences in false alarm non-search time between Block 1 and Blocks 3, 4, 5, and 6 ($p < 0.001$), Block 2 and Blocks 3, 4, 5 and 6 ($p < 0.01$), Block 3 and Blocks 5 and 6 ($p < 0.01$), and Block 4 and Block 6 ($p < 0.05$). A similar analysis on false alarm search time reveled significant differences between Block 1 and Blocks 3, 4, 5, and 6 ($p < 0.001$), Block 2 and Blocks 3, 4, 5, and 6 ($p < 0.001$), Block 3 and Blocks 5 and 6 ($p < 0.01$), and Block 4 and Blocks 5 and 6 ($p < 0.05$). In both cases, for search and non-search time, as the task proceeded, participants spent significantly less time in both the search and non-search aspects of the task.

DISCUSSION

Findings suggest little overall change in inspection performance with the image enhancements. The only significant differences in performance occurred between the original image and a negative image enhancement. Past research utilizing expert screeners has found suggestions of original images as being superior to most "enhancements" (Schwaninger, 2005; Michel et al. 2007). However, this was not found here for novice screeners. Novice screeners performed no differently under the original image than under either black and white, super enhancements, organic only, or organic stripping images. Only the negative image enhancement resulted in decreased performance. This enhancement required a further mental transformation (positive to negative) beyond that required for interpreting the original X-ray image, perhaps resulting in the decreased performance. Other image enhancements did not differ significantly from one another, but all present themselves on white backgrounds with various changes in luminance and hue. The novel presentation of the images may be to blame for the decreased performance.

Typical effects of threat type were found significant throughout this study. This phenomenon has been well documented in security inspection research (Drury et al. 2006, Michel et al. 2007) and was unsurprising. The novice inspectors' overall performance significantly increased from detecting IEDs to knives to guns. However, the detection ability was not influenced differentially by any image enhancements, suggesting that characteristic of the threat type, per se, produce the threat type effect. IEDs present themselves in various forms, and people generally are not familiar with their shapes. In contrast, knifes are encountered more on a daily basis (e.g. kitchen knife, pocket knife), the shape of gun is a well known image which often occupies more image space than a knife (e.g. Drury et al, 2006). The relative detectability of the three threat types remained intact regardless of the image enhancement.

Further analysis of the reaction times separating search and non-search times reveled little more about the overall effects of image enhancements, with the only differences occurring between the negative enhancement and organic stripping in the false alarm non-search time. A decrease in non-search time signals a change in the amount of time spent in either the decision making or response execution activities of the task (Ghylin et al., 2006). Participants spent less time in these activities using the organic stripping enhancement. However, this is only seen when compared to the Negative enhancement, providing little enhancement effect.

A substantial decrease in false alarm search and non-search time was found as the task progressed. There was clearly some learning in both the search and non-search aspects for reporting false alarms. However, it is important to note that this phenomenon was not found for hits or for overall performance. Learning effects were not present throughout this study. No change in performance, as measured by A' was found, indicating no overall learning took place. This is unsurprising as no feedback was given to the participants on the task.

CONCLUSION

Novice inspectors were found to have no performance differences when given the use of image enhancements. Performance was the same for the original image, a black and white image, super enhanced images, organic-stripping and organic only images. A slight decrease in performance was found for negative images. A typical effect of threat type, whether gun, knife, or IED, was found, but was not influenced by image enhancement. Conclusions suggest that the lack of image enhancement effects is not only due to the familiarity of the original image. The enhancements considered provided no benefit to the detection of threat objects even for novice users.

REFERENCES

Abidi, B., Liang, J., Mitckes, M., & Abidi, M. (2004). Improving the detection of low-density weapons in x-ray luggage scans using image enhancement and novel scene-decluttering techniques. *Journal of Electronic Imaging, 13(3), 523-538.*

Drury, C.G., Ghylin, K.M., & Holness, K. (2006). Carry-On Baggage Inspection. In *Proceedings of the 50th Human Factors and Ergonomics Society Conference,* October 16-20, 2006, San Francisco, CA.

Eibert, R.F. & Shi, S. (2004). Recent advances in imaging for x-ray inspection systems. *IEEE.*

Kase, K. (2002). Effective use of color in x-ray image enhancement for luggage inspection. Master's Degree thesis.

Klock, B.A. (2005). Test and evaluation report for X-ray detection of threats using different X-ray functions. *IEEE ICCST Proceedings*, 182-184.

Ghylin, K.M., Drury, C.G. & Schwaninger, A. (2006) Two-component model of security inspection: application and findings. *Proceedings of the 16th World Congress of the International Ergonomics Association*, *2006*. Maastricht, Holland.

Gonzalez, R.C., & Woods, R.E. (2002). Digital Image Processing (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Michel, S., Koller, S.M., Ruh, M., & Schwaninger, A. (2007). Do "Image Enhancement" functions really enhance x-ray image interpretation? In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1301-1306). Austin, TX: Cognitive Science Society.

Schwaninger, A. (2005). X-ray imagery: enhancing the value of the pixels. *Aviation Security International*, Oct 2005, 16-21.

Schwaninger, A. (2004). Computer based training: A powerful tool to the enhancement of human factors. *Aviation Security International,* Feb/2004, 31-36.

Treisman, A. (1998). Feature binding, attention and object perception. *Phil. Trans. Of the Royal Society of London, B,* 1295-1306.

APPENDIX 1

Hit

For Hit calculations, hit (h) and miss (m) data is utilized.

A cumulative probability of hit up to a variable time t can be written as:

P(hit(t)) = [1-exp((-STh) * (t – NSTh))]*Pdh(1)

Where:

STh represents search time and the exponential distribution parameter
NSTh represents non-search time
Pdh represents the probability of detection
t represents the various reaction times obtained from data

By ordering raw data by their response times, and assigning and ordinal number to each entry, the cumulative probability can be determined utilizing:

P(hit(t)) = Ordinal # / (#hit + # miss) (2)

By utilizing Equation 2 and the corresponding reaction times, Equation 1 can now be solved for STh and NSTh as:

Ln (Pdh – P(Hit(t)))/Pdh = -STh * (t – NSTh) (3)

This results in a linear equation of the form y = at + b where:
a = -STh
b = STh * NSTh, and therefore NSTh = b/ STh

False Alarm

Calculations for false alarms follow the same format as for hits.

For False Alarm calculations, false alarm (fa) and correct rejection (cr) data is utilized.

A cumulative probability of a false alarm up to a variable time t can be written as:

P(fa(t)) = [1-exp((-STfa) * (t – NSTfa))]*Pdfa (4)

Where:

STfa represents search time and the exponential distribution parameter
NSTfa represents non-search time
Pdfa represents the probability of detection
t represents the various reaction times obtained from data

By ordering raw data by their response times, and assigning an ordinal number to each entry, the cumulative probability can be determined utilizing:

P(fa(t)) = Ordinal # / (#false alarm + #correct rejection) (5)

By utilizing Equation 5 and the corresponding reaction times, Equation 4 can now be solved for STFA and NSTFA as:

Ln (Pdfa – P(fa(t)))/Pdfa = -STfa * (t – NSTfa) (6)

This results in a linear equation of the form y = at + b where:
a = -STfa
b = STfa * NSTfa, and therefore NSTfa = b/ STfa