Do "Image Enhancement" Functions Really Enhance X-Ray Image Interpretation?

Stefan Michel (s.michel@psychologie.unizh.ch)
University of Zurich, Department of Psychology, Binzmühlestrasse 14, 8050 Zürich, Switzerland, and
Max Planck Institute for Biological Cybernetics, Dept. Bülthoff, Spemannstraße 38, 72076 Tübingen, Germany

Saskia M. Koller (s.koller@psychologie.unizh.ch)
University of Zurich, Department of Psychology, Binzmühlestrasse 14, 8050 Zürich, Switzerland

Markus Ruh (m.ruh@psychologie.unizh.ch)
University of Zurich, Department of Psychology, Binzmühlestrasse 14, 8050 Zürich, Switzerland

Adrian Schwaninger (a.schwaninger@psychologie.unizh.ch)
University of Zurich, Department of Psychology, Binzmühlestrasse 14, 8050 Zürich, Switzerland, and
Max Planck Institute for Biological Cybernetics, Dept. Bülthoff, Spemannstraße 38, 72076 Tübingen, Germany

Abstract

State-of-the art x-ray screening systems offer a variety of so-called "image enhancement" functions (IEFs). Examples are color inversion, edge-enhancement, organic only, metal only etc. IEFs are often promoted because they would bring out detail that is obscured or highlight certain features, such as for example organic content. In this study, we investigated the usefulness of IEFs for cabin baggage screening (CBS) and hold baggage screening (HBS) in airport security. The results showed that the standard image provided the best detection performance. Some IEFs impaired detection performance substantially, which was also dependent on threat type (guns, knives, improvised explosive devices, other threat items). Together with previous work (Klock, 2005), these results highlight the importance of systematically studying the usefulness of IEFs in order to optimize human-computer interaction in x-ray screening.

Keywords: Aviation security; image enhancement filters; computer based training; image display technology; object recognition.

Introduction

In recent years, the importance of baggage x-ray screening at airports has increased dramatically. The image quality of older x-ray screening equipment was sometimes in need of improvement. For example an early version of a coloring algorithm as enhancement function did not serve the purpose of increasing detection performance of threat objects, actually it impaired it. This was due to the occlusion of object parts by the opaque coloring algorithm (Schwaninger, 2005a; Schwaninger, 2005b). But there has been much technological progress in the last years, especially regarding x-ray screening machines, which nowadays provide high image quality and various image enhancement functions (IEFs). The main objective of such functions is to process an image so that the result is more suitable than the original image for a specific application as for example x-ray screening at airports (Gonzalez & Woods, 2002). In x-ray images, the image enhancements might increase the visibility of objects within the bag and remove background noise. State-of-the-art x-ray machines provide many IEFs. The aim of this study is to investigate whether IEFs actually help human operators (screeners) to better detect threat items in x-ray images of passenger bags. Interestingly, reports regarding an evaluation of IEFs have not been publicly available except two recent publications (Klock, 2005; Schwaninger, 2005b). Klock (2005) examined whether IEFs increase screeners' threat detection performance when visually inspecting carry-on bags using a Rapiscan emulator. She found that high penetration, organic stripping and inorganic stripping functions resulted in decreased probability of detection (see below for more information on different IEFs). Crystal clear, black and white, and low penetration resulted in the best performance, while it should be noted that the original color image was not included in the analysis. Klock (2005) also found that these effects are dependent on threat type, i.e. whether guns, knives, improvised explosive devices (IEDs) or other prohibited items had to be detected. Schwaninger (2005b) reported a study on the effects of IEFs for the detection of IEDs in hold baggage. He found that the original image resulted in the best performance, whereas the organic stripping, organic only and luminance negative functions substantially impaired detection of IEDs. The purpose of this study is to extend previous research in order to evaluate the value of different IEFs. In addition, a comparison between IEFs used in cabin baggage screening (CBS) and hold baggage screening (HBS) was of interest.

The nine IEFs examined in this study can be applied to the x-ray images online when working at an aviation security checkpoint. Each pixel in the image format used in these x-ray machines has a material and a luminance value. To show the images on a screen, the pixel values are color coded using red for organic, blue for metallic and green for mixed organic/metallic material. The luminance value defines the luminance of the pixel. Table 1 gives an overview and description of all IEFs used in Experiment 1.

Table 1: Image Enhancement Filters

Grayscale (GR) 	The Grayscale filter removes the material information from the image and shows only the luminance value.
Luminance High (LH) 	In this filter, the luminance of the image is increased by applying a gamma correction (Pratt, 2001) to the luminance value. This allows the screeners to see details in dark areas of x-ray images, but as a consequence the visibility of details in light areas of the images is reduced.
Luminance Low (LL) 	As the opposite of the Luminance High filter, the luminance of the image is decreased. Details in light image is decreased. Details in light areas of the image become more visible, dark areas lose the details.
Luminance Negative (LN) 	In the Luminance Negative filter, the luminance of the image is inverted. The material value and therefore the hue of each pixel remains the same.
Metal Only (MO) 	Here, only the metallic parts of the image are shown in color. The organic parts are transformed to light gray with low contrast. The organic parts of the mixed organic/metallic pixels are removed as well, giving them a blue color similar to the all-metallic parts. The motivation for this filter is to allow the screeners to concentrate on the metallic objects perhaps leading to less search time for such objects.
Metal Stripping (MS) 	The Metal Stripping filter removes the metal from the image. Metallic parts are transformed to light gray and from the mixed organic/metallic pixels the metallic part is removed. As some mixed organic-metallic parts originate from metallic objects laying upon organic objects, this removal of metal sometimes shows the complete organic object without potentially distracting metallic parts.
Organic Only (OO) 	The Organic Only filter shows the organic parts of the image in color, while the metallic pixels are set to gray. The mixed organic/metallic pixels are assigned to the metallic or organic parts depending on the proportion of metallic and organic material. The difference to the Metallic Stripping filter is that less of the image remains visible and that the remaining mixed organic/metallic pixels are still green.
Original (OR) 	Original (OR) refers to the unaltered images as produced by the x-ray screening machine without applying any image enhancement filter.
Organic Stripping (OS) 	As the opposite to the Organic Only filter, the metallic parts of the image remain colored and the organic parts are shown in light gray with low contrast. The resulting image is similar to the Metal Only image, except that in this filter the mixed organic/metallic pixels are still green.
Super Enhancement (SE) 	The Super Enhancement filter adaptively adjusts the contrast of the image. Similar to a Local Histogram Equalization (Gonzalez & Woods, 2002) or an Adaptive Contrast Enhancement (Stark, 2000), the luminance of each pixel is adjusted to the luminance of its surrounding pixels. In the resulting image, each area has a medium average luminance.

Experiment 1

Experiment 1 was conducted to evaluate IEFs available in conventional cabin baggage screening (CBS).

Participants

A total of 443 airport security screeners of the CBS at a European airport participated in this study. All had on-the-job experience of at least 6 months. A between-subjects design was used to compare the effect of the IEFs on detection performance with each other. To this end, participants were randomly assigned to one of nine experimental groups, one group for each of the nine IEFs specified in Table 1. The control group was used for testing detection performance when images were displayed using the Original (OR) image type. The assignment of participants to groups was conducted so that the distribution of gender, age, and days on job were equal across groups. The ten groups showed an equal average of detection performance A', which was calculated using data of a separate test conducted prior to this study. The experimental groups varied in size between 37 (Luminance Negative filter) and 66 screeners (Grayscale filter); the control group consisted of 39 screeners. The difference in the group sizes is due to missing values (i.e. incomplete tests) for several screeners who originally were assigned to the study.

Method and Procedure

The X-Ray Competency Assessment Test (X-Ray CAT) was used in Experiment 1. This computer-based test contains 256 x-ray images of real passenger carry-on bags. Half of these images contain one prohibited item. The prohibited items have been selected by police experts to be representative for the variety of different threat types. The test contains 32 x-ray images of passenger bags with guns, 32 images with knives, 32 images with improvised explosive devices (IEDs), and 32 images with other prohibited items. For further details on the X-Ray CAT, see Koller and Schwaninger (2006). In order to create the stimuli for Experiment 1, the nine IEFs explained in Table 1 above were applied to the x-ray images. The participants' task is to visually inspect the images and to judge whether they are OK (contain no prohibited item) or NOT OK (contain a prohibited item). In this study, images disappeared after 10 seconds. The experiment consisted of two blocks. In block 1, each of the 9 experimental groups was tested with only one IEF and the control group was tested with the Original image (OR). The purpose of block 2 was to confirm that the participant groups are equivalent regarding their x-ray image interpretation competency. In block 2, all participants were tested again using the same bags as in block 1 but images were displayed in the OR format (see Table 1).

Results and Discussion

Detection performance was measured using A', a measure derived from hit and false alarm rates (Pollack & Norman, 1964; see Hofer & Schwaninger, 2004 for x-ray image interpretation competency). The hit rate refers to the proportion of all images containing a prohibited item that have been judged as NOT OK. The false alarm rate refers to the proportion of NOT OK judgments for harmless bags. A' scores were calculated for each block separately. Figure 1 shows means and standard errors of A' scores of block 1 broken up by IEF and pooled across threat categories, including the results of the control group (OR). The results in Figure 1 suggest that the OR image type results in the best performance, while some IEFs result in substantial impairment of detection performance. Note that due to security reasons, A' scores are not shown in the figures. To estimate effect sizes we employ effect size analysis and interpret the results based on Cohen (1988).

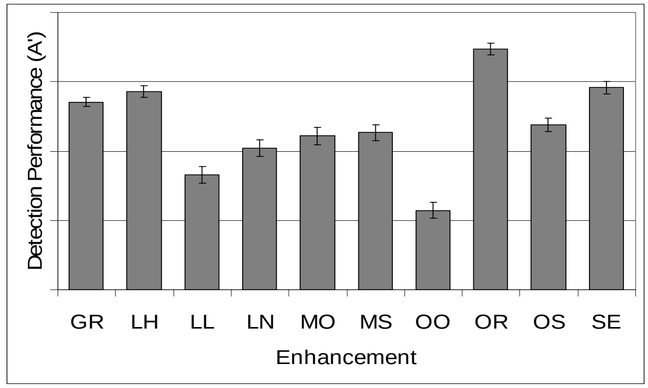

Figure 1: Detection performance Experiment 1, block 1, pooled across threat categories.
IEFs were tested between participant groups: GR = Grayscale, LH = Luminance High, LL = Luminance Low, LN = Luminance Negative, MO = Metal Only, MS = Metal Stripping, OO = Organic Only, OS = Organic Stripping, SE = Super Enhancement, OR = Original (control group).

An analysis of variance (ANOVA) with the between-participant factor IEF was carried out on individual A' scores averaged per screener across threat category. There was a main effect of IEF with a large effect size of $\eta^2 = .46$, $F(9, 433) = 41.67$, $p < .001$.

Figure 2 shows means and standard errors of A' scores of block 1 broken up by IEF and threat category. For all four threat categories, the OR image type resulted in the best performance. Again, some IEFs impaired detection performance substantially. Moreover, the results in Figure 2 suggest that the effects of IEFs on performance vary between threat categories. These results were confirmed by a separate ANOVA using individual A' scores calculated for each of the four threat categories (guns, knives, IEDs, other prohibited items). The ANOVA with the between-participant factor IEF and the within-participant factor threat category gave a large main effect of IEF with an effect size of $\eta^2 = .48$, $F(9, 433) = 43.66$, $p < .001$. There was also a large main effect of threat type with an effect size of $\eta^2 = .30$, $F(3, 1299) = 180.84$, $p < .001$. And there was also a large interaction between threat category and IEF with $\eta^2 = .32$, $F(27, 1299) = 22.91$, $p < .001$.

The same A' scores were subjected to one-way ANOVAs that were conducted separately for each threat category. There was a large main effect of IEF for all threat categories. For guns, there was an effect size of $\eta^2 = .64$, $F(9, 433) = 86.09$, $p < .001$, for IEDs $\eta^2 = .32$, $F(9, 433) = 22.38$, $p < .001$, for knives $\eta^2 = .32$, $F(9, 433) = 23.10$, $p < .001$, and for other prohibited items $\eta^2 = .43$, $F(9, 433) = 36.27$, $p < .001$.

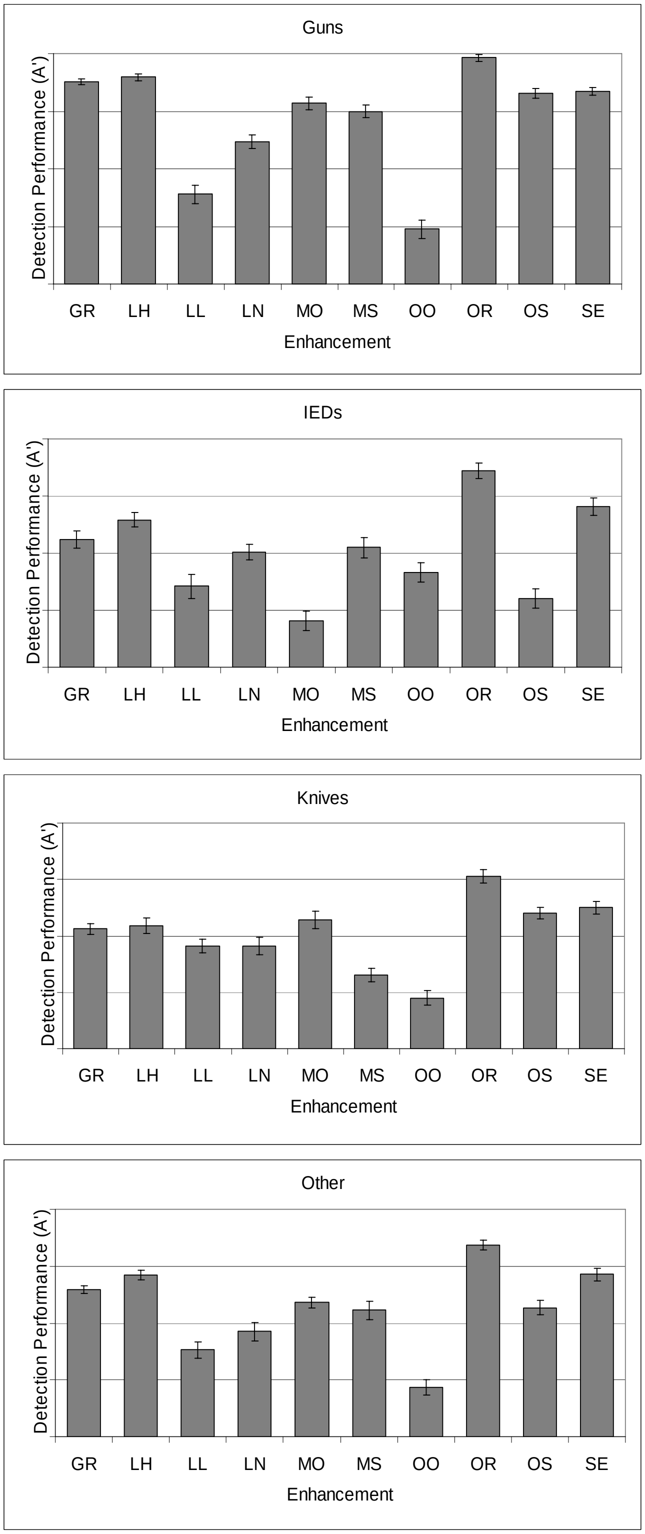

Figure 2: Detection performance in Experiment 1, block 1, broken up by threat category.
GR = Grayscale, LH = Luminance High, LL = Luminance Low, LN = Luminance Negative, MO = Metal Only, MS = Metal Stripping, OO = Organic Only, OR = Original, OS = Organic Stripping, SE = Super Enhancement.

In short, the OR image type resulted in the best performance for all threat categories. Moreover, some IEFs resulted in a substantial impairment which clearly depended on threat category. This interaction would be predicted if one takes into account that color information in x-ray images represents different materials and that different prohibited items vary in their material composition. For example, the Metal Only (MO) filter removes organic material from the x-ray image (see also Table 1). Since guns and knives usually consist of metallic material, their pixels in the filtered x-ray image remain largely unaffected when the MO filter is used. However, explosive material of IEDs is organic, thus it is not surprising that the MO filter results in a large impairment of IED detection (see Figure 2). A similar explanation applies to the effect of the Organic Stripping (OS) filter. When this filter is applied, all metallic parts of the image remain colored and the organic parts are shown in light gray with low contrast. The resulting image is similar to the MO image, except that for this filter the mixed organic/metallic pixels are still green. Since the Metal Stripping (MS) filter removes metallic information from the image, this IEF results in a substantial impairment of the detection of guns and knives, which usually contain much metal. Because organic explosive material in IEDs remains visible when the MS filter is used, IED detection is not affected substantially. The results in Figure 2 also indicate that the MS filter might be a better option than the Organic Only (OO) filter. As explained in Table 1, the MS filter includes information about organic material hidden behind metallic parts, whereas the OO filter simply removes these parts from the image. A comparison between the original image (OR) and the grayscale version gives some indications on the relevance of color information. The removal of the color-coded material information by the Grayscale filter (GR) does impair threat detection, while this effect is less pronounced for the detection of guns. Apparently, the luminance information seems to be more important than the material information. When inserting a threat object into a bag, the part of the bag with the object inside normally becomes darker than its surrounding. This is particularly the case for guns which contain much metallic material. Note however, that the removal of material information can conceal objects with the same luminance but different material than its surrounding. A similar problem appears when using the Super Enhancement (SE) filter. For this IEF, the material information remains the same, but the luminance contrast is slightly reduced which has a negative influence on detection performance. The Luminance High (LH) filter allows better threat detection than the Luminance Low (LL) filter. With the LL filter, most objects inside the bag have a luminance close to black, which generally reduces the differentiation of these objects. When using the Luminance Negative (LN) filter, material and luminance information remain in the image, but the luminance is inverted. The impairment of threat detection when using this IEF shows that screeners perform better with a dark object on a light background than if the luminance is inverted.

The results reported so far refer to block 1. As explained in the method section above, all participants conducted the X-Ray CAT again in block 2 using the original image type (OR). This was conducted to confirm post-hoc that the different participant groups are equivalent in terms of their x-ray image interpretation competency. This a prerequisite for the interpretation of the results reported above involving ANOVAs with IEF as between-participant factor. Separate ANOVAs of the data from block 2 confirmed that the 9 experimental groups and the control group were equivalent. Individual A' scores were calculated for each screener based on all trials of block 2. These data were subjected to a one-way ANOVA with participant group as between-participant factor. All groups were equivalent, since there was no effect of group, $\eta^2 = .02$, $F(9, 433) = 1.08$, $p = .38$. Individual A' scores were calculated also for each threat category separately and this data were then analysed using an ANOVA with participant group as between-participant factor and threat category as within-participant factor. Again, the results show that the participant groups were equivalent in terms of their x-ray image interpretation competency, since there was no main effect of participant group, and no interaction between participant group and threat category, $\eta^2 = .02$, $F(9, 433) = 1.04$, $p = .41$, and $\eta^2 = .02$, $F(27, 433) = 0.86$, $p = .63$, respectively.

Experiment 2

In hold baggage screening (HBS) x-ray images feature slightly different colors. Figure 3 shows examples of the stimuli used in Experiment 2. As explained in the introduction, screeners mainly search for IEDs, as other threat objects like for example knives do not pose a threat to the aircraft and passengers when placed in hold baggage. HBS screeners are often also more experienced screeners as it was the case in this participant sample. The main aims of Experiment 2 were to examine whether similar results are found in HBS regarding the effect of IEFs despite the operational and training differences between HBS and CBS.

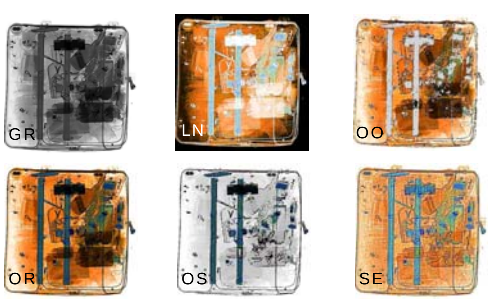

Figure 3: IEFs for HBS as used in Experiment 2. From top left to bottom right: GR, LN, OO, OR, OS, SE (see Table 1).

Participants

Data of 83 aviation security screeners of the HBS of the same European airport was analyzed.

As in Experiment 1, a between-subjects design was used to compare the effect of the IEFs. Due to the smaller sample size only 5 IEFs and the OR image could be tested. The 83 HBS screeners were randomly assigned to one of five experimental groups (GR, LN, OO, OR, OS, SE filters) or the control group (OR filter). The assignment of participants to groups was conducted so that the distribution of gender, age, and days on job was equal across groups. The six groups showed an equal average of detection performance A', which was calculated using data of a separate test conducted prior to this study. The number of screeners in each experimental group were between 10 (GR) and 17 (OO); the control group (OR) consisted of 15 screeners. As in Experiment 1, the difference in the group sizes is due to missing values (i.e. incomplete tests) for several screeners.

Method and Procedure

The Bomb Detection Test (BDT) was used in this study. This computer-based test contains 200 x-ray images of real hold baggage, whereas 100 images contain an IED. The IEDs were created by police experts. Participants were instructed to decide for each x-ray image whether it is OK (does not contain an IED) or NOT OK (contains an IED). Images disappeared after 10 seconds. As in Experiment 1, there were two blocks. In block 1, each of the 5 experimental groups was tested with their respective IEF. In block 2 all participants were then tested again using the same images but using the Original (OR) image function. The control group conducted the test twice using the OR image type in block 1 and block 2. As in Experiment 1, the purpose of block 2 was to confirm the comparability of the groups post hoc.

Results and Discussion

Analyses were similar to Experiment 1 but there was only one threat category, i.e. IEDs. Figure 4 shows means and standard errors of A' scores broken up by image enhancement function. As mentioned above, A' scores are not shown in the figure for security reasons. Effect sizes are calculated using effect size analysis and they are interpreted based on Cohen (1988). A one-way ANOVA with IEF as between-participant factor revealed a large main effect of IEF with an effect size of $\eta^2 = .26$, $F(5, 77) = 5.29$, $p < .001$. As in Experiment 1, the original image (OR) resulted in the best performance. Consistent with the results found in Experiment 1, we found in Experiment 2 that the Organic Stripping (OS) and Luminance Negative (LN) functions resulted in a substantial impairment of detection performance for IEDs.

All participants conducted the test again in block 2 using the original image type (OR). The aim was to confirm post-hoc that the different participant groups are equivalent in terms of their x-ray image interpretation competency. To this end, individual A' scores from block 2 were subjected to a one-way ANOVA with participant group as between-participant factor. There was no main effect of group, $\eta^2 = .05$, $F(5, 77) = 0.75$, $p = .59$, confirming that the six groups are equivalent regarding their x-ray image interpretation competency.

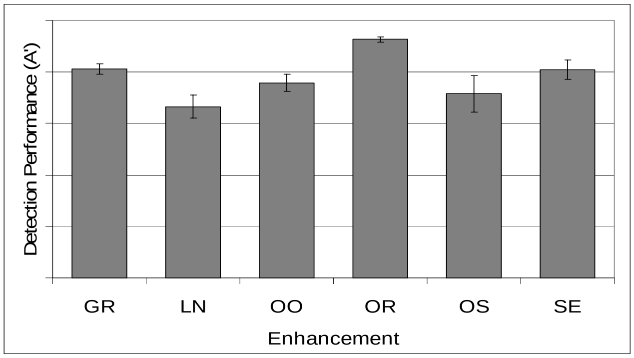

Figure 4: Detection performance Experiment 2, block 1. GR = Grayscale, LN = Luminance Negative, OO = Organic Only, OR = Original, OS = Organic Stripping, SE = Super Enhancement.

General Discussion

The aim of this study was to investigate the effect of image enhancement functions (IEFs) on x-ray detection performance of airport security screeners. Experiment 1 was conducted with cabin baggage and Experiment 2 with hold baggage. In both experiments the original image (OR) resulted in the best performance. One interpretation could be that for this manufacturer the default image is indeed the best image. However, since the OR image is the default image on the tested x-ray machine and since screeners received more training with OR images, further research is needed to clarify whether the benefit of the OR image type is due to expertise and training or whether it truly reflects better image quality. In both experiments, it was also found that some IEFs resulted in substantial impairments of detection performance. This general result is consistent with previous reports (Klock, 2005; Schwaninger, 2005b). The IEF effects are dependent on threat category; most likely due to differences in material properties of the different threat categories. For example, guns contain more metal than IEDs. Removing metallic content (MS function) therefore results in a larger impairment of detection performance for guns than for IEDs.

The main conclusions of this study are that user testing is crucial before implementing such filters into a system. Moreover, training when and how to use each of the filters is crucial to make effective use of them. We are conducting a set of additional experiments to further investigate the value of IEFs. For example, it could be that although on average certain IEFs impair detection, they could still be useful for detecting certain threat objects under certain conditions. Moreover, we are currently looking at CBT data where screeners have the possibility to choose a filter and to switch between filters. This allows investigating whether perhaps a certain combination and sequence of IEFs is useful for certain threat types and images. In addition, we are trying to clarify, whether IEFs actually do not improve detection of prohibited items or if however, when used according to individual preferences and to specific features of the image, they can improve the ability to locate targets. Finally, we also have implemented IEFs in a CBT system (X-Ray Tutor) to investigate potentially supporting effects that only can become manifest through training and familiarization.

Acknowledgments

This research was financially supported by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403). We thank four reviewers for valuable comments. Many thanks to Zurich State Police, Airport Division, for their help in creating the stimuli and the good collaboration for conducting parts of the study.

References

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* New York: Erlbaum, Hillsdale.

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

Hofer, F., & Schwaninger, A. (2004). Reliable and valid measures of threat detection performance in X-ray screening. *IEEE ICCST Proceedings*, 303-308.

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

Koller, S., & Schwaninger, A. (2006). Assessing X-ray image interpretation competency of airport security screeners. *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia and Montenegro, June 24-28, 2006,* 399-402.

Pollack, I. & Norman, D.A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, 1, 125-126.

Pratt, W. K. (2001). *Digital Image Processing: PIKS Inside, 3*rd *Edition*. John Wiley & Sons, Inc., New York, NY.

Schwaninger, A. (2005a). Objekterkennung und Signaldetektion: Anwendungen in der Praxis. In B. Kersten & M. Groner (Hrsg.), *Praxisfelder der Wahrnehmungspsychologie*. Bern: Huber.

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

Stark, J.A. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9, 5, 889 - 896.