

COMPUTER BASED TRAINING: THE ENHANCEMENT OF HUMAN FACTORS

In the last decade a large amount of money has been invested into new aviation security technology. State-of-the-art X-ray machines provide many image enhancement features, high resolution images and automatic explosive detection. However, the final decision rests with the human operator. It is being realised more and more that even the most expensive technology is of limited value if the humans who operate it are not trained well enough. This is of special importance for X-ray screening of passenger bags. **Dr. Adrian Schwaninger** shares some of his research with us and demonstrates just how important a training tool the computer is.

X-ray screening of passenger bags is a demanding task. In a few seconds human operators must decide whether a bag is harmless or whether it

contains prohibited items. Figure 1 shows two X-ray images produced by state-of-the-art X-ray technology. The bag on the left depicts an image of a harmless bag. The bag on the right con-

tains three lethal weapons. In the ideal case, a trained screener would be able to identify all threat items reliably within a few seconds. This article tells you how that can be achieved.

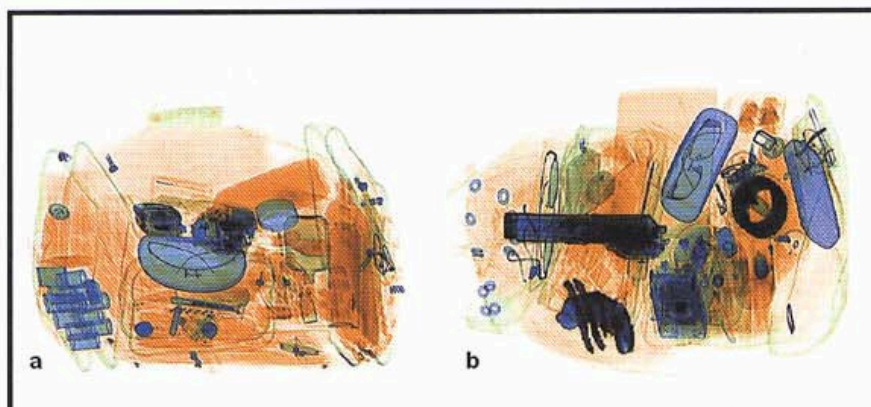


Figure 1: Detection of prohibited items is difficult without training.
a: Harmless passenger bag b: Bag with three threat items.

Visual Cognition

Visual cognition is the investigation of how the human brain processes visual information. It entails object recognition, perceptual learning, top-down processes, and visual attention. In the last decade, collaboration between psychologists, neuroscientists and computer vision specialists has resulted in a much better understanding of how the human brain recognises objects. Important aspects are visualised. (Figure 2). Essentially, object recognition means to match visual information from objects (Input) to visual representations stored

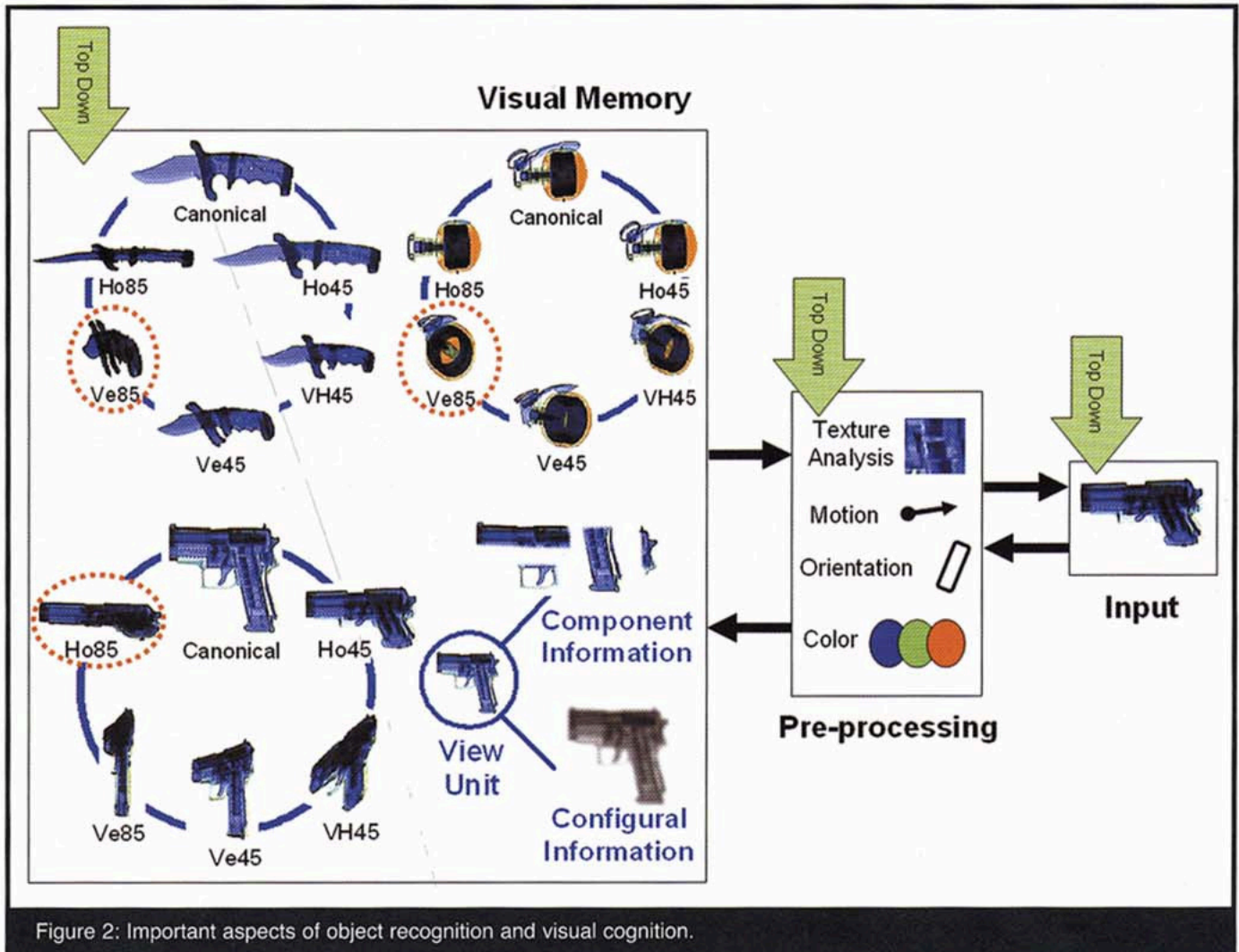


Figure 2: Important aspects of object recognition and visual cognition.

in memory (Visual Memory). A pre-processing stage extracts low-level features like colour, orientation, motion, texture and other properties. These features can be used to select possible candidate representations from visual memory in order to match them against the input. The matching can occur both on the basis of component and configural information. Component information refers to parts such as a gun barrel, a trigger, or a detonator. Configural information refers to the whole shape disregarding local details. The outputs of component and configural matching are combined by view-units. If there are enough different view-units for an object, it can be recognised from any viewpoint. However, the type and number of view-units is dependent on perceptual experience. If an object is never seen from a certain viewpoint it is difficult to develop

a corresponding view-unit. If the view is too different from stored views, recognition becomes a hard task if view-invariant features are not available. This is the reason why it is so difficult to recognise the three threat items in the X-ray image of Figure 1b. The depicted views correspond to the views in Figure 2, which are marked with red dotted circles. Because these views are rarely encountered, they have not been stored in visual memory. Since the views in Figure 1b are too different from stored views, it is quite difficult to recognise the threat items.

Recognition and categorisation can be influenced by top-down information. After carefully looking at Figure 2 it becomes much easier to recognise the threat items depicted in Figure 1b. The reason is that you create and select

view-based representations in visual memory (top down selection) and start looking for certain shapes and features when looking at Figure 1b.

View-Based Image Library

As explained previously, object recognition relies heavily on perceptual experience. Aviation security screeners should be trained to achieve a high detection performance within a few seconds no matter how the threat items are placed in the bag. In a close collaboration between vision scientists and aviation security specialists from Zurich State Police, Airport Division we have created a threat image library that currently contains 28,800 images. This library serves as the basis for adaptive computer based training and a new threat image projection system. The images are avail-



Effect of viewpoint

Effect of superposition

Effect of bag complexity

Figure 3: Effect of viewpoint, superposition, and bag complexity.

able in the Heimann image format (HIF) or in standard image formats such as JPG, TIFF or BMP.

Object recognition research indicates that 6 views are sufficient for most objects in order to capture the qualitatively different visual aspects. The canonical view refers to the view, which is easiest to recognise. It can be determined objectively using psychophysical procedures or subjectively using expert ratings. For creating a basic view set, the canonical view is rotated approximately 45° and 85° around the horizontal axis and around the vertical axis, resulting in four additional views. The sixth view corresponds to a rotation of the canonical view by about 45° around the vertical and about 45° around the horizontal axis. These views constitute the basic view set (see also Figure 2), which serves as a solid basis for training view-invariant threat detection. Several of our own studies as well as studies conducted in other labs have shown that the human recognition system can interpolate between stored views. This will allow the screeners to recognise views that are between the training views of the basic view set. Note however, that although the visual system is relatively good in performing a flip transformation it is conceivable to train screeners also on the recognition of mirror-reversed (or "flipped") views. Similarly, the threat items from the basic view set should also be displayed in plane rotated versions. The current version of the view-based library contains 400 prohibited items and each of them can be displayed in 6 basic views x 4 mirror reversals x 3 plane rotations = 72 views. This results in a total of 400 prohibited items x 72 views = 28,800 X-ray images. Of

course it is not necessary to show each screener all these images. But a good training system should determine for each screener which views are difficult. Based on an individual diagnosis, the system should create individual training sessions in order to enable screeners to achieve a reliable detection performance even if threat items are substantially rotated away from the easiest view.

Science vs. Application

In order to bridge the gap between science and application there has to be a close collaboration between scientists, airport security specialists and companies who develop X-ray equipment and training systems. Since June 2000 there is a close collaboration between vision scientists from the University of Zurich, APSS, and aviation security specialists from Zurich State Police, Airport Division. Several scientific studies have been conducted, which were financially supported by Zurich Airport and provided important insights for different human factor aspects in aviation security. Since 2002 there is a good cooperation with Smiths-Heimann and

Security Training International (STI), who provided useful technical information from the vendor side. The good relationship between these different partners made it possible to develop an adaptive training system ("X-Ray Tutor"), which is based on scientific findings from visual cognition, object recognition, and signal detection theory. It is operational at several international airports including Zurich Airport (since 2002) and 17 airports in Germany (since 2003).

Individually Adaptive Training

X-Ray Tutor is driven by software algorithms that monitor student performance and create training sessions adapted to each individual. The training system starts with easy images and increases in difficulty depending on the detection performance and difficulty ratings. Scientific studies conducted at our lab have shown that image difficulty depends on three image-related factors: viewpoint of the prohibited item, superposition by other objects and bag complexity. This is illustrated in Figure 3. Depending on perceptual experience and the ability to mentally rotate objects it is more difficult to recognise threat items when they are rotated (effect of viewpoint). Depending on their position in the bag threat items can be superimposed by other objects resulting in an impaired detection per-

formance (effect of superposition). Finally, bags differ substantially in the number and type of objects. This results in more or less "complex" X-ray images, which is a third determinant of detection performance (effect of bag complexity). Interestingly, screeners differ remarkably in their ability to cope with effects of viewpoint, superposition and bag complexity. (See Fig 3 on previous page)

The individually adaptive algorithm of X-Ray Tutor starts with easy X-ray images showing bags with threat items in canonical views with low superposition and low bag complexity. Then, view difficulty is increased and screeners learn to detect prohibited items in different viewpoints. Once a screener has been taught a representative sample of different viewpoints, he learns how to recognise them in increasing levels of bag

complexity and superposition. This is calculated for each screener using an individually adaptive algorithm, which uses objective measures of view difficulty, superposition and bag complexity. All objective measures have been shown to be correlated with human detection performance and they are based on several years of research involving vision scientists, image processing specialists, statisticians, and aviation security experts.

Individual Performance Feedback

Several studies have shown that the efficiency of a training system depends critically on the type of feedback. Most important for perceptual learning is immediate feedback. During training with X-Ray Tutor, screeners have to press an "OK" or "NOT OK" button and give a difficulty rating after each image



Figure 4 Interface of X-Ray Tutor showing the immediate feedback.

being presented. If an innocent bag has been shown the feedback message is "False Alarm" or "Bag OK". If a threat image has been presented, a hit or miss message is displayed ("Threat detected" or "Threat missed"). In both cases, the threat item is visually identified in the X-ray image by a flicker algorithm that helps to separate the figure (i.e. the threat item) from the background (i.e. the other contents of the bag). Screeners can display an information window that contains the canonical view of the threat item and a photograph. This helps to associate all views of a threat item to the same object concept in memory. After each training session (usually 20 minutes) two types of feedback are provided. The session feedback lists the absolute number of hits, false alarms, misses and correct identifications of clean bags of this session. Additionally, the system provides a feedback on detection performance across all sessions. This feedback specifies different difficulty levels and for each of them average hit rate, average false alarm rate, average response time for innocent bags and average response time for bags with a prohibited item. Whereas the immediate feedback is very important for perceptual learning, the session feedback and the individual performance reports are useful for visualizing training effectiveness. (See Fig 4 opposite)

Scientific Measurement of Effectiveness

In close collaboration between scientists from the University of Zurich and aviation security experts from Zurich State Police, Airport division the effectiveness of X-Ray Tutor was evaluated. A total of 72 screeners participated in this study. None of

them had used computer based training before. In the period between December 2002 and May 2003 each screener took 1-2 training sessions of 20 minutes per week using X-Ray Tutor HBS for training bomb detection. During the six months period four tests were conducted, in which improvised explosive devices (IEDs) had to be detected. In order to measure training effectiveness properly, only new IEDs were used that had not been shown previously during training. A Latin Square counterbalanced design

was used in which the four tests were of equal difficulty and the 72 screeners were distributed across four groups of equal average detection performance. As shown in Figure 5a detection performance (d') increased substantially during the six months of training. In order to evaluate training effectiveness, we calculated % increase in detection performance (d') relative to the first test. As you can see in Figure 5b training was very effective, leading to 71% relative increase of detection performance (d')

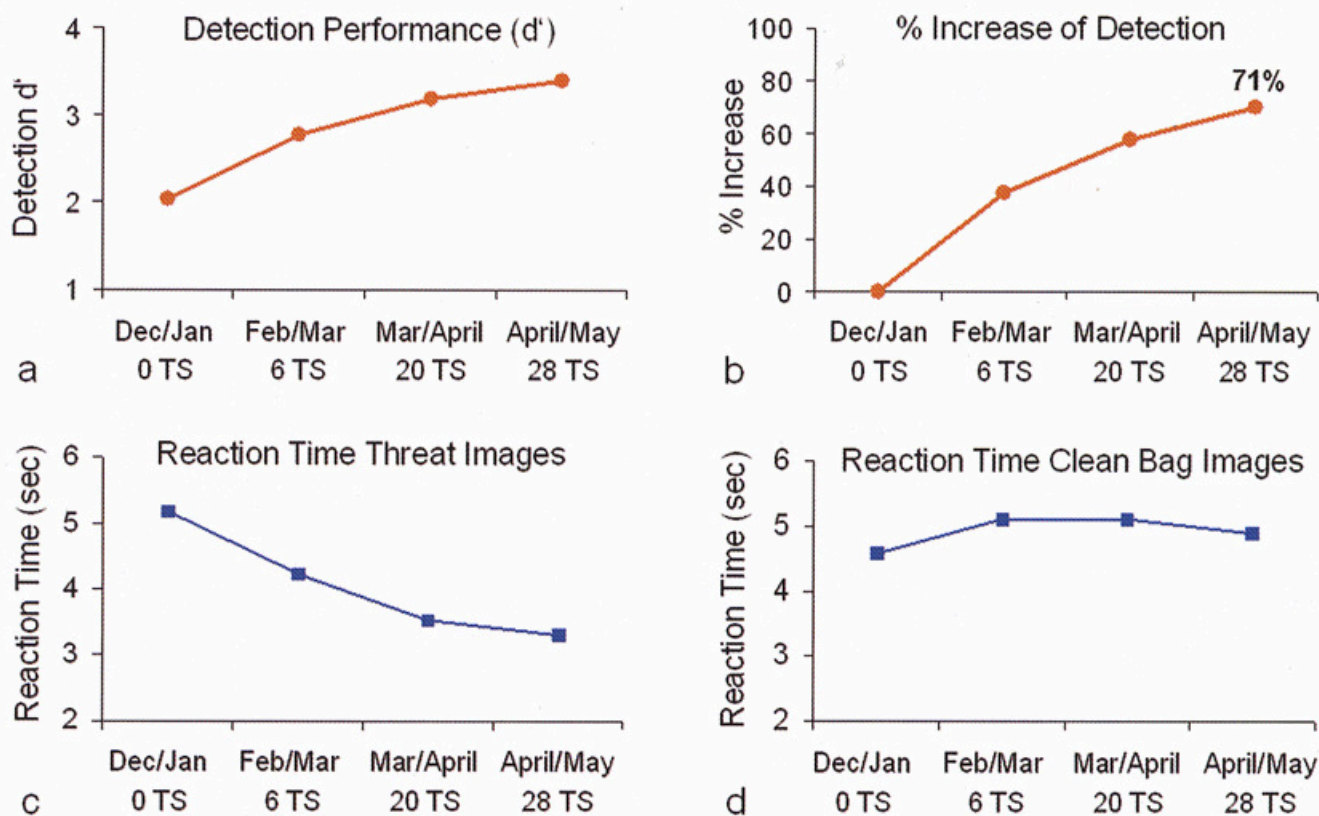


Figure 5 Results of a scientific study to evaluate training effectiveness of X-Ray Tutor. a detection performance (d'), b increase of detection performance (d') in % relative to the first measurement, c average correct reaction time for threat images, d average correct reaction time for clean bag images. TS: training session.

after six months of training. (See Fig 5 above)

This result is quite remarkable if it is taken into account that, on average, the 72 screeners took only 28 training sessions during the six-month period. For a subgroup of 52 screeners, who on average took 31 training sessions, the training effect was even higher; relative detection performance increase was 84%! As mentioned above, it should be stressed that in each test, threat items were shown that were new to the screeners. This is essential if training effectiveness should be measured properly.

Interesting results were also found for response times. Only data from the four tests are reported here (response times from training sessions were usually a bit longer). Generally, people responded faster when the bag contained a threat

item as opposed to clean bag images. It seems that if the screeners did not detect a threat item within the first few seconds, they continued to visually search the X-ray image in order to be sure that there was no prohibited item. This is not surprising, since qualitatively similar results have often been reported in visual search studies. A more interesting effect was the pattern of response time decrease across training. When a bag contained a threat item, average response times decreased from 5.2 to 3.3 seconds during the six months of training. In contrast, for clean bag images response times remained stable at around 5 seconds. These results show that training with X-Ray Tutor did not affect the search times when there was no threat item. But if a bag did contain a threat item, screeners became very fast in detecting it.

In short, the results of this study show that adaptive computer based training based on scientific research is a very powerful tool for the enhancement of aviation security human factors. ■

The author lectures at the University of Zurich and the Federal Institute of Technology Zurich. The main areas of his teaching and research activities are visual cognition, psychophysics, signal detection theory, research methods, object recognition, and visual attention. In close collaboration with Zurich State Police, Airport Division he has conducted several research projects, which were funded by Unique Zurich Airport, Safety and Security. He was the winner of Aviation Security International's Award of Excellence for the Enhancement of Human Factors in 2003.