

Steigerung der Erkennungsleistung von Cargo Screenern

Marcia MENDES, Stefan MICHEL, Adrian SCHWANINGER

Institut Mensch in komplexen Systemen (MikS), Hochschule für Angewandte Psychologie, Fachhochschule Nordwestschweiz (FHNW), Riggbachstrasse 16, CH-4600 Olten

Kurzfassung: Die visuelle Beurteilung von Röntgenbildern von Cargo Containern zur Sicherheitskontrolle und dem Auffinden von Schmuggelwaren hat in den letzten Jahren weltweit mehr und mehr an Bedeutung gewonnen. Die Erkennung von verbotenen Gegenständen in Röntgenbildern stellt für das Kontrollpersonal (Screener) jedoch keine leichte Aufgabe dar. Beispielsweise müssen Screener wissen, welche Gegenstände verboten sind und wie diese in Röntgenbildern aussehen. Ein entsprechendes Training ist aus diesem Grund von essentieller Bedeutung. In dieser Studie wurde erstmals, basierend auf einem existierenden computer-basiertem Training für Handgepäckkontrollen (X-Ray Tutor), ein computer-basiertes Trainingsprogramm für den Cargo-Bereich entwickelt und evaluiert. Ziel dieser Studie war es zu untersuchen, ob die Erkennungsleistung der Cargo Screener durch computer-basiertes Training verbessert werden kann. Um das Training zu evaluieren, wurden Röntgenbildinterpretationstests eingesetzt. Messungen erfolgten einmal vor Beginn des Trainings sowie erneut nach ungefähr drei Monaten Training. Zusätzlich wurde untersucht, ob die Analyse der Röntgenbilder in Pseudofarben, verglichen mit Graustufen, zu einer besseren Erkennungsleistung führt. Die Ergebnisse zeigten einen deutlichen Zuwachs der Erkennungsleistung von Cargo Screenern nach dem dreimonatigen Training. Kein signifikanter Unterschied bezüglich der Erkennungsleistung zeigte sich jedoch beim Vergleich der Bilddarstellung in Pseudofarben und Graustufen.

Schlüsselwörter: Cargo Screening, computer-basiertes Training, Erkennungsleistung, Mensch-Maschine-Interaktion, Röntgenbildanalyse

1. Einleitung

Die Sicherheitskontrollen an Flughäfen wurden in den letzten Jahren deutlich verschärft. Der Fokus lag dabei vor allem auf den Passagier- und Gepäckkontrollen. Trotz wirtschaftlicher Relevanz des Frachttransports und der Tatsache, dass Frachtgepäck häufig auch in Passagierfliegern transportiert wird, wurde diesem Bereich vergleichsweise wenig Aufmerksamkeit geschenkt. An vielen Flughäfen wurden in den letzten Jahren nun vermehrt Röntgenmaschinen zur Analyse von Cargo Containern installiert. Ein Vorteil der Analyse mittels Röntgenbildern ist, dass die Container nicht physisch auseinandergenommen und durchsucht werden müssen. Trotz des Einsatzes fortschrittlicher Technologien liegt die Entscheidung, ob ein Container einen verbotenen Gegenstand enthält oder nicht, bisher dennoch weitgehend beim Menschen. Die Erkennung von verbotenen Gegenständen in Röntgenbildern stellt für das Kontrollpersonal (Screener) jedoch keine leichte

Aufgabe dar. Zunächst müssen Screener lernen und wissen, welche Gegenstände verboten sind und wie diese in Röntgenbildern aussehen.

Wie frühere Studien im Bereich der Gepäckkontrollen (z.B. Schwaninger et al. 2005) zeigen konnten, hängt die Erkennung von verbotenen Gegenständen in Röntgenbildern von sogenannten wissensbasierten- und bildbasierten Faktoren ab. Erstere beziehen sich auf das zur Bildanalyse benötigte Wissen, also welche Gegenstände verboten sind und wie diese in Röntgenbildern aussehen. Hinsichtlich der bildbasierten Faktoren konnten Studien belegen, dass drei relevante Faktoren existieren, welche die Erkennungsleistung eines Screeners bei der Analyse von Röntgenbildern beeinflussen: Rotation des verbotenen Gegenstandes, Verdeckung des verbotenen Gegenstandes sowie die Komplexität des Gepäckstückes (Schwaninger et al. 2005). Ein computerbasiertes Training (CBT) mit individuell adaptiven Algorithmus (z.B. X-Ray Tutor, XRT) hat sich im Hinblick auf die wissens- und bildbasierten Faktoren als besonders effektiv erwiesen (z.B. Michel et al. 2007; Koller et al. 2008; Halbherr et al. 2013). Beim XRT bekommen Screener verbotene Gegenstände zu sehen mit welchen sie im Alltag sonst nicht oder nur selten konfrontiert werden (z.B. unkonventionelle Spreng- und Brandvorrichtungen (USBV)). Zusätzlich werden sie darin trainiert, diese in verschiedenen Rotationen, mit verschieden starker Verdeckung sowie in Taschen bzw. Containern mit unterschiedlicher Komplexität zu erkennen (Schwaninger 2004).

Basierend auf dem bereits existierenden CBT für Gepäckkontrollen (X-Ray Tutor for Cabin Baggage Screening, XRT CBS) wurde in dieser Studie ein Trainingsprogramm für den Cargo-Bereich entwickelt (Cargo X-Ray Tutor, C-XRT) und evaluiert. Ziel war zu untersuchen, ob die Erkennungsleistung der Cargo Screener durch den regelmäßigen Einsatz eines CBTs verbessert werden kann. Zusätzlich zur Trainingsevaluation wurde außerdem untersucht, ob eine höhere Erkennungsleistung erzielt werden kann, wenn die Bilder im Test in Pseudofarben eingefärbt dargestellt werden anstatt in Graustufen (Originaldarstellung).

2. Methoden

2.1 Versuchsteilnehmende und experimentelles Design

40 Angestellte einer Cargo Screening Organisation (davon 9 Personen aus der Administration (4 männlich) und 31 erfahrene Cargo Screener (25 männlich)) nahmen an der Studie teil. Drei Gruppen wurden gebildet (Abb. 1) – zwei Experimentalgruppen und eine Kontrollgruppe. Die Personen aus der Administration stellten die Kontrollgruppe dar. Um sicherzustellen, dass die Gruppen hinsichtlich der visuell kognitiven Fähigkeiten ausbalanciert waren, lösten alle Teilnehmenden vor Beginn der Studie einen Objekterkennungstest (X-Ray ORT, Hardmeier et al. 2005). Basierend auf diesen Testergebnissen wurden die Cargo Screener in die beiden Experimentalgruppen eingeteilt. Die durchschnittlichen Erkennungsleistungen, berechnet mit dem in der Signaldetektionstheorie verwendeten Mass A' (Pollack & Norman 1964), zeigten keine signifikanten Unterschiede zwischen den drei Gruppen (Graustufen-, Pseudo-, Kontrollgruppe) $F(2, 36) = 0.032$, $p = .969$, $\eta^2 = .002$. Beide Experimentalgruppen absolvierten regelmäßiges Training mit C-XRT (min. 20 Min./Woche) für einen Zeitraum von etwa drei Monaten (Gesamtrainingsdauer $M = 8.95$ h, $SD = 5.68$ h), während die Kontrollgruppe nicht trainierte. Die Erkennungsleistung wurde mit dem Cargo X-Ray CAT (C-CAT) gemessen.

Messungen erfolgten vor Einführung des Trainings und erneut nach den drei Monaten Training. Beide Experimentalgruppen absolvierten jeweils verschiedene Versionen des C-CAT (Graustufen: $n = 15$ vs. Pseudofarben: $n = 16$).

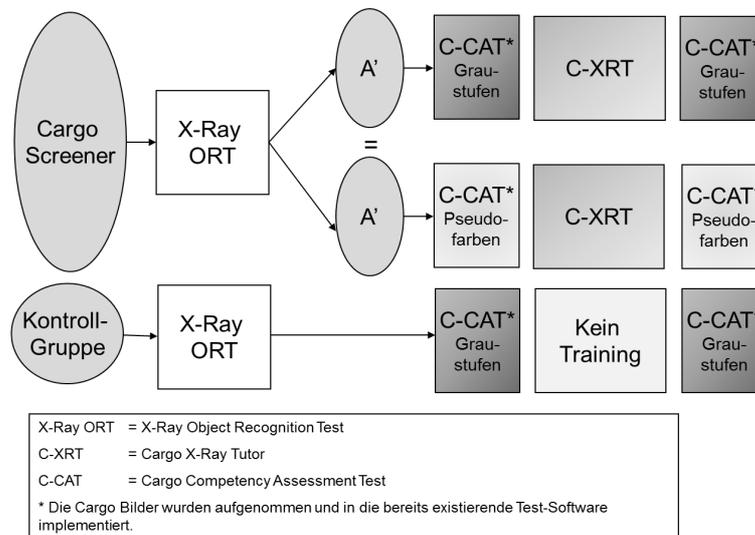


Abbildung 1: Experimentelles Design der Studie mit Experimental- und Kontrollgruppe (Michel et al. 2014).

2.2 Material

X-Ray Object Recognition Test (X-Ray ORT): Der X-Ray ORT ist ein Test der zur Unterstützung bei der Personalauswahl von Screenern eingesetzt wird. Nicht jeder Mensch besitzt die visuell kognitiven Fähigkeiten, welche zum Umgang mit den bildbasierten Faktoren Rotation von Gegenständen, Verdeckung durch andere Gegenstände und Gepäckkomplexität bei der Röntgenbildanalyse benötigt werden. Der X-Ray ORT misst wie gut eine Person mit diesen Faktoren umgehen kann, ohne über nötige Vorerfahrungen in der Röntgenbildinterpretation zu verfügen. Für weitere Informationen siehe Schwaninger et al. (2005) und Hardmeier et al. (2005).

Cargo X-Ray Tutor (C-XRT): XRT ist ein CBT zur Schulung der Röntgenbildinterpretationskompetenz von Screenern. Während des Trainings werden den Screenern Röntgenbilder von Gepäckstücken bzw. Cargo Containern gezeigt. Die Screener müssen diese visuell analysieren und entscheiden, ob ein verbotener Gegenstand enthalten ist oder nicht. Nach jeder Antwort erhalten die Screener Feedback zu ihrer Entscheidung. Falls ein verbotener Gegenstand im Gepäck vorhanden war, erscheinen Informationen darüber, wo dieser Gegenstand im Gepäck lokalisiert ist, um was für ein Objekt es sich handelt und wie dieses geröntgt aussieht (Schwaninger 2004). XRT ist individuell adaptiv, d.h. das Trainingsprogramm passt sich der individuellen Leistung und dem Lernfortschritt des Screeners an. Für diese Studie wurde eine neue XRT Version speziell für den Cargo-Bereich entwickelt (Cargo X-Ray Tutor, C-XRT).

Cargo X-Ray Competency Assessment Test (C-CAT): Der X-Ray CAT ist ein Testinstrument zur Messung der Röntgenbildinterpretationskompetenz von Screenern, welcher für den Bereich der Gepäckkontrollen entwickelt wurde (Koller & Schwaninger 2006). Während des Tests müssen Screener Röntgenbilder eines Gepäckstückes bzw. Cargo Containers visuell analysieren und entscheiden, ob ein verbotener Gegenstand enthalten ist oder nicht. Für diese Studie wurde der C-CAT in zwei Versionen entwickelt – Röntgenbilder in Graustufen vs. Pseudofarben.

Abgesehen von der Einfärbung der Bilder waren beide Testversionen identisch. Der C-CAT bestand aus 240 Cargo Röntgenbildern, wovon die Hälfte einen verbotenen Gegenstand beinhaltete. Die verbotenen Gegenständen umfassten dabei folgende Kategorien: Schusswaffen, Messer, unkonventionelle Spreng- und Brandvorrichtungen (USBV), explosives Material, Drogen, sowie weitere zollrelevante Gegenstände (z.B. chinesische Medizin, Uhren, etc.).

3. Resultate

Abbildung 2 stellt den Vergleich der Erkennungsleistung (A') aus dem C-CAT für die drei Gruppen und beide Messzeitpunkte dar. Beide Experimentalgruppen (Röntgenbilder in Graustufen vs. Pseudofarben) wiesen einen signifikanten Zuwachs in der Erkennungsleistung nach dem Training auf, während die Kontrollgruppe keine signifikante Verbesserung erzielte (Tabelle 1). Eine ANOVA mit Messwiederholungen mit dem intraindividuellen Faktor Testzeitpunkt und dem interindividuellen Faktor Gruppe (Röntgenbilder in Graustufen vs. Pseudofarben) ergab einen grossen signifikanten Effekt für Testzeitpunkt, und keinen Effekt für Gruppe oder der Interaktion zwischen Testzeitpunkt und Gruppe (Tabelle 2). Beide Gruppen erzielten einen ähnlichen Leistungszuwachs; es konnte kein Vorteil hinsichtlich einer der beiden Testversionen (Röntgenbilder in Graustufen vs. Pseudofarben) festgestellt werden. Der deutliche Zuwachs der Erkennungsleistung beider Experimentalgruppen deutet darauf hin, dass das dreimonatige Training mit C-XRT dazu beitragen konnte, die Röntgenbildinterpretationskompetenz der Screener zu verbessern.

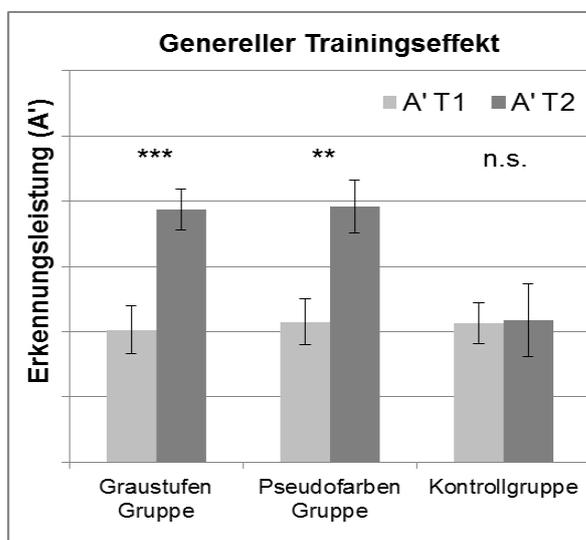


Abbildung 2: Erkennungsleistung (A')* mit Standardfehler der drei Gruppen vor (T1) und nach dem Training (T2) (Michel et al. 2014).

*aus Sicherheitsgründen sind die tatsächlichen A' Werte nicht aufgeführt.

Tabelle 1: Resultate der t -Tests zum Vergleich der Erkennungsleistung (A') vor (T1) und nach dem Training (T2) für alle drei Gruppen.

	$t(14)$	$t(15)$	$t(9)$	p Wert	D
Graustufen T1-T2	-5.38			<.001	1.39
Pseudofarben T1-T2		-3.96		<.01	1.16
Kontrollgruppe T1-T2			-.12	=.909	.04

Tabelle 2: Resultate der Varianzanalyse mit Testzeitpunkt und Gruppe als unabhängige Variablen (UVs) und Erkennungsleistung (A') als abhängige Variable (AV).

	df	F	η^2	p Wert
Testzeitpunkt (T) (T1, T2)	1, 29	40.46	.582	<.001
Gruppe (G) (Graustufen vs. Pseudofarben)	1, 29	.039	.001	=.845
T x G	1, 29	.021	.001	=.887

Detailliertere Analysen wurden für die einzelnen Kategorien der verbotenen Gegenstände durchgeführt (Abb. 3). Für jede der Kategorien war nach dem Training eine höhere Erkennungsleistung vorhanden. Die höchste Leistung erlangten beide Experimentalgruppen für die Erkennung von USBV, vor allem nach dem Training

(T2). Die niedrigste Leistung wurde bei zollrelevanten Gegenständen /Schusswaffen/Messer erzielt (Abb. 3). Signifikante Verbesserungen der Erkennungsleistung (A') konnten bei beiden Experimentalgruppen für die Kategorien Drogen, Explosives Material und USBV gefunden werden. Der grösste Trainingseffekt zeigte sich ebenfalls für die Erkennung von USBV (Tabelle 3). Wie Abbildung 3 zu erkennen gibt gab es kaum Unterschiede in der Leistung der beiden Experimentalgruppen bezüglich der einzelnen Kategorien und Messzeitpunkte. Die ANOVA mit Messwiederholungen (Tabelle 4) bestätigte diese Annahme. Es gab weder einen Haupteffekt noch signifikante Interaktionen welche den Faktor Gruppe (Röntgenbilder in Graustufen vs. Pseudofarben) involvierten. Somit konnte kein Vorteil für die Bilddarstellung in Pseudofarben nachgewiesen werden.

Weitere Analysen können in Michel et al. (2014) nachgelesen werden.

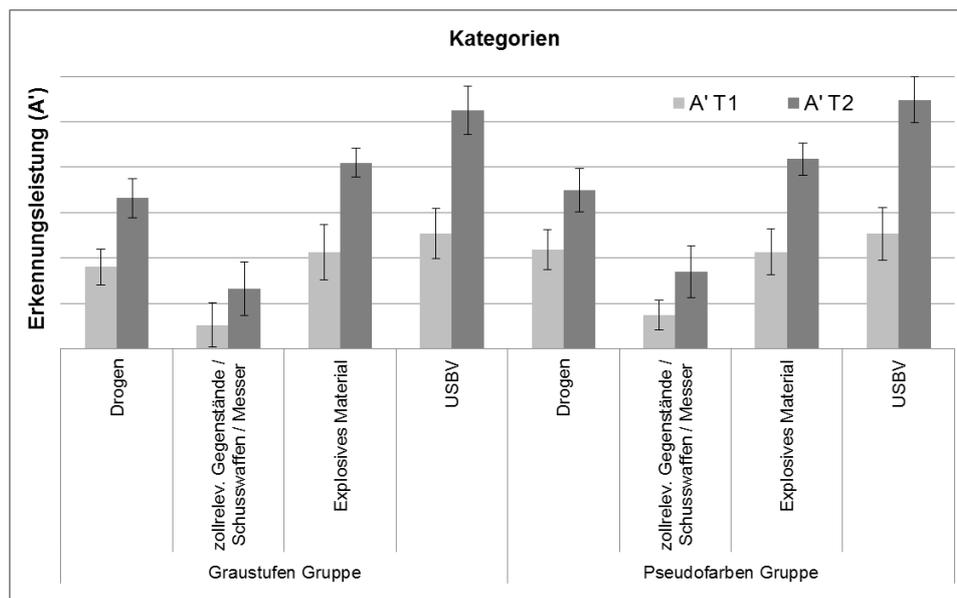


Abbildung 3: Erkennungsleistung (A')* mit Standardfehler der beiden Experimentalgruppen vor (T1) und nach dem Training (T2) für jede Kategorie (Michel et al. 2014).

Tabelle 3: Resultate der t-Tests zum Vergleich der Erkennungsleistung (A') vor (T1) und nach dem Training (T2) für jede Kategorie.

Gruppe	t(14)	t(44)	p Wert	d
Graustufen Gruppe				
Drogen	-3.54		<.01	.95
zollrelevante Gegenstände, Schusswaffen, Messer		-1.03	=.307	.21
Expl. Material	-3.67		<.01	1.06
USBV	-4.64		<.001	1.29
Pseudofarben Gruppe				
Drogen	-2.41		<.05	.71
zollrelevante Gegenstände, Schusswaffen, Messer		-1.47	=.147	.29
Expl. Material	-3.29		<.01	1.16
USBV	-4.18		<.001	1.36

Tabelle 4: Resultate der Varianzanalyse mit Testzeitpunkt, Kategorie und Gruppe als unabhängige Variablen (UVs) und Erkennungsleistung (A') als abhängige Variable (AV).

	df	F	η^2	p Wert
Testzeitpunkt (T) (T1, T2)	1, 29	37.90	.566	<.001
Kategorie (K)	2, 31, 67	33.45	.536	<.001
Gruppe (G) (Graustufen vs. Pseudofarben)	1, 29	.24	.008	=.626
T x G	1, 29	.01	.000	=.915
T x K	2, 18, 63	4.40	.132	<.05
K x G	3, 67	.09	.003	=.965
T x K x G	3, 63	.06	.002	=.982

4. Zusammenfassung und Schlussfolgerung

Ziel dieser Studie war herauszufinden, ob die Röntgenbildinterpretationskompetenz von Cargo Screenern durch den Einsatz von CBT verbessert werden kann. Übereinstimmend mit früheren Studien im Bereich der Handgepäckkontrolle (z.B. Michel et al. 2007; Koller et al. 2008; Halbherr et al. 2013) konnte ein signifikanter Zuwachs der Erkennungsleistung der Screener nach drei Monaten Training festgestellt werden. Die Erkennungsleistung der Screener war auch nach dem Training noch verbesserungswürdig, was jedoch auch auf die relativ kurze Trainingsdauer zurückzuführen ist. In anderen Studien trainierten die Screener über einen Zeitraum von min. 6-12 Monaten (Michel et al. 2007; Koller et al. 2008).

Des Weiteren konnte in dieser Studie kein Vorteil für die Darstellung mit Pseudofarben im Vergleich zu Graustufen nachgewiesen werden.

Abschliessend muss darauf hingewiesen werden, dass die in dieser Studie verwendeten verbotenen Gegenstände so in die Container eingebaut wurden, dass sie visuell erkennbar waren. In Realität können solche Gegenstände in Containern u.U. schwieriger versteckt sein. Weiterentwicklungen und Verbesserungen der Mensch-Maschine Schnittstelle und des Gesamtsystems sind aus diesem Grund wichtig. Diese sollten nicht nur das Training sondern auch technische Weiterentwicklungen und Unterstützungen, wie z.B. die automatisierte Erkennung verbotener Gegenstände (z.B. automatische Sprengstofferkennung), umfassen (siehe dazu z.B. Singh & Singh 2003; Eilbert 2009).

5. Literatur

- Eilbert RF (2009) Chapter 6 - X-ray technologies. In: Marshall MM, Oxley JC (Eds.) Aspects of Explosives Detection. Oxford: Elsevier, 89-130.
- Halbherr T, Schwaninger A, Budgell G, Wales A (2013) Airport security screener competency: A cross-sectional and longitudinal analysis. *International Journal of Aviation Psychology* 23(2): 113-129.
- Hardmeier D, Hofer F, Schwaninger A (2005) The X-ray object recognition test (X-Ray ORT) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. *IEEE ICCST Proceedings*, 39: 189-192.
- Koller S, Hardmeier D, Michel S, Schwaninger A (2008) Investigating training, transfer and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security* 1(2): 81-106.
- 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.
- Michel S, de Ruiter J, Hogervorst M, Koller S, Moerland R, Schwaninger A (2007) Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners. *Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa, October 8-11*.
- Michel S, Mendes M, de Ruiter J, Koomen G, Schwaninger A (2014) Increasing X-Ray image interpretation competency of cargo security screeners. *International Journal of Industrial Ergonomics* 44: 551-560.
- Pollack I, Norman DA (1964) A non-parametric analysis of recognition experiments. *Psychonomic science* 75:125-126.
- Schwanger A (2004) Computer based training: A powerful tool to the enhancement of human factors. *Aviation Security International* 2004(2): 31-36.
- Schwanger A, Hardmeier D, Hofer, F (2005) Aviation security screeners' visual abilities and visual knowledge measurement. *IEEE Aerospace Elct. Sys.* 20(6): 29-35.
- Singh S, Singh M (2003) Explosives detection systems (EDS) for aviation security. *Sign. Proc.* 83 (1): 31-55.