

An objectively-analyzed method for measuring the useful penetration of x-ray imaging systems

Jack L. Glover^{1,2}, and Lawrence T. Hudson¹

¹ National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899 USA

² Theiss Research, La Jolla, CA 92037 USA

Abstract—The ability to detect wires is an important capability of the cabinet x-ray imaging systems that are used in aviation security as well as the portable x-ray systems that are used by domestic law enforcement and military bomb squads. A number of national and international standards describe methods for testing this capability using the so called useful penetration test metric, where wires are imaged behind different thicknesses of blocking material. Presently, these tests are scored based on human judgments of wire visibility, which are inherently subjective. We propose a new method in which the useful penetration capabilities of an x-ray system are objectively evaluated by an image processing algorithm operating on digital images of a standard test object. The algorithm advantageously applies the Radon transform for curve parameter detection that reduces the problem of wire detection from two dimensions to one. The sensitivity of the wire detection method is adjustable and we demonstrate how the threshold parameter can be set to give agreement with human-judged results. The method was developed to be used in technical performance standards and is currently under ballot for inclusion in a US national aviation security standard.

Keywords— x-ray imaging, useful penetration, bomb squads, wire detection, aviation security, IEEE/ANSI standards, ASTM standards, image processing, object detection, portable x-ray systems

I. INTRODUCTION

SINCE Röntgen produced the first radiographs, one of the most useful properties of x rays has been their ability to penetrate through otherwise opaque objects. Indeed, when Röntgen reported his discovery of “a new type of ray” in 1895¹, he first considered “how far other bodies can be penetrated” by them. Röntgen systematically studied how the fraction of transmitted x rays changed as he varied the thickness, density, and type of material in their path [1, 2].

In medical physics, the most common way to quantify the penetration of a polychromatic x-ray beam is by measuring the half-value layer (*HVL*), which is the thickness of a material that reduces the intensity, usually measured in terms of air kerma, to 50 % of its initial value. Measurements of the *HVL* are one of the main metrological quantities used to characterize the spectral qualities of x-ray beams in many fields, such as radiation protection [3] and medical physics [4]. For a monochromatic x-ray beam of energy E , the *HVL* can be related to the linear-attenuation coefficient, $\mu(E)$, of the material using the equation,

$$\mu(E) = \frac{\log_e 0.5}{HVL}. \quad (1)$$

In x-ray imaging, penetration is sometimes defined in terms of a particular imaging task. For example, in the American National Standard for Determination of the Imaging Performance of X-Ray and Gamma-Ray Systems for Cargo

¹Röntgen's discovery was first published in German in 1895 [1], however the quote comes from the English translation that appeared in Nature the following year [2].

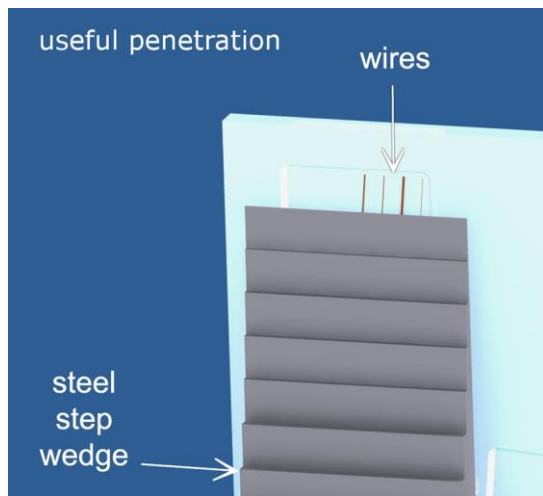


Fig. 1. Part of proposed ASTM F792-OE test object that can be used to measure useful penetration. Wires of various gauges are placed under steel steps of increasing thicknesses.

and Vehicle Security Screening, penetration is defined as the maximum thickness of steel through which the orientation of a steel arrow can be determined by a human operator [5]. The orientation of the steel arrow is varied randomly and the operator must correctly report the direction the arrow is pointing, based on the image. More and more steel blocking material is added until the operator can no longer determine the direction of the arrow. Other measures of penetration exist in other x-ray imaging subfields, such as the so called simple penetration test for checkpoint x-ray systems in which lead numerals are discerned through different thicknesses of steel and subsequently identified by an operator [6].

Useful penetration is a term that has been used in security x-ray imaging to mean the maximum thickness of material through which a wire of a particular gauge can be detected using a given imaging system and source. The results of useful penetration tests are generally expressed in terms of the amount of blocking material through which a given wire can be seen, for example, “system X can detect 20 gauge wire through up to 6 mm of steel.” One of the first useful penetration tests was created by the ASTM F792² working group in the 1980s. The test object consisted of a series of sinusoidal wires with gauges varying between AWG 22 and AWG 30 (diameters between 0.644 mm and 0.255 mm), under a step wedge of aluminum ranging in thickness from

approximately 1.5 mm to approximately 15.9 mm in 10 equal steps. The image was evaluated by an operator, who was to judge which wires were visible under which steps. The standard has been revised several times in recent decades but has always retained a useful penetration test. The standard is widely used, with many thousands of ASTM F792 test objects in use throughout the world [7].

The traditional useful penetration test, as well as most other image quality tests for security x-ray systems, fall into the class of subjective image-quality tests that are scored based on human judgments. While such tests are necessary for systems that use human operators, since they reflect end-to-end system performance that includes human factors such as operator performance and image display conditions [8], these tests also have a number of clear drawbacks. The results of such tests are always subjective, since they rely on human judgements of object visibility, which vary from person to person depending on eyesight, training, fatigue and numerous other factors. Since subjective results depend on these variable conditions, they do not accurately reflect the intrinsic image quality of the x-ray imaging system alone and are less useful for quantitative assessments such as comparing the technical performance of different systems, site acceptance testing, and the tracking system performance over time.

For these reasons, there has been a recent trend toward objectively evaluated image quality test methods in security imaging [9]. In objective image quality testing, automated analysis algorithms are applied to a set of digital images of a standard test object in order to measure a set of metrics that reflect the performance of the system. For example, the newly published IEEE/ANSI³ standard for portable x-ray imaging systems defines a set of objective image quality metrics and stands in contrast with earlier standards for these systems that relied entirely on subjective methods [10]. The American National Standard for Evaluating the Image Quality of X-ray Computed Tomography Security-Screening Systems, published in 2011, also relies entirely on objective methods and has been adopted by TSA in the United States as part of

²ASTM International is a standards development organization. ASTM originally stood for American Society for Testing and Materials.

³ IEEE is the Institute of Electrical and Electronics Engineers. ANSI is the American National Standards Institute.

factory- and site-acceptance testing.

In this paper, we propose an objectively evaluated method for measuring the useful penetration of an x-ray system. The proposed test object consists of a series of standard-gauge wires placed under a steel step wedge (see Figure 1, for example). Images of the wires under each successive step are analyzed until their visibility is reduced below a specified threshold. The proposed test is applied to three x-ray checkpoint systems currently in use at aviation checkpoints. The results of these tests are anonymized with respect to vendor.

The need for such a method was identified in a working group of the ASTM F792 [6], which is currently under revision. The standard has previously relied entirely on subjective image-quality test methods, however a standalone sub-working group was created under the current revision (co-chaired by the authors) to investigate the feasibility of introducing objective methods. The sub-working group developed a new test object and objective test methods for measuring the image quality of checkpoint x-ray systems. This paper reports the novel method for measuring useful penetration. The standard also includes methods for measuring the metrics steel differentiation, spatial resolution, organic differentiation, dynamic range, and frequency-dependent noise, but these are not discussed here.

II. AN OBJECTIVELY-ANALYZED METHOD FOR MEASURING USEFUL PENETRATION

For the purposes of testing, a prototype test object was constructed using unclad enameled round copper wires of gauges varying between 16 AWG (1.291 mm) and 30 AWG (0.255 mm) placed under a steel (UNS S30400) step wedge. The wires were oriented at an angle of approximately 5° to reduce effects caused by pixel aliasing. The prototype test piece was then imaged using three x-ray checkpoint systems,

referred to as systems A, B and C (pixel size approximately). Images of the test piece were acquired in four different image orientations, rotated nominally 90° from each other, and always with the long axis of the test piece nominally parallel to or normal to the belt direction. The surface of the thickest step of the step wedge should be normal to the direction of the x-ray beam.

A. Processing Images of Wires Using the Radon Transform

The Radon transform has been advantageously applied to curve parameter detection in digital images. In particular, it is well suited to perform line parameter extraction even in the presence of noise [11], in this case the x-ray image data of wires on the ASTM useful penetration test object that are blocked by increasing thicknesses of steel. Here the Radon transform will be used to map lines within the image spatial domain to points in the Radon domain of line parameters (*viz.*, the parameters *angle* and *distance*, see Figure 2). Various forms of this technique have found broad application, *e.g.* in the automated reading of barcodes and in reconstructing images in tomography. The use of the Radon transform here offers the following advantages: (1) a global detection problem is converted into an easier local task; (2) it is less sensitive to noise in the image domain where other methods are often challenged, for example trying to align individual noisy rows to perform a column sum of pixel values; (3) the actual parameters of the lines are recovered, providing a check on test-object alignment and a check against false positives since both the relative positions and approximate orientations of the wires are known *a priori*; (4) a single, ROI (Region of Interest) is selected for each step which allows for the consistent analysis of all wire gauges in one computational step; and finally, (5) since the test piece employs parallel wires, a single column of the Radon domain can be analyzed to determine the visibility of individual wires on a given step.

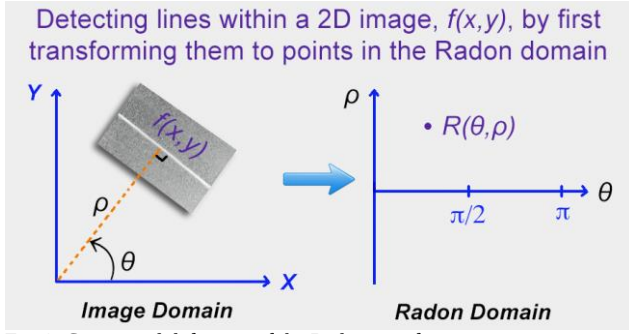


Fig. 2. Conceptual definition of the Radon transform.

The Radon transform can be defined in different ways, but is conventionally introduced as an integral transform of an integrable and/or continuous function in two dimensions [12]. Here the Radon transform $R(\theta, \rho)$ of $f(x, y)$ is computed by performing line integrals through $f(x, y)$ at different offsets ρ_i and angles θ_k from the origin, where ρ is the perpendicular distance from the line to the origin, and θ is the angle formed by the offset parameter ρ to the line (as shown in Figure 2). A mathematical expression for the Radon transform is given in Figure 3, where δ is the Dirac delta function. With respect to the argument of the delta function, it is noted that $\rho = x \cos \theta + y \sin \theta$ is simply the definition of a line in terms of the parameters ρ and θ . Also shown is a projection of a series of line integrals at θ_k which will form a single column in the Radon-domain image. The two peaks visible in this projection $R(\theta_k, \rho_i)$ illustrate “detection” of two lines that were oriented at θ_k in the synthetic image $f(x, y)$.

When the Radon transform is applied to discretized data, such as two-dimensional pixel arrays, one must choose a convention for evaluating the various line integrals. It is common to partition the Radon domain into two regions: one for more vertical lines and one for more horizontal lines. For the particular application considered in this document, only the integrals along more vertical lines are relevant (*i.e.* $|\sin \theta| < \sqrt{2}/2$).

Consider an M column by N row grid of pixel values $A[m, n]$ where m and n are the discrete row and column indices. We will calculate the discrete Radon transform using a sum over rows for near vertical line integrals or a sum over columns for near horizontal line integrals.

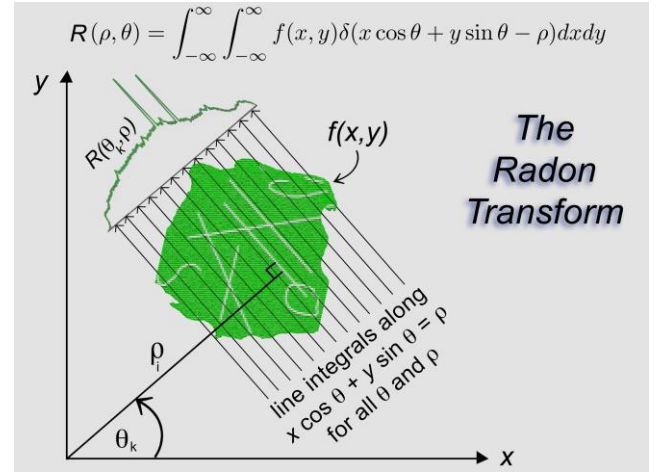


Fig. 3. A mathematical expression for the Radon transform, where δ is the Dirac delta function. With respect to the argument of the delta function, it is noted that $\rho = x \cos \theta + y \sin \theta$ is simply the definition of a line in terms of the parameters ρ and θ . The two peaks in the function $R(\theta_k, \rho)$ illustrate “detection” of two lines that were oriented at θ_k in the synthetic green image $f(x, y)$.

$$R[\rho, \theta] = \begin{cases} \frac{1}{|\sin \theta|} \sum_{m=1}^M A[m, \{n'\}], & |\sin \theta| > \frac{\sqrt{2}}{2} \\ \frac{1}{|\cos \theta|} \sum_{n=1}^N A[\{m'\}, n], & |\sin \theta| \leq \frac{\sqrt{2}}{2} \end{cases}, \quad (2)$$

where the curly brackets indicate that the value should be rounded to the nearest integer. The values of n' and m' are given by

$$n' = \frac{\rho}{\sin \theta} + \frac{m_c - m}{\tan \theta} + n_c, \quad (3)$$

$$m' = \frac{\rho}{\cos \theta} + (n_c - n) \tan \theta + m_c, \quad (4)$$

where (m_c, n_c) is the position of the origin about which the Radon transform will be computed. In this implementation, we choose this to be the center of the image, but this is not required.

In our application, the ROI is a rectangular area, given by $A[m, n]$, that is selected from each imaged step of the test object that includes the shielded wires. In practice, a discrete approximation of the Radon transform of the image minus its median pixel value will be evaluated over a regular grid with an angular step size of 1° for lines within plus or minus 20° of vertical and a step size of 1 pixel in ρ over the full range of lines that intersect the image. The discrete Radon transform has been implemented as a plug-in in many of the popular image or data analysis environments, some of which are open source. Off-grid values of $A[m, n]$, are handled with nearest-neighbor or linear interpolation methods. While the results are fairly similar, we propose the nearest-neighbor as the standard

method as it is computationally simpler. The parameters of the minimum pixel value in the Radon transform ($\theta_{\min}$, $\rho_{\min}$) should occur for the line integral that is most closely oriented along the length of the darkest of the wire signatures in the ROI (*i.e.* most x rays attenuated). The column within $R[\theta, \rho]$ where $\theta = \theta_{\min}$ is defined as the Wire Profile Function (WPF). Note that since all the wires are, by design, nominally parallel on the test object (same θ), it follows that (a) the WPF will include all detectable signatures from all the wires, and (b) the $\theta_{\min}$ from the 0 mm step should be used for analysis of the ROIs of other (more-noise-prone) steps. Figure 4 shows an example ROI, its Radon transform, and the WPF.

B. Testing the Statistical Significance of the Signal Due to the Wire

A wire is said to be visible if its signature in the WPF is different from the background level at a defined level of statistical significance. The following reviews the terminology and formulas of statistical hypothesis testing and applies them to the problem at hand.

Here the problem of peak detection is framed in terms *t*-testing. This begins with stating a null hypothesis, in this case, that any given point, WPF_i , is not significantly below (since the peaks are “loss” peaks) the mean of the background level (defined as a set of contiguous points in the WPF known *a priori* to not contain wire signatures). Next a *t*-statistic is computed, then an assumed probability density function (PDF) is integrated from *t* to infinity, to determine the probability that the non-background WPF_i is consistent with having been sampled from the same population as the background points. If this probability is less than a pre-specified level of statistical significance, called α , then the null hypothesis is rejected and the non-background point will be flagged as “significant” and therefore consistent with peak detection in the Radon domain corresponding to wire visibility in the checkpoint-image domain. Note that α is also the probability of falsely rejecting the null hypothesis, if the data conformed to the independency and distribution assumptions of *t*-testing. These data, however, are *not* assumed to be distributed normally *nor* sampled independently; nevertheless the use of the formula statistics of

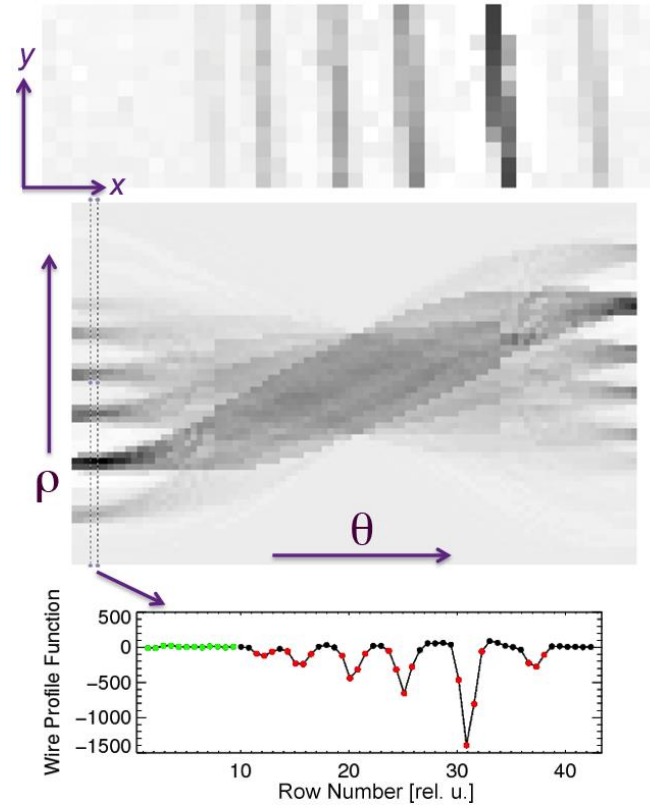


Fig. 4. TOP: The ROI from the x-ray image showing six wires oriented about 5 degrees from the vertical. MIDDLE: The Radon transform of the ROI minus its median pixel value, using the nearest-neighbor method and with the origin of the input image taken to be at the center. The column with the minimum value is highlighted. This column corresponds to a θ value of about 5 degrees, *i.e.* the angle of the wires when the test object is aligned along the image grid. This column of data is referred to as the Wire Profile Function, plotted at BOTTOM. The points in green were used as the background region. The points plotted in red were found to differ significantly from background levels using the criterion discussed below in section II B.

t-testing is motivated by its robustness against assumption violations, and is ultimately validated in our context only by comparing its performance against humans. Hence true significance levels are technically unknown, and the appropriate value of α will be established by comparison to human-judged results.

The test method first requires identification of a background region of at least 15 contiguous WPF points so that a representative mean, μ_{bkg} , and standard deviation, σ_{bkg} , can be computed. These points are shown with green plot symbols in the example of Figure 4. Next, the following *t*-statistic is computed for each point in the WPF:

$$t_i = \frac{\mu_{bkg} - WPF_i}{\sigma_{bkg}}. \quad (5)$$

The probability of the null hypothesis is then calculated by integrating the probability density function from *t* to +infinity. Student’s *t*-distribution for the appropriate number of degrees

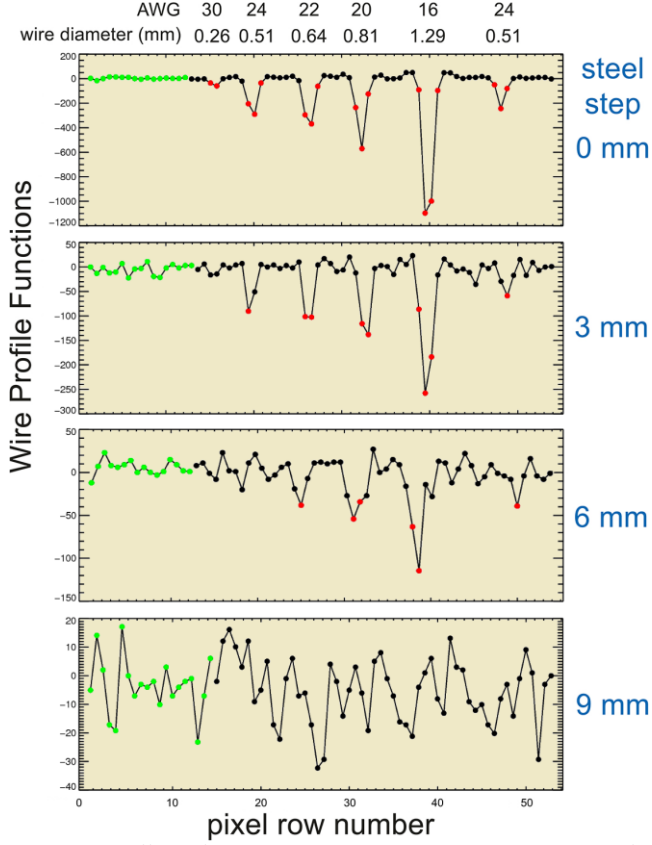


Fig. 5. The effect of increasing step thickness upon the wire profile function for System A, orientation 2. The green symbols indicate the background region and the red data show the points returned by the algorithm as statistically significant, and hence the wires that were scored as visible.

of freedom would normally be evoked for situations where the sample size is small and population standard deviation is unknown. As the number of degrees of freedom increases, the t -distributions approach the standard normal distribution. For simplicity, it is proposed to use the normal distribution as the PDF in all the t -testing proposed here. A so-called p -value is then calculated by integrating the standard normal distribution from t to $+\infty$, which results in:

$$p = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{t^2}{2}} = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{t}{\sqrt{2}} \right) \right], \quad (6)$$

where $\operatorname{erf}()$ is the usual error function.

A t -test should be calculated for each non-background WPF _{i} . For the example in Figure 4, there are 46 non-background data points in the WPF. If the total number t -tests is N_{tot} , then since $N_{tot} \gg 1$, the increased chance of false positives must be taken into account. The simplest and most conservative way is to

Image 11

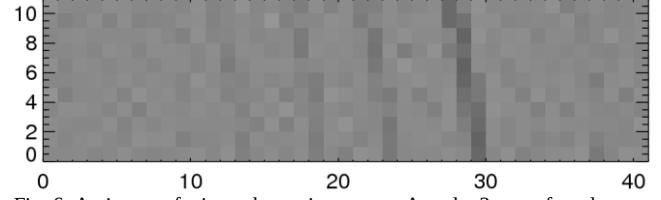


Fig. 6. An image of wires taken using system A under 3 mm of steel.

test each individual hypothesis at a statistical significance level of $1/N_{tot}$ times what it would be if only one hypothesis were tested (this is known as the Bonferroni correction/ p -value adjustment). So, if the desired significance level for the whole family of t -tests is (at most) α , then the threshold for each of the individual tests should be at a significance level of α/N_{tot} . For the analysis presented in this report, α was set to a value of 8.8×10^{-5} in order to best agree with results based on human observations using the same set of images (see section II C for a detailed account).

If one or more of the t -tests associated with a particular wire are found to be significantly different from the background level, as defined above, then that wire will be scored as detected. This procedure was applied to the data in Figure 4, and the WPF points shown as red dots in Figure 4 BOTTOM were found to be “significant.” Figure 5 similarly shows some example WPFs from images of wires beneath increasing thicknesses of the steel step wedge with the significant points shown as red dots. The reported result of the test for a given wire gauge is defined as the thickest step of steel under which the wire can still be detected in at least three out of the four image orientations. We require detection in only three out of four image orientations because we are working close to the limit of detection, where random noise fluctuations can push a wire above or below the threshold of visibility. Applying this

TABLE I
USEFUL PENETRATION RESULTS

Wire Gauge	Wire Diam. (mm)	System A	System B	System C
20 AWG	0.812	6 mm	0 mm	6 mm
24 AWG	0.511	3 mm	3 mm	3 mm
30 AWG	0.255	0 mm	0 mm	wire not visible

The useful penetration results for the three checkpoint systems. The values reported are the steel thickness under which images of the wires are visible to the significance specified in the test and in three of the four image orientations. The results correlate well with human perception testing using the same set of images (see Table II), including the unintuitive 0 mm for Sys B (20 gauge wire). The 30 gauge wire was not visible at the 0 mm “step” for Sys C.

methodology to the three checkpoint x-ray systems tested produced the results given in Table I. The three systems have an effective pixel size in the plane of the test object varies between 1.1 mm and 0.6 mm. The 20 AWG wire is, therefore, of the order of the pixel width while the 24 AWG and 30 AWG wires have diameters that are much smaller than the width of a pixel.

C. Setting the Threshold for Visibility

The performance of the wire-detection algorithm described above was adjusted by design to correlate well with the results of observations made by experts. Images of wires under the steel step wedge were distributed to a dozen NIST researchers to score for visibility. Figure 6 shows one of the 48 images (3 checkpoint systems x 4 orientations x 4 steps). The most common interpretation of this image was that there were four wires visible and that they intersected the x-axis at approximately $x = 13, 18, 23, 29$. Two participants also saw a wire intersecting at around $x = 37$.

Each image was created by imaging 6 wires of various gauges under a steel step of some thickness (including 0 thickness). The researchers identified between 38 and 69 wires. The useful penetration test can be scored using the results of the researchers' observations to judge which wires

TABLE II
THE RESULTS OF THE USEFUL PENETRATION TEST AS JUDGED BY THE
MEDIAN OF THE RESEARCHERS' OBSERVATION RESULTS

Wire Gauge	Wire Diam. (mm)	System A	System B	System C
20 AWG	0.812	6 mm	0 mm	6 mm
24 AWG	0.511	3 mm	3 mm	3 mm
30 AWG	0.255	0 mm	0 mm	Wire not visible

are visible. The median of all the observation results are shown below in Table II. The median number of false positives identified was 2.

A detailed comparison of the results of the researcher observations and the algorithmic results are given in table III. The table includes results for the algorithm using different threshold α values and the best agreement with the researcher observation results was for $\alpha = 8.8 \times 10^{-5}$. With this value of α , the number of wires detected by the algorithm matches exactly the median number detected using the researcher observations. It is noted that the algorithm was able to reach higher detection probability than almost all the human participants while maintaining a low number of false positives. It is also the case that the number of false positives corresponds very closely to the number predicted by a normal noise model (48 multiplied by α). This suggests that the model assumptions underlying t-testing were not strongly violated.

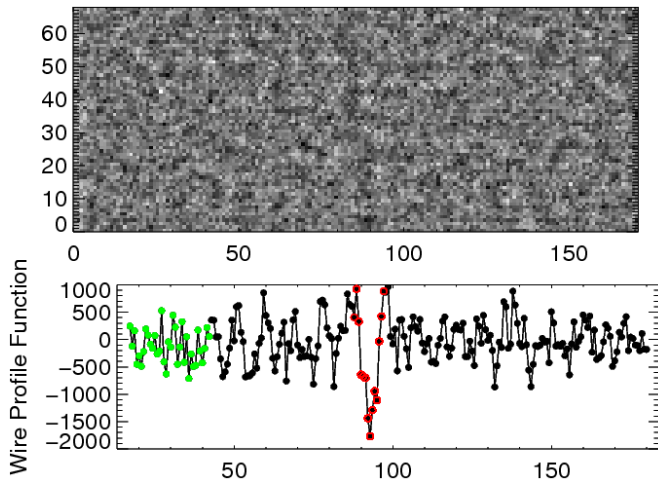


Fig. 7. The upper portion of the figure shows an image taken using a portable x-ray imaging system of a 20 AWG (diameter 0.812 mm) wire under a 15 mm steel step. The lower portion of the figure shows the derived WPF. The multi-point t -test allowed for the detection of a wire that would have gone undetected using the single-point t -tests.

It is interesting to consider to what critical t -statistic this α corresponds. The median N_{tot} value of the 48 images was 50. It follows that the t -statistic must be greater than 4.6 in order to be considered significantly below background.

III. EXTENSION TO HIGHER RESOLUTION SYSTEMS

The method described so far was designed for testing aviation checkpoint x-ray systems, for which the wires of interest had a diameters of the order of a pixel or less and so a single-point t -test was sufficient. However, the much-higher resolution and pixel density of bomb squad imaging systems recommends t -testing of sets of adjacent points from the test region, rather than just single points. For example, when imaging a thin wire with a high-resolution bomb squad system, one can obtain a signal from a wire that is spread over several pixels with values that are all only slightly different from background. On occasion it is possible that none of the individual points may be significant when judged using the single-point t -test, but when a series of adjacent points are considered collectively, wire visibility is attained. As such, multipoint (“two-sample”) t -testing may be necessary to detect a wire. An example of such a case is shown in figure 7.

TABLE III
COMPARISON OF USEFUL PENETRATION RESULTS: RESEARCHER OBSERVATION RESULTS AND THE PROPOSED ALGORITHM

Wire detection method	Wires detected	False positives	System A results	System B results	System C results
Human results	38 to 69	0 to 6	Median results	Median results	Median results
	median: 59	median: 2	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 0 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 6 mm AWG24: 3 mm AWG30: none
Algorithm results					
$\alpha = 0.1$	70	5	AWG20: 6 mm AWG24: 6 mm AWG30: 0 mm	AWG20: 3 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm
$\alpha = 0.01$	65	1	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 3 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm
$\alpha = 1E-4$	60	0	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 0 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm
$\alpha = 8.8E-4$	59	0	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 0 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm
$\alpha = 1E-5$	58	0	AWG20: 6 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 0 mm AWG24: 3 mm AWG30: 0 mm	AWG20: 3 mm AWG24: 3 mm AWG30: 0 mm

We will now generalize the statistical tests employed earlier for checkpoint systems. Consider a set of N_s consecutive data points in the WPF ($N_s > 1$). The null hypothesis is now that the set of points N_s and the set of background points come from distributions that possess the same mean and standard deviation. The first step in calculating the probability of this null hypothesis is to calculate the unpaired (independent) Student's t -statistic,

$$t = \frac{\mu_{bkg} - \mu_s}{\sigma_{bkg,s} \cdot \sqrt{\frac{1}{N_{bkg}} + \frac{1}{N_s}}}, \quad (7)$$

where the subscript bkg indicates background and the subscript S indicates the sample of N points under consideration. The quantity $\sigma_{bkg,s}$ is the pooled standard deviation (weighted average of σ_{bkg} and σ_s) and is given by:

$$\sigma_{bkg,s} = \sqrt{\frac{(N_{bkg}-1)\sigma_{bkg}^2 + (N_s-1)\sigma_s^2}{k}}, \quad (8)$$

where $k = N_s + N_{bkg} - 2$ is the number of degrees of freedom.

The probability of the null hypothesis is calculated by integrating the appropriate probability density function from t to ∞ :

$$p = F(t, k) = \int_t^\infty f(x, k) dx.$$

Student's t -distribution should technically be used for the function f , which itself is a function of k . As before, we approximate this by the normal distribution since the number of degrees of freedom is sufficiently large. Therefore, the p -values calculated here do not correspond to actual probabilities for our real, non-normally distributed data. As such, final thresholding to define wire visibility will be chosen as before, based upon comparison to researcher observations. The p -value is then calculated as before by integrating the standard normal distribution from t to infinity:

$$p = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{t^2}{2}} = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{t}{\sqrt{2}} \right) \right].$$

These p -values are compared to a threshold value of α/N_{tot} and the α should be set based on human perception tests performed by bomb squad personnel on images made using bomb squad systems.

This method was applied to measure the useful penetration

TABLE IV
BOMB SQUAD USEFUL PENETRATION RESULTS

Wire Gauge (diam.)	Bomb squad system A	Bomb squad system B	Bomb squad system C
20 AWG (0.812 mm)	9 mm	15 mm	3 mm
24 AWG (0.511 mm)	6 mm	9 mm	3 mm
32 AWG (0.202 mm)	0 mm	6 mm	3 mm
36 AWG (0.127 mm)	not visible	0 mm	0 mm
41 AWG (0.071mm)	not visible	0 mm	0 mm

The useful penetration of a number of bomb squad systems was determined for a selection of wire gauges. The pixel size for these systems was between 0.05 mm and 0.2 mm.

of three bomb squad systems for a variety of wire gauges. In the absence of a human perception results that are specific to bomb squad systems, the α value for checkpoint systems was used to measure the ability of the bomb squad portable x-ray system to provide images of wires under different thicknesses of steel. The test results are given in table IV.

IV. CONCLUSION

An image processing algorithm has been developed for the determination of the x-ray image quality metric, useful penetration. Useful penetration is evaluated, in x-ray imagery, by determining the visibility of wires through different thicknesses of steel. The image processing algorithm for computing useful penetration allows for its objective evaluation. The sensitivity of the algorithm can be adjusted depending on the application. The algorithm was developed for testing of cabinet x-ray systems used in aviation security and has been introduced in an upcoming revision of the ASTM F792 performance standard. An extension of the method was also described that could be used to test higher-resolution systems, such as the portable x-ray systems used by bomb squads.

ACKNOWLEDGMENT

We gratefully acknowledge helpful discussions with Hariharan K. Iyer of the NIST Statistical Engineering Division. This material is based upon work supported by the Science and Technology Directorate of the U.S. Department of Homeland Security under Awards # HSHQPM015-T-00021.

REFERENCES

- [1] W. C. Röntgen, "Ueber eine neue Art von Strahlen: Vorläufige Mittheilung," *Sitzungsberichten der Würzburger Physik.-medic. Gesellschaft*, vol. 137, pp. 132-141, 1895.
- [2] W. C. Röntgen, "On a New Kind of Ray," *Nature*, vol. 53, pp. 274-276, 1896.
- [3] S. C. Klevenhagen, R. J. Aukett, R. M. Harrison, C. Morett, A. E. N. and K. E. Rosser, "The IPEMB code of practice for the determination of absorbed dose for x-rays below 300 kV generating potential (0.035 mm Al–4 mm Cu HVL; 10–300 kV generating potential)," *Physics in Medicine and Biology*, vol. 41, no. 12, p. 2605–2625, 1996.
- [4] R. L. Kruger, C. H. McCollough and F. E. Zink, "Measurement of half-value layer in x-ray CT: A comparison of two noninvasive techniques," *Medical Physics*, vol. 27, pp. 1915-1919, 2000.
- [5] *American National Standard for Determination of the Imaging Performance of X-Ray and Gamma-Ray Systems for Cargo and Vehicle Security Screening*, IEEE/ANSI Standard N42.46-2008, 2008.
- [6] *Standard Practice for Evaluating the Imaging Performance of Security X-Ray Systems*, ASTM F792-08, 2008.
- [7] Personal communication with Larry Copello, a manufacturer of x-ray test objects, who reports having sold 2000 to 5000 ASTM F792 test objects at the time this paper was prepared.
- [8] K. K. Shieh and C. C. Lin, "Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference," *International Journal of Industrial Ergonomics*, vol. 26, no. 5, pp. 527-536, 2000.
- [9] L. Hudson, F. Bateman, P. Bergstrom, F. Cerra, J. Glover, R. Minniti, S. Seltzer and R. Tosh, "Measurements and standards for bulk-explosives detection," *Applied Radiation and Isotopes*, vol. 70, no. 12, pp. 1037-1041, 2012.
- [10] *American National Standard for Evaluating the Image Quality of X-ray Computed Tomography (CT) Security-Screening Systems*, IEEE/ANSI Standard N42.45-2011, 2011.
- [11] J. Zhang, B. Xin and X. Wu, "Density measurement of yarn dyed woven fabrics based on dual-side scanning and the FFT technique," *Measurement Science and Technology*, vol. 25, no. 11, p. 115007, 2014.
- [12] P. Toft, *The Radon Transform - Theory and Implementation*, Technical University of Denmark, Ph.D. Thesis, 1996.