

A Clinical Evaluation of the Image Quality Computer Program, CoCIQ

E. Norrman,¹ M. Gårdestig,² J. Persliden,^{1,2,3} and H. Geijer⁴

To provide an objective way of measuring image quality, a computer program was designed that automatically analyzes the test images of a contrast-detail (CD) phantom. The program gives a quantified measurement of image quality by calculating an Image Quality Figure (IQF). The aim of this work was to evaluate the program and adjust it to clinical situations in order to find the detectable level where the program gives a reliable figure of the contrast resolution. The program was applied on a large variety of images with lumbar spine and urographic parameters, from very low to very high image qualities. It was shown that the computer program produces IQFs with small variations and there were a strong linear statistical relation between the computerized evaluation and the evaluation performed by human observers ($R^2 = 0.98$). This method offers a fast and easy way of conducting image quality evaluations.

KEY WORDS: CDRAD phantom, image quality, computed calculated image quality figure, contrast-detail curves

INTRODUCTION

In the process of optimization in radiology, the dose administered to the patient shall be as small as possible and the image quality good enough for the diagnostic task to be solved. In this process, the image quality is often subjectively judged by radiologists. There are different methods used to determine the image quality of an imaging system. One such method is the Receiver Operating Characteristic (ROC) technique.¹⁻³ Other methods used to evaluate images are the Visual Grading Analysis Score (VGAS) and the Image Criteria Score.^{1,4} In the methods mentioned above, the whole imaging chain (including the radiologist) is evaluated. However, it is often found that there can be large interobserver variability in the opinion of what a good image looks like. A way to more objectively assess image quality is to use

contrast-detail phantoms (CD phantoms). In the literature, there are several examples where CD phantoms have been used.^{5,6} The evaluation is performed by letting human observers read and evaluate images of the phantom. From these evaluations, a graphical description of image quality, a contrast-detail curve (CD curve), can be obtained and a quantified measurement of image quality, the Image Quality Figure (IQF), can be calculated. This value is a measure of the just-visible contrasts and details in the image. It has been observed that, with CD phantom, there can also be large variations in subjective opinion regarding image quality between different observers.^{6,7} At our Radiology department, the x-ray images are totally digitized. In the quality control program and optimization processes of this department, a matter of concern has been the ability to conduct image quality evaluations that would exclude human variations in the results. A computer program, which reads and analyzes the x-ray images of the CDRAD phantom (Instrumentele

¹From the Department of Natural Sciences, Örebro University, S-70182 Örebro, Sweden.

²From the Department of Radiation Physics, IMV, Linköping University, S-58185 Linköping, Sweden.

³From the Department of Medical Physics, Örebro University Hospital, S-70185 Örebro, Sweden.

⁴From the Department of Radiology, Örebro University Hospital, S-70185 Örebro, Sweden.

Correspondence to: Eva Norrman, Department of Medical Physics, Örebro University Hospital, 70185 Örebro, Sweden; tel: +46-19-6022433; fax: +46-19-106738; e-mail: eva.norrman@orebroll.se

Copyright © 2005 by SCAR (Society for Computer Applications in Radiology)

Online publication 19 April 2005

doi: 10.1007/s10278-004-1036-0

Dienst, Nijmegen, The Netherlands),⁵ was designed to perform objective image quality evaluations. The program generates the contrast-detail curve and calculates the IQF. In this study, the program was applied on images exposed with varying parameters used in urographic and lumbar spine examinations, and the question addressed was: Can the computer program evaluate the images and give similar but more objective results as compared to human observers? Computerized evaluations of image quality have previously been performed in the field of mammography.^{8,9}

MATERIALS AND METHODS

The CDRAD Phantom

A radiograph of the CDRAD phantom is shown in Figure 1. The phantom consists of a PMMA plate $256 \times 256 \times 10$ mm, divided into 225 squares of equal sizes. In each square, except at the top three rows where only one hole is drilled, two holes of identical sizes are drilled—one in the center and the other randomly located in one of the four corners. In each column the diameter of the holes decreases in 15 steps, representing spatial variation. In each row the depth decreases, showing contrast variation. The reader tries to visualize the just-visible hole in each column. The quadrants containing the just-visible holes

can be combined to form the CD curve. An evaluation of a high-quality image where small holes and low contrasts can be seen results in a CD curve in the lower left part of the image. To avoid adaptation to the drill pattern, there are four test phantoms with different patterns that differ in the random distribution of the hole in the corner.

Radiographic Images

The x-ray images of the CDRAD phantom were acquired with tube potential settings as used in the clinics for lumbar spine and urography examinations. A variety of mA s values (tube charges) was used to determine if the program could work with different levels of noise. This would also indicate that the results could be valid for different patient thicknesses and different doses in the optimization process, neglecting the change of scattered radiation with patient thickness. We used 15 cm of PMMA with a CDRAD phantom in the middle to simulate the scattering effect created by a standard human body. To compensate for random variations in noise, three images were exposed at each setting. Image processing was performed according to the manufacturer and as used in clinics (urology, lumbar spine). The x-ray system used a flat-panel detector (Pixium 4600, Trixell, Moirans, France) mounted in a bucky table with a moving grid (Digital Diagnost, Philips Medical Systems, Best, The Netherlands). Images were stored in the PACS (Sectra, Linköping, Sweden/Philips Medical Systems).

From the PACS, the images were exported to a local computer server where our MATLAB (The MathWorks, Natick, MA, USA) computer program, CoCIQ, could read the images into MATLAB. The images were evaluated by the computer and

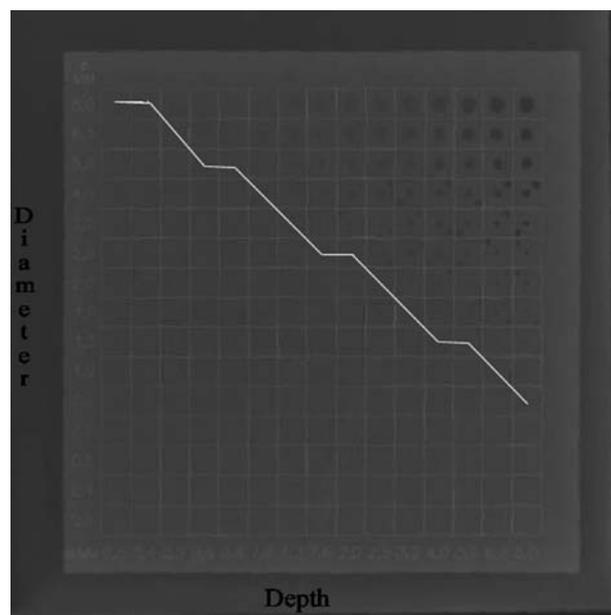


Fig 1. Radiographic image of the contrast detail phantom with a schematic illustration of the contrast detail curve connecting the just visible holes.

by human observers. The human observers were allowed to adjust the gray scale and to zoom the images to their preference.

Two sets of images were evaluated.

- (1) The tube potential was set to 77 kV, which is the standard clinical setting for lumbar spine examinations. The images were exposed in a wide range of mA s values: 2, 5, 10, 16, 20, 25, 32, 40, 63, 80, and 100 mA s. The images were evaluated by the computer and by three human observers (one radiologist and two medical physicists). It has been reported that, for images of a contrast-detail phantom, there are no significant differences between radiologists and non-radiologists.⁷
- (2) Images used in a previous study,⁶ which were evaluated by four human observers, two radiologists and two medical physicists. The evaluations performed by the human observers were compared to the computer evaluation. The mA s values used were 10.3, 6.8, and 4.5 mA s. The tube potential was set to 70 kV as for urographic examinations.

The Computer Program, CoCIQ

Computer Calculated Image Quality (CoCIQ), which is written in Matlab, reads and analyzes the x-ray images of the CDRAD phantom.¹⁰ The program steps through the squares of the image (Fig 1) column by column until no hole can be detected. One extra square is examined in each column for analysis later. In order to correctly position the randomly positioned holes, each square is divided into nine subsquares of equal sizes. The subsquares were labeled 1–9 from left to right, where 1 is in the upper left corner and 9 is in the lower right corner.

Four background levels are determined in the subsquares 2, 4, 6, and 8, where no hole is positioned and the background level standard deviation is calculated as the mean value of the standard deviations in these four subsquares. The pixel values in each corner that might contain a hole 1, 3, 7, or 9 are then registered. The lowest pixel value is compared to the mean value from the other subsquares (1, 3, 7, 9). The difference in pixel values are recognized as a signal by the program. In order for a hole to be registered as “detected,” the contrast difference has to be equal or has to exceed a multiple of the standard deviation in background level according to Equation (1).

$$D \geq k \times \sigma \quad (1)$$

where D is the difference in pixel value, k is a constant set by the user, and σ is the standard deviation in background level. The constant k , set by the user, is called threshold.

The CoCIQ indicated corners are compared to the true hole positions and corrected according to a correction scheme.¹¹ In the correction scheme, there are three possibilities for each observation:

- T: The eccentric hole was indicated at the true position.
- F: The eccentric hole was indicated at the false position.
- N: The eccentric hole was not indicated at all.

A True detection is valid only if true detections exist in at least two of the four neighboring regions (nearest horizontal and vertical squares). A False detection or no detection is converted to a True detection if true detections exist in three neighboring regions. There are some rule exceptions that apply in the four

corners where there are only two neighboring regions. A True detection is valid if at least one of the neighbors is detected as true. A False or No detection is converted into a True detection if True exists in both neighboring regions. The corrections result in a more even CD curve.

Summation over all contrast columns yields the IQF, defined in Equation (2).

$$IQF = \sum_{i=1}^{15} c_{i,th} \times \phi_i \quad (2)$$

where $c_{i,th}$ is the threshold depth of the just-visible hole and ϕ_i is the diameter in contrast column i . In a column where no hole can be seen, the depth is set to 10 mm. A lower IQF indicates better image quality.

EVALUATION OF THE PROGRAM

The ability of the program to obtain stable results was tested by evaluating the images repeatedly at different thresholds.

The variation of IQF as a result of different threshold settings was investigated with two objectives: (1) to find operating levels where the program detects as much as possible; (2) to find the operating level that reflects the results as performed by human observers.

The variation of the IQF with mA s value was studied to see if the program could work with different levels of noise and to find out if it reflects the variation in image quality due to different noise levels.

The computer results were compared to the results achieved by the human observers. The comparison was performed for the images exposed at levels between 2 and 100 mA s. The IQF values at thresholds 0.015, 0.1, and 0.2 were calculated and compared to the mean value of the IQF achieved by the three human observers.

The correlation between the results obtained by the program and the human observers were determined by linear regression. To see the variation between the IQF as performed by the human observers and the computer, the computer evaluation of the images exposed between 4 and 10 mA s at threshold 0.2 were compared with the IQF achieved by four human observers.

RESULTS

Reproducibility

By testing the same image repeatedly at different thresholds, it was shown that the variations

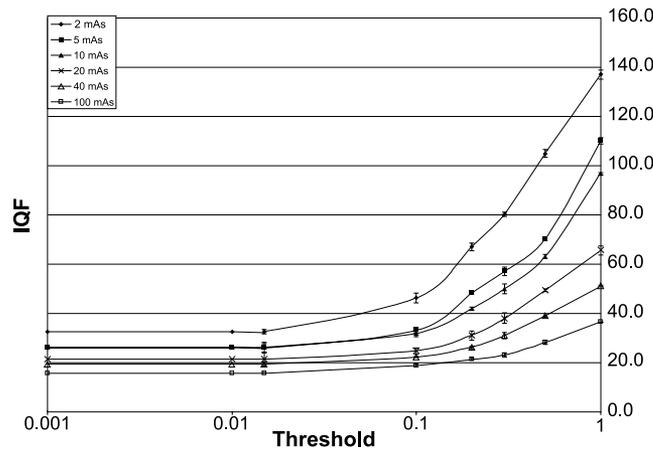


Fig 2. Image quality figures (IQF) vs. threshold for images exposed with tube charges between 2 and 100 mA s. The values are shown with \pm one standard deviation.

in IQF were small (within 5%) at all threshold and noise levels. Hence the results presented by the computer were regarded as stable.

IQF continued to decrease down to approximately threshold 0.015.

IQF vs. mA s Value for Human Observers and the Computer

Threshold

Figure 2 shows the IQF plotted as a function of different thresholds for the images exposed with lumbar spine parameters. The values are shown with one standard deviation. We can see that for all mA s value levels except for the lowest, the IQF decreased with the decreasing level of threshold down to 0.1, after which the IQF became constant. For the lowest mA s value, the

In Figure 3, the IQF for the images exposed with lumbar spine parameters is plotted as a function of mA s value for the computer and the human observers. The computer-calculated IQF was generated at the three lowest threshold levels 0.015, 0.1, and 0.2. The average IQF value for the human observers were used and shown with one standard deviation. The IQF generated by the computer and the human observers follow the

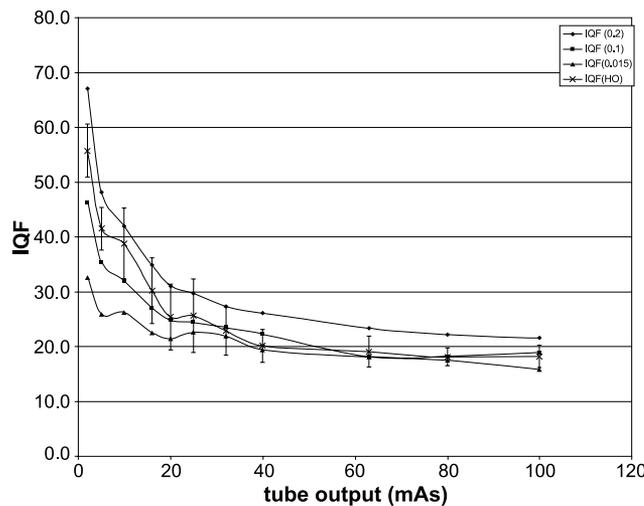


Fig 3. IQF as a function of tube output for the computer program at threshold 0.2, 0.1, 0.015, and the mean value in IQF by the human observers (HO) shown with \pm one standard deviation.

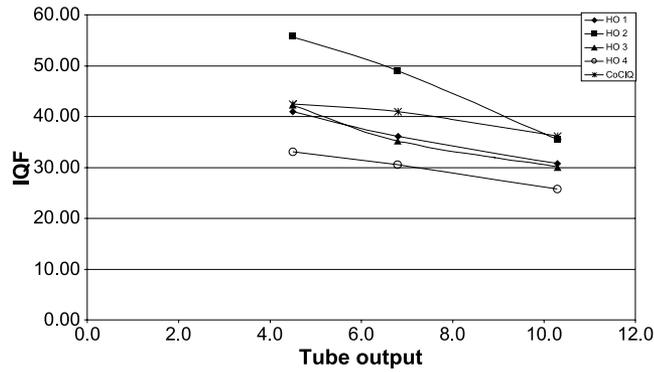


Fig 4. IQF performed by four human observers (HO) and the computer program with the threshold set to 0.2 (CoCIQ 0.2).

same trend, which was a rapid decrease in IQF down to approximately 20 mA s, then the decrease was slower to a certain level after which the IQF was constant. With the threshold set to 0.2, the IQF calculated by the computer was somewhat higher than the IQF achieved by the human observers (HO). The lower thresholds 0.1 and 0.015 resulted in equivalent IQF values for images exposed with medium to high mA s values, while for images exposed with low mA s values the computer-calculated IQFs were lower than the IQF attained by the human observers. Figure 4 shows the IQF for the images exposed with urographic parameters achieved by the human observers. The variations between the four human observers were substantial, showing differences in IQF of almost a factor of 2. The IQF generated by the computer at the threshold 0.2 appeared in the

middle of the IQF performed by the human observers.

Statistical Correlation

Figure 5 shows the IQF calculated by the CoCIQ program at threshold 0.2 compared to the average IQF by the three human observers. There was a strong linear statistical relation ($R^2 = 0.987$), and the regression line $IQF_c = 1.1415 \times IQF_{HO} + 1.1926$ shows that the CoCIQ program scores the images to a slightly lower image quality than the human observers. IQF_c and IQF_{HO} is the IQF obtained by the computer and the human observer, respectively. Corresponding relation when the lowest threshold 0.015 were used resulted in the regression line $IQF_c = 0.3824 \times IQF_{HO} + 11.26$ ($R^2 = 0.933$) (Fig 6).

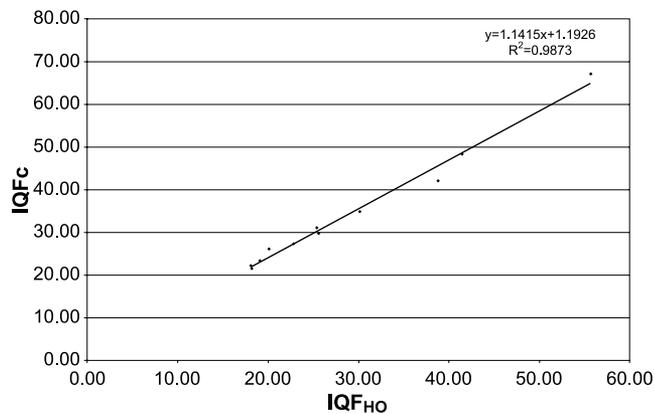


Fig 5. Linear statistical relationship between the IQF calculated by the computer program at threshold 0.2 and the mean value of the IQF performed by the human observers.

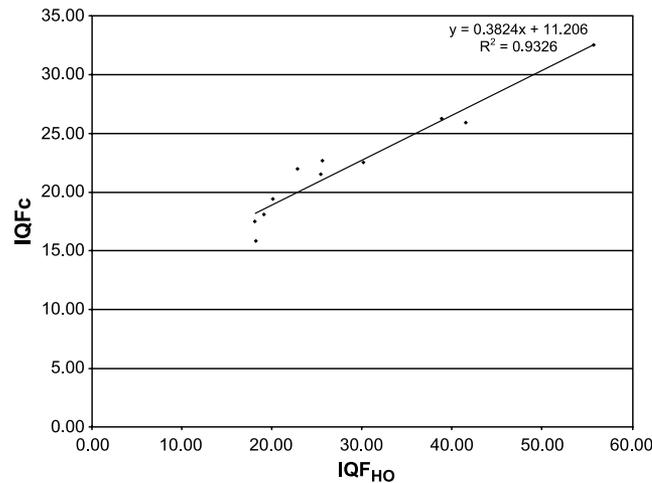


Fig 6. Linear statistical relationship between the IQF calculated by the computer program at threshold 0.015 and the mean value of the IQF performed by the human observers.

DISCUSSION

Figure 3 shows that when threshold 0.2 was used, the shape of the IQF curves generated by the computer and the human observers were similar. This is verified by Figure 5, which shows a strong linear statistical relation between the human observer and the computer when the threshold was set to 0.2. It can be mentioned that low–medium–high mA s value reflects high, medium, and low noise levels, respectively. At threshold 0.1, comparative results from the program and the human observers were equivalent for the images with a low to medium noise level. For the images containing a high level of noise, it was easier for the program to detect the signal compared to the human observers; hence, for those images the program produced a slightly lower IQF (better image quality). It is also seen that the lowest threshold value (0.015) resulted in an IQF rate that was almost independent of noise and did not reflect the human observer's results. The weaker statistical relation between the computer and the human observer at the lowest threshold is shown in Figure 6. It also shows that the CoCIQ program evaluates the images to a better image quality than the human observers. If the intention is to achieve similar results in image quality with human observers, the threshold level should be set to 0.2. But if the intention is for the program to detect as much as possible, the lowest threshold 0.015 should be used. This will result in

lower IQF rates (better image qualities) than those assigned by human observers for images containing a high level of noise, because the program's ability to detect low-contrast objects is less affected by noise.

The CoCIQ program makes it possible to conduct image quality evaluations that are comparable to evaluations performed by human observers, with the advantage of eliminating the human variations in the results. Thus it offers the opportunity of comparing image quality between different departments and different hospitals while disregarding inter-reader variations. It is also a faster way of getting a measure of image quality, hence a substantial amount of time can be saved.

The program is now used in the quality research program of the departments where an image of the CDRAD phantom is taken and analyzed and stored in the PACS. In a recently performed study,¹² a phantom-based evaluation was made to investigate the relationship between tube potential and image quality at a constant effective dose. The aim was to reach the optimal image quality for a given effective dose. Images of a humanoid Alderson phantom were evaluated by eight radiologists in a Visual Grading Analysis (VGA). Images of the CDRAD phantom was also used and evaluated by the computer program. A very good consistency between human observers with an Alderson phantom and the computer reading with a contrast detail phantom was seen.

CONCLUSION

The computer evaluation of the CDRAD phantom makes it possible to conduct image quality evaluations objectively. Moreover, it can be calibrated to produce results comparable to those achieved by human observers.

REFERENCES

1. Tingberg A: Quantifying the quality of medical x-ray images. In: Department of Radiation Physics. Malmö: Lund University, 2000, p 54
2. Metz CE, Wagner RF, Doi K, Brown DG, Nishikawa RM, Myers KJ: Toward consensus on quantitative assessment of medical imaging systems. *Med Phys* 22(7):1057–1061, 1995
3. Kheddache S, Mansson LG, Angelhed JE, Denbratt L, Gottfridsson B, Schlossman D: Effects of optimization and image processing in digital chest radiography: an ROC study with an anthropomorphic phantom. *Eur J Radiol* 13(2):143–150, 1991
4. Sund P, Bath M, Kheddache S, Mansson LG: Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography. *Eur Radiol* 14(1):48–58, 2004
5. Thijssen MAO, THOM, Merx JL, Lindeijer JM, Bijkerk KR: A definition of image quality: the image quality figure. *BIR Rep* 20:29–34, 1989
6. Geijer H, Beckman KW, Andersson T, Persliden J: Image quality vs. radiation dose for a flat-panel amorphous silicon detector: a phantom study. *Eur Radiol* 11(9):1704–1709, 2001
7. Rong XJ, Shaw CC, Liu X, Lemacks MR, Thompson SK: Comparison of an amorphous silicon/cesium iodide flat-panel digital chest radiography system with screen/film and computed radiography systems—a contrast-detail phantom study. *Med Phys* 28:2328–2335, 2001
8. Castellano Smith AD, Castellano Smith IA, Dance DR: Objective assessment of phantom image quality in mammography: a feasibility study. *Br J Radiol* 71(841):48–58, 1998
9. Jansen JT, Zoetelief J: Computer aided assessment of image quality for mammography using a contrast detail phantom. *Radiat Prot Dosim* 90:181–184, 2000
10. Norrman E, Gårdestig M, Geijer H, Persliden J: A clinical evaluation of the image quality computer program, CCIF (abstract). In: 21st Meeting of the Society for Computer Applications in Radiology. Vancouver Convention and Exhibition Centre, Vancouver, Canada, 2004
11. Thijssen MAO, Bijerk KR, van der Burght RJM: Manual CDRAD phantom type 2.0. Project Quality Assurance in Radiology. Section Clinical Physics, Dept. of Radiology. St. Radboud, The Netherlands: University Hospital Nijmegen, 1998
12. Geijer H, Persliden J: Effects of varying the tube potential with constant effective dose at lumbar spine radiography using a flat-panel digital detector—lower kV is better (abstract). In: Second Malmö Conference on Medical X-ray Imaging. Optimisation Strategies in Medical X-ray Imaging. Malmö, Sweden, April 2004, pp 23–25