Evaluation of a histogram-based image quality measure for X-ray computed tomography

Michael Reiter¹, Daniel Weiß², Christian Gusenbauer¹, Marco Erler², Christoph Kuhn², Stefan Kasperl³, Johann Kastner¹

¹University of Applied Sciences Upper Austria - Wels Campus, Stelzhamerstrasse 23, 4600 Wels, Austria, e-mail: michael.reiter@fh-wels.at, christian.gusenbauer@fh-wels.at, johann.kastner@fh-wels.at
²Carl Zeiss IMT GmbH, Carl-Zeiss-Straße 22, 73446 Oberkochen, Germany, email: christoph.kuhn@zeiss.com, marco.erler@zeiss.com, daniel.weiss@zeiss.com
³Fraunhofer IIS - EZRT, Dr.-Mack-Str. 81, 90762 Fürth, Germany, e-mail: stefan.kasperl@iis.fraunhofer.de

Abstract
This paper presents the evaluation of an image quality measure for X-ray computed tomography (XCT) scans. This quality measure can aid XCT users in selecting scans with optimised quality, which finally objectify their selection. The proposed measure is calculated on the base of a grey value histogram which leads to a global quality criterion that is able to rate scans of single- as well as multi-material objects.

Several micro-XCT scans of different specimens with varying image quality are used to verify the proposed image quality measure.

Keywords: X-ray computed tomography, image quality measure, histogram,

1 Introduction

X-ray computed tomography (XCT) scans are nowadays optimised by the system operator, which leads to subjective and often non-ideal scan results. If preliminary information of the specimen geometry and material is available, XCT simulation tools can optimise the acquisition parameters automatically [1-3]. Otherwise, the X-ray spectrum can be selected based on the minimal X-ray transmission occurring during the scan as proposed by standards for XCT [4-6].

A common way to perform and validate such optimizations is to calculate either signal-to-noise ratio (SNR) or contrast-to-noise ratio (CNR) values in reconstructed XCT images for user defined regions. The specification of these regions can be very tedious and may need expert knowledge to avoid misleading results due to the local appearance of artefacts.

Contrary to that, this work presents a quality measure for reconstructed XCT images that does not necessarily need an experienced user that defines regions for the optimisation.

An objective of the presented work is to develop a quality measure that gives a global impression of the image quality by considering quality reducing effects like noise and all kinds of artefacts.

In Chapter 2 we propose a quality measure for the given scope, which will be evaluated by a series of scans in Chapter 3. Final conclusions are listed in Chapter 4.

2 Discussion of the image quality measure $Q$

The proposed image quality measure is calculated on the base of a grey value histogram and it is a measure for the degree of separation of two material classes in the analysed image. Since histograms contain cumulated information of the image this measure gives a global impression of the image
quality by considering statistically relevant effects. Such an effect is relevant, if it significantly influences the histogram shape.

Figure 1 shows a slice and a histogram of an exemplary XCT image, which contains one class for material and one for air. The distance between the two distributions can be interpreted as contrast between the two classes. Furthermore, the width of the distributions is a combined measure for the strength of artefacts, noise and image blur. The most significant artefact types of industrial XCT are beam hardening, scattering, ring and partial volume artefacts. Noise is mainly caused by photon noise and image blur mainly originates from geometric or motion blurring effects occurring during the XCT acquisition.

Figure 1: Exemplary slice and histogram of an XCT scan.

The combination of contrast and distribution widths leads to the proposed definition of the quality measure $Q$ shown in Equation 1. $Q$ seems to be similar to typical CNR definitions [4,5,10], but it is calculated on the base of histograms and takes into account both distribution widths by their standard deviation $\sigma$. Furthermore the distribution widths in the denominator represent all kinds of width defining effects mentioned above and not only noise.

$$Q = \frac{|\mu_2 - \mu_1|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$  

The calculation of the mean values $\mu_1$ and $\mu_2$ as well as the estimated standard deviations $\sigma_1$ and $\sigma_2$ for the two selected class distributions requires a classification of the histogram. Note that classes can be limited by histogram thresholds and histogram borders since the number of classes $M$ can be greater than two.

If the image contains several material classes ($M>$2) two relevant classes need to be selected either automatically or manually by a user. In this work we focus on an automatic selection based on the occurrence probability $p$ of the classes. The class with highest probability to occur in the scanned volume will be compared to the class corresponding to air. Certainly other class selections would be possible.

Unfortunately, in most of the cases more than $M$ normal distributions would be necessary to approximate the shape of XCT image histograms. Hence, threshold methods based on fitting normal distributions are not feasible [7]. Common alternatives are the histogram minimum, “iso-50%” [8] and Otsu’s threshold [9]. Figure 2 shows a comparison of the mentioned histogram thresholds applied on a mixture of normal distributions which are typically for industrial XCT scans. The discrete probability
functions \( p \) (blue graphs) are the sum of two given class probabilities \( p_1 \) and \( p_2 \). According to Equation 2 the sum of both probabilities is 1. The summed probability \( e(t) \) of erroneously classifying a voxel as class 1 instead of class 2 and vice versa depends on the used threshold index \( t \) and is given by Equation 3. In Figure 2 the red graphs show the trend of \( e(t) \) for every histogram. Furthermore, the global minimum of \( e(t) \) is marked with a red dot and is referred to as the minimal error threshold \( e_{\text{min}} \) [7].

\[
\sum_{i=0}^{\text{max}} \{ p_1(i) + p_2(i) \} = 1 \tag{2}
\]

\[
e(t) = \sum_{i=0}^{t} p_2(i) + \sum_{i=t}^{\text{max}} p_1(i) \tag{3}
\]

For equal class distributions there is no difference between the thresholding methods (Figure 2a), but if the class probabilities are unequal, in terms of probability and standard deviation, especially Otsu’s threshold is deviating from \( e_{\text{min}} \) (Figure 2b and 2c). For clearly separated distributions (Figure 2d) the absolute values of Otsu’s and the “iso-50%” threshold show a strong deviation from \( e_{\text{min}} \), but the absolute values of \( e(t) \) are very low.

Figure 2: Histogram thresholds for typical bimodal histogram distributions of XCT scans. The number of used histogram bins is 128.
Since real XCT scan histograms contain more complex shapes than a normal distribution, the minimal error threshold is not calculable. Nevertheless it seems beneficial to use a histogram minimum threshold to minimize the classification error, since it is closest to the minimal error threshold for a wide variety of bimodal normal distributed histograms.

### 2.1 Calculation of the quality measure $Q$

All necessary steps to calculate the proposed quality measure $Q$ are listed in the flowchart of Figure 3.

![Figure 3: Calculation pipeline of the image quality measure $Q$.](image)

The first step is to generate a grey value histogram $h$ of the reconstructed XCT image which should be analysed regarding quality. Without making assumptions about the shape of distributions in the histogram a common rule to calculate the number of histogram bins is shown in Equation 4, where $N$ is the number of image voxels and $C$ is a constant equal 1. Usually $C$ needs adjustments to optimize subsequent histogram analysis. The resulting bin width is given by Equation 5, where the minimal grey value $g_{v_{\text{min}}}$ and the maximal grey value $g_{v_{\text{max}}}$ of the image define the histogram range. The bin width is a trade-off between histogram resolution and noise which may interfere with subsequent analysis. For the analysis of images with sizes from $512^3$ to $2048^3$ voxels we suggest to use Equation 4 with $C$ ranging from 1/4 to 1/8. This reduces the number of local extreme values due to statistical noise in the histogram for typical industrial XCT scans.

\[
\text{bins} = C \cdot \sqrt{N} \quad (4) \quad \text{bin width} = \frac{g_{v_{\text{max}}}-g_{v_{\text{min}}}}{\text{bins}} \quad (5)
\]

All kinds of preliminary data type mappings to reduce data size are only allowed without truncations of the value range to maintain the comparability of different histograms and extracted quality measures.
Furthermore, image regions outside reconstructable regions of a scan, which are filled with values corresponding to an attenuation coefficient of zero, have to be excluded from the statistical analysis. The second step is to find local maxima in the histogram that correspond to a specific material in the scanned image. A local maximum is found when the first derivative of the histogram $h'$ is zero and the second derivative $h''$ is smaller than zero. To suppress local maxima caused by noise, we suggest to calculate $h'$ and $h''$ by convolution of $h$ with a derivated discrete Gaussian function $g'$. The standard deviation $\sigma_{\text{derivate}}$ of the Gaussian function $g$ should be chosen in relation to minimal possible distribution widths in the histogram. For industrial XCT images we suggest $K_{\text{derivate}}$ values between 64 and 512, which define the smoothing via Equation 6 in relation to the number of histogram bins.

$$\sigma_{\text{derivate}} = \frac{\text{bins}}{K_{\text{derivate}}} \quad (6)$$

Although the histogram has been smoothed by previous steps there still might be more local maxima in the histogram than real material classes, because local maxima can be induced by artefacts. The number of material classes $M$ (including air), which is assumed as an user input, can be used to preselect $M-1$ local maxima with the highest relevance. A peak seems relevant if its occurrence in the image is high. So we propose to select the $M-1$ highest peaks, beside the peak corresponding to air. The local maximum caused by air can be detected in the histogram by finding the maximum nearest to an attenuation coefficient of zero. Furthermore, the detection of local maxima beyond attenuation coefficients of zero should be avoided, since local maxima with negative attenuation coefficients typically originate from artefact mechanisms like partial volume effects or beam hardening near highly absorbing materials. Nevertheless, such effects will be considered by $Q$ as an increased width of the distribution corresponding to air.

The fourth step is to classify the histogram $h$ by thresholds $t$. Due to preliminary work presented in Chapter 2 we suggest to use histogram minima as thresholds between local maxima. Typically, for a robust detection of local minima smoothing is mandatory. Gaussian filtering with a standard deviation $\sigma_{\text{minima}}$ described by Equation 7 suppresses local minima that do not correspond to the overall minimal value between two local maxima. The constant $K_{\text{minima}}$ should be typically in the range from 8 to 64 with $\text{bins}_{\text{dist}}$ being the distance in histogram bins from maximum to maximum.

$$\sigma_{\text{minima}} = \frac{\text{bins}_{\text{dist}}}{K_{\text{minima}}} \quad (7)$$

After the classification the mean values $\mu$, standard deviations $\sigma$ and probability values $p$ are calculated for each class $m$ using Equations 8-10. The histogram indices $t_{m,l}$ and $t_{m,r}$ limit a class to the left and right, whereat a limit can be either a histogram border or another threshold if $M>2$.

$$\mu_m = \frac{\sum_{i=t_{m,l}}^{t_{m,r}} h(i) \cdot g(v(i))}{\sum_{i=t_{m,l}}^{t_{m,r}} h(i)} \quad (8)$$

$$\sigma_m = \frac{\sum_{i=t_{m,l}}^{t_{m,r}} h(i) \cdot (g(v(i)) - \mu_m)^2}{\sum_{i=t_{m,l}}^{t_{m,r}} h(i)} \quad (9)$$

$$p_m = \frac{1}{N} \sum_{i=t_{m,l}}^{t_{m,r}} h(i) \quad (10)$$

The final step is the calculation of $Q$ with Equation 1. As discussed in Chapter 2 the quality measure in this work is automatically calculated between air and the material with the highest occurrence probability $p_{\text{max}}$ calculated via Equation 10.
3 Evaluation of the image quality measure $Q$

3.1 Description of used XCT images

Scans for the evaluation of $Q$ are done on a RayScan 250E cone beam XCT device which consists of a Viscom 225 kV micro-focus tube XT9225-D with a tungsten reflection target and a Perkin Elmer flat panel detector XRD 1620 AN14 (2048x2048 pixels, pixel size 200 µm). The scans of three different specimens listed in Table 1 have been originally chosen to allow conclusions for transmission based optimisation of XCT scan [6]. In this work these scans are used with special interest in the various image qualities to evaluate the proposed image quality measure. The selected specimens fit into the typical part spectrum of micro XCT devices:

- Part A is an injection moulded part made of rubber with a wall thickness of about 2 mm.
- Part B is an aluminium step cylinder consisting of 5 steps, each 10 mm in height. The outer diameter is increasing from 15 mm to 55 mm in steps of 10 mm. The central drill hole has a diameter of 8 mm.
- Part C is an injection-moulded multi-material component consisting of plastic, rubber seals and metallic pins.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Part A</td>
<td>82</td>
<td>40</td>
<td>0</td>
<td>1000</td>
<td>3900</td>
<td>1</td>
<td>94</td>
<td>0.142</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>0</td>
<td>500</td>
<td>1900</td>
<td>1</td>
<td>46</td>
<td>0.241</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>0</td>
<td>333</td>
<td>1500</td>
<td>1</td>
<td>36</td>
<td>0.291</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160</td>
<td>0</td>
<td>250</td>
<td>1350</td>
<td>1</td>
<td>33</td>
<td>0.317</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0</td>
<td>200</td>
<td>1260</td>
<td>1</td>
<td>31</td>
<td>0.337</td>
<td>-</td>
</tr>
<tr>
<td>Part B</td>
<td>45</td>
<td>80</td>
<td>0</td>
<td>500</td>
<td>1900</td>
<td>1</td>
<td>46</td>
<td>0.016</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>0</td>
<td>333</td>
<td>1500</td>
<td>1</td>
<td>36</td>
<td>0.031</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160</td>
<td>0</td>
<td>250</td>
<td>1350</td>
<td>1</td>
<td>33</td>
<td>0.042</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0</td>
<td>200</td>
<td>1260</td>
<td>1</td>
<td>31</td>
<td>0.050</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.5</td>
<td>200</td>
<td>3300</td>
<td>1</td>
<td>80</td>
<td>0.081</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.0</td>
<td>200</td>
<td>3200</td>
<td>2</td>
<td>154</td>
<td>0.096</td>
<td>-</td>
</tr>
<tr>
<td>Part C</td>
<td>45</td>
<td>80</td>
<td>0</td>
<td>500</td>
<td>1900</td>
<td>1</td>
<td>46</td>
<td>0.030</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td>0</td>
<td>500</td>
<td>1900</td>
<td>1</td>
<td>46</td>
<td>0.030</td>
<td>MAR</td>
</tr>
</tbody>
</table>

Table 1 lists scans parameter that cause varying image qualities for part A, B and C. Further parameters: 1440 projections, P=40 W, detector gain 0.5 pF, FDK reconstruction with Shepp-Logan filtering, volume data type float32, distance source-detector=1530 mm.
3.2 Part A: rubber gaiter

Figure 4 shows slice cut-outs of part A acquired with different image qualities, whereat Figure 4a corresponds to a scan done with an acceleration voltage of 40 kV and Figure 4b has been acquired with 200 kV. Part A consists of two material classes: air and rubber. The plot of the image quality $Q$ over the voltage confirms the visual impression of a slight degradation of the image quality towards higher transmission values caused by the reduced absorption of X-rays with higher voltages. $Q$ correlates with CNR values calculated in user defined regions including almost the complete rubber part and air shown in [6], with the advantage that $Q$ does not need region definitions, leading to a more objective measure.

![Figure 4: Evaluation of $Q$ for part A. (a) Slice cut-out at 40 kV, (b) Slice cut-out at 200 kV, (c) Q plot for different image qualities achieved for test part A with 40 to 200 kV (Table 1).](image)

3.3 Part B: aluminium step cylinder

The image quality of the aluminium step cylinder scans (Table 1) is mainly limited by inhomogeneous grey value distributions caused by beam hardening artefacts. $Q$ in Figure 5c depicts that an increasing minimal transmission, equivalent to a higher mean effective spectrum energy, reduces these artefacts. Note that the quality increase from data point 4 to 5 is caused by the use of copper prefilter plates. Furthermore, Figure 5a and 5b show a slice comparison of the worst and the best scan by means of $Q$, and Figure 5d illustrates the histograms of these two scans. Although 200 kV with 1.0 mm copper prefilter offers a decreased contrast between aluminium and air compared to 80 kV, the disproportionally high width reduction of the class distributions leads to an overall gain in image quality, which is detected by $Q$.

3.4 Part C: multi-material component

The connector is made of plastics equipped with rubber seals and metallic contacts, which would require three histogram thresholds to separate all materials beside air. Unfortunately, the rubber seal is statistically not relevant and no distinct distribution will be visible in the histogram of the complete XCT scan. Generally, it is difficult to predict which material will cause a distinct distribution in the histogram. That is why we suggest $Q$ for multi-material components only with cross-checking of the user or additional algorithms that detect issues of histogram-based classification.

Preliminary investigations on part C showed that the CNR between plastics and rubber is best for voltages in the range of 80 kV without any prefiltering on a typical industrial micro XCT device [6]. Unfortunately, low voltages enforce metal artefacts that are mainly caused by beam hardening. Figure 6a shows typical streaking artefacts that occur near metallic parts. Since the amount of disturbing metal in the scanned volume is relatively low, the application of a metal artefact reduction (MAR) algorithm [11,12] seems promising. The gain in image quality by applying an MAR is
documented by Figure 6b. Further analysis shows the global effect of an MAR on the complete grey value histogram. Distributions for air, plastics and metal become narrower (Figure 7). Especially, below the used MAR threshold a drastic improvement for the separation of plastics and metal is achieved. These image enhancements can be quantified by an improvement of $Q$ from 2.7 to 5.9.

Figure 5: Side by side comparision of reconstructed slices of part B acquired with (a) 80 kV without prefilter and (b) 200 kV with 1.0 mm Cu prefilter. (c) Plot of Q over the minimal transmission for part B shows. (d) Histograms with 8192 bins of scans (a) and (b).

Figure 6: Reconstructed slice of part C a multi-material component acquired with (a) 80 kV without prefilter and (b) with an additionally metal artefact reduction. The slice is showing classes of air, plastics and metal.
Figure 7: Histograms of part C scans at 80 kV without MAR (Q=2.8) and with MAR (Q=5.9), 8192 histogram bins, classification of three material classes. (a) Logarithmic count axis. (b) Detailed view starting at grey values of 0.05 and a linear count axis.

4 Conclusions

We proposed a histogram-based measure for the quality analysis of XCT images. The measure $Q$ is able to capture statistically relevant scanning mechanisms (artefacts, blur and noise) that determine the image quality of industrial XCT. Typically, the scan quality is influenced by beam hardening and scatter artefacts, noise and image blur.

$Q$ has been applied on scan series of two single-material components and is able to automatically quantify the image quality of every scan with less user interactions compared to region-based CNR calculations.

If $Q$ is calculated for scans of multi-material components we suggest using $Q$ only with cross-checking of a user or additional algorithms that detect classification issues.

Beside the sensitivity of $Q$ to XCT scan parameters, we demonstrated the possibility of quantifying image improvements achieved by correction algorithms like the metal artefact reduction.

Acknowledgments

This work was supported by the “K-Project for non-destructive testing and tomography” grant number 820492. Furthermore we want to thank G. Schmee (Delphi Connection Systems Austria GmbH) for providing real world test parts and the Siemens AG for providing the metal artefact reduction module of CERA.

References


