ISBN 978-82-326-3378-4 (printed version)
ISBN 978-82-326-3379-1 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:294

David Völgyes

Image quality in forensic CT imaging

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2018:294

NTNU

NTNU
Norwegian University of
Science and Technology

David Völgyes

Image quality in forensic CT imaging

Thesis for the degree of Philosophiae Doctor

Gjøvik, October 2018

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-3378-4 (printed version)
ISBN 978-82-326-3379-1 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:294

Printed by Skipnes Kommunikasjon as

Contents

Declaration of Authorship

I, David Völgyes, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed: 

(David Völgyes)

Date: 6th August 2018

Acknowledgements

This thesis couldn't be created without tremendous help I have got. I would like to thank for the supervision from my supervisors: Anne Catrine Trægde Martinsen, Arne Stray-Pedersen, Dag Waaler and Marius Pedersen. Their collaboration not only allowed me to deepen in medical imaging but gave a unique opportunity to work with cutting edge devices and otherwise rarely accessible data.

I express my gratitude to my wife, who has been patient with me when I needed it, and her support and critical feedbacks were invaluable.

Finally, I would like to thank all of my fellow Ph.D. students who inspired and supported me.

Abstract

What factors affect the diagnostic image quality in forensic computed tomography (CT) imaging? In this thesis I briefly present the main steps from CT scanning to diagnostics.

The thesis is built on four core papers. The first paper focuses on reconstructions of a quality control phantom measurement series. Objective image quality descriptors (contrast to noise ratio (CNR), mean CT numbers) are used for comparisons. Besides the reported objective measurements, some unexpected properties are identified, for instance, decreasing CNR when low level of iterative reconstruction is used.

The second paper evaluates the applicability of an in vivo CT protocol in post mortem imaging. Radiographers and radiographer students are asked to score nine reconstruction alternatives from three post mortem cardiac cases. The results indicate that in this case the in vivo protocol is applicable in post mortem conditions.

Images not only have to be measured, but also have to be presented. The third paper proposes a tone mapping approach which can compress the dynamic range of the CT image while preserves local contrast. Traditionally, a CT volume is read several times using different intensity windows. This contribution might reduce the number of required readings of the same volume, and could improve visualization of multiple pathologies in a single image.

The fourth paper tackles a classic image processing problem, the de-quantization problem. The appearance of false contours might be both disturbing and misleading. If false contours are already present in an image, e.g. due to aggressive denoising, this approach might reduce or eliminate them.

These four contributions are the core of this thesis, augmented with non peer reviewed contributions.

Abbreviations and notations

α, s	sinogram-space axes
$\delta(i,j)$	Dirac-function
DFT	Discrete Fourier Transform
$\mathcal{F}$, $\mathcal{F}^{-1}$	Fourier Transform and its inverse
$\hat{f}$	Fourier transform of f
HU	Hounsfield unit
μ	linear attenuation coefficient [$mm^{-}1$]
∇, Δ, Δ^n	gradient operator, Laplace operator, polyharmonic operator
ν	frequency
$\overline{f}$, $\langle$expression$\rangle$	mean value of f, averaging an expression
$\mathcal{R}$, $\mathcal{R}^{-1}$	Radon Transform and its inverse
$\mathrm{TV}(f)$	Total variation functional of f
$\vec{x}$, $\vec{y}$, $\underline{\underline{A}}$	linear vector and matrix notation
$f \star g$	convolution of f and g
f, g, u	functions
$H(\ldots)$	histogram
$I(\ldots)$	2D or 3D discretized image
N_x, N_y, N_z, N	number of pixels/voxels
v_x, v_y, v_z	pixel/voxel sizes in [mm]
x, y, z	spatial axes
AUC	area under the curve
CSF	Contrast sensitivity function
CT	computed tomography
DEI	Dual energy index
DET	detection error trade-off
DSDE CT	Dual source dual energy CT
FBP	filtered back projection

FOV	field of view
IR	iterative reconstruction
MTF	modulation transfer function
NPS	noise power spectrum
PSF	point spread function
ROC	receiver operating characteristic
TMO	tone mapping operator
USM	unsharp masking

Chapter 1

Introduction

Medical imaging is an multi-disciplinary field which builds not only on medicine but also on physics, computer science and color science, among others. This thesis is written as part of a computer science Ph.D. programme, but interested readers from other fields may find it useful too.

My personal motivation originates from the fact that in previous tomography projects I took part in detector design and image reconstruction algorithm developments but our work ended when the scanner was operational. What happens afterwards? What are the contributing factors to diagnostic image quality beyond the purely technical parameters?

Diagnostic image quality is result of a long, strongly interconnected chain of building blocks, see Fig. 1.1, but other categorizations also exist [1]. The ultimate goal is to improve the accuracy of the diagnostics while reducing the risks and expenses. This is a task-specific, multivariate optimization problem which has no unique, optimal solution, however, it has many good approximate solutions. One of the challenges of the medical imaging field is its rapid development which renders this optimization problem into a moving target where the achievable best compromise changes continuously.

The focus of this Ph.D. was on the perceived diagnostic image quality which aims to bridge the purely technical side of tomography with daily diagnostics. I had the unique opportunity to take part in a collaboration between Norwegian University of Science and Technology (NTNU), Oslo University Hospital Department of Diagnostic Physics and Oslo University Hospital, Department of Forensic Sciences aiming to improve post mortem CT imaging performance using a dual source dual energy CT scanner.

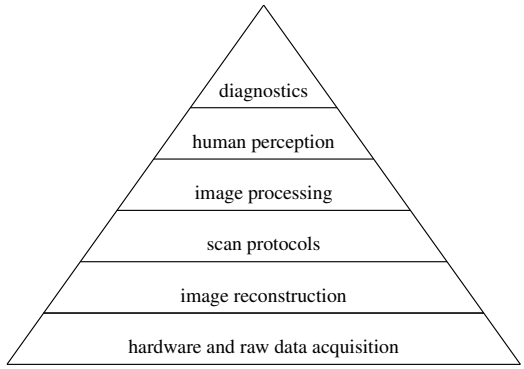


Figure 1.1: Diagnostic image quality depends on several interconnected building blocks.

The scanners are treated gray-boxes where some parameters are known or measurable, for instance, dose and image quality, but there are also unknown parameters such as the technical details of reconstruction algorithms. The other end of the chain is the radiologists, who use the images for diagnostics. But experts' performance depends on the quality of the presented information, hence the need for optimization.

During my work I studied the contributing factors of diagnostic image quality, and I tried to contribute to the relevant fields. Four peer reviewed papers give the core of this theses. In the following chapters the context and relevant background will be introduced briefly, then contributions will be presented and discussed in context. The thesis ends with outlook to current open issues and what can be anticipated in the near future.

List of the core contributing papers:

[1] D. Völgyes et al. "How Different Iterative and Filtered Back Projection Kernels Affect Computed Tomography Numbers and Low Contrast Detectability". In: *Journal of Computer Assisted Tomography* 00.00 (2016), p. 1. ISSN: 0363-8715. DOI: 10.1097/RCT.0000000000000491.

[2] D. Völgyes et al. "Applicability of a clinical cardiac CT protocol in post mortem studies". In: *Journal of Forensic Radiology and Imaging* 12 (2018), pp. 25–30. ISSN: 2212-4780. DOI: 10.1016/j.jofri.2018.01.003.

[3] D. Völgyes et al. "A Weighted Histogram-Based Tone Mapping Algorithm for CT Images". In: *Algorithms* 11.8 (July 2018), p. 111. ISSN: 1999-4893. DOI: 10.3390/a11080111.

[4] D. Völgyes et al. "Image De-Quantization Using Plate Bending Model". In: *Algorithms* 11.8 (July 2018), p. 110. ISSN: 1999-4893. DOI: 10.3390/a11080110.

Chapter 2

Background

The main focus of this thesis is the contributing factors to diagnostic image quality in a post mortem CT imaging using a dual source dual energy CT scanner. This is a quite inter-disciplinary topic and even interested readers might only have partial background. The thesis is written in a computer science programme, therefore, a little bit more computer science literacy is assumed, but interested radiologists and physicists are also aimed. Hopefully, any reader will find something applicable for his/her field, and maybe get interested in the other fields too.

As it will be presented later, the main contributions are related to CT scan protocols, image processing of the scans, and presentation, visualization of the results to the diagnostician. In the following sections brief history and main concepts of CT imaging is presented, followed by a short overview about some most important aspects of the human visual system. The chapter ends with brief image processing overview.

2.1 Diagnostic image quality

Image quality is a highly subjective term. CT imaging uses ionizing radiation which should be minimized. However, too low dose might lead to diagnostically unacceptable image quality. From two alternatives, the better one achieves the same diagnostic performance with lower dose, or at the same dose level, it yields higher diagnostic performance [2]. The famous ALARA principle states that the dose should be "**a**s **l**ow **a**s **r**easonably **a**chievable" [3]. Instead of image quality, often the term *diagnostically acceptable image quality* [4] is used.

Defining the image quality through the diagnostic performance is intentional. Without well defined diagnostic criteria, image quality becomes a highly sub-

jective term, sometimes referred as *beauty contest* [5]. In order to avoid this, the European Commission released guidelines for computer tomography quality criteria [6].

The diagnostic performance can be measured [7] with accuracy (sensitivity and specificity), predictive values, likelihood ratios, receiver operating characteristic (ROC) curves, among others. An important realization is that there is no universal image quality. Different tasks require different images. The diagnostic image quality is context dependent, belongs to a specific task or protocol, and it is optimized through maximizing the diagnostic performance.

2.1.1 Subjective evaluation

The recommended way for optimizing medical images is to use well defined criteria for a specific task. The diagnostic task depends on the pathology to be detected, and it is important to use the same approach, otherwise results from different studies wouldn't be comparable.

There are wide range of diagnostic tasks, e.g. localization objects, or sharpness or clear visibility of structures. For localization or binary classification tests, specificity and sensitivity, also known as true positive rate and true negative rate, are used to characterize the diagnostic performance, and ROC curves [8] are used to analyze the trade-off between them.

Fulfillment of criteria is often not binary, but defined on a scale. Visual grading analysis (VGA) uses numeric scoring to describe the fulfillment [9]. While this method is quite straightforward, it has challenges. First, if two alternatives have similar scores, it might be hard or impossible to determine which option is better by the criteria[1]. Second, if there is a statistically significant difference in the VGA scores, it does not mean that there will be statistical difference in the diagnostic interpretation. Third, the interpretation of the scoring is always a bit subjective, despite all the efforts to make it standardized, and the readers might score the images differently. This makes the aggregation and analysis of the scores challenging [5][10].

One way to overcome the challenge of different internal scales, is to use pairwise comparisons [9][11]. When the better of two images should be selected (by a given criterion), the internal scales of the readers are less important, as long as they are consistent. The disadvantage of this method is the larger number of required comparisons. Typically, VGA is linear in the number of alternatives, while pairwise comparisons are quadratic[2]. ROC analysis, VGA, pairwise comparisons are only a

[1] E.g. the study might require larger number of readers than available.

[2] There are alternative approaches, e.g. sorting options with pairwise comparisons.

few examples from the large number of approaches to evaluate image quality.

All of the evaluation methods suffer from three issues. First, the intra-observer differences [12]. The readers might remember previous images and decisions. As a result, the order of the alternatives might affect the evaluation, and presenting the same option several times might produce different results [13]. Second, the inter-observer differences [14], e.g. the different internal scales, might lead to significant disagreement between the readers. For instance, a structure might be clearly visible for one reader, but not for another. Third, the number of observers are often limited, and their reading time must be limited in order to keep the reading session short, which limits the available data for numerical analysis.

Some of these issues could be minimized during experiment design, e.g. using randomized order for the images, or requiring certain time to pass between reading sessions.

2.1.2 Objective evaluation

Mathematical metrics could be used for image quality description. The obvious advantage is their objective nature. However, while there are published observer models for specific tasks, e.g. low contrast detectability [15], and the correlation between human and model observers performance could be high [16]; every model has limitations, and there is no universal observer model.

Despite these challenges, objective metrics are important. They can be used as a first step to pre-select a shorter list of candidates which can evaluated using expert readers. Second, objective metrics are important for quality assurance [17]. Third, objective metrics can be used to compare solutions from different vendors and find the closest matches [18].

Another issue arises when several numerical parameters are measured. In a diagnostic task the reader can decide with pairwise comparison which alternative is better for a certain task. The aggregating numerical descriptors into a single score would require an observer model. The lack of this model makes the interpretation complex.

In the next sections the history and technical details of CT imaging will be introduced along with the most important image descriptors.

2.2 Roots of forensic CT imaging

Two dimensional X-ray images play a central role in diagnostic imaging ever since W. C. Röntgen discovered X-rays in 1895 [19]. An X-ray image is a two dimensional image which records the attenuation of X-rays passing through an

object, for instance, a patient. These two dimensional images are often called projections, as they can be seen as a 2D projection of a 3D distribution.

Just two weeks after Röntgen held his speech at the meeting of the Physical-Medical Society, an X-ray was taken to localize bullet in a firearm victim [20] in Montreal, Canada. Later this X-ray was used in the trial as evidence. A few weeks later a musket ball needed to be located in a broken arm [21]. According to the author, the difference between the bullet's location based location X-ray and based on surgery was no more than an eight of an inch. This is also probably the first reported case when three dimensional position was determined from two 2D X-ray projections. Despite its short length, this paper also mentioned a technicality, grounding the cathode, which shows the importance of stable, high quality X-ray beam generation.

Ever science the discovery of X-rays, radiology and forensic radiology continuously improve and keen to implement new technical solutions and inventions. But despite the rapid start, until quite recently radiological imaging had only a supporting role in forensic examinations [22]. It was not part of the protocol, autopsy rooms were not augmented with a nearby CT rooms.

The breakthrough came with the Virtopsy research project in the mid-nineties, which proved the importance of post mortem scans in forensics, and established radiology as part of the daily forensic workflow [23]. Since then forensic radiology went through a revolution, and dedicated forensic scanners are installed all over the world.

The non-invasive nature of forensic CT, the lack of radiation risk in post mortem cases and the available ground truth information from autopsies not only allowed to improve forensic documentation but it can contribute to clinical radiology.

Finally, laymen often find radiological images less disturbing and more understandable then autopsy-findings, which eases communication with next of kin or with laymen during a court trial [24]. The possibility of second opinion from an another expert makes the forensic findings more reliable and more trusted.

2.2.1 Post mortem CT scans

CT examination of cadavers is significantly different from a clinical examination. The most obvious technical challenge is the lack of possibility to use contrast agents[3]. This not only limits soft tissue contrast enhancement, but also prevents functional imaging, and renders many clinical protocols inapplicable in post mortem conditions. Consider for instance, detection of cerebral stroke or heart infarction.

[3] Technically, there are experimental methods, but their applicability is limited, see [25].

The three main post mortem changes altering CT imaging are gravity dependent changes, decomposition and rigor mortis (muscle stiffening) [26]. Clotted blood has slightly higher CT numbers, than non-clotted blood, after some time gas volumes appear due to decomposition, rigor mortis changes the muscles, which for instance leads to dilation of the right atrium and ventricle of the heart. Lungs collapse and gravity changes shape of organs and fluid distribution.

In clinical practice, the presence of foreign objects in the body can be seen as source of artifacts, for example, a dental implant. These objects are only rarely in focus of the examination, e.g. when the actual implant is examined. In forensic practice, the mentioned dental implants, bullet fragments, or any other foreign object might be significant evidence. This might cause two major challenges: First, these objects are usually denser than the human body, and they might cause serious image artifacts, for instance metal artifact. To avoid metal artifact, the kVp might be increased. Using high energy CT scan protocol, which records these dense objects, might degrade the contrast between the tissues. Second, the CT images already have high dynamic range, but presence of foreign objects increases this range even further. This makes visualization even more challenging.

2.3 Details of CT imaging

2.3.1 Basic principles

After the quick glance of applications, it is time to describe the mathematical and technical foundations of CT imaging.

The Radon transform [27], introduced by Johann Radon in 1917, describes a mapping from N dimensional objects to series of N-1 dimensional projections. The main idea of the CT imaging is that these N-1 dimensional projections can be measured as X-ray images, and applying inverse Radon transform the N dimensional distribution could be reconstructed.

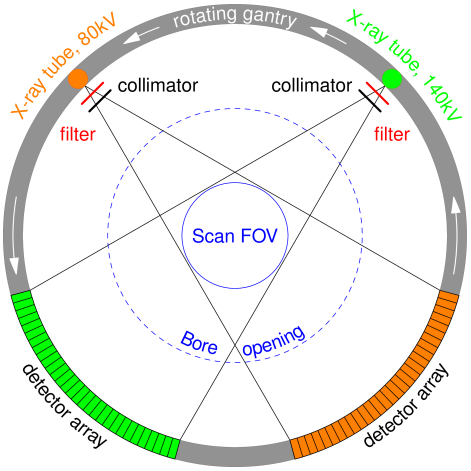
The transform [28] is depicted in Fig. 2.2. The straight lines in the original domain can be parameterized as follows:

$$x(z), y(z) = z\sin(\alpha) + s\cos(\alpha), -z\cos(\alpha) + s\sin(\alpha) \tag{2.1}$$

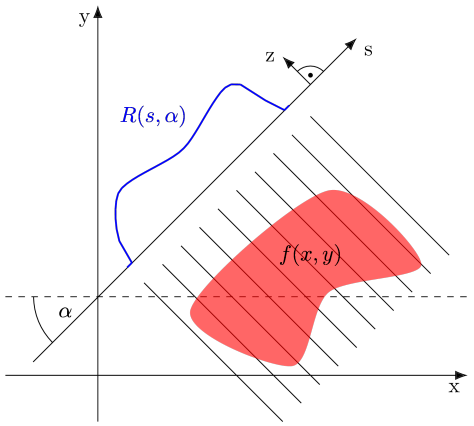
For every such line the transformed function takes the line integral of the original function:

$$R(\alpha, s) = \mathcal{R}\{f(x, y)\} = \int_{-\infty}^{\infty} f(x(z), y(z))\mathrm{d}z \tag{2.2}$$

(a) Photo of a CT scanner



(b) Schematic diagram of a dual source CT scanner



(c) Illustration of the Radon transform

Figure 2.1: CT scanner: a) photo, b) schematic diagram, c) model

For a fixed α value, $R_\alpha(s)$ is the projection of $f(x, y)$ to the axis s. The projection-slice theorem connects Fourier transform with Radon transform. This connection can be easily described if a few new operators are introduced. $F_{x,y}$ denotes the the 2D Fourier transform of the original function, and S_1 denotes the extraction of the central slice parallel to s. From the Radon transform part, let R_α denote the projection of the function, and finally, F_s denotes a Fourier transform along s axis. Using this notation, the projection-slice theorem is as follows:

$$S_1 F_{x,y} = F_s R_\alpha \tag{2.3}$$

Applying the formula for all angles, the whole 2D Fourier signal can be identified using the 1D Fourier transform of the Radon transformed signal. From the 2D Fourier signal the original signal can be easily reconstructed.

In practice, the measured signals are discretized, and calculations are performed using discrete approximations. The discrete approximation of the Fourier transform will be used frequently, and the $\mathrm{DFT}\{\}$ notation will be used for it.

The discretized function which represents the object is called image, and consists of voxels or pixels, depending if it is three or two dimensional. The Radon-transform of an image is called sinogram. This name comes from the fact that an off-center point has a sinusoid form after applying the transform 2.2. The domain of the image is often called image space, while the domain of the transformed image is called as sinogram space.

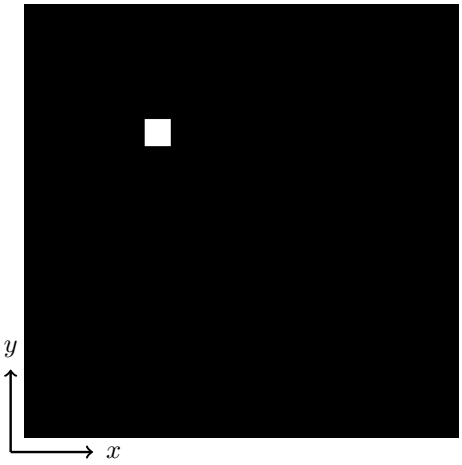
3D objects can be seen as series of transformed 2D objects, or 2D objects can be seen as 1 pixel tall 3D objects. Due to this simple connection, reconstruction theory is explained in 2D but it is trivial to generalize it to 3D.

The numerical inversion requires large amount of calculations, and it only became feasible with the spread of digital computers, and this fact gave the name ("computed") of the field. After Allan McLeod Cormack laid the theoretical foundation of CT imaging in the early 1960s, the first Computed Tomography scanner was built by Godfrey Hounsfield at EMI [30]. The scanner was officially introduced to medical practice on 1st October 1971 in Atkinson Morley Hospital in Wimbledon, London, United Kingdom [31]. For their contributions Cormack and Hounsfield were awarded with the Nobel Prize in Medicine in 1979 [32].

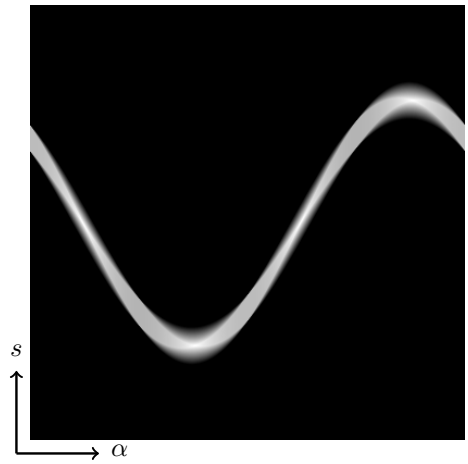
2.3.2 Image reconstruction

The inverse Radon transform is most frequently done with *filtered back projection* (FBP) [33]. It has two parts: filtration and back projection. Often the first step is the filtration. First, 1D Fourier transform applied to the sinogram in every angle, filter applied in the Fourier space, then inverse transform is performed. The function used in the filter is called kernel. It is also possible to first back-project and then applying filtration as second step, this is called *backproject-filter* (BPF) method.

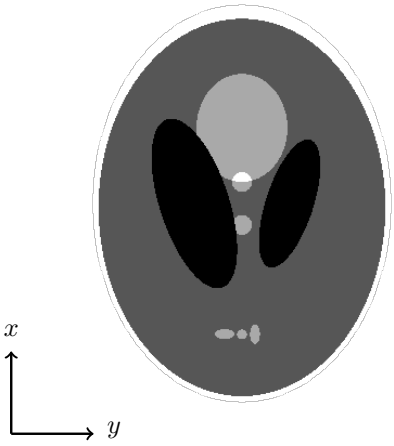
$$\hat{f}(\alpha, s) = \mathcal{F}_{s'}^{-1}\left\{\mathrm{filter} \cdot \mathcal{F}_s\left\{f(\alpha, s)\right\}\right\} \tag{2.4}$$



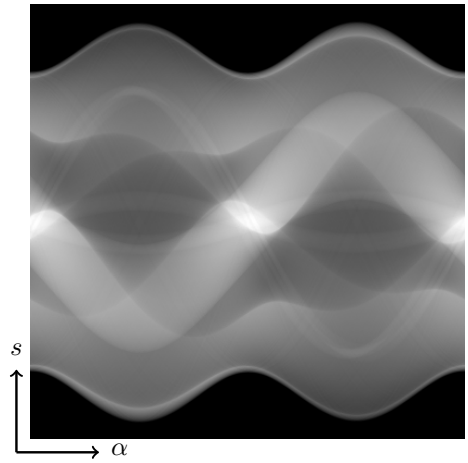
(a) Image of a point source



(b) Sinogram of a point source



(c) Image of a Shepp-Logan phantom[29]



(d) Sinogram of a Shepp-Logan phantom

Figure 2.2: Example images and corresponding sinograms.

The filtered function can be back-projected into the image space:

$$f(x,y) = R^{-1}\{f(\alpha,s)\} = \int_0^{\pi} \hat{f}(\alpha,s)\delta(x\cos\alpha, y\sin\alpha)\mathrm{d}x\mathrm{d}y \qquad (2.5)$$

The main drawback of FBP is the fact that it doesn't take statistical properties of the measurements into account. For instance, some projections have higher noise than others, but the algorithm gives equal weight to all of them. This affects the image quality of the reconstruction seriously.

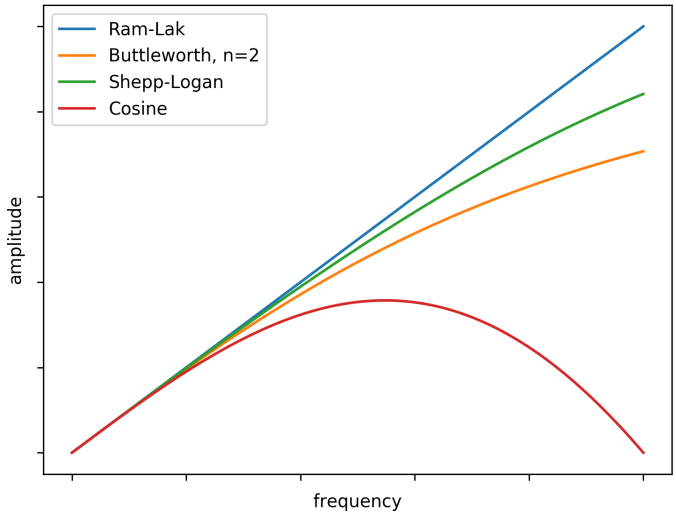


Figure 2.3: Common FBP filters.

One possible solution is to manipulate the filter before the back-projection. Choosing a kernel with a cut-off frequency blurs the reconstructed image but also limits the high frequency noise. The kernel selection affects the the point spread function (PSF), see some examples in Fig. 2.3- 2.4, and also affects the noise power spectrum (NPS). The magnitude of the PSF's Fourier transform is called modulation transfer function (MTF), while the NPS is defined as the average noise power carried by a frequency [34]:

$$MTF = |\mathrm{DFT}\{PSF\}| \tag{2.6}$$

$$NPS(\nu_i) = \frac{v_x v_y}{N_x N_y} \left\langle |\mathrm{DFT}\{f - \bar{f}\}| \right\rangle \tag{2.7}$$

The most simple filter is the Ram-Lak filter which simply cuts the frequency at a given cut-off frequency (k_cut). More advanced filters were developed to mitigate high frequency noise and emphasize medium frequencies [35]. The effect of reconstructions on PSF, MTF and NPS are depicted in Fig. 2.4, 2.6 and 2.7, respectively.

Alternatively, the Radon transform can be seen as a linear transform($\underline{\underline{A}}$) between an

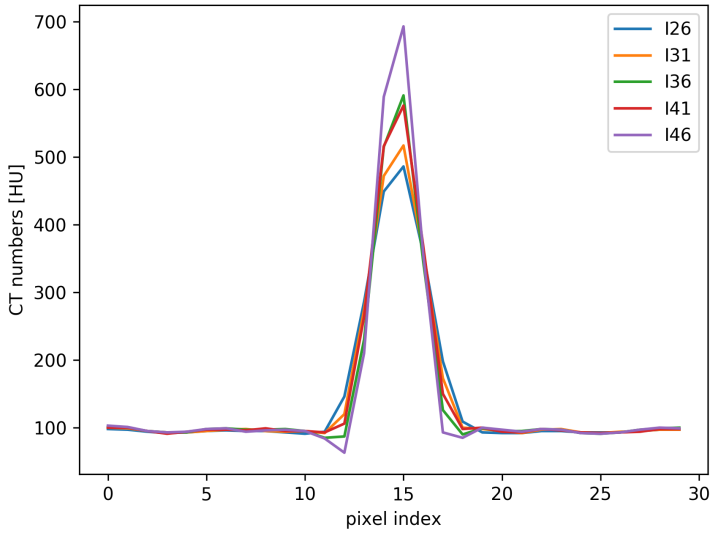


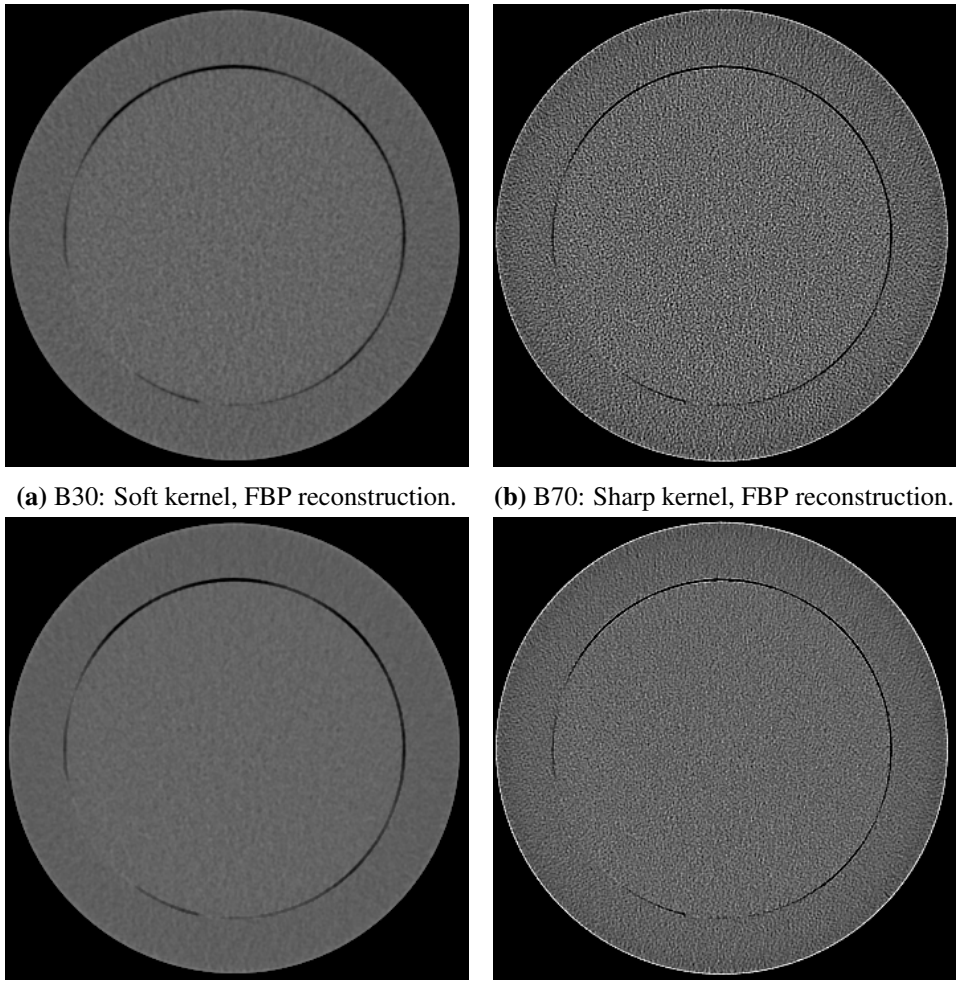
Figure 2.4: Point spread functions of different single energy reconstructions in a Siemens Somatom Definition Flash CT.

image space vector($\vec{x}$) and the projection space vector($\vec{y}$).

$$\vec{y} = \underline{\underline{A}}\vec{x} \tag{2.8}$$

The transform matrix is called *system matrix*, and it is a huge, sparse matrix, calculation of its inverse matrix is infeasible for realistic problem sizes. The solution could be found with direct algebraic methods (algebraic reconstruction technique (ART) [36],simultaneous ART (SART) [37]) or statistical methods (e.g. transmission maximum likelihood (TRML) [38] and its faster variant with ordered subsets (TRML-OS) [39]. These algorithms are called iterative reconstructions (IR) [40][41], because they approximate the solution iteratively.

There are two main advantage of iterative reconstructions. First, they can incorporate statistical models about the measured data [40], and priors about the image. For instance, the measured projections might vary in relative noise, and their contribution to the final image can be weighted taking the noise into account. Image priors are assumptions about the final image which can be incorporated into the model, e.g. using a total variation minimization prior [42] which from all of the equally possible images prefers the one which has the smallest integrated gradient



(a) B30: Soft kernel, FBP reconstruction.

(b) B70: Sharp kernel, FBP reconstruction.

(c) I30: Soft kernel, iterative reconstruction (SAFIRE 3).

(d) I70: Sharp kernel, iterative reconstruction (SAFIRE 3).

Figure 2.5: Noise structure for different reconstructions at 120kVp in a Catphan 600 phantom.

magnitude.

$$\mathrm{TV}(I(x,y)) = \iint |\nabla I(x,y)| \mathrm{d}x\mathrm{d}y \tag{2.9}$$

Second, IR can model physical phenomena which were ignored in classic FBP, e.g. the X-ray spectrum is polyenergetic [43], and the attenuation is a simplification behind absorption and scattering [44]. Ignoring these phenomena could lead to artifacts or inferior image quality. Artifacts are detailed in the next subsection.

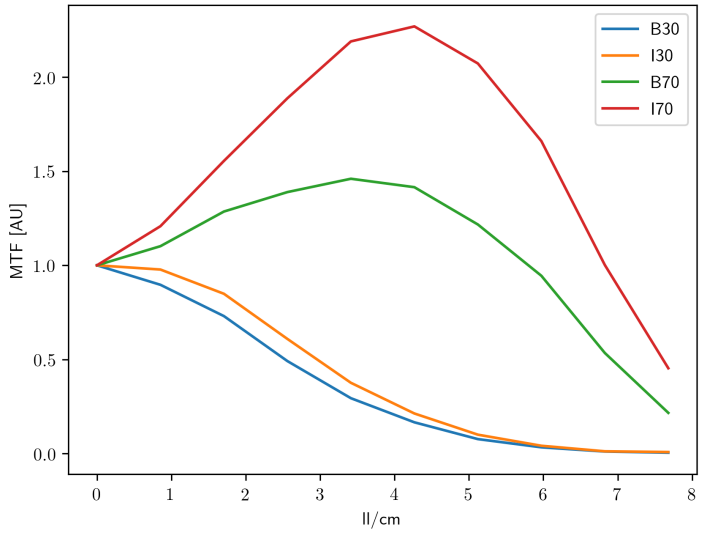


Figure 2.6: MTFs for different reconstructions.

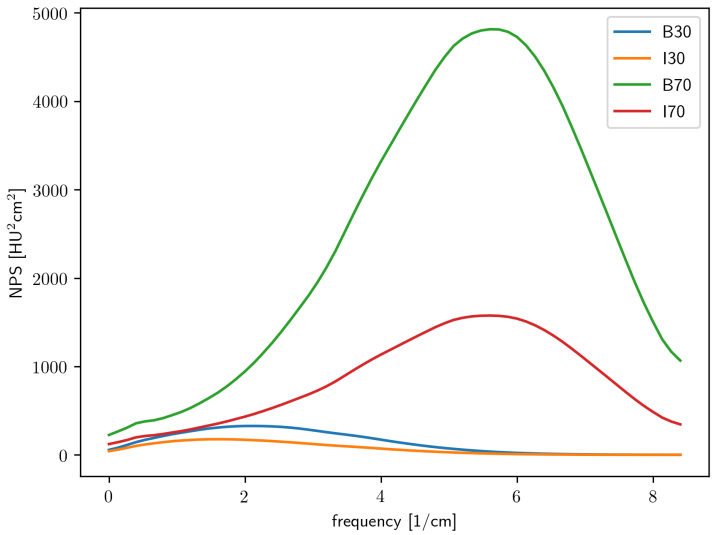


Figure 2.7: NPSs for different reconstructions.

There are also drawbacks of iterative reconstructions. First, they are non-linear which makes image quality analysis challenging, and could lead to artifacts. Second, IR requires much more computational resources than FBP [45].

2.3.3 Image reconstruction and artifacts

Image artifacts can originate from various sources: noise in the measured data, detector imperfections, polyenergetic X-ray spectrum, among others. FBP reconstructions use different filters to reduce the noise and optimize perceived image quality, as it was mentioned in the previous subsection. Statistical algorithms are more suitable for advanced modelling where properties of the X-ray tubes, object models and detector models can be incorporated. Due to the different compromises, hybrid approaches are often used, for instance, when FBP and iterative techniques are blended into a composite image, or when FBP provides the first approximation in an iterative scheme [41].

Regardless of the algorithm and the source of the noise, there is a very general property of the tomographic imaging: a point in sinogram space affects many points in image space, and one image space pixel is reconstructed from many sinogram points. This means that a single pixel noise in sinogram space generates correlated noise in image space. Similarly, there are typical artifact patterns, e.g. horizontal (constant radius) sensitivity homogeneity in sinogram space causes ring artifacts in image space, while inhomogeneities by angles contribute to radial artifacts such as photon starvation or metal artifacts, depending on the source of the artifact. Barrett and Keat have published an excellent overview about the of the CT artifacts [46].

This property makes CT image processing challenging because some effects are easier to handle in the sinogram domain while others are more suitable for handling in the image domain. While transforms between the domains is technically feasible, it is resource-demanding. First, vendors implemented algorithms which minimized the number of back and forward projections, and noise or artifact reduction was mostly done in one of the domains, e.g. Iterative Reconstruction in Image Space (IRIS), Adaptive Iterative Dose Reduction (AIDR) [41].

Recently IR schemes started to address challenges in both spaces, e.g. Sinogram-affirmed iterative reconstruction (SAFIRE), Advanced modeled iterative reconstruction (ADMIRE) and VEO [41], but the development of new algorithms and addressing all of the issues are far from over.

2.3.4 X-ray physics

Until now, the mathematical foundation of tomography has been introduced. Understanding the underlying physical phenomena is at least as important as the

mathematical foundation.

When a target object is bombarded with high-speed electrons, X-rays are produced. Electrons in X-ray tubes are accelerated with high electric fields, usually in the range of 80-140kV, therefore the electrons have 80-140keV peak kinetic energy when they hit the target. This kinetic energy can generate X-rays through deceleration radiation, also known as *bremsstrahlung*, while for other electrons the kinetic energy can be lost in different processes, e.g. in scatter.

As a result, small portion of X-rays have the theoretical peak energy, and large amount has lower energies Some electrons might be kicked out from the electron shells of the target atoms. When these holes in the shell are filled again, characteristic X-ray radiation is emitted [47]. The final X-ray spectrum is a continuous spectrum with a well defined upper limit, and with some characteristic peaks, depending on the target material, see Fig. 2.8. The tube voltage determines the peak, and denotes as kVp referring to the peak, while kVe is used to describe the effective kV. Besides kVp, anode and filtration effect kVe.

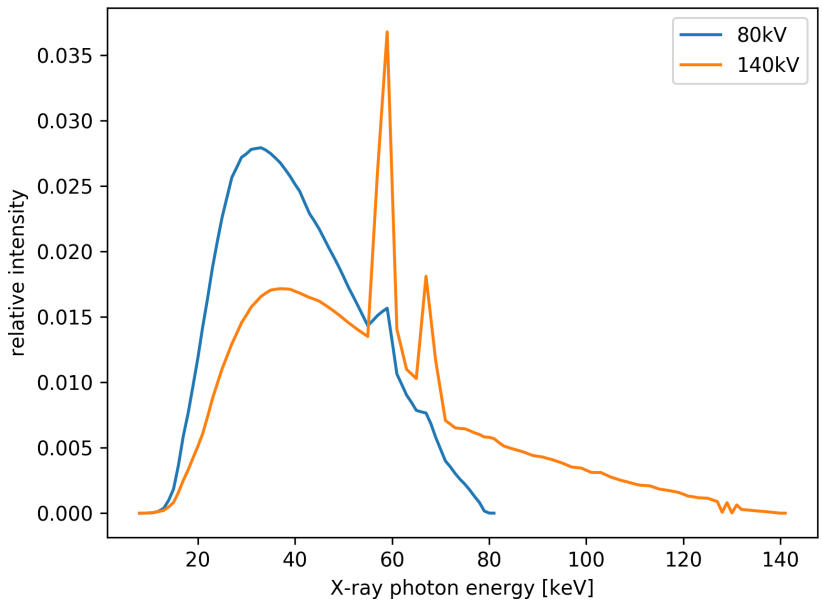


Figure 2.8: X-ray tube spectra with Tungsten anodes, simulated with [48].

CT maps the attenuation of the X-ray beam through a slice of an object. Attenuation consists of several physical phenomena, most importantly photoelectric effect and Compton scatter. The probability of these interactions depends on the energy of the photon and the electron density and the atomic number of the medium. Therefore,

attenuation in the medium changes the spectrum, and it attenuates lower energy photons more, see Fig. 2.9. The X-ray spectra for Fig. 2.8 and 2.9 were simulated with [48].

This energy and material specific attenuation lies behind the so called beam hardening effect [49], but it also enables spectral CT imaging [50], and K-edge imaging [51].

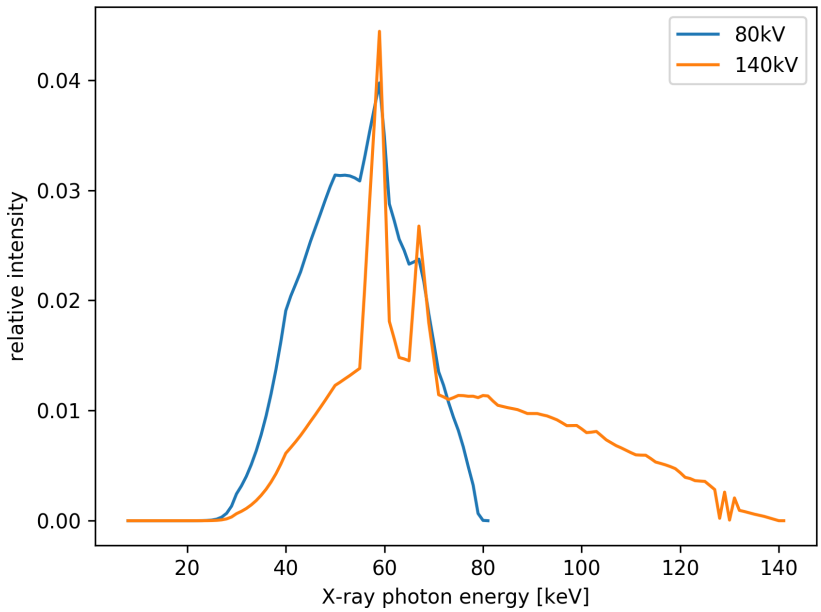


Figure 2.9: X-ray tube spectra with Tungsten anodes, after filtered by 20cm of water, simulated with [48].

CT numbers and dual energy index

The quantity recorded in CT images is often referred to as *CT numbers*, and its unit is the Hounsfield unit(HU). CT numbers are meant to measure the linear attenuation coefficient ($\mu[mm^{-1}]$) of the medium relative to water and air as reference points. By definition water should have 0HU, and air should have -1000HU [52].

$$\text{CT number} = \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}[\text{HU}] \tag{2.10}$$

The measured CT numbers depend on the X-ray beam spectrum, because the materials have different cross-sections depending on the X-ray photon energy and the (effective) atomic number (Z) of the materials. Because the spectrum depends

on the tube voltage, the same object can be scanned with different beam spectra. Most of the time this means two beams, and known as dual energy CT imaging.

The dual energy index [53] (DEI) measures the relative linear attenuation coefficient ratio, which is the following expression assuming two scans of the same region (x_i), one with 80kVp and another with 140kVp:

$$\mathrm{DEI} = \frac{x_{80kV} - x_{140kV}}{x_{80kV} + x_{140kV} + 2000} \tag{2.11}$$

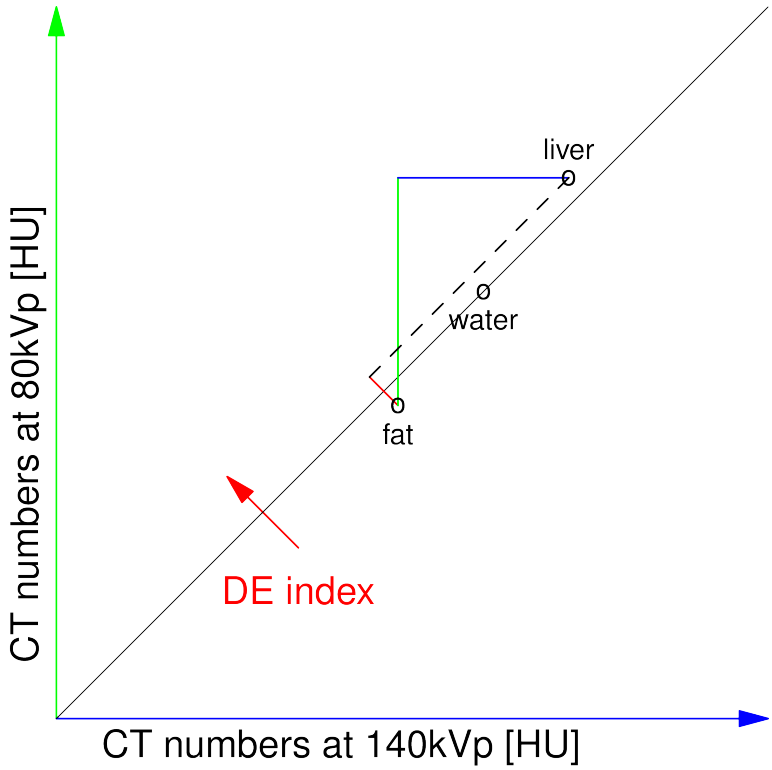


Figure 2.10: Dual energy imaging: CT numbers depend on the X-ray beam spectrum. Using two scanning energies instead of one makes the tissue and material separation more reliable.

DEI depends on all of the factors as CT numbers in single energy scan, and in addition it depends on the relative differences of the two X-ray spectra.

CT numbers are used for identifying tissues and diagnosing pathologies. Unfortunately, the differences between soft tissue densities are small and the reported

Table 2.1: DEI index for typical materials seen in CT colonoscopy, reproduced from [56]

Material	DEI Value
Adipose tissue	-0.0242
Soft tissue	-0.0039
Water	0
Muscle, body tissue	0.0022
Carbon dioxide	0.0025
Air	0.0063
Bone, femur	0.2313
Iodine	0.5686

numbers also depend on the manufacturer, the CT model, the reconstruction algorithms and individual differences between patients [54][55]. Using only absolute thresholds for diagnostics and/or segmentation is an error-prone approach due to these variations. One possible methodological improvement is to use reference points in the body, for instance, a liver metastasis should be compared to the background healthy tissue. Another direction would be to exploit the dual energy index. Due to the difference term in the DEI formula, DEI has large relative error, see Fig. 2.11, and while it yields helpful extra information, it might be challenging to utilize this information.

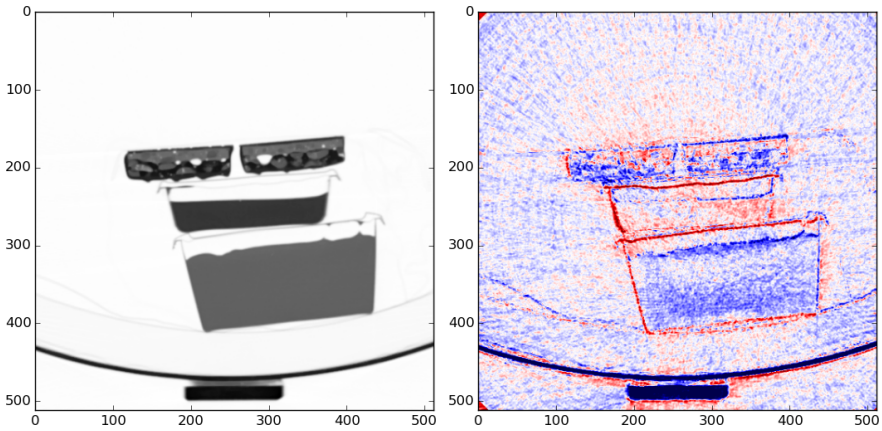
The magnitude of the relative DEI error can be approximated easily from Eq. 2.11. Assuming that the material has around $x \approx 0HU$ average density, and the standard deviation of the CT numbers is around $\sigma = 10HU$, the standard deviation for DEI is approximately:

$$\sigma_{\mathrm{DEI}} \approx \frac{(\sigma_{x_{80kV}} + \sigma_{x_{140kV}})/\sqrt{2}}{2000} \approx \frac{14.1}{2000} = 0.00705 \tag{2.12}$$

DEI for typical materials seen in CT colonoscopy was published in [56]. For demonstration purposes, these values are reproduced in Table 2.1.

2.3.5 Dual source CT

While dual energy CT imaging dates back to the 1970s [57][58], the proliferation of the technology in clinical practice is quite recent [50]. Dual energy CT could be realized with single source using kilovoltage switching or using two separate X-ray sources simultaneously, or using special detectors which are either capable of measuring photons individually (photon counting detectors [59] or have a special pixelated filter in front of the detector pixels giving different spectral sensitivity



(a) Single energy CT scan

(b) Dual energy index image, positive values are red

Figure 2.11: DE scan of Snickers bars (top), cheese (middle) and butter (bottom).

to them. Photon counting detectors are not widespread (yet) due to high cost and engineering challenges, but it is subject of intense research. Filter-based detectors are relatively cheap and simple but due to the post-patient filtering, it uses higher dose than regular detectors. The combination of the two interlaced detector pixel type yields a lower sampling density in both energy spectra which leads to lower effective resolution. Due to these reasons, the two most frequently used spectral CT imaging approaches are the dual source and the kilovoltage switching approaches.

2.4 The human visual system

Ever since Isaac Newton discovered that white light can be disassembled into colors using a prism, color vision is subject of intense research.

2.4.1 Processing of the visual information

The human brain transforms the individual activations into information. This is an extremely complex process, on various levels. Intuitively the most important features are edge detection, color vision, movement/change detection, and brightness detection. In nature, the illumination of a scene depends on the intensity of the light, the spectrum of the light, its direction, whether it is spot light or ambient, among others. The distance of the objects, texture, and many other parameters are also not controlled.

The human visual system is evolved to recognize patterns, and to categorize the same object similarly, regardless of the lightning conditions. As a profound example,

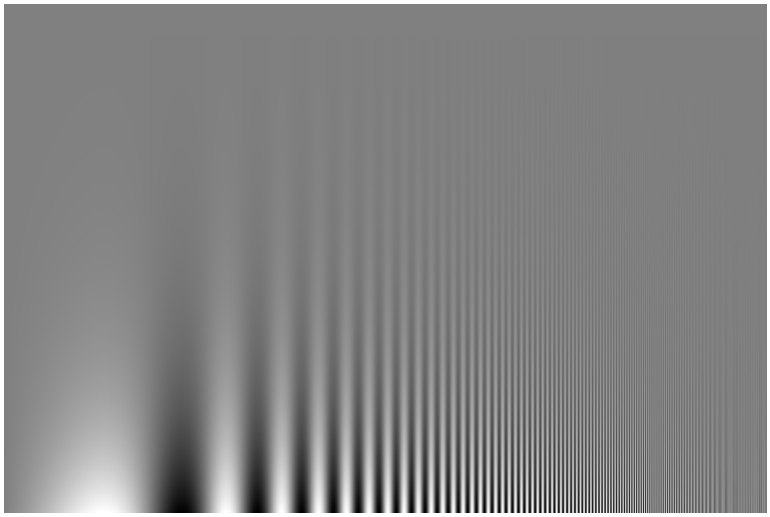


Figure 2.12: Campbell-Robson contrast sensitivity chart [61]

it is very useful to recognize a striped tiger both in the morning lights and in a shadow in the afternoon.

Physically, the eye sees different color from a given object, if the conditions change. But the brain tries to compensate these changes using context information. The opposite happens when the context information transforms close properties to be perceived as different ones. These are the optical illusions.

In the following some important properties of the human visual system and related optical illusions will be presented. While there are biological models to explain why a given illusion occurs, these models are not crucial for the rest of the thesis. But the illusions itself are important, and they are considered as external factors which must be taken into account. These factors could both improve or degrade the perceived image quality which could affect the diagnostic performance.

Contrast sensitivity function

The first and most important observation is that the perceived contrast in an image depends on the frequency. In Fig. 2.12 the frequency changes along the x-axis, while the contrast changes along the y-axis. It is immediately obvious that at constant contrast, the frequency affects the detectability. The chart in 2.12 is often called as Campbell-Robson contrast sensitivity chart [60]. The examples in this thesis use Ohzawa's public domain reproductions [61].

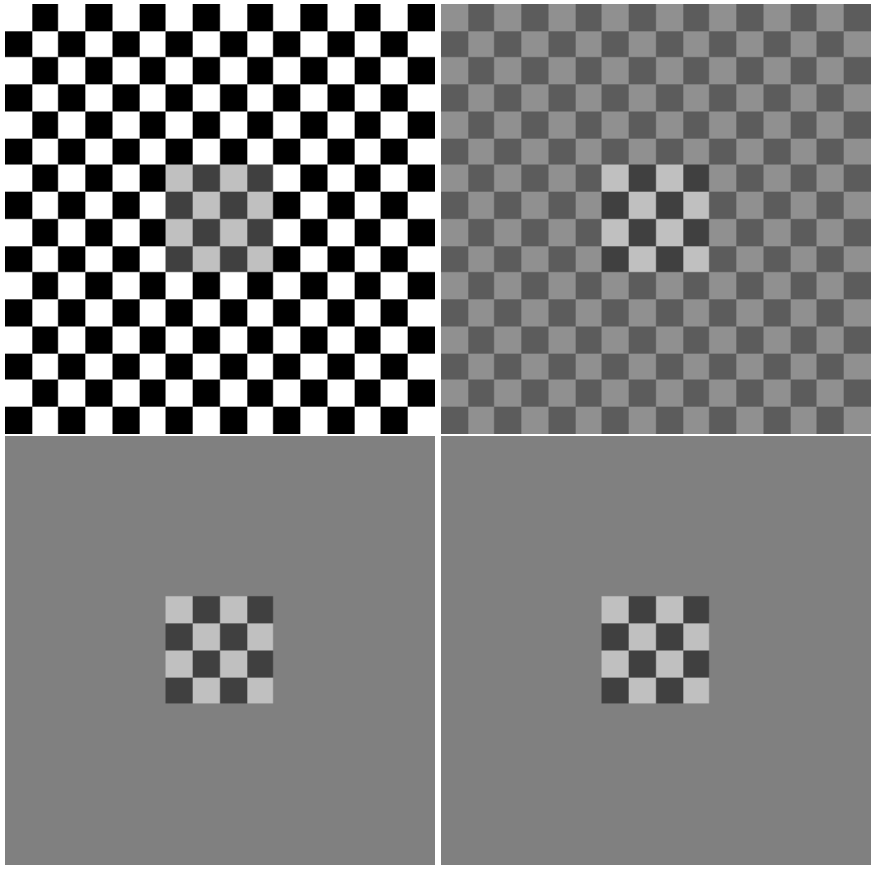


Figure 2.13: Chubb's illusion: perceived contrast depends on the neighborhood[62]

Local contrast

The perceived contrast also depends on the environment. If the neighborhood has high contrast elements, then the perceived contrast is lower in the region of interest than if the neighborhood has low contrast. This is called Chubb's illusion [62].

Color maps and false contours

Mapping densities to display colors is also a tedious task. First, the perceived contrast can be very different using different color maps, see Fig. 2.15. But good local contrast isn't the only requirement for displaying medical images. Often two other properties are required: avoiding false contours [63] and ability to produce linear mapping between the perceived contrast and the measured physical quantity in the scan. A false contour might be perceived as a border between anatomical structures, and mislead the diagnostician. Similarly important to keep the linear

mapping between the physical quantity and the perceived colors difference. The infamous rainbow color map exhibits both issues, see Fig. 2.16 and 2.17, but even the PET and PET20 color maps from the DICOM standard [64] can lead to confusion for a color deficient reader. The perceived color difference is calculated in a uniform color space (CAM02-UCS) [65] based on the CIECAM'02 color appearance model [66]. Simulation of the color deficient vision is based on [67]. Fig. 2.17 is part of an unpublished, in-progress work of the author.

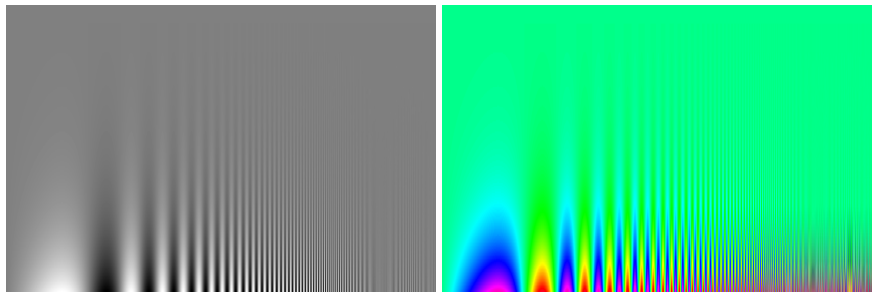


Figure 2.14: Contrast sensitivity plot with rainbow color map

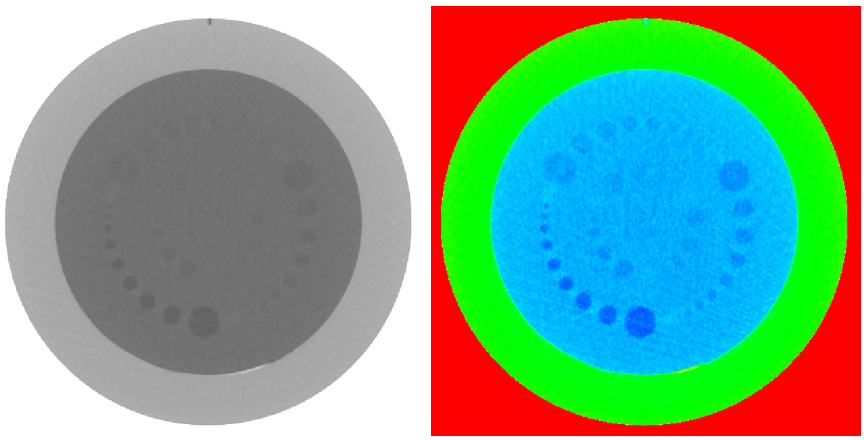


Figure 2.15: Low contrast detection test volume visualized with gray and rainbow color maps.

2.4.2 Color and brightness constancy

One final important property of the human visual system is the fact that colors and brightness are measured as relative values compared to the neighborhood. This is extremely useful considering the fact that illumination conditions in nature can be very different, while the same physical object should produce similar perceived colors. On the other hand, in radiological settings, where both the environment

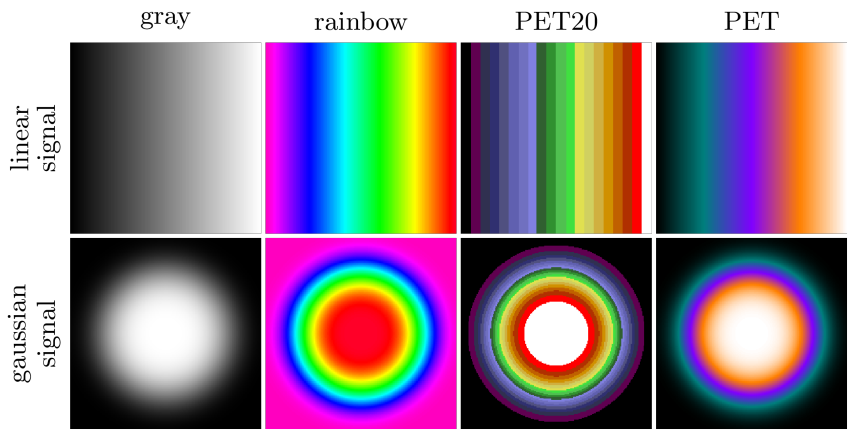


Figure 2.16: False contours: edges might be perceived with some color maps even in slowly changing continuous signals

and the displays are strictly controlled, this adaption can be misleading. Fig. 2.18 shows an example where the background changes but the foreground has constant color. But due to the slowly changing background, the perceived brightness of the foreground object is also slowly changing in the opposite direction as the background does. This is called simultaneous contrast [68].

2.4.3 Illusions and CT images

It is obvious that anything which changes the perception of an image might change the diagnostic value of this image.

Contrast sensitivity function plays central role in low contrast detectability tasks, see e.g. the CTP515 module in the Catphan 600 quality control phantom. The same task is also very sensitive to color map selection, see Fig. 2.15. Chubb's illusion could play significant role in reconstruction kernel selection. Sharp kernel produces noisier images, and the higher noise act as a high contrast background which reduces the perceived contrast of the already low contrast patches. This explains why soft kernels are recommended for such diagnostic task. Through carefully chosen FBP filters, the modulation transfer function might emphasize some lower frequencies. In this case, the approximate expected size of the low contrast object, the contrast sensitivity function and the noise level together should be taken into account for an optimal low contrast enhancement filter.

The simultaneous contrast illusion might appear when the image contains artifacts, e.g. beam hardening or strike artifacts. Beam hardening slowly changes the baseline, and should be considered as an artifact which mostly affect low frequency

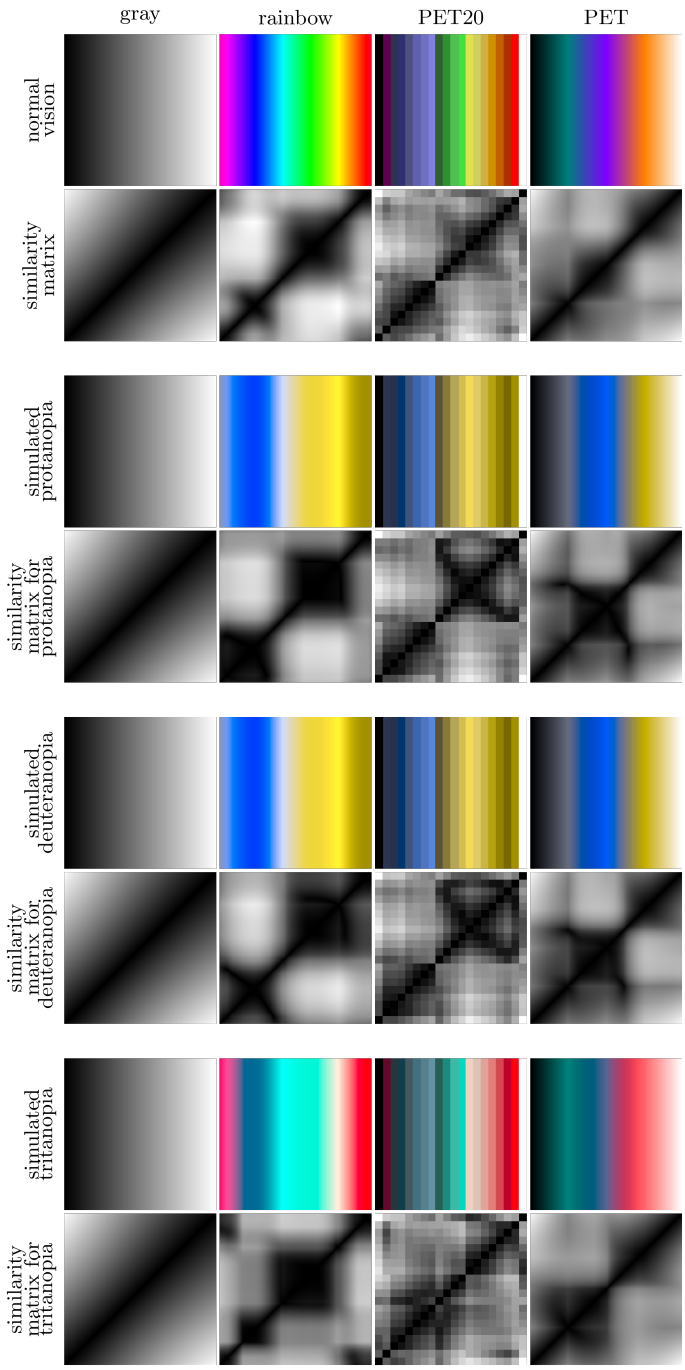


Figure 2.17: Color deficiency might affect perceived color difference and lead to confusion, see e.g. the PET20 color map with simulated deuteranopia. (Unpublished, in-progress work of the author.)

Figure 2.18: Simultaneous contrast illusion. Gradient in the background distorts the perception of the foreground object: despite foreground object's constant brightness, a slow change is perceived.

components of the image, see [49]. While it can contribute to simultaneous contrast, it could also generate false contours if an improper color map is used. In this case, a discontinuity in the color map would render close values as very distinct colors or shades which would give the illusion of a tissue border.

In addition to simultaneous contrast illusion and perceptually linear color map, the perceived brightness of a region is not independent from the edges of the region. Craik–O'Brien–Cornsweet illusion, depicted in Fig. 2.19, might render two equally bright region perceived as different. Similarly, simultaneous brightness contrast illusion, Mach bands and Chevreul illusions might change the perception [69].

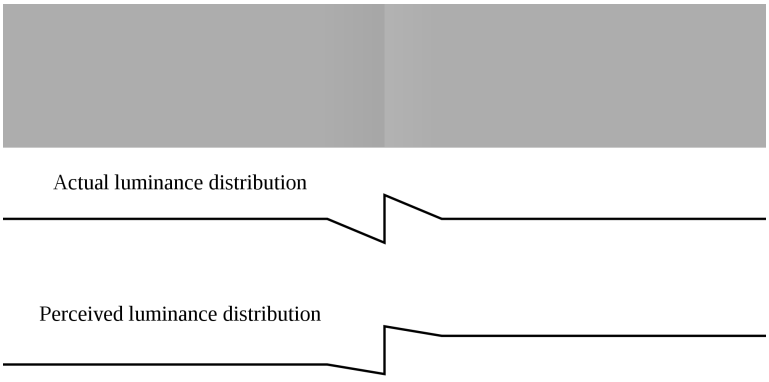


Figure 2.19: Craik–O'Brien–Cornsweet illusion [69].

The reader might have the impression that these illusions should be avoided at all

cost for good diagnostic image quality. This is not necessarily the case. Local contrast could be enhanced and the perceived dynamic range of the image could be extended if the properties of the human visual system is exploited carefully.

However, exploiting these illusions might lead to false positive diagnosis, while not using them might reduce the sensitivity of the diagnostic protocols. The optimal protocol should be determined through phantom studies and eventually clinical trials.

For further details, Chesters' excellent article [70] is recommended about the human visual perception in medical context.

2.5 Image processing

The two major purpose of medical image processing are quality improvement and enhanced visualization, but sometimes these categories overlap.

The quality assurance phantoms were specifically developed to ensure high quality. Arguably, the most notable parameters are resolution, noise level, noise texture, and contrast. But all of these should be interpreted as *perceived*. For instance, it was shown that the contrast perception depends both on the colors and on the frequency of the changing pattern.

Unfortunately, any processing step affects all of the descriptors.

2.5.1 Noise reduction

One of the most simple imaging problems is random noise. Noise is an unwanted signal mixed to the signal meant to be measured. Noise is often described by its distribution, e.g. Gaussian, and how it affects the signal, e.g. additive, multiplicative, quantization.

Every digital representation of analog signals contains quantization noise. CT images also have Poisson noise from the photon distribution, Gaussian noise from electronics, among others. While there are attempts to model the noise sources and the noise propagation in the CT data acquisition, this is mostly given as an unchangeable property of the device. However, the actual distribution of the noise can be measured, see NPS before.

The simplest noise model is assuming additive white noise.

$$f(i,j) = g(i,j) + \epsilon(i,j) \tag{2.13}$$

The white noise have uniform power at all frequencies, while about the signal it can be assumed that it has diminishing contribution at high frequencies.

If the measured image is disassembled into spectral components, using Discrete Fourier Transform (DFT), then in this domain, the signal-to-noise ratio (SNR) is a function of the frequency. At low frequencies, signal dominates, at high frequencies, noise dominates. Suppressing high frequencies increases the overall SNR. Using weights in frequency space ($w(u, v)$), the filtered image ($f_r(i, j)$) is an approximation of the noise free image ($g(i, j)$), and calculated as follows:

$$g(i,j) \approx f_r(i,j) = \mathrm{DFT}^{-1}_{u,v} \left\{ \mathrm{DFT}_{i,j} \left\{ f(i,j) \right\} \cdot w(u,v) \right\} \tag{2.14}$$

While it might not be obvious, this noise reduction is built on an assumption: the spectral distribution of the noise and the signal. Often, this is the only information about the noise component, but there can be much more assumed about the signal.

The most important assumption comes directly from diagnostics: what is the expected dominant frequency of the signal to be measured, and what is the expected signal contrast. For instance, lung structures are small, their visualization requires high frequency components, but the contrast between regions with air and with vessels is good, and good SNR is expected. On the other hand, liver lesions might be large in physical size, but due to their low contrast, might be hard to detect.

Image priors

Noise reduction is the art of exploiting every available information in order to reconstruct the noise-free signal. One of the recent direction in noise reduction is having a prior assumption about the signal. For instance, if the image mostly contains piecewise constant regions with some sharp transitions, then the total number of edge pixels could be minimized while the signal should remain close to the measurement. Mathematically, this conditions are summarized as functional minimization problem with two terms: an image prior term (J) and a data fidelity term (E). From the following, the pixel indices are omitted, and f is used instead of $f(i, j)$ for the reconstructed signal, and similarly, f_0 for the noise measured signal. λ balances between the noise reduction and data fidelity term.

$$\min_f \left(\lambda J(f) + E(f, f_0) \right) \tag{2.15}$$

Defining the L_p-norm will help to formulate this minimization problem.

$$||x||_p = \left(\sum_i \mathrm{abs}(x_i)^p \right)^{1/p} \tag{2.16}$$

If $p = \infty$ is defined as maximum-norm, and $p = 0$ is defined as the number of non-zero elements.

J takes a function as a parameter, it is called functional, and the optimization problem belongs to the field of variational calculus. Having this notation, the Total Variation (TV) minimization problem can be defined as follows:

$$\min_f \left(\lambda||\nabla f||_1 + ||f - f_0||_2^2\right) \tag{2.17}$$

As mentioned, λ balances between noise reduction and fidelity. Intuitively, $\lambda = 0$ means that the prior part is ignored, and only the fidelity part is enforced. In this case, the solution is $f = f_0$ with zero error (but with all the noise). If $\lambda = \infty$, then only the prior is taken into account. The signal with smallest number of transitions without any constraint, is a constant function. This is obviously noise free solution, but the signal is lost. $0 < \lambda < \infty$ yields a non-trivial, edge-preserving denoising algorithm.

CT images are often assumed to be piecewise constant images, where the only source of local texture is noise, the transitions are sharp. This simple model performs surprisingly well in many cases, but too strong denoising leads to stair-casing artifact. Technically, this is the result of violated assumptions. The underlying signal has a slight gradient, but the denoising converges to piecewise constant solution, and breaks the gradient into several smaller constant regions.

A slightly modified version of the prior term could prevent this stair-casing:

$$\min_f \left(\lambda||\nabla f||_2 + ||f - f_0||_2^2\right) \tag{2.18}$$

However, if gradients are not punished, the image will be blurred. As a matter of fact, this above functional is equivalent with Gaussian smoothing, which is also equivalent with Gaussian smoothing in frequency space. This is an equivalent formulation of Eq. 2.14, if $w(u, v)$ has Gaussian weights.

Modern noise reduction

The introduced TV denoising is a relatively widely used algorithm. The algorithm can be generalized to dual energy images, when a matrix norm has to replace the vector norm [71], e.g. using nuclear norm.

Edge preservation is not a unique feature of the TV-filter family. The simple median filter also preserves edges, and many more have been developed in the past decades. Similarly to TV, partial differential equation based Perona-Malik [72] diffusion

and the generalized anisotropic diffusion are also edge preserving filters. They are based on the idea that edges separate regions from each other, and smoothing an image parallel to an edge preserves the edge but effectively reduces the noise.

Bilateral filtering [73] resembles TV denoising in two ways, it is edge preserving, and might cause stair-casing artifacts. Its main contribution is the idea that both spatial and photometric distance can be taken into account. While it is a widely used elegant filter, it requires significant computational resources. Recently, guided image filter [74] was introduced as a very fast alternative, fast enough for real-time video processing. The guided image filter also generalizes the idea of weighting, because some properties of the guiding image is transferred to the filtered image. If the guide is a Gaussian, then it is only an edge preserving noise reduction. But in case of dual energy CT scan, the images from two different scans can be each others guide.

The recent research directions are focused on finding better image priors, for instance, a representation space where the signal is sparse, and the denoising can be formulated as an L_0 or L_1 problem. This is the so called compressed sensing. Until now, the presented denoising worked on a single image or image pair. However, sometimes the prior and the fidelity term do not use the same space. In CT imaging, the primary data is the measured projection, the fidelity term should ensure the minimal change of the projections. The assumptions about piecewise constant regions describe the reconstructed image:

$$\min_f \left(\lambda J(f) + E(p, p_0)\right) \tag{2.19}$$

This can be easily solved for some functionals but not for all.

Finally, the noise structure is usually assumed to be Gaussian, but this is just an approximation.

2.5.2 Contrast enhancement

Contrast is the difference of two signal levels. The perceived detection highly depends on how fast the signal changes, on the magnitude of the difference, and the contrast of the neighborhood.

Histogram methods

The most simple global contrast enhancement is the contrast stretching. Assuming that the display can reproduce N colors/grey shades, the dynamic range will be [0,N]. If the image has a smaller effective range [A,B], then this could be stretched, and mapped to the closest color.

$$f_s = \text{round}\left(N \cdot (f - A)/(B - A)\right) \tag{2.20}$$

In more general sense, contrast stretching is a simple histogram manipulation method, where only the smallest and largest non-zero bins are used for stretching.

This still might be imperfect use of the dynamic range, e.g. if a few shades represent most of the pixels. Histogram equalization builds on the idea that every shade should have approximately the same number of pixels.

There is no locality information in the histogram, nothing guarantees good local contrast. The natural generalization is the adaptive histogram equalization (AHE) [75]. Sliding window is used for histogram calculation, and the local histogram is to determine the pixel intensities. With too small sliding window, this method yields very stretched local contrast, and might strongly amplify noise. The amplification can be controlled, if histogram peaks above a threshold are cut, and the cut area evenly redistributed to every histogram bin before equalization. In practice, this sets a limit for the maximal amplification, but do not affect sufficiently flat histograms. This method is the contrast limited adaptive histogram equalization (CLAHE) [76].

These methods are used for CT image processing since the 1980s, but there is nothing CT specific in the algorithms. Domain specific information could be incorporated in several ways. One is to split the histogram into two or more parts, and equalize the parts independently. This is the bi-histogram equalization [77]. In this way, given gray-shade ranges could be assigned to lung tissue, soft tissues, and bones. But histogram methods are not the only ways to enhance local or global contrast.

Local enhancements

A general framework for local contrast enhancement is extracting local structures, e.g. edges, and adding them back in amplified form [78].

The most simple way is the amplification of the high frequencies. This is the opposite of the frequency domain noise suppression, and as any linear filter, this could be also expressed with DFT operations.

A more general form of edge enhancement builds on the separation of the image (I) into a base (B) and a detail (D) layer.

$$I = B + D \tag{2.21}$$

A sufficiently good filter leaves the base layer intact, but removes the details.

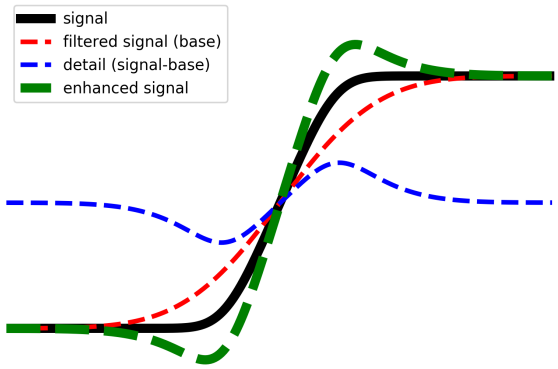


Figure 2.20: Unsharp masking. The detail information is the difference of the original signal and its blurred version. Adding back this detail to the original signal enhances the local contrast.

$$B = \mathcal{F}(I) \tag{2.22}$$

And the detail layer is the difference between the original image and the base:

$$D = I - B = I - \mathcal{F}(I) \tag{2.23}$$

The amplified detail can be added back:

$$I = I_0 + \alpha D \tag{2.24}$$

Fig. 2.20 depicts these above steps. This is a simple but very effective solution. If the filter is a Gaussian, then this method is called Unsharp Masking (USM). But any filter could be used for base-detail separation, for instance, median filter,edge preserving filters [79], or any noise removal algorithm, e.g. total variation denoising [80]. Edge enhancement is a huge topic, further details can be found in the literature [81].

Tone mapping

The local edge enhancement could be approached from a different point of view. If the source image has high dynamic range, and the display has low dynamic

range, then a simple linear rescaling could map the intensities to the display without burned out pixels, but the local contrast will be poor. Or if the local contrast is kept, then some pixels will not fit into the dynamic range.

The idea of tone mapping is to somehow compress the dynamic range and preserve local contrast at the same time. The previous base-detail layer idea could be re-used here. The important information is carried in the detail layer, and the base layer only sets the mean values for regions. An effective dynamic range compression method could compress the base layer, and the detail layer back to get the tone mapped image (TMI) [78]:

$$TMI = \text{compress}(B) + D \tag{2.25}$$

As it was seen at unsharp masking, this approach could lead to over- and undershoots. Using edge preserving filters reduces this issue, and multi-level processing can ensure that details are enhanced at all scales. Due to the proliferation of digital cameras, the field of tone mapping algorithms is a very intensive research area. Recent algorithms model some aspects of the human visual system, for instance, light adaptation. Also finding a good measure for the quality of tone mapping operator is a recent research direction [82][83][84]. Presenting the most information might not be visually appealing (digital photography) or not yield the best diagnostic performance (medical images).

2.5.3 Further literature

Radiology, forensic sciences and image processing are all enormous fields on their own. The goal of this chapter was to give context before the contributions and discussion. Further details can be read in excellent books [81], [85] and [86].

Chapter 3

Summary of contributions and results

This thesis is based upon four papers, describes different parts of the imaging workflow. The first paper investigates the quantitative properties of CT scans and reconstruction protocols using an image quality phantom. The second paper addresses the perceived image quality, and whether or not a clinical protocol is acceptable for post mortem scans. The third paper takes the reconstructed image given, and propose a visualization approach in order to maximize the amount of presented information. The fourth paper addresses quantization artifacts, staircasing artifacts, which might be a result of too strong denoising.

These contributions are only a step towards improved image quality at different stages in the imaging chain. Every step of the CT imaging workflow is subject of intensive research. The presented contributions were incremental, but together they cover a larger part of the workflow. The primary goal of the papers are presented below, but the secondary aim was to deepen the understanding of the interconnectedness of the steps. We believe that the context, the pre- and post-processing steps are important, and future contributions should take these into account besides the immediate problem in the given task.

In the following sections the primary goals and direct results of the papers are explained. Wider context and future directions are considered in the discussion chapter. Source codes, non-peer-reviewed contributions, for instance, experimental softwares, were published in order to facilitate reproducibility, prepare and pave future experiments and studies. Fig. 3.1 depicts the relevant contributions and refers to the papers in the 'Complete list of publications' section.

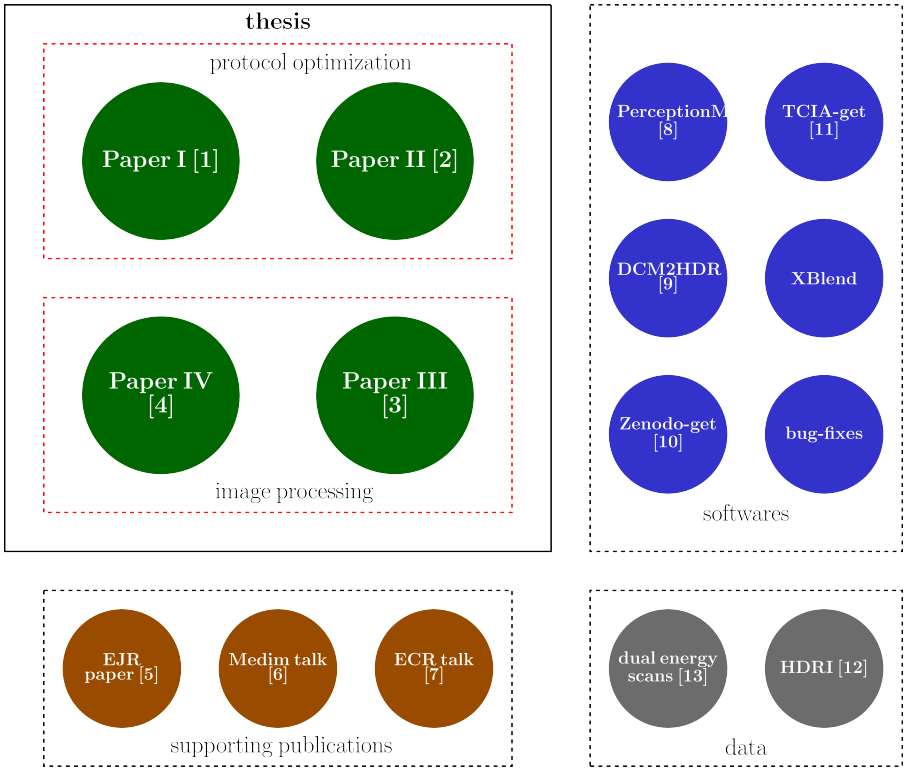


Figure 3.1: Contributions. Paper I-IV (green nodes) form the core of the thesis. Supporting publications, softwares and data supplement augment the core papers but they are not part of the thesis. Citation indices refer to the papers in the 'Complete list of publications' section.

3.1 Paper I: How Different Iterative and Filtered Back Projection Kernels Affect Computed Tomography Numbers and Low Contrast Detectability

While every patient, every case is different, quality assurance requires standardized protocols. Vendors developed a set of available options for scan and reconstruction, but the number of available options are enormous, and also vendor and model specific. Performing clinical trials is only feasible for a limited number of variations. This set of candidate protocols were designed based on the diagnostic requirements of the pathology, e.g. required low contrast visibility, or required sharpness, etc.

Some of the scan parameters, e.g. dose and pixel size depend on the patient, and determined by the automatic dose modulation and the field of view. For the most

popular choices, several parameters were reported in the literature (NPS, MTF, CNR, etc.). In this paper, 313 variations of scan and reconstruction protocols were compared in order to detect subtle differences.

The approach was kept as simple as possible to eliminate as many variables as possible. A Catphan 600 phantom was scanned using 70kVp, 80kVp, 100kVp, 120kVp and 140kVp x-ray tube voltage using the same field of view. These five scans were reconstructed with every available standard reconstruction kernel and iterative reconstruction strength.

The general assumption could be summarized briefly: at a given tube voltage for a reasonable sized region of interest, the mean of the CT numbers should be independent of the kernel selection and the iterative reconstruction. Also the CNR is expected to increase using stronger iterative reconstruction. CT numbers depend on the tube voltage, except for water and air. This was taken into account in the fact that a selected kernel was fixed as a baseline, and every other reconstruction was compared to this baseline.

Our results show that these are reasonable assumptions but not always true. Two major deviations are discovered. First, the mean CT numbers are systematically deviated between some reconstruction kernels, e.g. B26/B36. This fact should be taken into account during optimization. E.g. if the mean CT number is used for deciding between benign and malignant tumor, then these differences might be important. Even if very light (air) and very dense (teflon) materials or ignored in the study, the measured CT number differences using B26 and B36 kernels were in the -9,+10HU range.

Second, arguably more surprising result affects the iterative reconstructions. CNR is calculated from the ratio of signal difference and noise standard variation. If the signal difference is unaffected by iterative reconstruction, then CNR is expected to increase because noise is expected to decrease. However, in case of two kernels, the first IR option increased the variance. This happens only a very small subset of the investigated options, but it contradicts expectations.

3.2 Paper II: Applicability of a clinical cardiac CT protocol in post mortem studies

Are post mortem scans different from in vivo scans? They are. Post mortem alterations cause visible differences, for instance, atrium dilatation, blood clotting and appearance of gas volumes. The severity of these alterations depends on the time difference between the death and the scan.

Forensic CT as a sub-field of CT imaging is relatively young, and it is reasonable

to re-use experience from clinical scan protocol optimizations, if the investigated pathology is similar to the clinical pathologies. Investigating special forensic issues which has no clinical counterpart requires specialized protocols.

A clinical protocol is result of an earlier optimization. Any change in an optimal protocol would lead to inferior performance. Paper II summarizes a study where a protocol from clinical practice was tried in a forensic settings, and slightly changed reconstruction kernels and iterative reconstruction strength were applied. Three post mortem cardiac scans were selected and nine different reconstruction were performed for each case. The 27 volumes were presented to three radiographers and three radiographer students. These readers scored the volumes for seven criteria, using 5-scale visual grading.

Acknowledging the limitations, including the limited number of cases and readers, and the inter-observer differences, the conclusion of the paper that no alternative reconstruction yield better results than the baseline did, therefore, in this case, the clinical protocol seems to be applicable in forensic setup too.

3.3 Paper III: A weighted histogram-based tone mapping algorithm for CT images

CT images have high dynamic range, while the medical displays have lower. Even if a display was able to reproduce the whole dynamic range, the local contrast might be imperfect.

The focus of this paper is dynamic range compression in order to present as much information as possible. This is a bit vague goal, and unfortunately, we are not aware any image metric which takes the properties of CT imaging into account.

The paper presents a histogram equalization based tone mapping approach for CT images. When the human visual system determines the relative local brightness of a patch, it takes every other patch into consideration with a decreasing, distance-dependent weight, which follows power law. The human vision is able to handle huge dynamic range differences, and the idea is to mimic this property. The presented histogram equalization method is built on a power-law based weighting for local adaptive histogram equalization with a contrast limit. This contrast limit ensures that the image distortion and noise amplification are controlled.

While direct implementation of the idea requires large amount of memory and computation time, presented optimization approaches, including FFT based contribution calculation, down-sampling in discretization levels, spatial down-sampling and dithering, make the problem not only tractable, but near-realtime (few sec/image), if every optimization is utilized.

The results are compared to popular tone mapping algorithms, both by image metrics and visually. While the presented method performs well against the alternatives, clinical validation is not yet performed.

3.4 Paper IV: Image De-Quantization Using Plate Bending Model

Paper III utilized an observation about the human visual system. In paper IV, a challenge from image processing is addressed, namely the de-quantization problem. This arises when an analog signal is discretized with insufficient number of discretization levels. While CT images themselves have high bit depth, post-processing of the images might lead to a so-called stair-casing artifacts, for instance, in case of extreme denoising.

The main idea of the paper is based on a physical model. The signal is modeled as a bended plate. The motivation is two-fold. First, the signal might actually originate from a bended surface. Second, from mathematical point of view, this model leads to a high order smooth approximation which is able to reproduce signal in local maximal/minimal areas and yield smooth surface. Alternatively, the problem could be approached through splines or using the thin plate energy functional. The main difference between these approaches is the ability to allow or to ban saddle points on the reconstructed surface.

The paper presents solutions for analytical signals (Gaussian and Rosenbrock functions) using structural similarity metric (SSIM), peak signal to noise ratio (PSNR) and normalized mean square error (NRMSE). A low bit depth photo is also used as demonstration.

As it was mentioned in the beginning, extreme denoising is an application area for this kind of algorithm in CT imaging. The discussion section of this thesis contains a dual-energy CT denoising example, but this example was not part of the original publication.

3.5 Non peer reviewed contributions

Not every contribution results in an article. For instance, custom softwares, bug fixes in widely used algorithms are useful, but do not necessarily yield a peer-reviewed paper. These contributions are not core to the thesis, but as artefacts of the project they are worth mentioning, and also indicate the efforts for future studies.

PerceptionMD

One of the main drawbacks of visual grading analysis becomes apparent when close options are compared. For instance, if two alternatives have image quality score of 3.1 and 3.2 by a criterion, it might be impossible to show the difference using VGA. On the other hand, pairwise comparison of the volumes might quickly reveal if the difference is subtle but clinically significant.

PerceptionMD[87] open source software for conducting pairwise observer studies in radiology. The latest version is available at:
`https://gitlab.com/dvolgyes/perceptionmd`

We are not aware of freely accessible open source program for pairwise comparison of radiological volumes. PerceptionMD meant to extend the current options for observer studies.

Key features:

- Open source program, implemented in Python2/3[88].
- Studies are defined in a simple domain specific language (DSL).
- Randomized and sequential studies are both supported.
- Both DICOM and RAW volumes can be displayed.
- Screens can be described using ReStructured Text[89].

XBlend

While PerceptionMD can be used for pairwise comparison studies, it forces the observers to make a binary decision. In some cases, it would be beneficial to blend two images, and optionally compare the result to a reference image. Nor visual grading analysis, neither constant stimuli pairwise comparison are efficient for this task. Determining the optimal blending ratio can be efficiently determined if the ratio can be changed in continuously, or using several discrete levels.

The following three optimization problem fit well into this profile:

- dual energy mixing ratio,
- mixing ratio between FBP and IR reconstruction,
- optimal kernel sharpness selection.

XBlend has been written to support such experiments. It allows full reference comparison, where a blend of two images is compared to a reference, and it also allows no reference optimization, when the observer has to choose the mixing ratio which maximizes the given criterion for the task.

Optimizing a continuous parameter is harder than selecting the best of a few discrete options. For instance, the median of the preferred mixing ratio could be interpreted as the result of the optimization. However, the selected option could and should be cross-checked in follow-up experiments, for instance, using pairwise comparison, visual grading analysis, or conducting ROC study.

The latest version is available at:
`https://gitlab.com/dvolgyes/xblend`

DCM2HDR

DCM2HDR [90] is another open source program which was developed as a utility for the paper III. It allows conversion of DICOM images to high dynamic range traditional image formats. These images can be further processed with image processing softwares and/or with our proposed algorithm in the paper. The latest version is available at:
`https://gitlab.com/dvolgyes/dcm2hdr`

Zenodo-get and tcia-get

Archiving large amount of research data is essential, but interacting with data repositories could be tedious. Zenodo is a open science data repository operated by CERN. Zenodo-get[91] is a tool to download huge datasets from Zenodo in a single step, using only the DOI or the record ID of the data. The tcia-get is a very similar tool, except it interacts with The Cancer Imaging Archive database.

For instance, example data for the previously mentioned PerceptionMD is stored at Zenodo, see next section, and this software makes the access to it simple.

The latest versions are available at:
`https://gitlab.com/dvolgyes/zenodo_get` and
`https://gitlab.com/dvolgyes/tcia_get`.

Data publication

There are two competing issues in science: reproducibility which requires openly accessible data, and privacy, which is especially challenging in clinical context. However, not every data becomes part of articles, and not every data carries personal information. Even data with absolutely no novelty could be very useful for algorithm testing purposes, as a visualization example, or a baseline.

We published two major datasets: dual energy CT scans of some selected ordinary objects showing various effects[92], for instance, metal artifacts, or the significance of dual energy scans for material identification. Finally, as an artifact of some experiments, large amount of high dynamic range panoramic photos[93] were collected from public domain images. These three datasets are published at Zenodo.org in order to ease further research.

Bug fixes and open source contributions

This thesis wouldn't exist without open source softwares. However, even the best softwares have bugs and need contributions. During the thesis, contributions were made to several open source softwares, most notably, scikit-image, where the Shannon entropy calculation needed a bug fix. Unsharp masking implementation, mentioned briefly in the introduction, was also contributed to the same project.

Most notable open source contributions:

Project	Contribution	URL
Scikit-image	Unsharp masking	`https://github.com/scikit-image/scikit-image/pull/2772`
Scikit-image	Entropy bug fix	`https://github.com/scikit-image/scikit-image/pull/2749`
STIR	Appveyor integration	`https://github.com/UCL/STIR/pull/102`
STIR	Travis integration	`https://github.com/UCL/STIR/issues/21`
FFX	Travis integration	`https://github.com/natekupp/ffx/pull/24`

Chapter 4

Discussion

The contributions of this thesis are incremental, and in some cases, specific to a scanner, to a protocol, or to an artifact. However, together they cover a large part in the imaging workflow, and while individual results might be overcome soon, the approaches and the applied principles are generic, and could be applied to new scanner generations and new problems.

The thesis can be organized into two major sections: Paper I and II cover scan protocol optimization steps with objective (paper I) and subjective (paper II) image quality evaluation. Paper III and IV present two image processing algorithms for tone mapping and image de-quantization, respectively.

4.1 Paper I

It is a common approach to address image quality first in phantom scans, and later refine the results in clinical experiments.

Paper I compares 313 reconstructions of a Catphan 600 phantom, using a single-day scan at a single dose level with a single phantom. This might seem as a serious limitation, but this approach rules out all of these variables affecting the reconstruction, and leaves the reconstruction parameters and the kilovoltage the two determining parameters. This gives a detailed picture about the reconstruction algorithm.

Using phantoms for investigating reconstruction algorithms is a frequent approach [94][95], and fixing parameters and focusing on a sub-problem is necessary in order to avoid combinatorical explosion of the problem space.

The fixed parameters shadow important factors. Is there any difference between

scanners from the same vendor? How do the scanners change in time? Are there inter-phantom differences? What are the similar reconstruction options between two scanners from different vendors? How does dose affect the results?

These questions were investigated before. Inter-phantom differences were reported in [96], comparison of noise texture across CT scanners was published in [18], the effect of dose is studied in [97]. Stability of the image quality parameters were investigated in [98] using 100 scans over a period of six months.

The volumes could be evaluated both with objective metrics and subjective experiments. Our choice for objective evaluation originates from the large number of options. It would be very demanding task to score 313 volumes, and it would require large number of observers to reveal significant differences between the options, if there is any. Our choice for simpler metrics, namely contrast-to-noise ratio (CNR) and mean values, is motivated to produce results which are objective and easily interpretable for a reader.

Both phantom studies [99][94] and clinical experiments [100] use similar approach, but these papers investigate the effect of low tube voltage on the image quality, dose, and the effect of iterative reconstruction, respectively. The last paper is especially interesting for comparison purposes. It found that CNR is always increased when iterative reconstruction was used. But the presented approach used only a limited set of reconstruction options. Our investigation had the same conclusion for this parameter subset (B40/I40 reconstruction kernels). However, for the B26/I26 and B36/I36 reconstruction family our study found a slight decrease in CNR when SAFIRE 1 is used, and SAFIRE 2 was found to produce very close results to FBP. It was also reported that SAFIRE 2 and 3 are the most preferred reconstruction options [101] and it was suggested that these options might be the optimal choices for any body part being imaged. Our findings challenge this opinion. CNR does not always increase with higher level of SAFIRE reconstruction, and it might be premature to conclude that intermediate SAFIRE levels are optimal. They are certainly good starting points for optimization, but their effectiveness must be validated. There is also a disagreement whether CNR affects the preference of the readers or not. While in [101] it was found that there was no correlation between CNR and the preferred option, it is known that CNR is a good predictor for low contrast detectability [102]. The explanation is most likely the different image quality criteria / diagnostic task, see the Background chapter.

4.2 Paper II

The second paper focused on a selected clinical protocol. CT imaging in forensic radiology is a relatively new development, and it began to spread around 2000. Due

to the post mortem alterations [26], in vivo scan protocols might not be applicable for post mortem scans. If the protocol is not applicable, there should be a better performing alternative. The selected nine protocols were derived from a cardiac protocol with modifying the reconstruction kernel and the iterative reconstruction. The results show that the original in vivo protocol did not perform worse, than any of the alternatives.

What does this mean clinically? Can an in vivo protocol used in post mortem imaging without modification? We do not imply this. First of all, there were limitations in the study, namely, the number of criteria was limited (7), the number of observers was low (6), there were significant inter-observer differences, and the study only focused on cardiac imaging.

In general, modeling a cadaver and a living patient in the same way is a model assumption, and every model has limitations. But cadavers are frequently used to bridge the gap between phantom and in vivo scans [103][104].

Our results imply that in this specific example, the in vivo protocol was performing reasonable well. We recommend that post mortem protocol optimization should begin with an adapted in vivo protocol, if applicable, and refine it, if necessary.

Possibly this has already been the approach in several hospitals, and our results give support to this approach. While the literature enumerates post mortem alterations which are visible in post mortem CT imaging [26], we are not aware of publications which explicitly argue for the inapplicability of clinical protocols in post mortem setups.

In fact, a recent publication discusses the adjustable technical parameters [105], but does not mention any necessary deviation from in vivo protocols, except when high density objects, e.g. dental implants, are the subject of investigation. The lack of radiation risk and lack of body movements enable further noise reduction, but they do not imply change in the reconstruction parameters.

The two papers together contribute to wider field of protocol optimization. With the advances of machine learning, the medical images play a role not only in the direct diagnostic decisions, but as historical data, they might be basis of future training data. One of the main challenge of machine learning is to obtain high quality training data. Besides the optimized protocols, the presented anomalies in paper I might be useful to avoid sub-optimal training data. One also can assume that post mortem scans, as long as the post mortem alterations are not severe, might be good data source for machine learning algorithms.

4.3 Paper III

Local contrast enhancements are successfully applied to CT images since the 1980s [76]. However, unlike other modalities, CT numbers represent a calibrated absolute physical property. This enables radiologist to learn absolute values, the same density could have the same gray level, regardless of the CT scanner or the patient. This is both an advantage, an expert can gain decades of experience, and a disadvantage, because every new method has to outperform this decades old experience.

Tone mapping is intensive subject of research for at least two decades, but it is still often referred as a beauty contest. Only recently were objective image quality metrics introduced to quantify tone mapping operator performance [106][83][82][84]. The two major components are structural similarity to the high dynamic range source image, and optionally a naturalness component.

Any visible property could be used to measure image fidelity, e.g. edges, relative brightness of regions, etc., and the human visual system takes them into account but with unknown weight. Image quality metrics were developed to measure distortions, e.g. compression artifacts, compared to a reference image. Tone mapping is special, because while there is a reference image, it could not be displayed directly due to the display properties.

Tone mapping, local contrast enhancements distort the image, but not necessarily decrease the diagnostic image quality. One approach is to take the high dynamic range image, and locally compare it to the tone mapped image. If the features are similar, then it is a good mapping. The problem with this approach is the assumption that the locally perfect mapping yields the best image. This assumption does not hold in digital photography, but in CT imaging faithful representation is not the measure of image quality. The measure is diagnostic accuracy. Even otherwise disturbing artifacts are acceptable, if the diagnostic accuracy is improved.

While it is claimed in [107] that monotonic global operators are preferred for tone mapping of medical images, we respectfully disagree. The brightness and color perception is context dependent, and the same displayed color might be perceived differently due to its neighborhood. As a matter of fact, many non-monotonic algorithms were proposed for or demonstrated in medical context [76][108][109]. It was also demonstrated that using CLAHE reduced the reading time and improved detectability [110], but also made the readers less comfortable with specific clinical settings [111]. We argue that brightness perception is not different when medical images are presented, and global monotonic operators should not be prioritized, and quality comparison between algorithms should be based on objective metrics and/or on subjective clinical experiments. The proposed method has adjustable

local contrast enhancement, and might very close to a global monotonic operator, or could be parameterized to be local and non-monotonic.

There are two reasons to use distortion based metrics instead of subjective experiment. First, we are not aware of any tone mapping metric which is specifically developed for CT imaging. Second, using intensity windows (IW) has a long history in CT, and it is considered as gold standard. If a tone mapped image is locally similar to a windowed representation, then it is close to a standard representation. This might not be the best visualization, but it is still the standard. Distortion based metrics reasonably well measure the similarity between the tone mapped image and the windowed image(s).

The main question is why and where tone mapping should be used? Technically, any image could be displayed several times, using different windows. This way the whole dynamic range could be displayed at any display. There are two arguments for using tone mapping. First, one of the bottlenecks in modern radiology is the time of the radiologists. If the number of readings of the same slice could be reduced, then the throughput of the radiologist could be improved. Long reading is not only a cost issue, but might lead to decreased attention, and inferior diagnostics. Second, presentations of multi-trauma cases are simplified if the compressed dynamic range allows to present multiple findings in the same image. In a forensic setting, whole body CT images are often grossly evaluated by the forensic pathologist prior to performing the autopsy. Only selected cases or findings of dubious significance are later evaluated by a clinical radiologist. Enabling visualization of both soft tissue and bone on the same window may reduce typical interpretation errors called "search satisfaction" - the tendency to focus on the initial, very obvious finding and pay less attention to smaller or more subtle findings due to the assumption that "the search has been fulfilled" [112].

As it was discussed previously, there is no universal diagnostic image quality model. Therefore, the metrics based comparison is built on an assumption, namely that similarity to linear mapping is a good enough measure. Clinical validation of the method would require subjective experiments with expert readers. Contrast enhancement methods were shown to decrease evaluation time and increase diagnostic performance in some experiments. However, whether this proposed algorithm performs better than the alternatives, is an open question. Such experiment is not easy to perform, due to several factors. First, the task based evaluation needs criteria. Second, scan protocols are not only different in display window, but could be different in several other parameters, e.g. slice thickness or kernel sharpness. The improved local contrast might or might not compensate for these changes.

There are also methodological challenges. Standard protocol optimization com-

pares alternatives with similar properties: similar number of slices, similar display window, etc. A tone mapped volume substitutes several volumes, and not only diagnostic performance but the required time should be considered. However, when a volume is evaluated, the reader remembers some details, which affects the reading time of the same volume with different display window. Therefore, the order of the presentation affects the outcome. Designing experiment for a localization task is even more complicated, due to the fact that the locations of the objects is remembered too.

4.4 Paper IV

De-quantization originates from image processing where low bit depth images lose details due to the quantization, and false contours give disturbing appearance. However, quantization could be replaced with any process which yields larger flat regions, for instance, some types of de-noising which are prone to produce stair-casing artifact [113][114], e.g. total variation minimization, or bilateral filters [73].

There are two directions in CT noise reduction: improved filters which do not exhibit stair-casing artifact, and mixing of the iterative reconstructed or denoised image with the unprocessed FBP image. Image mixing is a workaround which is unnecessary, if the filtering does not cause artifacts.

The technical details of the iterative reconstructions are often not published in details, only generalities are known. But even the type of reconstruction is indicative. For instance, reconstruction which iterates in image space, has to use assumptions about the image structure, and one such assumption could be that the image has piecewise constant regions. Such model would be prone to stair-casing artifacts.

In some cases, the image reconstruction itself does not produce stair-casing artifact, but a derived quantity is needed, e.g. dual energy index, which has very high relative noise. As it was shown in the Background chapter, the relative noise of DEI might be several times higher than the DEI difference between different tissue types. Suppressing noise in such situation might require extremely strong denoising, which might lead the aforementioned artifact.

Recently noise reduction is achieved using model based iterative reconstruction which is superior in image quality compared to alternative approaches, and avoids stair-casing artifacts. This might render de-quantization unnecessary, but handling historical data might require it on longer term.

From technical point of view, there are several de-quantization approaches, built on various assumptions, from image smoothness to machine learning. The optimal

approach depends on the underlying signal, and larger scale study could decide which approach performs best in given circumstances. Due to the nature of the signals, different approaches might be optimal for different subtasks.

The partial differential equation based approaches are the closest to the proposed method, such as the Laplace equation and spline approximation. The order of a polynomial approach always could be increased, but it is known that high order approximations are more sensitive to noise [115]. Using polyharmonic equation is a intermediate approximation which minimizes the mean curvature of the surface. Mean curvature is used as smoothness notion in [116] to model liver surface.

4.5 Limitations

Every paper has limitations, including the four paper this thesis is based on. The most obvious one is the limited amount of data. Specifically, only one CT scanner was used for the data collection, therefore scanner, model and vendor variations cannot taken into account.

Second, the number of test cases and the number of readers are both limited in paper II. Using five-six readers [102][101] are common practice, but from statistical point of view, it is a very low number.

Third, even expert readers have limited attention span, and a reading session cannot be arbitrary long, which sets a limit on the number of criteria for visual grading analysis. These limitations must be taken into account when the results interpreted.

The main limitation of the tone mapping paper is the evaluation. The lack of dedicated medical tone mapping image quality metrics means that only general metrics and observer studies could be used to validate the results. Observer studies are especially challenging in this context because the readers got training for the baseline images but did not get for the tone mapped images. Also the tone mapping is more similar to multi-trauma imaging. These makes an observer study closer to a beauty-contest which is inherently subjective. The promise of the image quality metrics is its objective nature. Unfortunately, another factor plays important role here: the field of view. Metrics in general images use every pixel in order to determine the quality of tone mapping. However, pixels from outside of the body do not carry any diagnostic information. Metrics should be aware of the context, but they do not take context into account.

The de-quantization paper has two major limitations. First, the assumption that signals can be approximated well with a bended plate, which minimizes the mean surface curvature. There might be cases where e.g. total curvature should be minimized. Second, de-quantization itself is a special problem. There are denoising

algorithms which are prone to stair-casing artifacts, e.g. total variation denoising or bilateral filtering. But one of the major research direction is to design filters which do not exhibit such artifacts. Using such filters makes de-quantization unnecessary and/or inapplicable.

4.6 Future perspectives

There are two kind of challenges in CT optimization. The first group consists challenges where the parameters are known but there is a trade-off between them, e.g. dose and noise level. These are expected to improve gradually as the underlying technology improves. For instance, faster read-out, higher spatial and temporal resolution in the detector pixels, more monochromatic X-ray sources could all incrementally improve image quality. Arguably, the two most promising developments are the photon counting detectors and the laser-driven monochromatic X-ray sources.

The other type of challenges is where there is no reliable model of the human observers. First major step could be the proliferation of automatic segmentation methods. The segmentation information could give a context which is not present at this moment. For instance, 2D projections are used for dose modulation. Context could fundamentally change how the scan is planned, which pathology could benefit from dual energy scan, which one is more sensitive to noise. Context also could contribute to noise reduction and visualization.

Even when the 2D/3D images are segmented and context information is available, it is hard to diagnostic image quality. There are carefully developed protocols, but they give a method and not an explanation. Machine learning could be used to model the human readers preferences. Theoretically, neural networks could approximate arbitrary functions, while alternative methods could find directly explainable models. For instance, the symbolic regression [117][118] fits the structure of the function to the data, not only the coefficients of the components.

Not only the human reader's preferences are partially unknown, but also the exact structure of the data. Recently compressed sensing and image constraints achieved very promising results. They build on a priori information about the signal. However, these are based on assumptions. Machine learning holds the potential to learn the structure of the data, the best representation space, the most successful image constraints.

More advanced models and machine learning both require large amount of high quality, annotated data both for model creation and validation. While data collection is performed in every hospital, sharing data, especially internationally, raises privacy and data protection issues, while restricting models to local data carries the danger

of local anomalies, lack of generality and irreproducibility.

In the near future, the author anticipates more advanced observer models, both about the human visual system and about the medical context. When reliable image quality metrics appeared in other fields, they always lead to objective, metrics based optimization with reproducible results. Similar leap is anticipated if or when objective metrics can substitute observer studies.

The deeper understanding of the image structure also might lead to new constraints in image space or using basis functions where sparsity constraints can more effectively exploited.

New metrics and new constraints would directly affect image quality. While context and more accurate observer models could improve visualization.

Perspectives are always based on predictions, and inherently unstable. However, metrics, machine learning, observer models have a successful history in digital photography, and while medical imaging has its specialties, there is no reason to assume that these factors would not be transformative in this field.

Developing such advanced machine learning models and image quality metrics requires high quality data to which this thesis meant to contribute both with protocol optimization and with image processing algorithms.

Chapter 5

Conclusion

The first paper presented anomalies which might affect protocol optimization. Especially interesting result was the decreasing contrast-to-noise ratio in some special cases despite the enabled iterative reconstruction. This should be taken into account during protocol optimizations.

The second paper concluded that clinical protocols might be applicable in post mortem investigations, despite the post mortem alterations. This might not be generally true due to the post mortem alterations, but the related in vivo protocols should be considered during protocol optimization for post mortem imaging.

The third paper focused on tone mapping. Applying a human visual system inspired power-law weighting function effectively balances between local features and artifacts, most importantly noise amplification and halos. Using slight approximations, the presented algorithm scales well in term of execution time and memory requirements too, which makes the method fast enough for experimentation. The presented algorithm can contribute to better visualization, and reduced number of reading sessions, but these advantages require confirmation by readers.

The fourth paper presented a de-quantization algorithm which could decrease, mitigate stair-casing artifacts. While the importance of the proposed method is expected to decrease with the proliferation of more advanced filters, it fits well into the partial differential equation based de-quantization approaches.

Bibliography

[1] D. G. Fryback and J. R. Thornbury. “The Efficacy of Diagnostic Imaging”. In: *Medical Decision Making* 11.2 (1991), pp. 88–94. ISSN: 0272-989X. DOI: 10.1177/0272989X9101100203.

[2] J. Y. Vaishnav et al. “Objective assessment of image quality and dose reduction in CT iterative reconstruction.” In: *Medical physics* 41.7 (July 2014), p. 071904. ISSN: 0094-2405. DOI: 10.1118/1.4881148.

[3] C. H. McCollough et al. “Strategies for Reducing Radiation Dose in CT”. In: *Radiologic Clinics of North America* 47.1 (2009), pp. 27–40. ISSN: 0033-8389. DOI: 10.1016/j.rcl.2008.10.006.

[4] L. W. Goldman. “Principles of CT: Radiation Dose and Image Quality”. In: *Journal of Nuclear Medicine Technology* 35.4 (Dec. 2007), pp. 213–225. ISSN: 1535-5675. DOI: 10.2967/jnmt.106.037846.

[5] M. Båth and L. G. Månsson. “Visual grading characteristics (VGC) analysis: A non-parametric rank-invariant statistical method for image quality evaluation”. en. In: *British Journal of Radiology* 80.951 (Mar. 2007), pp. 169–176. ISSN: 0007-1285. DOI: 10.1259/bjr/35012658.

[6] H. Menzel, H. Schibilla and D. Teunen. “European guidelines on quality criteria for computed tomography”. In: *Luxembourg: European Commission* (2000).

[7] A.-M. Šimundić. “Measures of Diagnostic Accuracy: Basic Definitions”. In: *Journal of the International Federation of Clinical Chemistry* 19.4 (Jan. 2009), pp. 203–211. ISSN: 1650-3414.

[8] T. Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (June 2006), pp. 861–874. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2005.10.010.

[9] L. G. Månsson. “Methods for the Evaluation of Image Quality: A Review”. In: *Radiation Protection Dosimetry* 90.1-2 (Aug. 2000), pp. 89–99. ISSN: 0144-8420, 1742-3406. DOI: 10.1093/oxfordjournals.rpd.a033149.

[10] Ö. Smedby and M. Fredrikson. “Visual grading regression: analysing data from visual grading experiments with regression models”. en. In: *The British Journal of Radiology* 83.993 (Sept. 2010), pp. 767–775. ISSN: 0007-1285. DOI: 10.1259/bjr/35254923.

[11] L. L. Thurstone. “A law of comparative judgment”. In: *Psychological Review* 34.4 (1927), pp. 273–286. DOI: 10.1037/h0070288.

[12] K. D. Hopper et al. “Analysis of interobserver and intraobserver variability in CT tumor measurements.” In: *American Journal of Roentgenology* 167.4 (Oct. 1996), pp. 851–854. ISSN: 0361-803X. DOI: 10.2214/ajr.167.4.8819370.

[13] W. M. Reed et al. “Can Prevalence Expectations Drive Radiologists’ Behavior?” In: *Academic Radiology* 21.4 (Apr. 2014), pp. 450–456. ISSN: 1076-6332. DOI: 10.1016/j.acra.2013.12.002.

[14] K. A. Chun et al. “Interobserver Variability in the Assessment of CT Imaging Features of Traumatic Brain Injury”. In: *Journal of Neurotrauma* 27.2 (Feb. 2010), pp. 325–330. ISSN: 1557-9042. DOI: 10.1089/neu.2009.1115.

[15] H. H. Barrett et al. “Model observers for assessment of image quality”. In: *Proceedings of the National Academy of Sciences* 90.21 (1993), pp. 9758–9765. ISSN: 0027-8424. DOI: 10.1073/pnas.90.21.9758.

[16] I. Hernandez-Giron et al. “Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms”. In: *The British Journal of Radiology* 87.1039 (July 2014), p. 20140014. ISSN: 0007-1285. DOI: 10.1259/bjr.20140014.

[17] D. Goodenough. “Catphan 500 and 600 manual”. In: *Greenwich, NY: The Phantom Laboratory* (2014).

[18] J. B. Solomon, O. Christianson and E. Samei. “Quantitative comparison of noise texture across CT scanners from different manufacturers”. In: *Medical Physics* 39.10 (Sept. 2012), pp. 6048–6055. ISSN: 2473-4209. DOI: 10.1118/1.4752209.

[19] W. Röntgen. “Über eine neue Art von Strahlen: vorläufige Mitteilung Sitzungsber”. In: *Phys. Med. Gesell.* (1895). DOI: 10.1007/978-3-662-13247-0.

[20] R. C. Kirkpatrick. “The New Photography: with Report of a Case in which a Bullet was Photographed in the Leg”. In: *Montreal Medical Journal* (1896). URL: https://archive.org/details/cihm_01784.

[21] C. L. Norton. “The X-rays in medicine and surgery”. In: *Science* 3.72 (1896), pp. 730–731. ISSN: 0036-8075. DOI: 10.1126/science.3.72.730.

[22] S. A. Bolliger et al. “Virtual autopsy using imaging: bridging radiologic and forensic sciences. A review of the Virtopsy and similar projects”. In: *European Radiology* 18.2 (Aug. 2007), pp. 273–282. ISSN: 1432-1084. DOI: 10.1007/s00330-007-0737-4.

[23] W. Schweitzer et al. “Virtopsy approach: Structured reporting versus free reporting for PMCT findings”. In: *Journal of Forensic Radiology and Imaging* 2.1 (Jan. 2014), pp. 28–33. ISSN: 2212-4780. DOI: 10.1016/j.jofri.2013.12.002.

[24] S. A. Bolliger and M. J. Thali. “Imaging and virtual autopsy: looking back and forward”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1674 (June 2015), p. 20140253. ISSN: 0962-8436. DOI: 10.1098/rstb.2014.0253.

[25] S. Grabherr et al. “Postmortem Angiography: Review of Former and Current Methods”. In: *American Journal of Roentgenology* 188.3 (Mar. 2007), pp. 832–838. ISSN: 0361-803X. DOI: 10.2214/ajr.06.0787.

[26] I. M. Wagensveld et al. “Total-body CT and MR features of postmortem change in in-hospital deaths”. In: *PLOS ONE* 12.9 (Sept. 2017). Ed. by A. Churg, e0185115. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0185115.

[27] J. Radon. “On the determination of functions from their integral values along certain manifolds”. In: *IEEE Transactions on Medical Imaging* 5.4 (Dec. 1986), pp. 170–176. ISSN: 1558-254X. DOI: 10.1109/tmi.1986.4307775.

[28] S. Deans. *The radon transform and some of its applications*. Mineola, NY: Dover Publications, 2007. ISBN: 9780486462417.

[29] L. A. Shepp and B. F. Logan. “The Fourier reconstruction of a head section”. In: *IEEE Transactions on Nuclear Science* 21.3 (June 1974), pp. 21–43. DOI: 10.1109/tns.1974.6499235.

[30] R. E. Alexander and R. B. Gunderman. “EMI and the First CT Scanner”. In: *Journal of the American College of Radiology* 7.10 (Oct. 2010), pp. 778–781. ISSN: 1546-1440. DOI: 10.1016/j.jacr.2010.06.003.

[31] E. C. Beckmann. “CT scanning the early days”. In: *The British Journal of Radiology* 79.937 (Jan. 2006), pp. 5–8. ISSN: 0007-1285. DOI: 10.1259/bjr/29444122.

[32] G. Di Chiro and R. Brooks. “The 1979 Nobel prize in physiology or medicine”. In: *Science* 206.4422 (1979), pp. 1060–1062. ISSN: 0036-8075. DOI: 10.1126/science.386516.

[33] R. A. Brooks and G. Di Chiro. “Theory of Image Reconstruction in Computed Tomography”. In: *Radiology* 117.3 (Dec. 1975), pp. 561–572. ISSN: 0033-8419. DOI: 10.1148/117.3.561.

[34] S. N. Friedman et al. “A simple approach to measure computed tomography (CT) modulation transfer function (MTF) and noise-power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom”. In: *Medical Physics* 40.5 (Apr. 2013), p. 051907. ISSN: 2473-4209. DOI: 10.1118/1.4800795.

[35] D. Dance et al. *Diagnostic Radiology Physics*. Vienna, Austria: International Atomic Energy Agency, 2014, p. 710. ISBN: 9789201310101. URL: http://www-pub.iaea.org/books/IAEABooks/8841/Diagnostic-Radiology-Physics.

[36] R. Gordon, R. Bender and G. T. Herman. “Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography”. In: *Journal of Theoretical Biology* 29.3 (Dec. 1970), pp. 471–481. ISSN: 0022-5193. DOI: 10.1016/0022-5193(70)90109-8.

[37] A. H. Andersen and A. C. Kak. “Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm”. In: *Ultrasonic Imaging* 6.1 (Jan. 1984), pp. 81–94. ISSN: 0161-7346. DOI: 10.1177/016173468400600107.

[38] K. Lange, M. Bahn and R. Little. “A Theoretical Study of Some Maximum Likelihood Algorithms for Emission and Transmission Tomography”. In: *IEEE Transactions on Medical Imaging* 6.2 (June 1987), pp. 106–114. ISSN: 1558-254X. DOI: 10.1109/tmi.1987.4307810.

[39] S. H. Manglos et al. “Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT”. In: *Physics in Medicine and Biology* 40.7 (July 1995), pp. 1225–1241. ISSN: 0031-9155. DOI: 10.1088/0031-9155/40/7/006.

[40] M. Beister, D. Kolditz and W. A. Kalender. “Iterative reconstruction methods in X-ray CT”. In: *Physica Medica* 28.2 (Apr. 2012), pp. 94–108. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2012.01.003.

[41] L. L. Geyer et al. “State of the Art: Iterative CT Reconstruction Techniques”. EN. In: *Radiology* 276.2 (Aug. 2015), pp. 339–57. ISSN: 1527-1315. DOI: 10.1148/radiol.2015132766.

[42] Z. Tian et al. “Low-dose CT reconstruction via edge-preserving total variation regularization”. In: *Physics in Medicine & Biology* 56.18 (2011), p. 5949. ISSN: 0031-9155. DOI: 10.1088/0031-9155/56/18/011.

[43] B. De Man et al. “An iterative maximum-likelihood polychromatic algorithm for CT”. In: *IEEE Transactions on Medical Imaging* 20.10 (Oct. 2001), pp. 999–1008. ISSN: 0278-0062. DOI: 10.1109/42.959297.

[44] G. Poludniowski et al. “An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT”. In: *Physics in Medicine & Biology* 54.12 (2009), p. 3847. ISSN: 0031-9155. DOI: 10.1088/0031-9155/54/12/016.

[45] C. Kamphuis and F. Beekman. “Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm”. In: *IEEE Transactions on Medical Imaging* 17.6 (1998), pp. 1101–1105. ISSN: 1558-254X. DOI: 10.1109/42.746730.

[46] J. F. Barrett and N. Keat. “Artifacts in CT: recognition and avoidance”. en. In: *Radiographics* 24.6 (Jan. 2004), pp. 1679–1691. ISSN: 0271-5333. DOI: 10.1148/rg.246045065.

[47] J. Hsieh. *Computed tomography : principles, design, artifacts, and recent advances*. Bellingham, WA: SPIE, 2015. ISBN: 9781628418255.

[48] S. H. GmbH. *Simulation of X-ray Spectra*. https://www.oem-products.siemens.com/x-ray-spectra-simulation. Accessed: 2016-10-6.

[49] R. A. Brooks and G. D. Chiro. “Beam hardening in X-ray reconstructive tomography”. In: *Physics in Medicine & Biology* 21.3 (1976), p. 390. ISSN: 0031-9155. DOI: 10.1088/0031-9155/21/3/004.

[50] M. Karçaaltıncaba and A. Aktaş. “Dual-energy CT revisited with multidetector CT: review of principles and clinical applications”. In: *Diagnostic and Interventional Radiology* 17.3 (Sept. 2011), pp. 181–194. ISSN: 1305-3612. DOI: 10.4261/1305-3825.DIR.3860-10.0.

[51] S. J. Riederer and C. A. Mistretta. “Selective iodine imaging using K-edge energies in computerized x-ray tomography”. In: *Medical Physics* 4.6 (Nov. 1977), pp. 474–481. ISSN: 2473-4209. DOI: 10.1118/1.594357.

[52] G. N. Hounsfield. “Computerized transverse axial scanning (tomography): Part 1. Description of system”. In: *The British Journal of Radiology* 46.552 (1973). PMID: 4757352, pp. 1016–1022. ISSN: 0007-1285. DOI: 10.1259/0007-1285-46-552-1016.

[53] B. Krauss, B. Schmidt and T. G. Flohr. “Dual Source CT”. In: *Dual Energy CT in Clinical Practice*. Ed. by T. Johnson et al. Heidelberg, Germany: Springer Berlin Heidelberg, 2011, pp. 11–20. ISBN: 9783642017407. DOI: 10.1007/174_2010_44.

[54] E. P. S. Sande et al. “Interphantom and interscanner variations for Hounsfield units–establishment of reference values for HU in a commercial QA phantom”. In: *Physics in Medicine & Biology* 55.17 (2010), p. 5123. ISSN: 0031-9155. DOI: 10.1088/0031-9155/55/17/015.

[55] R. Lamba et al. “CT Hounsfield Numbers of Soft Tissues on Unenhanced Abdominal CT Scans: Variability Between Two Different Manufacturers’ MDCT Scanners”. In: *American Journal of Roentgenology* 203.5 (Nov. 2014), pp. 1013–1020. ISSN: 0361-803X. DOI: 10.2214/ajr.12.10037.

[56] W. Cai et al. “Informatics in Radiology: Dual-Energy Electronic Cleansing for Fecal-Tagging CT Colonography”. In: *RadioGraphics* 33.3 (May 2013), pp. 891–912. ISSN: 0271-5333. DOI: 10.1148/rg.333125039.

[57] H. K. Genant and D. Boyd. “Quantitative bone mineral analysis using dual energy computed tomography”. In: *Investigative Radiology* 12.6 (1977), pp. 545–551. ISSN: 0020-9996.

[58] F. L. Roder. “Explosives Detection By Dual-Energy Computed Tomography (CT)”. In: ed. by R. P. Kruger. International Society for Optics and Photonics, Oct. 1979, pp. 171–178. DOI: 10.1117/12.957385.

[59] J. S. Iwanczyk et al. “Photon Counting Energy Dispersive Detector Arrays for X-ray Imaging”. In: *IEEE Transactions on Nuclear Science* 56.3 (June 2009), pp. 535–542. ISSN: 1558-1578. DOI: 10.1109/tns.2009.2013709.

[60] J. Robson. “Spatial and temporal contrast-sensitivity functions of the visual system”. In: *Journal of the Optical Society of America* 56.8 (1966), pp. 1141–1142. ISSN: 0030-3941. DOI: 10.1364/JOSA.56.001141.

[61] I. Ohzawa. *Make Your Own Campbell-Robson Contrast Sensitivity Chart*. Accessed: 2018-8-1. URL: http://ohzawa-lab.bpe.es.osaka-u.ac.jp/ohzawa-lab/izumi/CSF/A_JG_RobsonCSFchart.html.

[62] C. Chubb, G. Sperling and J. A. Solomon. “Texture interactions determine perceived contrast”. In: *Proceedings of the National Academy of Sciences* 86.23 (1989), pp. 9631–9635. ISSN: 0027-8424. DOI: 10.1073/pnas.86.23.9631.

[63] D. Borland and R. T. Ii. “Rainbow Color Map (Still) Considered Harmful”. In: *IEEE Computer Graphics and Applications* 27.2 (Mar. 2007), pp. 14–17. ISSN: 1558-1756. DOI: 10.1109/mcg.2007.323435.

[64] N. E. M. A. of Radiology and A. College. *Digital imaging and communications in medicine (DICOM)*. 1998. URL: http://dicom.nema.org/standard.html.

[65] M. R. Luo, G. Cui and C. Li. “Uniform colour spaces based on CIECAM02 colour appearance model”. In: *Color Research & Application* 31.4 (2006), pp. 320–330. ISSN: 1520-6378. DOI: 10.1002/col.20227.

[66] N. Moroney et al. “The CIECAM02 color appearance model”. In: *Color and Imaging Conference*. Vol. 2002. 1. Society for Imaging Science and Technology. 2002, pp. 23–27.

[67] G. M. Machado, M. M. Oliveira and L. A. F. Fernandes. “A Physiologically-based Model for Simulation of Color Vision Deficiency”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (Nov. 2009), pp. 1291–1298. ISSN: 1077-2626. DOI: 10.1109/TVCG.2009.113.

[68] H. J. M. Gerrits and A. J. H. Vendrik. “Simultaneous contrast, filling-in process and information processing in man’s visual system”. In: *Experimental Brain Research* 11.4 (Nov. 1970), pp. 411–430. ISSN: 1432-1106. DOI: 10.1007/BF00237914.

[69] Z.-L. Lu and G. Sperling. “Second-order illusions: Mach bands, chevreul, and Craik-O’Brien-Cornsweet”. In: *Vision Research* 36.4 (Feb. 1996), pp. 559–572. ISSN: 0042-6989. DOI: 10.1016/0042-6989(95)00139-5.

[70] M. S. Chesters. “Human visual perception and ROC methodology in medical imaging”. In: *Physics in Medicine & Biology* 37.7 (1992), p. 1433. ISSN: 0031-9155. DOI: 10.1088/0031-9155/37/7/001.

[71] J. Duran et al. “Collaborative Total Variation: A General Framework for Vectorial TV Models”. In: *SIAM Journal on Imaging Sciences* 9.1 (Jan. 2016), pp. 116–151. ISSN: 1936-4954. DOI: 10.1137/15m102873x.

[72] P. Perona and J. Malik. “Scale-space and edge detection using anisotropic diffusion”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.7 (July 1990), pp. 629–639. ISSN: 0162-8828. DOI: 10.1109/34.56205.

[73] C. Tomasi and R. Manduchi. “Bilateral filtering for gray and color images”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Jan. 1998, pp. 839–846. ISBN: 81-7319-221-9. DOI: 10.1109/ICCV.1998.710815.

[74] K. He, J. Sun and X. Tang. “Guided Image Filtering”. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 35 (June 2013), pp. 1397–1409. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.213.

[75] S. Pizer et al. “Adapative histogram equalization and its variations”. English (US). In: *Computer Vision, Graphics, and Image Processing* 39.3 (Jan. 1987), pp. 355–368. ISSN: 0734-189X. DOI: 10.1016/S0734-189X(87)80186-X.

[76] K. Zuiderveld. “Graphics Gems IV”. In: ed. by P. S. Heckbert. San Diego, CA: Academic Press Professional, Inc., 1994. Chap. Contrast Limited Adaptive Histogram Equalization, pp. 474–485. ISBN: 0123361559. URL: http://dl.acm.org/citation.cfm?id=180895.180940.

[77] Y.-T. Kim. “Contrast enhancement using brightness preserving bi-histogram equalization”. In: *IEEE Transactions on Consumer Electronics* 43.1 (Feb. 1997), pp. 1–8. ISSN: 0098-3063. DOI: 10.1109/30.580378.

[78] G. Eilertsen. “The high dynamic range imaging pipeline : Tone-mapping, distribution, and single-exposure reconstruction”. PhD thesis. Linköping, Sweden: Linköping UniversityLinköping University, Media and Information Technology, Faculty of Science & Engineering, 2018, p. 132. ISBN: 9789176853023. DOI: 10.3384/diss.diva-147843.

[79] Z. Farbman et al. “Edge-preserving Decompositions for Multi-scale Tone and Detail Manipulation”. In: *ACM Trans. Graph.* 27.3 (Aug. 2008), 67:1–67:10. ISSN: 0730-0301. DOI: 10.1145/1360612.1360666.

[80] L. Xu et al. “Image Smoothing via L0 Gradient Minimization”. In: *ACM Trans. Graph.* 30.6 (Dec. 2011), 174:1–174:12. ISSN: 0730-0301. DOI: 10.1145/2070781.2024208.

[81] T. Acharya and A. K. Ray. *Image processing: principles and applications*. Hoboken, New Jersey, USA: John Wiley & Sons, 2005. ISBN: 978-0-471-71998-4.

[82] T. O. Aydin, K. Mantiuk Rafałand Myszkowski and H.-P. Seidel. “Dynamic Range Independent Image Quality Assessment”. In: *ACM Trans. Graph.* 27.3 (Aug. 2008), 69:1–69:10. ISSN: 0730-0301. DOI: 10.1145/1360612.1360668.

[83] K. Ma et al. “High Dynamic Range Image Compression by Optimizing Tone Mapped Image Quality Index”. In: *IEEE Transactions on Image Processing* 24.10 (Oct. 2015), pp. 3086–3097. ISSN: 1941-0042. DOI: 10.1109/tip.2015.2436340.

[84] H. Z. Nafchi et al. “FSITM: A Feature Similarity Index For Tone-Mapped Images”. In: *IEEE Signal Processing Letters* 22 (8 Aug. 2015), pp. 1026–1029. ISSN: 1558-2361. DOI: 10.1109/LSP.2014.2381458.

[85] M. Michael J. Thali, M. D. Viner and B. Brogdon. *Brogdon’s Forensic Radiology, Second Edition*. Boca Raton, FL: CRC Press, Nov. 2010. ISBN: 9781420075625.

[86] E. Seeram. *Computed Tomography - E-Book: Physical Principles, Clinical Applications, and Quality Control*. St. Louis, MO, USA: Elsevier, 2013. ISBN: 9780323277518.

[87] D. Völgyes. *PerceptionMD*. Zenodo. July 2018. DOI: 10.5281/zenodo.1324409.

[88] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2009. ISBN: 978-1441412690.

[89] B. N. Miller and D. L. Ranum. “Beyond PDF and ePub: Toward an Interactive Textbook”. In: *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education*. ITiCSE ’12. Haifa, Israel: ACM, 2012, pp. 150–155. ISBN: 9781450312462. DOI: 10.1145/2325296.2325335.

[90] D. Völgyes. *DCM2HDR: DICOM to HDR image conversion*. Zenodo. June 2018. DOI: 10.5281/zenodo.1290825.

[91] D. Völgyes. *Zenodo-get: a downloader for Zenodo records*. Zenodo. June 2018. DOI: 10.5281/zenodo.1261813.

[92] D. Völgyes. *Dual energy CT scan of ordinary objects*. Zenodo. June 2018. DOI: 10.5281/zenodo.1253035.

[93] D. Völgyes. *HDRIHaven: Public domain high dynamic range image dataset*. Zenodo. June 2018. DOI: 10.5281/zenodo.1285800.

[94] Y. Funama et al. “Radiation Dose Reduction without Degradation of Low-Contrast Detectability at Abdominal Multisection CT with a Low-Tube Voltage Technique: Phantom Study”. In: *Radiology* 237.3 (2005), pp. 905–910. ISSN: 1527-1315. DOI: 10.1148/radiol.2373041643.

[95] M. J. Siegel et al. “Radiation Dose and Image Quality in Pediatric CT: Effect of Technical Factors and Phantom Size and Shape”. In: *Radiology* 233.2 (2004), pp. 515–522. ISSN: 1527-1315. DOI: 10.1148/radiol.2332032107.

[96] K. Gulliksrud, C. Stokke and A. C. T. Martinsen. “How to measure CT image quality: Variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom”. In: *Physica Medica* 30.4 (June 2014), pp. 521–526. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2014.01.006.

[97] J. A. Seibert. “Tradeoffs between image quality and dose”. In: *Pediatric Radiology* 34.S3 (Oct. 2004), S183–S195. ISSN: 0301-0449. DOI: 10.1007/s00247-004-1268-7.

[98] H. Ellen et al. “100 days with scans of the same Catphan phantom on the same CT scanner”. In: *Journal of Applied Clinical Medical Physics* 18.6 (), pp. 224–231. ISSN: 1526-9914. DOI: 10.1002/acm2.12186.

[99] K. Tang et al. “Effect of Low Tube Voltage on Image Quality, Radiation Dose, and Low-Contrast Detectability at Abdominal Multidetector CT: Phantom Study”. In: *Journal of Biomedicine and Biotechnology* 2012 (2012), pp. 1–6. ISSN: 1110-7251. DOI: 10.1155/2012/130169.

[100] J. Greffier et al. “Which dose for what image? Iterative reconstruction for CT scan”. In: *Diagnostic and Interventional Imaging* 94.11 (Nov. 2013), pp. 1117–1121. ISSN: 2211-5706. DOI: 10.1016/j.diii.2013.03.008.

[101] A. D. Hardie et al. “What is the preferred strength setting of the sinogram-affirmed iterative reconstruction algorithm in abdominal CT imaging?” In: *Radiological Physics and Technology* 8.1 (Jan. 2015), pp. 60–63. ISSN: 1865-0341. DOI: 10.1007/s12194-014-0288-8.

[102] M. E. Baker et al. “Contrast-to-noise ratio and low-contrast object resolution on full- and low-dose MDCT: SAFIRE versus filtered back projection in a low-contrast object phantom and in the liver”. In: *American Journal of Roentgenology* 199.1 (July 2012), pp. 8–18. ISSN: 0361-803X. DOI: 10.2214/ajr.11.7421.

[103] A. De Crop et al. “Correlation of clinical and physical-technical image quality in chest CT: a human cadaver study applied on iterative reconstruction.” In: *BMC medical imaging* 15.1 (Jan. 2015), p. 32. ISSN: 1471-2342. DOI: 10.1186/s12880-015-0075-y.

[104] A. Padole et al. “CT Radiation Dose and Iterative Reconstruction Techniques”. en. In: *American Journal of Roentgenology* (Mar. 2015). ISSN: 0361-803X. DOI: 10.2214/AJR.14.13241.

[105] D. Gascho, M. J. Thali and T. Niemann. “Post-mortem computed tomography: Technical principles and recommended parameter settings for high-resolution imaging”. In: *Medicine, Science and the Law* 58.1 (Jan. 2018), pp. 70–82. ISSN: 2042-1818. DOI: 10.1177/0025802417747167.

[106] H. Yeganeh and Z. Wang. “Objective Quality Assessment of Tone-Mapped Images”. In: *IEEE Transactions on Image Processing* 22.2 (Feb. 2013), pp. 657–667. ISSN: 1941-0042. DOI: 10.1109/tip.2012.2221725.

[107] N. Nikvand, H. Yeganeh and Z. Wang. “Adaptive windowing for optimal visualization of medical images based on normalized information distance”. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2014. DOI: 10.1109/icassp.2014.6853787.

[108] R. Fattal, D. Lischinski and M. Werman. “Gradient domain high dynamic range compression”. In: *Proceedings of the 29th annual conference on Computer graphics and interactive techniques - SIGGRAPH’02*. ACM Press, 2002. DOI: 10.1145/566570.566573.

[109] C.-K. Tran, C.-D. Tseng and T.-F. Lee. “Improving Diagnostic Viewing of Region of Interest in Lung Computed Tomography Image Using Unsharp Masking and Singular Value Decomposition”. In: *2016 3rd International Conference on Green Technology and Sustainable Development (GTSD)*. IEEE, Nov. 2016. DOI: 10.1109/gtsd.2016.11.

[110] E. D. Pisano et al. “Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms”. In: *Journal of Digital Imaging* 11.4 (Nov. 1998), p. 193. ISSN: 1618-727X. DOI: 10.1007/BF03178082.

[111] *Comparison of clinical findings between intensity-windowed versus CLAHE presentations of chest CT images*. Vol. 1653. 1992, p. 13. DOI: 10.1117/12.59496.

[112] K. M. Schartz et al. “Trauma in CT: The Role of Severe Injury on Satisfaction of Search Revised”. In: *Journal of the American College of Radiology* 13.8 (Aug. 2016), 973–978.e4. ISSN: 1546-1440. DOI: 10.1016/j.jacr.2016.04.014.

[113] T. Chan, A. Marquina and P. Mulet. “High-Order Total Variation-Based Image Restoration”. In: *SIAM Journal on Scientific Computing* 22.2 (Jan. 2000), pp. 503–516. DOI: 10.1137/s1064827598344169.

[114] A. Buades, B. Coll and J. M. Morel. “The staircasing effect in neighborhood filters and its solution”. In: *IEEE Trans Image Process* 15.6 (June 2006), pp. 1499–1505. ISSN: 1941-0042. DOI: 10.1109/TIP.2006.871137.

[115] L. N. Trefethen. *Approximation Theory and Approximation Practice*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2012. ISBN: 9781611972399.

[116] R. Palomar et al. “A novel method for planning liver resections using deformable Bézier surfaces and distance maps”. In: *Computer Methods and Programs in Biomedicine* 144 (2017), pp. 135–145. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2017.03.019.

[117] M. Schmidt and H. Lipson. “Distilling Free-Form Natural Laws from Experimental Data”. In: *Science* 324.5923 (2009), pp. 81–85. ISSN: 0036-8075. DOI: 10.1126/science.1165893.

[118] T. McConaghy. “FFX: Fast, Scalable, Deterministic Symbolic Regression Technology”. In: *Genetic Programming Theory and Practice IX*. Ed. by R. Riolo, E. Vladislavleva and J. H. Moore. New York, NY: Springer New York, 2011, pp. 235–260. ISBN: 978-1-4614-1770-5. DOI: 10.1007/978-1-4614-1770-5_13.

List of Figures

List of Tables

Original articles

This chapter consists the reproduction of the articles which this thesis is built on. They are referred in the text as Paper I to IV.

[1] D. Völgyes et al. “How Different Iterative and Filtered Back Projection Kernels Affect Computed Tomography Numbers and Low Contrast Detectability”. In: *Journal of Computer Assisted Tomography* 00.00 (2016), p. 1. ISSN: 0363-8715. DOI: 10.1097/RCT.0000000000000491.

[2] D. Völgyes et al. “Applicability of a clinical cardiac CT protocol in post mortem studies”. In: *Journal of Forensic Radiology and Imaging* 12 (2018), pp. 25–30. ISSN: 2212-4780. DOI: 10.1016/j.jofri.2018.01.003.

[3] D. Völgyes et al. “A Weighted Histogram-Based Tone Mapping Algorithm for CT Images”. In: *Algorithms* 11.8 (July 2018), p. 111. ISSN: 1999-4893. DOI: 10.3390/a11080111.

[4] D. Völgyes et al. “Image De-Quantization Using Plate Bending Model”. In: *Algorithms* 11.8 (July 2018), p. 110. ISSN: 1999-4893. DOI: 10.3390/a11080110.

PAPER I

OPEN

How Different Iterative and Filtered Back Projection Kernels Affect Computed Tomography Numbers and Low Contrast Detectability

David Völgyes, MSc, Marius Pedersen, PhD,* Arne Stray-Pedersen, PhD, MD,†‡ Dag Waaler, PhD,§ and Anne Catrine Trægde Martinsen, PhD||¶*

Objective: The aim of this study was to evaluate how different iterative and filtered back projection kernels affect the computed tomography (CT) numbers and low contrast detectability.

Methods: Five different scans were performed at 6 different tube potentials on the same Catphan 600 phantom using approximately the same dose level and otherwise identical settings. The scans were reconstructed using all available filtered back projection body kernels and with iterative reconstruction techniques.

Results: The CT numbers and the contrast-to-noise ratios were reported and how they are affected by the kernel choice and strength of iterative reconstruction.

Conclusions: Iterative reconstruction improved contrast-to-noise ratio in most cases, but in certain situations, it decreased it. Variations in CT numbers can be large between kernels with similar sharpness for certain densities.

Key Words: CT, mean CT numbers, contrast-to-noise ratio, image quality

(*J Comput Assist Tomogr* 2016;00: 00–00)

Computed tomography (CT) image quality is essential for high-quality diagnostics. Soon after the first commercial CT scanners had appeared, the need for standardized quality assurance (QA) tests arose.[1] The main goal of the QA measurements is to ensure that CT image quality and dose are in agreement with specifications and international recommendations. Subjective evaluations with expert readers supplement these objective criteria, but due to the time-consuming nature of these evaluations, they are mostly used in optimization of image quality for one specific examination protocol.[2–4]

Diagnostic image quality cannot be fully assessed without the knowledge of the anatomical area of interest and pathology to be searched for. Due to that, many different anatomical phantoms have been developed, such as cardiac, liver, lung, thorax phantoms, among others.[5] These are more or less anthropomorphic, with difference in texture, density, size, and complexity. Reading conditions, such as ambient lighting or display window settings used, also affect the reader's performance, and thus diagnostic quality.

However, there are basic criteria which must be fulfilled by all CT scanners. Therefore, some general purpose image quality QA phantoms and test methods have been developed. These are meant to be used for daily, weekly, and longer term QA tests. One of the most widely used such phantom is the Catphan 600 (Phantom Laboratories, New York), which has a modular structure.[1] Each module was designed for specific measurements. Performing well with this phantom is required, but not sufficient per se for accurate diagnostics.

Dose reduction and image quality are currently in the center of research. This research typically targets a few selected kernels or few selected parameters, and optimize these for dedicated diagnostic purposes.[6–10] Iterative reconstructions (IRs) have just started to gain wider acceptance in clinical practice in recent years, with a research focus on specific applications with respect to increased diagnostic accuracy and/or dose reduction.[11,12] In contrast, this article focuses on the basic criteria and evaluates them in large number of combinations of scan and reconstruction parameters. This exhaustive approach also allows discovering rare effects which are otherwise easily overlooked.

MATERIALS AND METHODS

The Catphan 600 phantom is a widely used general purpose phantom for CT image quality evaluation.[1] It is a modular phantom where individual modules are used for specific tests. In our work, the CTP404 module was used for linearity tests to measure mean CT numbers, and the CTP515 module for contrast-to-noise measurements. The 2 modules and the performed measurements are depicted in Figures 1 and 2, respectively.

In the experiment, the same phantom was scanned 5 times with the exact same scanning parameters except for 2 parameters: peak tube voltage [kilovolt (peak), kV(p)] and effective tube current (mAs). Effective mAs was selected to produce the same CTDIvol dose level (10.0 mGy) for all of the measurements. The peak voltages used were 70, 80, 100, 120, and 140 kV(p) with 280, 613, 292, 178, and 122 mAs tube current, respectively. The common parameters are presented in Table 1. The applied dose was close to the level normally used for abdominal CT with IR. The potential for dose reduction using IR algorithms was not evaluated in this study.

From the *Faculty of Computer Science and Media Technology, Gjøvik University College, Gjøvik; †Department of Research and Development in Forensic Pathology, The Norwegian Institute of Public Health; ‡Department of Forensic Medicine, University of Oslo, Oslo; §Faculty of Health, Care and Nursing, Gjøvik University College, Gjøvik; ||The Intervention Centre, Oslo University Hospital Rikshospitalet; and ¶Department of Physics, University of Oslo, Oslo, Norway.

Received for publication December 22, 2015; accepted May 23, 2016.

Correspondence to: David Völgyes, MSc, Ragnhild Schibbyes vei 16, 0968, Oslo, Norway (e-mail: david.volgyes@hig.no).

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jcat.org).

This research has been funded by the Research Council of Norway through project no. 221073 "HyPerCept—Colour and quality in higher dimensions."

The authors declare no conflict of interest.

DOI: 10.1097/RCT.0000000000000491

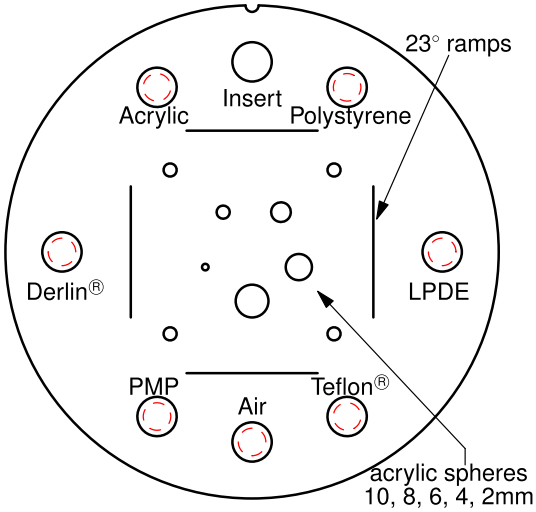


FIGURE 1. Main components of the CTP404 module. CT numbers are measured in 7 positions. Figure 1 can be viewed online in color at www.jcat.org.

CT Scanner and Reconstruction Kernels

All scans were performed on a Siemens Somatom Definition Flash dual-source multi-slice CT scanner (Siemens AG, Forchheim, Germany; http://www.healthcare.siemens.com/computed-tomography/dual-source-ct/somatom-definition-flash). The scanner provides both application specific and general purpose reconstruction kernels. A summary of the kernels is presented in Table 2. In this article, we will refer to filtered back projection (FBP) and the related IRs as a kernel family or reconstruction family. Some FBP kernels have no iterative counterpart (Table 2), and some of them (B22f, B23f) are not present at 70 kV(p).

In this study, the general body kernels were investigated both with FBP and IR where IR was available. The IR is called Sinogram Affirmed Iterative Reconstruction (SAFIRE), and it operates in the projection domain in addition to the image domain to reduce noise and artifacts. The parameter choice of SAFIRE is its *level* or *strength*, which can be varied from 1 to 5.

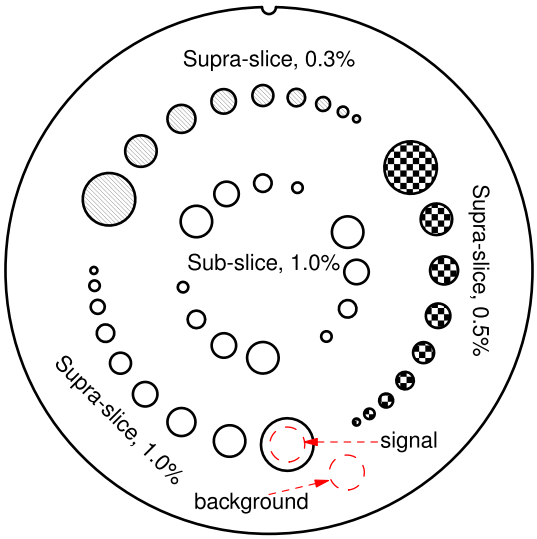


FIGURE 2. Main components of the CTP515 module. Contrast-to-noise measurement is performed using 2 ROIs, one in the largest 1% cylinder and one close to it in the background. Figure 2 can be viewed online in color at www.jcat.org.

TABLE 1. Scan and Reconstruction Parameters

CTDIvol	10.0 mGy
Data collection diameter	500 mm
Reconstruction diameter	300 mm
Single collimation width	0.6 mm
Total collimation width	38.4 mm
Pitch	0.6 mm
Rotation speed	0.5 s
No. axial slices	103
Slice thickness	2.0 mm
Matrix size	512 × 512
In plane voxel size	0.586 mm

Minimizing External Effects

All of the measurements were performed in the exact same patient-table positions without modifying anything in the setup except the tube voltages which yielded one scan for every tube voltage. Therefore, at a given tube voltage, every reconstruction used the exact same raw data. Both positioning and inter-phantom differences were supposed to be eliminated with this approach.

Linearity

One of the most important image quality features is CT number linearity. Computed tomography images are graphical

TABLE 2. Reconstruction Kernels

Kernel Family	FBP	IR	Sharpness, Purpose
10	B10f	—	Very smooth
20	B20f	—	Smooth
22	B22f	—	Smooth, quantitative
23	B23f	—	Smooth, quantitative with dedicated iodine beam hardening correction
26	B26f	I26	HeartView smooth
30	B30f	I30	Medium smooth
31	B31f	I31	Medium smooth+
35	B35f	—	HeartView medium
36	B36f	I36	HeartView medium
40	B40f	I40	Medium
41	B41f	I41	Medium+
45	B45f	—	Medium, designed for Ca-scoring
46	B46f	I46	HeartView sharp
50	B50f	I50	Medium sharp
60	B60f	—	Sharp
70	B70f	I70	Very sharp
75	B75f	—	Very sharp
80	B80f	—	Ultra sharp

TABLE 3. Attenuation Values for Phantom Materials and Typical Tissues

Material or Tissue	Attenuation, HU
Air	−1046:−986
PMP	−220:−172
LDPE	−121:−87
Water	−7:7
Polystyrene	−65:−29
Acrylic	92:137
Delrin	344:387
Teflon	941:1060
Lung	−600:−400
Fat	−100:−60
Soft tissue	40:80
Bone	400:1000

representation of the linear x-ray attenuation coefficient (μ) of an object. Computed tomography numbers are measured in Hounsfield units (HU), where HU for water is 0, and HU for air is −1000. However, the relation between CT numbers and μ is not unambiguous.

A scanner can map the same physical object into slightly different CT numbers depending for instance on the spectrum of the x-ray tube, reconstruction kernel, or correction algorithms such as a dedicated beam hardening correction. Computed tomography numbers are sometimes directly used in diagnostics; therefore, it is of utmost importance that these values are accurate. If 2 kernels give different mean CT numbers for the same area, then any comparative study between them should take this fact into account.

The CTP404 module (Fig. 1) of Catphan 600 phantom contains 6 cylindrical inserts filled with solid materials and 1 with air as reference materials, and an optional water insert which was not used in this study. The reference materials [acrylic, polystyrene, low density polyethylene (LDPE), polymethylpentene (PMP), Delrin (DuPont's registered trademark), and Teflon (DuPont's registered trademark)] were selected to ensure that, for the most important density regions, the CT scanner produces the expected image. The references objects are cylinders with 10-mm diameter. To avoid edge effects radially, a centered circular region of interest (ROI) with 5-mm diameter were used for the measurements. Along the axial direction, measurements from 7 slices (2 mm each) were used to reduce the effect of statistical fluctuations. The ROI of the measurements consisted of 399 voxels. Nominal values are reported in Table 3 for the phantom materials and for some similar typical tissues, based on the phantom reference manual[13] and Holmes et al.[14]

Contrast-to-Noise Ratio

Low contrast detectability is an important CT image quality descriptor. The CTP515 module supports both psychophysical tests and numerical comparisons. The numerical evaluation requires the calculation of contrast-to-noise ratio (CNR). Many different definitions exist and any of them can be used as long as it is used for relative comparisons. Contrast-to-noise ratio is calculated as follows[15]:

$$\mathrm{CNR} = \frac{2(S_A - S_B)^2}{\sigma_A^2 + \sigma_B^2}$$

Here S_A and S_B are the mean values for signals with 2 ROIs, and σ_A^2 and σ_B^2 are the variances of these signals, respectively.

The CNR value for each reconstruction was calculated from the same ROI pixels using the low contrast detectability cylinder with the largest diameter (15 mm) with 1% (10 HU) density as the signal, and a corresponding region of the same size just outside as the background. The central 10 mm area in 7 slices (2 mm each) was used for the measurement, which comprised 1575 voxels. The setup is depicted in Figure 2.

Data Collection

Each reconstruction provided 8 data points: 7 mean CT numbers for different materials, and 1 CNR. The materials had a nominal CT density to which the measurements should be compared during a QA test. Deviation from the nominal values should be in a certain range.[13] The available combinations of tube voltages and kernels (FBP and SAFIRE) yielded 313 reconstructions, which resulted in 2504 data points for CT numbers and CNR.

RESULTS

Linearity

Figures 3 to 9 visualize the measured mean CT numbers in the ROIs. The coloring in these figures is aimed to visualize the distribution of the negative (blue), close to zero (gray), and positive (red) relative differences. It is clear from these figures that there are kernels with similar properties. The kernels 22, 31, 35, 36, 41, 45, 46, and 50 are all within the ±3 HU range of the reference kernel for which kernel 36 was chosen. This group can be extended with kernels 60, 70, and 75, if the range is increased to ±4 HU, and air measurements are excluded.

Also, a second, smaller, weaker group of kernels can be identified. The core of this group consists of kernels 26 and 30 where the reported mean values differ in less than ±1 HU. Kernel 40 shows similar properties for medium attenuations (±2.5 HU) but for Teflon and air the differences are larger, up to 8.1 HU. Summarizing these results, the 2 groups consist of these reconstructions, where parenthesis shows the eased conditions for group 1:

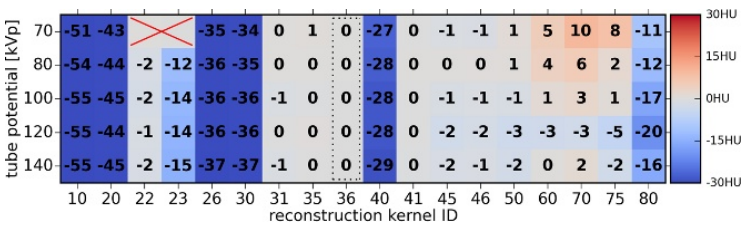


FIGURE 3. Relative CT number differences for air. The values are relative to the B36f kernel at the given energy. B22f and B23f kernels cannot be used for 70-kV(p) scans. Figure 3 can be viewed online in color at www.jcat.org.

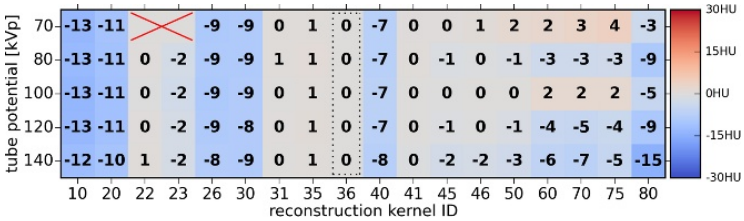


FIGURE 4. Relative CT number differences for PMP material relative to the B36f kernel. Figure 4 can be viewed online in color at www.jcat.org.

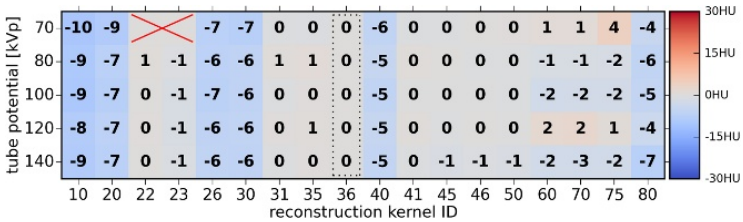


FIGURE 5. Relative CT number differences for LDPE material relative to the B36f kernel. Figure 5 can be viewed online in color at www.jcat.org.

- 22, 31, 35, 36, 41, 45, 46, 50, (60, 70, 75)
- 26, 30, 40

The CT number uniformity tests allow a ±4 HU range for water or water equivalent material only.[2] This kind of uniformity is not required for air and dense materials where noise and artifacts might play significant roles. The previously mentioned 2 groups yield lower inhomogeneities for medium dense materials than the requirement for homogeneity. However, if a material has to be excluded (air, Teflon) although it is not tested in standard homogeneity tests, then it shows a weaker connection.

The rest of the kernels (10, 20, 23, and 80) sometimes show similarities to other kernels but not strong enough to associate them with one of the groups. The unique behavior of kernel 23 can be explained by the fact that it applies a dedicated iodine beam hardening correction.

Low Contrast Detectability

One of the important measures in low contrast detectability is CNR. In general, IRs reduce the noise level and improve CNR. This does not, however, necessarily improve the diagnostic image quality, and often a medium noise suppression is preferred.[16]

Three relations were examined as follows:

- correlation between CNR and SAFIRE level for a given kernel,
- CNR as function of SAFIRE level at given tube voltages,
- relative CNR improvement as function of SAFIRE level at given tube voltages.

First, sharper kernels benefited relatively more from IR, as it is demonstrated for 120 kV(p) in Figure 10. Other tube voltages produced similar results. Second, 2 of the 3 cardiac kernels (26, 36) show lower CNR for low level IR (SAFIRE 1) than for FBP. This was found for all tube voltages. Figure 11 depicts kernel 26 with this strange result, and Figure 12 shows the general case. Figure 13 shows all kernels with IR at 120 kV(p). Third, the same 2 kernels (26, 36) produced lower CNR values with IR than their slightly sharper general body kernel versions. This means that, while with FBP, kernel 26 is smoother than kernel 30; 36 is smoother than 40 and 41. This reverses at SAFIRE 3, kernel 30 became smoother than 26 whereas kernel 41 and 40 became smoother than 36 (Fig. 11). [CNR figures of the remaining kernels and peak tube voltages are present online as Supplemental Digital Content (see Figures, Supplemental Digital Content 1, http://links.lww.com/RCT/A56).]

DISCUSSION

CT Numbers

Reported mean CT numbers might be affected by the size and the position of the ROIs. One particular example is beam-hardening

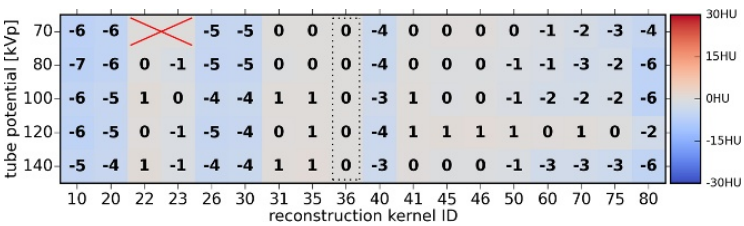


FIGURE 6. Relative CT number differences for polystyrene material relative to the B36f kernel. Figure 6 can be viewed online in color at www.jcat.org.

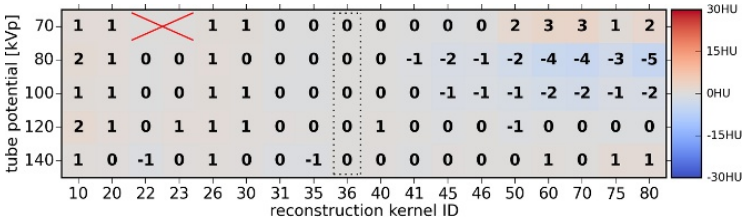


FIGURE 7. Relative CT number differences for acrylic material relative to the B36f kernel. Figure 7 can be viewed online in color at www.jcat.org.

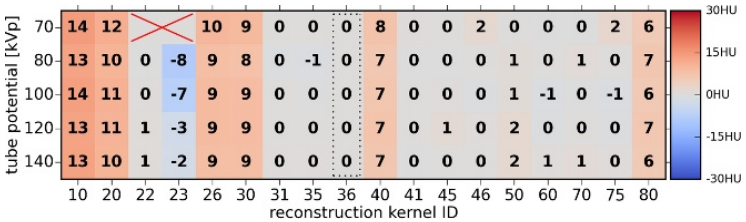


FIGURE 8. Relative CT number differences for Delrin material relative to the B36f kernel. Figure 8 can be viewed online in color at www.jcat.org.

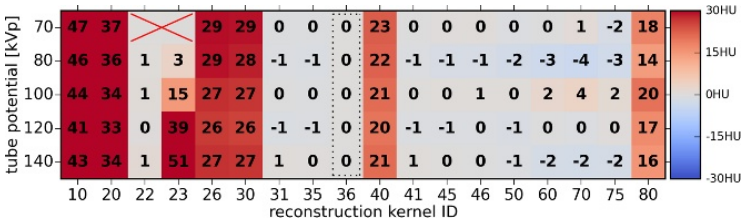


FIGURE 9. Relative CT number differences for Teflon material relative to the B36f kernel. Figure 9 can be viewed online in color at www.jcat.org.

artifact, which is a low-frequency artifact, and thus its appearance is affected by the low-frequency part of the modulation transfer function. Beam hardening is also sensitive to the x-ray spectrum, peak tube voltages, and patient size, among other factors. Therefore, linearity check alone cannot claim equivalence of 2 kernels. However, it is enough to claim that 2 kernels produce different results with a ROI size used in QA tests. The relatively small phantom size (20-cm diameter) and the fact that the results show no clear pattern when tube potential changes, imply that beam hardening is not the reason for the presented results. Images of some selected slices are presented in the Supplemental Digital Content (see Figures, Supplemental Digital Content 1, http://links.lww.com/RCT/A56).

There are 3 cases where similar behavior can be assumed for the kernels, see the kernel overview in Table 2. The cardiac kernels (26, 36, and 46) are different, not only in sharpness, but also kernel 26 yields different CT numbers from the 36's and 46's results. According to the application guide, kernels 30 and 31 (medium smooth and medium smooth+ kernels) should have the same visual sharpness, although with a slightly different noise structure. Therefore, it could be assumed, incorrectly, that they also produce similar mean CT numbers. The exact same pattern repeats with kernels 40 and 41 (medium and medium+ kernels).

These differences among the kernels should be taken into account during protocol optimization because they might affect the HU values and potentially the diagnostics. The assumption that kernels with similar purpose (eg, cardiac, same sharpness but different noise structure) yield similar mean CT numbers can be misleading.

In this study, no effect on the mean HU numbers was seen for SAFIRE compared to the corresponding FBP kernels. This is in accordance with other studies reporting for some selected kernels that they are not significantly affected by SAFIRE.[17]

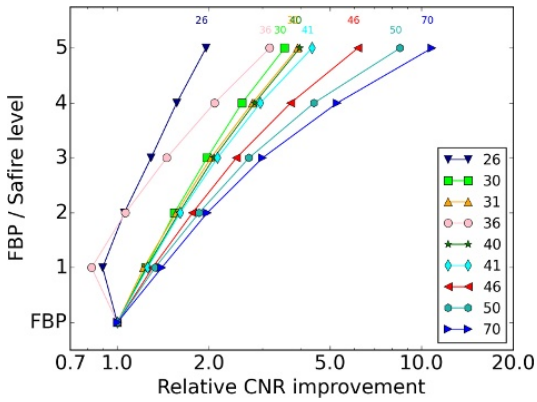


FIGURE 10. Relative CNR changes for various kernels at 120 kV(p). Figure 10 can be viewed online in color at www.jcat.org.

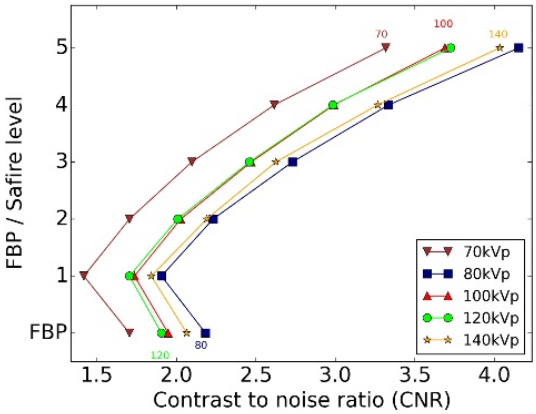


FIGURE 11. Contrast-to-noise ratio for B26f/I26 kernel at different tube potentials. Figure 11 can be viewed online in color at www.jcat.org.

CNR and IR Level

It is important to note that CNR depends on dose level, patient size, tube voltages, among other factors, and therefore, the findings in this article might not be universal. Sharper kernels produce higher noise levels, and the main advantage of IRs is the reduction of the noise level. This implies that the sharper kernels benefit more from IRs. However, the CNR decrease for the relatively smooth kernels 26 and 36 at SAFIRE 1 is unexpected. The CNR decrease could originate from signal change or from noise level change. In both cases, the cause of the lower CNR was consistently the higher noise level at SAFIRE 1. For these 2 kernels, changes in the CNR changes are decomposed into a signal and a noise part in the Supplemental Digital Content (see Figures, Supplemental Digital Content 1, http://links.lww.com/RCT/A56).

If one kernel is smoother and has higher CNR than another one (eg, B26f and B36f), then this order is expected to remain unchanged even if IR is applied (eg, I26 and I36 at SAFIRE 5). The mentioned break in the CNR curves invalidates this assumption for kernels 26 and 36. However, for the rest of the kernels, the assumption remains true. Contrast-to-noise ratio is only one of the

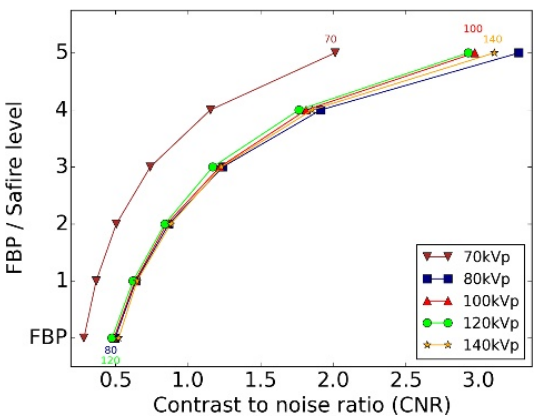


FIGURE 12. Contrast-to-noise ratio for B46f/I46 kernel at different tube potentials. Figure 12 can be viewed online in color at www.jcat.org.

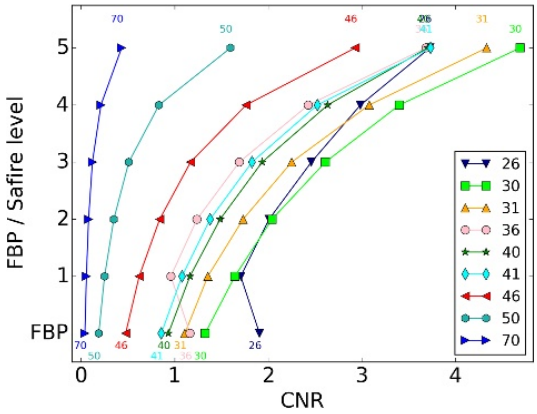


FIGURE 13. Contrast-to-noise ratio for various kernels at 120 kV(p). Figure 13 can be viewed online in color at www.jcat.org.

many aspects of image quality. Despite the lower CNR results, these kernels might remain favorable for specific applications due to, for example, their different noise structure, but caution is recommended.

CONCLUSIONS

Unexpected results were found both for mean CT numbers and for low contrast detectability. Although kernels are manufacturer-specific, the conclusion is general: even with widely used kernels, differences can easily be overlooked. Therefore, any protocol optimization effort should devote extra attention to this detail.

This study shows that for kernels normally used for soft tissue, the HU values will be minimally shifted for tissue densities close to zero. The HU shifts were, however, observed for tissue densities in the higher and lower part of the HU scale. The results show that it is important that radiologists use absolute HU values with care for diagnostic purposes.

It is worth mentioning that the unexpected results are related to the arguable most frequently used medium soft and medium kernels (26, 36, 30, 31, 40, and 41). Future work should settle the question whether modulation transfer function and noise power spectrum provide any similar results. Dose level, patient or phantom size, material-specific modulation transfer function, and kernel-specific correction algorithms make the optimization task even more complex.

REFERENCES

1. Goodenough DJ. Development of a phantom for evaluation and assurance of image quality in CT scanning. *Opt Eng*. 1977; 16:160152.
2. International Electrotechnical Commission (IEC). IEC 61223-3-5: Evaluation and routine testing in medical imaging departments—Part 3–5: Acceptance tests—Imaging performance of computed tomography X-ray equipment. 1.0. 2004:65.
3. International Electrotechnical Commission (IEC). Evaluation and routine testing in medical imaging departments. Part 2-6: Constancy tests – Imaging performance of computed tomography X-ray equipment. IEC publication No. 61223-2-6. Second edition. Tech. rep. IEC 61223-2-6. 2nd ed. Geneva: IEC. 2006;1–67:2006.
4. Hiles PA, Mackenzie A, Scally A, et al. *Recommended Standards for the Routine Performance Testing of Diagnostic X-ray Imaging Systems:*

Report 91. 2nd ed. York, UK: Institute of Physics and Engineering in Medicine; 2005.

5. DeWerd LA, Kissick M. *The Phantoms of Medical and Health Physics: Devices for Research and Development*. Biological and Medical Physics. New York: Springer; 2014.
6. Fernandez A, Greffier J, Langard E, et al. Database to CT scan to reduce doses with iterative reconstructions (SAFIRE). *Phys Med*. 2013;29:e12.
7. Sieren JP, Hoffman EA, Fuld MK, et al. Sinogram Affirmed Iterative Reconstruction (SAFIRE) versus weighted filtered back projection (WFBP) effects on quantitative measure in the COPDGene 2 test object. *Med Phys*. 2014;41:091910.
8. Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. *Med Phys*. 2014;41:091908.
9. Straten M, van Mendrik A, Schaap M, et al. Are iterative reconstruction techniques better than filtered backprojection? Quantitative evaluation on a CT phantom. In: *Proceedings, Radiological Society of North America, 97th Scientific Assembly and Annual Meeting*. Chicago, IL: Radiological Society of North America; 2011.
10. Jo JK, Cheol KD, Jong-Woong L, et al. Measurement of image quality in CT images reconstructed with different kernels. *J Korean Phys Soc*. 2011;58:334.
11. Korn A, Fenchel M, Bender B, et al. Iterative reconstruction in head CT: image quality of routine and low-dose protocols in comparison with standard filtered back-projection. *AJNR Am J Neuroradiol*. 2012;33: 218–224.
12. Löve A, Olsson ML, Siemund R, et al. Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. *Br J Radiol*. 2013;86:20130388.
13. Goodenough DJ. *Catphan 500 and 600 Manual*. Greenwich, NY: The Phantom Laboratory; 2014.
14. Holmes EJ, Forest-Hay AC, Misra RR. *Interpretation of Emergency Head CT: A Practical Handbook*. Cambridge: Cambridge University Press; 2008.
15. Thitaikumar A, Krouskop TA, Ophir J. Signal-to-noise ratio, contrast-to-noise ratio and their trade-offs with resolution in axial-shear strain elastography. *Phys Med Biol*. 2007;52:13–28.
16. Hardie AD, Nelson RM, Egbert R, et al. What is the preferred strength setting of the sinogram-affirmed iterative reconstruction algorithm in abdominal CT imaging? *Radiol Phys Technol*. 2015;8:60–63.
17. Ghetti C, Palleri F, Serreli G, et al. Physical characterization of a new CT iterative reconstruction method operating in sinogram space. *J Appl Clin Med Phys*. 2013;14:4347.

PAPER II

Journal of Forensic Radiology and Imaging 12 (2018) 25–30

Contents lists available at ScienceDirect

Journal of Forensic Radiology and Imaging

journal homepage: www.elsevier.com/locate/jofri

Applicability of a clinical cardiac CT protocol in post mortem studies

David Völgyes[a,*], Marius Pedersen[a], Arne Stray-Pedersen[b,c], Dag Waaler[d], Anne Catrine Trægde Martinsen[e,f]

[a] Norwegian University of Science and Technology, Department of Computer Science, Norway
[b] Oslo University Hospital, Department of Forensic Sciences, Norway
[c] University of Oslo, Faculty of Medicine, Institute of Clinical Medicine, Norway
[d] Norwegian University of Science and Technology, Faculty of Medicine and Health Sciences, Department of Health Sciences in Gjøvik, Norway
[e] University of Oslo, Department of Physics, Norway
[f] Oslo University Hospital, Department of Diagnostic Physics, Norway

ARTICLE INFO

Keywords:
Image quality
Observer study
CT
Cardiac
Post mortem

ABSTRACT

Objective: Confirmation whether an optimized clinical cardiac CT scan protocol is also optimal for post mortem cardiac CT scans without iodine contrast or the reconstruction parameters should be changed.
Materials and methods: 27 CT volumes (three cases for three recon- struction kernel with three different iterative reconstruction settings) were graded by six readers in order to find the optimal reconstruction parame- ters. The scans were performed on a Siemens Definition Flash CT scanner using 120 kV tube potentials.
Results: The study has shown that from the investigated options the softest cardiac kernel with the strongest iterative reconstruction were pre- ferred by the readers (I26 Safire 3).
Conclusion: The results indicate that the scan protocol which was adopted from clinical practice is applicable in forensic radiology too even though iodine contrast agent was not administered.

1. Introduction

The number of performed forensic computed tomography (CT) scans is con- siderable smaller than in general clinical practice. While special circumstances might require specialized methodology [1], clinical scans form the basis for recommended protocols for routine post mortem scans. The post mortem alterations [2] and the general lack of iodine contrast administration might introduce a need for scan protocols optimized for forensics medicine. Cardiac pathology is the most common cause of sudden unexpected death, and dedicated post mortem cardiac CT (PMcCT) is an experimental attempt to better visualize pathology when angiography has not been performed.

The aim of this study was to evaluate whether readers find the same cardiac CT scan protocol best for post mortem investigation as is recommended for clin- ical scans even though inferior vena cava (IVC) contrast was not administered. Filtered back projection (FBP) is considered the gold standard reconstruction for decades in CT imaging. The main drawback of FBP is the fact that it cannot take noise properties of the signal into account. Iterative reconstruc- tion (IR) was introduced to more accurately model the physical and statistical

phenomena. IR achieves similar-to-FBP image quality at reduced dose, or less noise at same dose level. However, iterative reconstruction also changes the noise power spectrum, and the textures in the image. Depending on the modeling strategy, whether it is image-space only, sinogram-space, or both, and what physical phenomena are included in the model, finding the optimal parameters are still an open question. FBP has more limited options to mitigate noise. Different kernel-families were developed by the vendors to find a good balance between sharpness and noise. Often the iterative reconstructions are derived from the FBP reconstruction, and they try to yield the same sharpness as their FBP counterparts but with reduced noise level. Detailed review about the state of the art iterative algorithms was published by Geyer et al. [3]

Image quality (IQ) assessment is a major challenge in radiology due to the multivariate optimization problem such as dose, sharpness, noise, among others [4–6]. Several methods are applied in order to determine the best clinical settings, as it presented in a review from Månsson et al. [7].

While special circumstances, e.g. presence of foreign objects, might cause image artifacts [8], arguable post mortem alterations mostly affect the subjective image quality. If readers evaluate several different reconstructions, would they prefer the same reconstruction settings as it was recommended in the clinical practice? This question is the core of this paper, and a visual grading study was performed to answer it.

* Corresponding author.
E-mail address: david.volgyes@ntnu.no (D. Völgyes).

https://doi.org/10.1016/j.jofri.2018.01.003
Received 21 August 2017; Received in revised form 13 January 2018; Accepted 29 January 2018
Available online 12 February 2018

Arguably, one of the most frequently used methods for subjective IQ assessment in radiology is the visual grading (VG) with or without reference image [7]. It has advantages, it scales well with the number of images, and has disadvan- tages, e.g. the difficulty of choosing criteria. The European Commission (EC) released guidelines for image quality criteria for computed tomography [9] which helps to standardize them. Similarly, the statistical analysis of VGA studies went through a long evolution from t-tests to visual grading characteristics (VGC) [10] and visual grading regression (VGR) [11].

This paper focuses on two choices for protocol optimization: the reconstruction kernel selection and the use of iterative reconstruction. These options were investigated using published IQ criteria, and analyzed with VGR.

Many other parameters could be investigated, including dose, slice thickness, tube voltage, but in this study these parameters were kept constant, and only the effect of the various kernels and iterative reconstruction parameters was studied.

2. Materials and methods

2.1. Data collection and readers

Three cardiac scans, later referenced as *cases*, performed on a Siemens Somatom Definition Flash dual-source multi-slice CT scanner (Siemens AG, Forchheim, Germany) [1] were selected for the study using three different reconstruction kernels with filtered back projection (FBP) and two iterative reconstruction (Safire 2 and 3). B26, B36 and B46 kernels were used for FBP, and the closely related I26, I36, I46 kernels were used for Safire reconstructions. The three scans with three kernels and the three reconstruction options yielded 27 CT volumes in total.

FBP and Safire 2 and 3 were chosen because in general these are the gold standard and the most preferred options [12], respectively. Scan and reconstruction parameters are presented in Table 1. Automatic dose modulation were used, and the table contains the effective tube current.

In this paper, B26/I26 notation refers to the group of B26 with FBP and I26 with Safire 2 and Safire 3. B36/I36 and B46/I46 follow the same logic.

These CT volumes were presented to six readers on calibrated 10 bit medical display (EIZO Radiforce G22[2]) from approximately 60 cm distance and using the display windows set by the scanner. No time constraints were given. The volumes were presented blinded, anonymized and in randomized order for each reader, both with respect to the reconstruction parameters and to the cases.

The three of the six readers were experienced radiographers, and three of the six were second year radiographer students. At the beginning of the reading session the objectives of the study and the use of the evaluation software were explained to the readers and they could ask as questions if they wished.

However, one of the readers (reader 4) made a mistake during recording responses, and was allowed to restart the study. While including or excluding this reader does not change the outcome, later we decided to exclude the reader's responses from the study. However, the responses are available in the online dataset.

2.2. Visual grading

Seven criteria were presented to the readers in their native language, and they were asked to score the CT volumes on a 5-grade scale where 1 was the worst and 5 was the best. The translation of the criteria and the specific meaning of the grades were presented in Tables 2, 3, respectively. The criteria C1-C4 are based on EC guidelines [9], and C5-C7 are meant to represent general perceived image quality.

ViewDEX presentation software [13] were used to display the volumes and criteria, and record the responses and the response times. The response times were only used for quality control (see in Discussion part), and were not taken into account as image quality descriptors.

Table 1
Scan and reconstruction parameters.

Scan parameter	Value
Tube potential	120 kV
Tube current	400mAs (effective)
Total detector collimation	38.4 mm
Dose modulation	on
Slice thickness	1 mm
Pitch	0.6
Window center	200 HU
Window width	600 HU

Table 2
Criteria for visual grading.

Identifier	Criterion
C1	Visually sharp reproduction of the heart
C2	Reproduction of the left venticular
C3	Visually sharp reproduction of the pleuromediastinal border
C4	Sharp/clear demarcation of the aortic wall
C5	General impression of contrast
C6	General impression of noise
C7	General impression of artifacts

Table 3
Scoring levels for visual grading.

Criteria	Score	Confidence level
C1, C2, C3, C4	5	Very well reproduced – the structure had completely distinct shape
(see Table 2)	4	Well reproduced – the structure was clearly reproduced
	3	Adequate reproduced – the structure was moderately reproduced
	2	Poorly reproduced - the structure was vaguely reproduced
	1	Not reproduced - the structure could not be discerned
		Contrast with regards to diagnostic
C5	5	Very good
	4	Good
	3	Medium
	2	Low
	1	Unacceptable
		Noise level and artefact with regards to diagnostics are
C6, C7	5	Not disturbing at all
	4	Barely disturbing
	3	Moderately disturbing
	2	Quite disturbing
	1	Highly disturbing

2.3. Data analysis

Visual grading uses ordinal scale which means that higher score belongs to better results but the difference between 2 and 3 is not necessarily the same as between 3 and 4. The aim is to determine which kernel and which iterative reconstruction perform the best while the external factors (e.g. reader differences) are taken into account.

For this purpose, visual grading regression [11] was developed, which is model based on ordinal logistic regression. This model predicts the logarithm of odds ratio (OR) changes if a risk factor is present. These odds ratios are a bit harder to interpret than differences between

[1] https://www.healthcare.siemens.com/computed-tomography/dual-source-ct/somatom-definition-flash, visited on 21st April 2017.
[2] http://www.ampronix.com/eizo-radiforce-g22.html, visited on 21st April 2017.

mean values, but they are mathematically more founded. For details of VGR we refer back to the original publication [11]. However, the mean VGA score (VGAS) [7] is still widely used, and despite their shortcomings [14], shows the magnitude of the difference between options, and makes the results more comparable to other VGA stu.

VGAS is calculated as follows:

$$VGAS = \frac{\sum_{I,N} S_C}{N_I N_R} \quad (1)$$

where S_c is the given score, and it should be averaged for the number of readers(N_R) and the number of images (N_I).

For using ordinal scale and taking into account the different factors (kernels, iterative reconstruction, readers differences, etc.), and deciding whether an effect is significant, one should use the regression model.

The data analysis were conducted with the R open source statistical langauge [15]. The ordinal logistic regression model is calculated with the MASS package [16] from R. Using R's notation, the model is the following:

$$polr(score \sim IR + Kernel + ReaderID + CaseID) \quad (2)$$

where *polr* denotes the ordinal logistic regression function, score is the predicate, and IR, Kernel, ReaderID and CaseID are the non-numeric groups for iterative reconstructions, reconstruction kernels, readers and the scanned volume, respectively. This model takes into account the iterative reconstructions, the kernels and also the differences between the readers and between the cases. These later two factors might influence the scores, therefore, they should be taken into account.

For each criterion, the odds ratios with corresponding p-values were deter- mined. The odds ratios in the ordinal logistic regression model are multiplicative factors, if more than two options are present.

Tables 4, 5 report the result of the statistical model of VGR for different iterative reconstruction options and for the different kernels, respectively. Note that the model uses B26 with FBP as baseline, and results show the increased or decreased odds if a parameter was changed. To ease the interpretation, the most preferred options are presented in the last column of the tables.

All criteria were related to an overall image quality. The line *overall* refers to the mean of the scores given averaging the scores over the criteria. As mentioned before, calculating VGAS is a very basic approach, but it is still frequently used and to give a complete picture Table 6 presents the mean values for this overall image quality score. Due to the shortcomings of this simple model, it would be misleading to present p-values and confidence intervals for this model.

3. Results

Tables 4, 5 show that I26 and Safire 3 were the most preferred reconstruction options. Safire 3 performed significantly better than Safire 2 or the baseline FBP reconstruction. On the other hand, B26/I26 performed best in all but one case. However, difference between B26/I26 and B36/I36 was not statistically significant in any of the cases. In general B46/I46 performed worse than the other two options, and in half of the cases it was statistically significantly worse. These results are in agreement with the simpler VGAS in the Table 6.

The logistic model assumes that there is no interaction between the covariates. Due to the number of observations in this study, interaction between the covariates could not be investigated to get statistically significant results for the large number of free parameters.

In addition to the regression model, Spearman's rank order correlation was used to investigate the responses to the different criteria and to analyze the agreement between readers. Correlation plot in Fig. 1 shows that scores for all of the criteria are strongly significantly correlated ($p < 0.001$) with each other. However, the linear correlation coefficients between the criteria are varying, and especially low for criterion C7.

Similarly, the Spearman's rank order correlation between the readers' responses are presented in Fig. 2.

4. Discussion

4.1. Image quality

I26 with Safire 3 was the most preferred technique in this study for all criteria and for the overall image quality. The difference between Safire 3 and other options were statistically significant for some criteria (C1,C2,C3, C6, and overall), the difference between the B26/I26 and B36/I36 were not statistically significant. The magnitude of the differences are easier to observe in Table 6 for VGAS. The difference in favor of Safire 3 is consistent but small. Evaluation of the clinical relevance of this difference is beyond the scope of this paper, and might be minimal.

The importance of sharpness and noise level depends on the diagnostic task. For the criteria in this study B26/I26 was slightly, but not significantly, better than the sharper B36/I36 kernel. For other criteria B36/I36 might perform better than B26/I26.

This study did not aim to determine the best possible kernel for a general post mortem cardiac study, but rather confirm the applicability of an existing clinical protocol in a forensic setup. If the baseline method would perform worse than any of the alternatives, then it would indicate this protocol couldn't be derived from clinical protocol, and definitely would need a dedicated protocol optimization study.

The results imply that in this case there is no indication of suboptimal performance. Presence of foreign objects, sever tissue degradation or other substantial differences from clinical volumes might require different protocols.

4.2. Response time analysis

There was no time constraint set for the reading session, but response times were recorded for each displayed volume each readers. and second-round response times for reader #4, who wanted to restart

Table 4
Ordinal logistic regression results for iterative reconstruction options. Most preferred option is summarized in the last column. Significant ($p < 0.05$) values are indicated with bold font.

Criterion	FBP < Safire 2			Safire 2 < Safire 3			Preferred
	Odds ratio	p	95% CI	Odds ratio	p	95% CI	
C1	1.35	0.458	0.61–3.03	3.04	**0.009**	1.34–7.06	Safire 3
C2	1.23	0.614	0.55–2.74	2.52	**0.025**	1.13–5.70	Safire 3
C3	1.01	0.977	0.42–2.42	4.21	**0.002**	1.71–10.80	Safire 3
C4	1.20	0.666	0.52–2.79	1.88	0.141	0.81–4.38	Safire 3
C5	0.77	0.521	0.34–1.71	1.35	0.461	0.61–3.03	Safire 3
C6	1.20	0.658	0.53–2.76	3.32	**0.004**	1.47–7.66	Safire 3
C7	0.69	0.380	0.30–1.57	1.15	0.734	0.51–2.62	Safire 3
Overall	1.01	0.974	0.48–2.15	2.67	**0.011**	1.26–5.73	Safire 3

Table 5
Ordinal logistic regression results for reconstruction kernels. Most preferred option is summarized in the last column. Significant ($p < 0.05$) values are indicated with bold font.

Criterion	**B26/I26 < B36/I36**			**B36/I36 < B46/I46**			Preferred
	Odds ratio	p	95% CI	Odds ratio	p	95% CI	
C1	0.87	0.722	0.39–1.92	0.48	**0.080**	0.21–1.09	B26/I26
C2	0.58	0.180	0.26–1.28	0.68	0.341	0.30–1.51	B26/I26
C3	0.55	0.190	0.22–1.33	0.27	**0.005**	0.11–0.67	B26/I26
C4	0.86	0.725	0.37–2.00	0.27	**0.003**	0.11–0.63	B26/I26
C5	0.64	0.281	0.29–1.43	0.44	0.053	0.19–1.01	B26/I26
C6	0.47	0.073	0.21–1.07	0.12	**< 0.001**	0.05–0.29	B26/I26
C7	0.83	0.655	0.37–1.86	0.81	0.609	0.36–1.83	B26/I26
Overall	0.59	0.150	0.28–1.21	0.28	**0.001**	0.13–0.59	B26/I26

Table 6
Overall mean scores for the different kernels and reconstruction options. Highest value is indicated with bold font.

	FBP	**Safire 2**	**Safire 3**
B26/I26	3.12	3.22	**3.46**
B36/I36	3.20	2.97	3.29
B46/I46	2.74	2.93	3.15

the study after a few volumes to correct her answers.

The response time graph (Fig. 3) is included for two reasons. First, it is a legitimate concern that readers became tired during the study and they lost attention. A linear fit after the fifth response shows that average the response times slightly decreased as the study proceeded. This change is approximately 0.5 s/volume which is statistically not significant, and arguably negligible. Second reason for the graph is to demonstrate that the readers used similar amount of time even though they were not constrained. The response time for the very first task was excluded for all readers because it contained the preparation time too.

4.3. Outliers

One reader gave counter-intuitive responses for the last criterion. Does inclusion or exclusion of these responses change the final results? The above analysis included all the responses. Excluding these presumed outliers would slightly decrease the p-values, but would not change significance. We did not consider them influential and rather kept the original data set.

4.4. Inter-observer differences

Both the response time analysis and the outlier points indicate some inter- observer differences. This difference might originate from the challenging evaluation of forensic volumes, but it also might be an

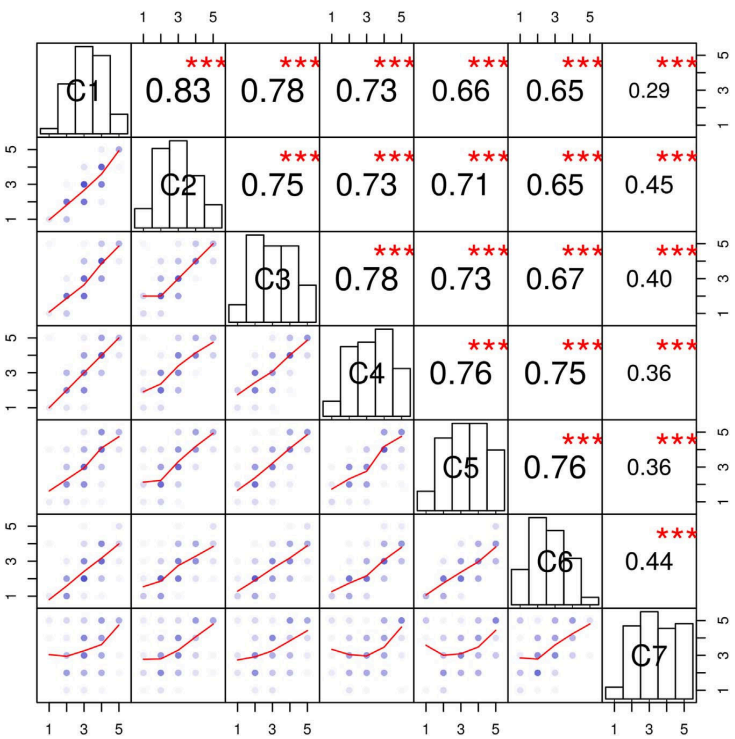


Fig. 1. Spearman correlation of the scores between the criteria. Distribution of the scores are plotted with blue density maps with local curve fit. Significance levels are: $p < 0.05$(*), $p < 0.01$(**), $p < 0.001$(***).

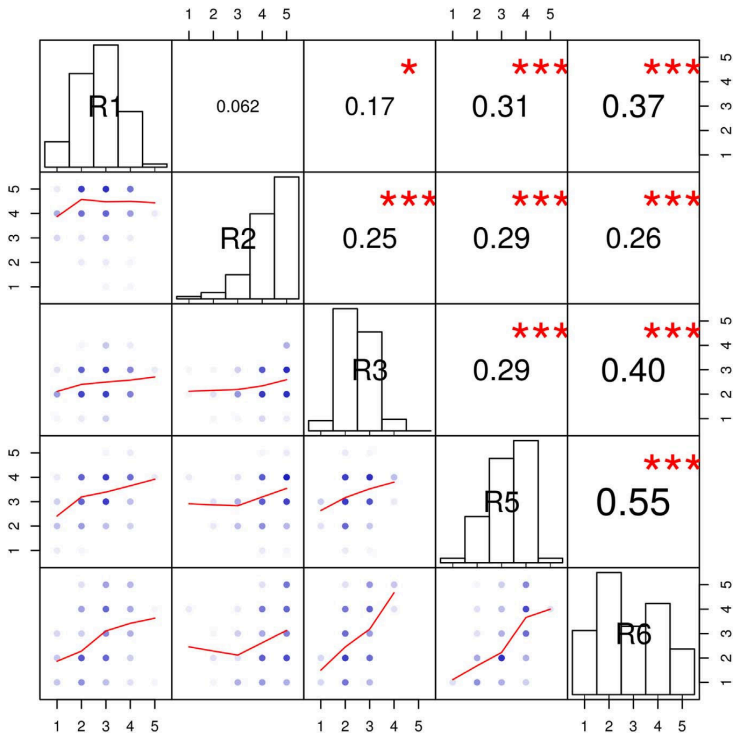


Fig. 2. Spearman correlation of the scores between the readers. Distribution of the scores are plotted with blue density maps with local curve fit. Significance levels: $p < 0.05$(*), $p < 0.01$(**), $p < 0.001$(***).

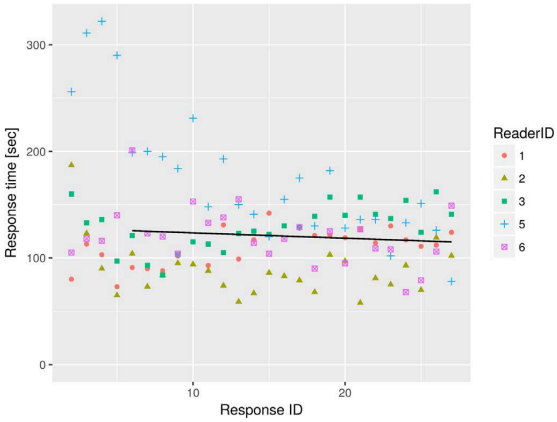


Fig. 3. Readers' response time. After a few try initial grading, the response times stabilized. The decrease in response time is not significant ($p = 0.321$).

underlying difference between the readers. We felt important to disclose this limitation of the study, and presented it in the most concise form in Fig. 2. Note that reader 4 was removed from study, as it was mentioned in the Materials and methods section.

4.5. Future work

Visual grading analysis has known short-comings. Most notably, VGA requires large number of responses to differentiate between similar quality images. The readers also might change their preference scale during the reading session which is known as adaptation [17]. To overcome these limitations, pairwise comparisons [18] (PC) might follow VG sessions. While PC makes easier to choose the best from two options, the number of comparisons increase quadratically with the number of options. Similarly to VG, strict criteria are required for PC to avoid being a „beauty contest'' and other systematic errors.

5. Conclusion

The study found the same kernel and iterative reconstruction (I26 Safire 3) optimal in forensic CT as in clinical use, despite the differences between the clinical and forensic setups.

Acknowledging the limitations of the study in the discussion part, there is no indication that applying a clinical protocol in post mortem scans would change the readers' preferred reconstruction, as long as the criteria are the same and the post mortem alterations are not sever. Statistically significant difference in preference does not necessarily mean difference in diagnostic performance.

Acknowledgements

This research has been funded by the Research Council of Norway through project no. 221073 'HyPerCept – Colour and quality in higher dimensions'.

References

[1] C. O'Donnell, N. Woodford, Post-mortem radiology—a new sub-speciality? Clin. Radiol. 63 (2008) 1189–1194, http://dx.doi.org/10.1016/j.crad.2008.05.008.
[2] C. Jackowski, W. Schweitzer, M. Thali, K. Yen, E. Aghayev, M. Sonnen- schein, P. Vock, R. Dirnhofer, Virtopsy: postmortem imaging of the hu- man heart in situ using MSCT and MRI, Forensic Sci. Int. 149 (2005) 11–23, http://dx.doi.org/10.

1016/j.forsciint.2004.05.019.
[3] L.L. Geyer, U.J. Schoepf, F.G. Meinel, J.W. Nance, G. Bastarrika, J.A. Leipsic, N.S. Paul, M. Rengo, A. Laghi, C.N.D. Cecco, State of the art: iterative CT reconstruction techniques, Radiology 276 (2015) 339–357, http://dx.doi.org/10.1148/radiol.2015132766.
[4] X. Zheng, T.M. Kim, R. Davidson, S. Lee, C. Shin, S. Yang, CT x-ray tube voltage optimisation and image reconstruction evaluation using visual grading analysis, in: B.R. Whiting, C. Hoeschen (Eds.), Proceedings of SPIE - The International Society for Optical Engineering, 2014, p. 903328. doi:10.1117/12.2043201.
[5] M.K. Kalra, M.M. Maher, T.L. Toth, L.M. Hamberg, M.A. Blake, J.-A. Shepard, S. Saini, Strategies for CT radiation dose optimization, Radiology 230 (2004) 619–628, http://dx.doi.org/10.1148/radiol.2303021726.
[6] F. Zarb, L. Rainford, M.F. McEntee, Developing optimized CT scan pro- tocols: phantom measurements of image quality, Radiography 17 (2011) 109–114, http://dx.doi.org/10.1016/j.radi.2010.10.004.
[7] L.G. Månsson, Methods for the evaluation of image quality: a review, Radiat. Prot. Dosim. 90 (2000) 89–99, http://dx.doi.org/10.1093/oxfordjournals.rpd.a033149.
[8] J.F. Barrett, N. Keat, Artifacts in CT: Recognition and Avoidance., Radiographics: A Review Publication of the Radiological Society of North America, Inc. 24, 2004, pp. 1679–1691. doi:http://dx.doi.org/10.1148/rg.246045065.
[9] H. Menzel, H. Schibilla, D. Teunen, European Guidelines on Quality Criteria for Computed Tomography, European Commission, Luxembourg, 2000.
[10] M. Båth, L.G. Månsson, Visual grading characteristics (VGC) anal- ysis: a non-parametric rank-invariant statistical method for image quality evaluation, Br. J. Radiol. 80 (2007) 169–176, http://dx.doi.org/10.1259/bjr/35012658.
[11] Ö. Smedby, M. Fredrikson, Visual grading regression: Analysing data from visual grading experiments with regression models, Br. J. Radiol. 83 (2010) 767–775, http://dx.doi.org/10.1259/bjr/35254923.
[12] A.D. Hardie, R.M. Nelson, R. Egbert, W.J. Rieter, S.V. Tipnis, What is the preferred strength setting of the sinogram-affirmed iterative reconstruction algorithm in abdominal CT imaging? Radiol. Phys. Technol. 8 (2015) 60–63, http://dx.doi.org/10.1007/s12194-014-0288-8.
[13] M. Håkansson, S. Svensson, S. Zachrisson, A. Svalkvist, M. Båth, L.G. Månsson, ViewDEX: an efficient and easy-to-use software for ob- server performance studies, Radiat. Prot. Dosim. 139 (2010) 42–51, http://dx.doi.org/10.1093/rpd/ncq057.
[14] M. Båth, Evaluating imaging systems: practical applications, Radiat. Prot. Dosim. 139 (2010) 26–36, http://dx.doi.org/10.1093/rpd/ncq007.
[15] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014⟨http://www.r-project.org/⟩.
[16] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Fourth, Springer, New York, 2002⟨http://www.stats.ox.ac.uk/pub/MASS4⟩.
[17] H. Helson, Adaptation-level Theory, Harper & Row, New York, 1964⟨http://psycnet.apa.org/psycinfo/1965-00332-000⟩.
[18] L.L. Thurstone, A law of comparative judgment, Psychol. Rev. 34 (1927) 273–286, http://dx.doi.org/10.1037/h0070288.

PAPER III

algorithms

MDPI

Article

A Weighted Histogram-Based Tone Mapping Algorithm for CT Images

David Völgyes [1,*], Anne Catrine Trægde Martinsen [2,3], Arne Stray-Pedersen [4,5], Dag Waaler [6] and Marius Pedersen [7]

1 Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway
2 Department of Physics, University of Oslo, 0316 Oslo, Norway
3 Department of Diagnostic Physics, Oslo University Hospital, 0424 Oslo, Norway; uxneti@ous-hf.no
4 Department of Forensic Sciences, Oslo University Hospital, 0424 Oslo, Norway
5 Institute of Clinical Medicine, University of Oslo, 0318 Oslo, Norway; arne.stray-pedersen@medisin.uio.no
6 Department of Health Sciences in Gjøvik, Norwegian University of Science and Technology, 2803 Gjøvik, Norway; dag.waaler@ntnu.no
7 Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway; marius.pedersen@ntnu.no
* Correspondence: david.volgyes@ntnu.no; Tel.: +47-4717-0724

Received: 16 June 2018; Accepted: 20 July 2018; Published: 25 July 2018

Abstract: Computed Tomography (CT) images have a high dynamic range, which makes visualization challenging. Histogram equalization methods either use spatially invariant weights or limited kernel size due to the complexity of pairwise contribution calculation. We present a weighted histogram equalization-based tone mapping algorithm which utilizes Fast Fourier Transform for distance-dependent contribution calculation and distance-based weights. The weights follow power-law without distance-based cut-off. The resulting images have good local contrast without noticeable artefacts. The results are compared to eight popular tone mapping operators.

Keywords: tone mapping; histogram equalization; CT

1. Introduction

Global and local contrasts are often imperfect in images. The available dynamic range of the display is either not utilized, or, on the contrary, the range is wider than the low dynamic range display.

Medical images, particularly Computed Tomography (CT) images, are challenging to visualize. First, artefacts could lead to inferior diagnostic performance [1]. Second, a lack of good local contrast could also limit diagnostic performance; local contrast enhancement can improve diagnostic efficiency [2] or significantly reduce interpretation times [3]. Third, CT images have low soft tissue contrast, a relatively high noise level compared to magnetic resonance imaging (MRI), and they have a high dynamic range (approximately 12 bits) [4].

CT images represent absolute tissue densities using CT numbers that are measured in Hounsfield units (HU) [5]. This calibrated nature of the CT imaging makes it appealing to use global, monotonic tone mapping operators (TMO) which assign the same color to pixels representing the same tissue density, regardless of the location of the pixels.

Specialized protocols are developed for diagnosing specific pathologies. However, the need for image post-processing in order to obtain better local contrast [6–8] dates back decades.

In the following, we briefly overview the most important approaches for global and local contrast enhancement approaches for CT images before we present our proposed local method.

1.1. Histogram Methods

Histogram equalization [9] is one of the simplest contrast adjusting algorithms, and it has been applied to CT images since the mid-1980s [6]. The main idea is that every shade available should be used for approximately the same number of pixels. Assuming N different shades, the frequency of a shade is then:

$$p_i = \frac{1}{N}. \tag{1}$$

Histogram equalization is a kind of maximum-entropy approach because Shannon-entropy Equation (2) has its maximum when the probabilities are equal:

$$S = -\sum_i p_i \log p_i. \tag{2}$$

Brightness is assigned to a given shade using, F, the cumulative distribution function of the pixel values and U as the cumulative distribution function of a uniform distribution:

$$\text{shade} = U(F^{-1}(x)). \tag{3}$$

While histogram equalization has some very useful properties, it does not guarantee good local contrast. One approach could be to minimize normalized Shannon information distance between the source distribution and the displayed image, which is the basis of the CT windowing algorithm from Nikvand et al. [10]. Despite its adaptive nature, this algorithm belongs to the global operators, and might not yield good local contrast, and its generalization seems to be non-trivial.

Local histogram equalization (LHE) [11] is meant to solve the problem of local contrast using a sliding window histogram and equalizing the local histograms. This algorithm could over-amplify local noise, especially for homogeneous regions which are larger than the window size. Contrast-limited adaptive histogram equalization (CLAHE) [7] uses an upper limit for the histogram bins. If a bin has higher counts than this number, the peak is truncated at the limit, and the extra counts are distributed to all of the bins uniformly. This effectively prevents high peaks in the histogram from over-stretching. In the case of an extremely small limit, all of the bins become truncated, and then the cut area is redistributed uniformly, which leads to a flat histogram. However, CLAHE has a limited window size which leads to halos, while larger windows limit the locality and adaptiveness of the algorithm.

Not only spatial adaptation, but also histogram processing is the subject of intensive research. The contrast limitation in CLAHE can be seen as an early approach. Another approach could be splitting the histogram into two or more parts, and equalizing them into pre-determined ranges. This technique is known as bi-histogram equalization [12]. This ensures that every important part of the histogram gets enough dynamic range. For instance, in CT images, the histogram could be split into three parts, representing lung tissue, soft tissues, and bones.

A generalized form is prescribing the shape of the histogram [13], using cumulation functions [14], or requiring the extreme of a mapping descriptor, for instance, the above-mentioned normalized information distance [10].

One of the most appealing features of the histogram-based methods is the fact that the individual processing methods could be combined easily. For instance, the CLAHE algorithm could easily be combined with the bi-histogram equalization approach.

1.2. Tone Mapping

CT images are often *windowed*, which means that only part of the dynamic range is displayed, and this part is processed with histogram equalization-based methods.

The problem can be seen as a high dynamic range imaging issue: the high dynamic range CT image should be tone mapped to be displayed on a low dynamic range display while preserving local contrast.

High dynamic range images are often processed with tone mapping operators to keep or enhance local contrast while compressing the dynamic range, to fit into the displaying medium's dynamic range.

Besides histogram equalization, the simplest dynamic range compression algorithm is gamma compression which maps source image intensity to a target image intensity in the following way:

$$I = I_0^{\gamma}, \tag{4}$$

where I_0, I are the source and target images, respectively. $I_0(x, y)$ and $I(x, y)$ are used to refer to the pixel values of these images.

While this effectively reduces the dynamic range if $\gamma < 1$, it does not ensure good local contrast. A popular choice to keep or increase local contrast is the base layer—detail layer separation. The base layer (B) of an image consists of larger structures and has a high dynamic range, while the detail layer (D) is the difference between the original image (I_0) and the base layer:

$$B = \mathcal{F}(I_0), \tag{5}$$

$$D = I_0 - B. \tag{6}$$

Any filter which removes small details could be used for this separation, including a Gaussian filter, and edge preserving denoising filters (median, total variation minimization, bilateral filter, among others).

After the dynamic range compression of the base layer, the details should be added back using a scale factor (α) and can even be enhanced. This could be simple multiplication ($\alpha > 1$), or more sophisticated edge enhancement where the detail layer is pre-processed before it is combined with the compressed base layer into a tone mapped image (TMI):

$$TMI = \text{compress}\{B\} + \alpha \cdot D. \tag{7}$$

This base-detail layer separation technique is the basis of numerous algorithms. Using a Gaussian filter for detail separation and leaving the base layer uncompressed yields the unsharp masking algorithm, which belongs to the wider class of local contrast enhancement techniques. Applying any compression on the base layer, even simple linear rescaling, yields a tone mapping algorithm. The approach can be generalized into multi-layer separation using linear [15] or nonlinear decomposition [16]. The quality of these approaches depends on the separation filter and the base layer compression. One very active research area is to find good, edge preserving filters avoiding such artefacts. For instance, the following filters were proposed for base-detail separation: Gaussian filter, anisotropic diffusion, weighted least squares (WLS) [16], total variation (TV) based filters using L_1 [17] and L_0 [18] norms and bilateral filter [19], among others. While this separation is more like a framework than an algorithm, it is an important building block of tone mapping operators and detail manipulation algorithms.

Besides the layer separation approach, many other tone mapping approaches have been developed. Reinhard'02 operator simulates the effect of photographic zones with an additional dodging-and-burning step [20]. Fattal's approach [21] is based on gradient attenuation. The smaller than threshold gradients are slightly magnified while larger gradients are suppressed. The attenuated gradients lead to a Laplace–Poisson problem which can be solved iteratively. Mantiuk presented a perceptual framework for contrast processing, which is related to the gradient suppression approach but models perceptual contrast. This is not the only way to model the human visual system (HVS). Drago's method is built on the logarithmic compression of luminance values [22], and the Reinhard'05 [23] operator models the photoreceptor adaptation on local and global levels. The approach by Ferradans et al. [24] models visual adaptation for global tone mapping which is followed by a second stage local contrast enhancement. The Mantiuk'08 operator [25] iteratively

minimizes the visible distortion of the image measured by an error metric, and takes into account both the display properties and the properties of HVS. Color perception plays a central role in retinex-based TMO [26] and in the so-called iCAM model which is a 'next-generation color appearance model' [27]. For further details, we refer the interested reader to the literature [28,29].

Our proposed method builds on spatially weighted histogram equalization. It can be seen as an effective tone mapping approach or as a method for local contrast enhancement. Our aim was to avoid halos and artefacts of local histogram methods and ensure good local contrast. While histogram methods have been used for medical image processing since the 1980s, the effectiveness of tone mapping operators has also been demonstrated on medical images, for instance Fattal's method [21].

The literature on histogram methods and tone mapping operators is vast; further details can be found, for instance, in [30].

2. Problem Statement

Good local contrast has a very important role in computed tomography (CT) images used for diagnosing pathologies [31]. Unlike traditional photographs, CT images contain a measurement of material densities, and these are unaffected by irradiation. Traditional tone mapping assumes that the image is the product of illumination and reflectance [32].

CT images also have huge dynamic range, and a global histogram equalization would not yield good enough local contrast, while local methods might yield unwanted halos. Our main goal is to develop an algorithm which is able to compress high dynamic range images into a low dynamic range while presenting as much local information as possible, preserving the main structures, not exhibiting strong halos, and more or less keeping pixel intensity ordering.

Theoretically, any tone mapping operator could be used for medical images. However, the required contrast strongly depends on the tissue, and the absolute contrast is the smallest for medium CT densities, for instance kidney (20–45 HU), muscle (35–55 HU) [33], while the absolute contrast is large for small CT densities, for instance gas volumes (−1000 HU) and lung tissue (−700–500 HU) [34] and also large for dense parts, such as calcification (>150 HU) and bones [35]. Traditional tone mapping operators do not necessarily perform well for these regions, global histogram equalization lacks locality, while local histogram equalization methods (LHE/CLAHE) often lead to halos.

In practical terms, the aim of this paper is to develop a local tone mapping operator which lies somewhere between global and local histogram equalization and combines their advantages and avoids their shortcomings.

3. Theory

The main drawback of local histogram equalization methods is the often visible halos around strong edges. This issue originates from the limited size of the locality window. However, applying a large window limits the locality of the method, and gradually converges towards a global method, while it requires more computational power. Using adjacent or slightly overlapping blocks instead of sliding windows could effectively reduce the required resources, but could lead to blocking artefacts.

Our approach is built on the following ideas:

- Local neighborhood is important in order to determine a given pixel's intensity.
- Neighborhood should not have a strong cut-off; the weighted contribution of the whole image should be taken into account.
- The contribution is a decreasing function of the distance.
- The contribution can be calculated using a Fast Fourier transform (FFT) [36].
- The intensity of the pixel is determined based on the local relative intensities in the source image.
- Locality and noise tolerance are equally important.

The concept of using convolution to calculate the sliding window local histogram was established decades ago [11]. Initially, the rectangular window function was used; later, this was generalized to

other shapes and functions. However, to the best of our knowledge, the weighting function always had a cut-off, either determined by the window size, or by the fast decreasing weights, such as Gaussian weighting.

We argue that a power function should be used as the weighting function for two reasons. First, according to Blommaert [37], normalized local contributions to perceived brightness follow a power-law with a negative exponent:

$$W(r) \sim \frac{1}{r^a}. \tag{8}$$

Using this function family means that our proposed algorithm uses a similar weighting approach as the human visual system.

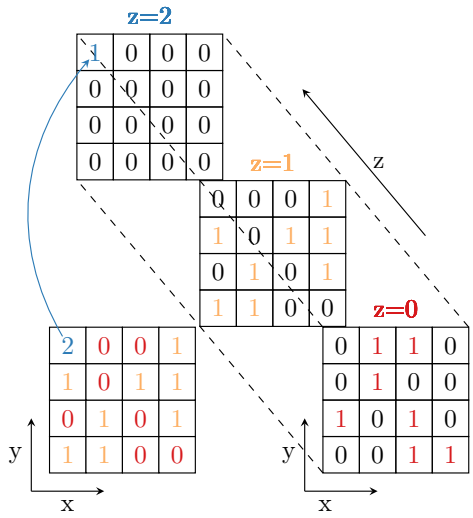
Second, intuitively, this function shape is required for good local contrast without creating halos: if the function does not decrease fast enough, the weighting will resemble a global histogram equalization and will not sufficiently prioritize local information. On the other hand, if the weighting function has a small effective kernel, either because of the window size, or because of the too fast decreasing weights, then the further lying pixels will not contribute enough to the local histograms which leads to halos around sharp edges. However, this effect can be controlled using contrast stretching limitations, such as in CLAHE.

The FFT-based contribution calculation [11] and contrast-limited equalization [7] are known methods, but they are important for the proposed algorithm, and are thus briefly reviewed in the following sections.

3.1. Indicator Array

First, the 2D image is transformed into a 3D indicator array, as depicted in Figure 1. The main steps are:

- create a 3D array putting a column over every pixel in the 2D image,
- the height of the column equals the number of discrete pixel value levels,
- the cells are filled with zeros, except the ones where the z coordinate of the cell equals the pixel value in the same (x,y) position in the image.



$$ID(x, y, z) = \delta(I_{(x,y)}, z)$$

Figure 1. Indicator array generation: z coordinates are calculated from the pixel value of the 2D image.

Using a Dirac function defined in Equation (9), the $ID(x, y, z)$ indicator array is as follows:

$$\delta(a, b) = \begin{cases} 1, \text{if } a = b, \\ \text{otherwise, } 0. \end{cases} \tag{9}$$

$$ID(x, y, z) = \delta(I(x, y), z). \tag{10}$$

3.2. Weighted Contribution

The simplest weighting function is a constant value; this gives global ranking. A limited weighting function is often used having 1 as the weight for a small area around the selected pixel, and 0 elsewhere. This weighting function is called a 'window', and the algorithm is a local histogram equalization. These histogram calculations can be implemented very efficiently. However, the most general form of weighting means evaluation of every pixel pair. If the image has n times n pixels, then the asymptotic complexity of this calculation is $\mathcal{O}(n^4)$.

There is a special case when the weighting only depends on the distance of the pixel pairs.

$$\mathrm{W}\left(x, x', y, y'\right) = f(d(x - x', y - y')), \text{where} \tag{11}$$

$$d(x - x', y - y') = \sqrt{(x - x')^2 + (y - y')^2}. \tag{12}$$

In this special case, every pixel gives $f(r)$ contribution at distance r.

Only pixels with this z_0 intensity value can contribute to a histogram bin z_0. These pixels are recorded in the plane z_0 in the indicator array. Due to the translation invariant weights, the weighted histograms can be calculated convolving the indicator array in the plane z with the weighting function:

$$H(x, y, z) = ID(x, y, z) \underset{x,y}{\star} W(x, y). \tag{13}$$

Convolutions can be efficiently calculated in the Fourier domain using fast Fourier transform (FFT), and its inverse (iFFT):

$$\tilde{f}(s, t) = FFT_{x,y}\left\{f(x, y)\right\}, \tag{14}$$

$$f(x, y) = iFFT_{s,t}\left\{\tilde{f}(s, t)\right\}, \tag{15}$$

$$H(x, y, z) = iFFT_{u,v}\left\{FFT_{x,y}\left\{ID(x, y, z)\right\} \cdot FFT_{x,y}\left\{W(x, y)\right\}\right\}. \tag{16}$$

The result is a 3D array which contains locally weighted histograms along the z-direction for every corresponding pixel (see Figure 2). For further details, we refer the interested reader to [11].

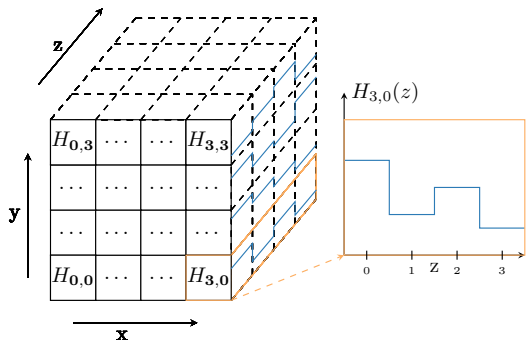


Figure 2. Columns in the z-direction contain the weighted histograms for corresponding pixels. Every pixel has its own local weighted histogram.

Due to the nature of FFT, it is important to note that we used the 'mirrored image' boundary condition, which is a frequently used condition for image boundaries when Fourier methods are used.

3.3. Relative Intensity

We define the relative intensity of a pixel in (x, y) position based on the cumulative distribution function (CDF) of the local histograms:

$$\text{intensity}(x, y) = \text{CDF}_{x,y}(f(x, y)). \tag{17}$$

This intensity should be converted to a pixel value using a perceptually linear color map.

3.4. Contrast Limit

Contrast-limited adaptive histogram equalization (CLAHE) [7] introduced the idea of clipping histogram peaks and redistributing the clipped area to all histogram bins, as is depicted in Figure 3.

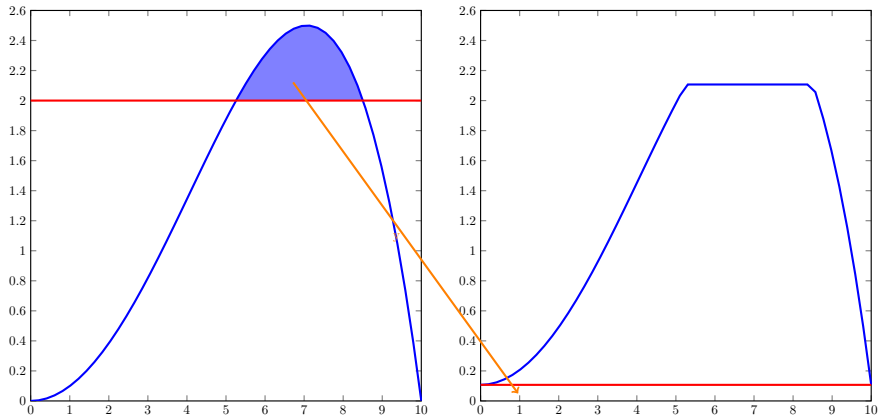


Figure 3. Local histograms might be clipped to reduce noise over-amplification.

In our approach, a radially decreasing weight function yields good locality and avoids halos due to the large weighting kernel, but it still may over-stretch contrast between similar shades. Limiting contrast stretching effectively mitigates this issue, and avoids over-amplification of noise.

3.5. Algorithm Summary

The core algorithm can be summarized as follows:

- read data $\rightarrow I_0$,
- reduce bit depth with dithering $\rightarrow I$,
- generate $W(r)$ weight array,
- loop over pixel values (z),
  - $ID(x, y, z) = \delta(I(x, y), z)$
    - use superpixels, if downscaling is required,
  - convolve ID in x,y plane with W in order to get H,
- clip H peaks,
- redistribute clipped areas along the z-axis,
- determine local intensity from the local histogram and the original image
  - use bilinear interpolation, if superpixels were defined,
- convert the final result from float to integer with Floyd-Steinberg dithering.

4. Materials and Methods

The results of our proposed algorithm are demonstrated in two ways. First, it has been compared to 8 different tone mapping operators, Ferradans [24], Drago [22], Durand [19], Fattal [21], Mantiuk '06 [38] and '08 [25], and Reinhard '02 [20] and '05 [23], using two post-mortem CT examples, a head and a chest CT. Second, the effects of changes of the algorithm parameters are depicted in two image montages, using the above-mentioned chest CT and a color photo of a ship, in Figure 4. The purpose of the ship image is to demonstrate that the algorithm is not specific to CT images. The image is an RGB image which was transformed to HSV color space [9], and the V component was processed using 256 discretization levels.

The presented CT examples are anonymized post mortem forensic CT images. The data collection was performed by the Oslo University Hospital, Department of Forensic Sciences, and was approved by the Attorney General of Norway. Next of kin were informed of the right to opt out from use of data in research projects. The ship photo was taken by the first author.

The CT image uses Hounsfield units; the pixel value range was -1024, +3069 HU for the chest, and -1024, +1935 HU for the head CT scan, using 1 HU step. The resolution of the ship image is 1024×1024 pixels, while the CT images natively have 512×512 pixel resolution.

Figure 4. Harbour in sunset, taken by the first author. The fine details of the deck and the buildings are hidden in the shadow.

4.1. Evaluation

Three evaluation metrics are used to present our results: the structural part of the tone mapped image quality index (TMQI) [39], image entropy, and gradient magnitude. TMQI is a metric for determining image quality of tone mapped images using a high dynamic range original as a reference image. It correlates the pixels of the original image and the tone mapped image using Pearson correlation. The correlation is calculated at five different scales, and weighted combination of the coefficients gives the structural similarity component of the TMQI score. The score should be between -1 and 1, and the higher value indicates better structural match. Entropy and gradient magnitude are used as image descriptors. Entropy measures how well the dynamic range is utilized. It gives a high value when the levels in the dynamic range are uniformly utilized. While entropy measures a global property of the image, gradient magnitude measures a local one. Gradient magnitude measures the averaged amount of edges in an image. A higher value indicates more or larger edges, which indicates more details in the image. Neither entropy nor gradient magnitude determine perceived image quality,

as they do not take the human visual system into account; however, as a frequently used image, they can contribute to the characterization of the proposed algorithm.

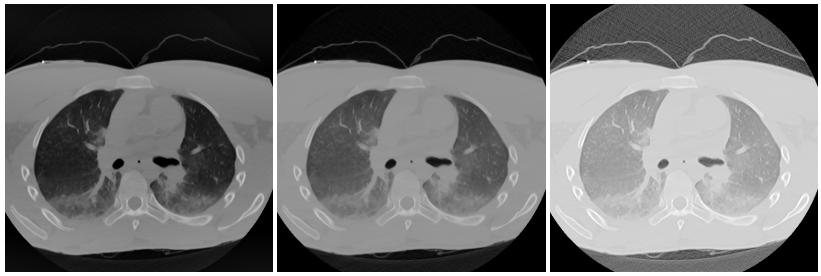
These descriptors above are used to compare the tone mapping operators to the proposed algorithm in two cases: a chest CT where the low tissue densities (lungs) are significant, and a head CT where the medium densities (white and gray matter) play an important role.

The parameters for the algorithms are presented in Table 1. They were selected using a grid-search algorithm around the recommended parameter settings. The parameter combination was selected which yielded the highest combined TMQI score for the two CT images. The proposed algorithm has a relatively wide parameter range where the TMQI score is stable. From this range, $a = 1$, $clip\_limit = 6$ were selected, as convenient integer parameters.

Table 1. Structural similarity scores from the TMQI algorithm, gradient magnitudes and image entropies. Bold text indicates the highest score.

Algorithm	Chest CT TMQI	Head CT TMQI	Chest CT grad.mag.	Head CT grad.mag.	Chest CT Entropy	Head CT Entropy
Drago	0.904	0.794	0.208	0.202	5.96	5.85
Durand	0.899	0.749	0.212	0.207	6.12	6.18
Fattal	0.971	0.836	0.217	0.202	6.89	6.15
Ferradans	0.916	0.835	0.014	0.019	6.71	7.08
Mantiuk'06	0.918	0.828	**0.226**	**0.231**	**7.38**	**7.10**
Mantiuk'08	**0.976**	0.783	0.216	0.207	6.77	6.26
Reinhard'02	0.898	0.803	0.206	0.202	5.81	5.86
Reinhard'05	0.908	0.798	0.208	0.200	6.00	5.80
proposed	0.957	**0.949**	**0.226**	0.224	7.03	6.73

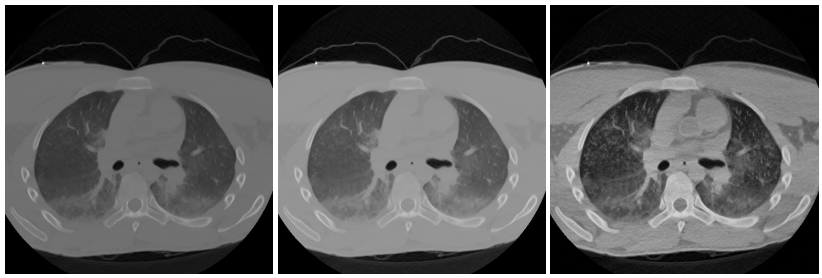
Next to the structural similarity, gradient magnitude and image entropy are also reported. These three quantitative results are presented in Table 2, while the tone mapped images are shown in Figures 5 and 6.



(**a**) Ferradans (**b**) Drago (**c**) Durand

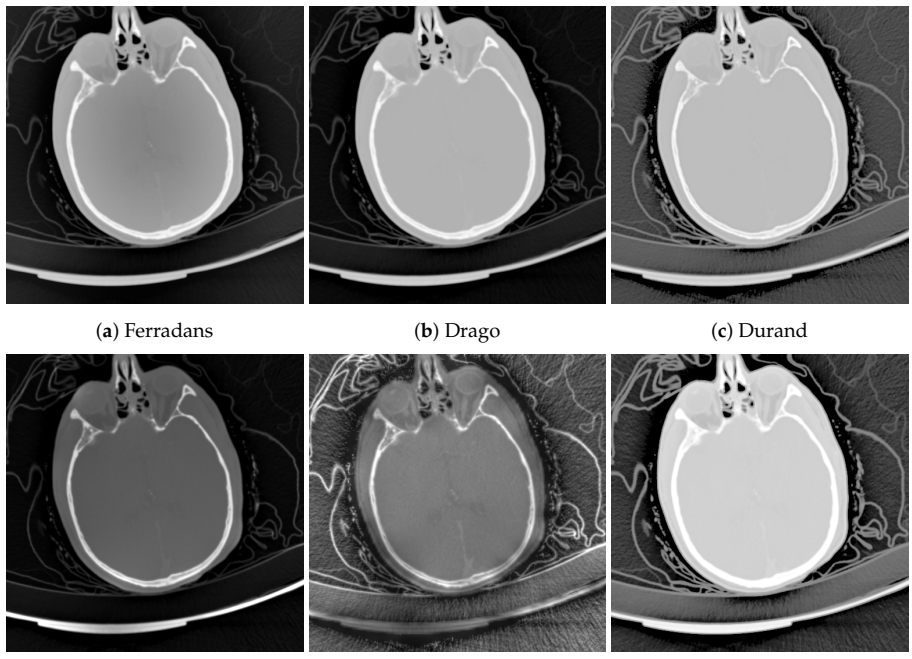
(**d**) Fattal (**e**) Mantiuk'06 (**f**) Mantiuk'08

Figure 5. *Cont.*



(**g**) Reinhard'02 (**h**) Reinhard'05 (**i**) proposed

Figure 5. Tone mapped chest CT scan with eight common operators and the proposed method. Parameters are summarized in Table 1; a quantitative comparison is presented in Table 2.



(**a**) Ferradans (**b**) Drago (**c**) Durand

(**d**) Fattal (**e**) Mantiuk'06 (**f**) Mantiuk'08

(**g**) Reinhard'02 (**h**) Reinhard'05 (**i**) proposed

Figure 6. Tone mapped head CT scan with eight common operators and the proposed method. Parameters are summarized in Table 1; a quantitative comparison is presented in Table 2.

Table 2. Parameter sets for the tone mapping operator.

Algorithm	Parameters			
Durand	σ_s = 7	σ_r = 1.5	base contrast = 4	
Drago	bias = 1.0			
Fattal	alpha = 0.5	beta = 0.95	saturation = 1	noise = 0.002
Ferradans	rho = 0.4	invAlpha = 5.5		
Mantiuk'06	scaleFactor = 0.25	saturationFactor = 0.5	detailFactor = 7.0	
Mantiuk'08	saturation = 1	contrast enhancement = 4.3		
Reinhard'02	key = 0.02	phi = 1.0	no scales used	
Reinhard'05	brightness = 7	lightness adapt. = 1	chromatic adapt. = 1	
proposed	exponent = 1.0	contrast limit = 6		

The effect of the two control parameters, the exponent of the power-law, and the contrast limit is visualized in Figures 7 and 8. Faster decreasing weights further enhance the local features, including noise. The contrast limit from CLAHE effectively regulates this issue.

Figure 7. The effect of the $1/r^a$ weighting function and clipping. Rows from top to bottom have a = 0.7, 1.0, 1.5, 2.0, respectively, and the clip limits in the columns from left to right are 1, 5, 10 and 20, using $1/N$ units where N is the number of histogram bins.

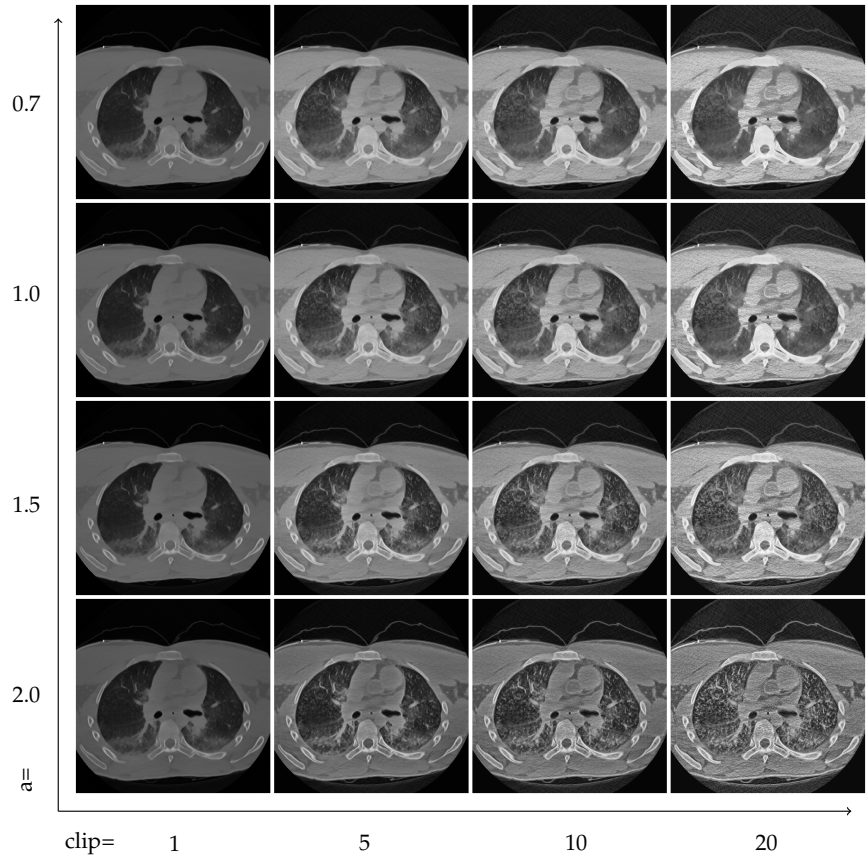


Figure 8. The effect of the $1/r^a$ weighting function and clipping. Rows from top to bottom have $a = 0.7$, 1.0, 1.5, 2.0, respectively, and the clip limits in the columns from left to right are 1, 5, 10 and 20, using $1/N$ units where N is the number of histogram bins.

4.2. Implementation

Our algorithm is implemented in Python3.6 [40] using Numpy 1.13.3 [41], Scipy 1.0.1 [42], and Numba 0.38.0 [43]. For color space transformation, scikits-image 0.14.0 [44] was used. The source code follows PEP8 [45] recommendations and is part of the supplementary material. Luminance HDR 2.5.1 [46] was used to generate the reference TMO images, using the Ubuntu 18.04 operating system. The TMQI score is calculated using the original source code from the authors [39], using Octave 4.2.2 [47]. All of the custom calculation codes, as well as the example images, are available as a supplement to the paper.

The calculations were performed on a Dell Latitutde E7440 notebook (Round Rock, TX, USA) which was equipped with an Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz (Santa Clara, CA, USA), 8GB of DDR3 RAM at 1600MT/s transfer speed. The data was stored on a 240 GB SAMSUNG SSD PM85 drive (San Jose, CA, USA), using btrfs filesystem. The operating system was Ubuntu 18.04 64bit. The computation time is dominated by the FFT calculation for which we used the Scipy implementation, but this part could be replaced with other numerical libraries, or could be executed on GPUs. However, some constant overhead is inevitable due to disk reading, data conversion, initializations, etc.

5. Discussion

The main difference between the proposed algorithm and traditional histogram equalization methods is the power-law-based distance-weighted contribution to the local histograms. LHE and CLAHE are special cases in our proposed generalized framework with constant spatial weights with radial cut-off (CLAHE) or without it (HE). While a similar approach is used for other types of weighting functions, we are not aware of power-law-based weighting. However, power-law was proposed as part of a generalized cumulation function-based histogram equalization [14]. This method is orthogonal to our approach because the Fourier series-based convolution approach is used on the intensity axis of the 3D histogram, while our method uses convolution along the spatial axes.

Many novel histogram-based methods try to modify or prescribe the shape of the local histograms but leave the spatial part intact. These methods, for instance, the bi-histogram equalization, could be easily combined with our approach, and therefore, we see these algorithms as complementers.

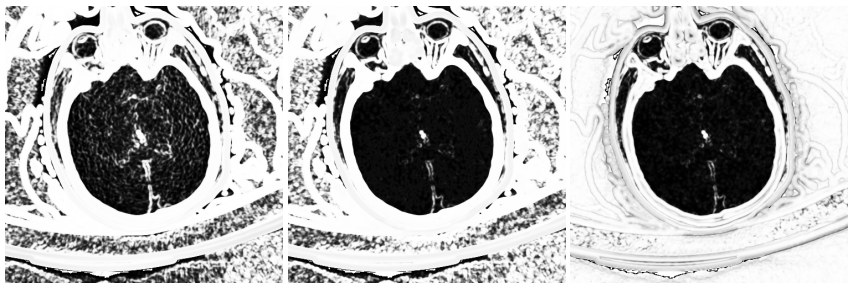
Traditional TMOs offer dynamic range compression with low noise but good local contrast. These operators perform well for selected sub-problems, usually either for low densities (lung) or for high densities (bones). The challenge for these algorithms is to reproduce good local contrast for medium densities (soft tissues) where the contrast is already low in the source. This situation is common in head CTs where the soft tissue contrast is poor.

As can be seen in Table 2, the proposed algorithm yields good structural similarity scores. The structural similarity score is from the TMQI algorithm's structural similarity part. TMQI was developed for natural images, and it contains a 'naturalness' component. This does not seem applicable to medical images, and we only present the structural score. However, if naturalness was taken into account, it would not change the conclusions.

A shortcoming of all of the scores, not only TMQI, is that they do not take into account the image content. Photographic images usually use the whole pixel domain to present information. In CT scans, information from outside of the human body has a limited role, and pixels outside of the scan field of view do not contribute to the image quality, but they are present in the DICOM files as zero pixels. Neither TMQI score nor the tone mapping algorithms take this into account.

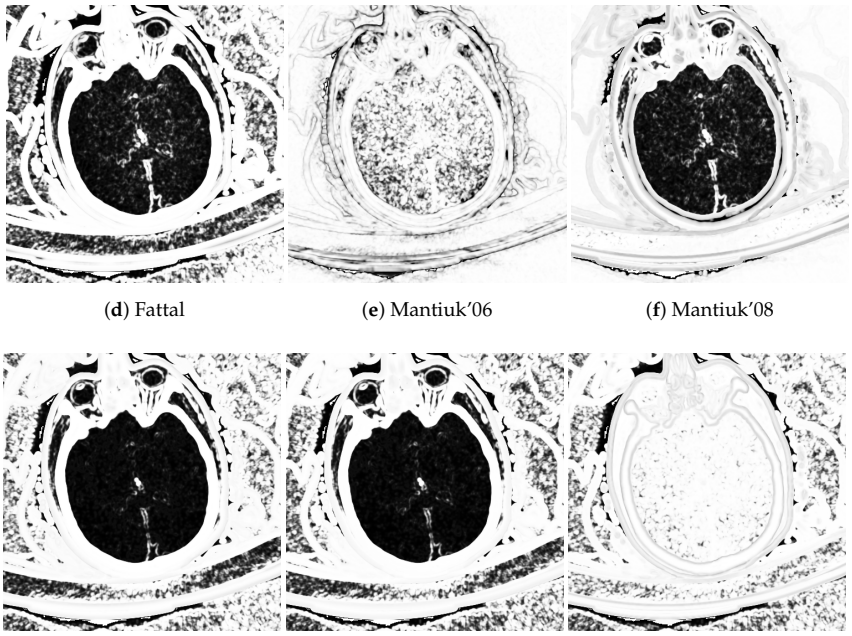
Structural similarity is not equally important in all parts of CT scans. It can be argued that a content-aware structural similarity metric should be developed in the future which takes the field-of-view and tissue properties into account. We are not aware of such a content-dependent performance metric for CT images, and it is beyond the scope of this paper to introduce one, but we recognize this limitation.

TMQI offers structural similarity maps at various resolution levels. Figure 9 shows the finest resolution structural similarity maps of the different tone mapping operators using the head CT example. The advantage of the proposed method is that it has better structural similarity for medium densities, e.g., white and gray matter in the brain, than the alternative methods which mostly perform well outside of the skull.



(**a**) Ferradans (**b**) Drago (**c**) Durand

Figure 9. *Cont.*



(**d**) Fattal (**e**) Mantiuk'06 (**f**) Mantiuk'08

(**g**) Reinhard'02 (**h**) Reinhard'05 (**i**) proposed

Figure 9. Structural similarity map for the head CT example. Brighter shades belong to higher local structural similarity (white = 1.0, black = 0.0).

5.1. Distance Metric

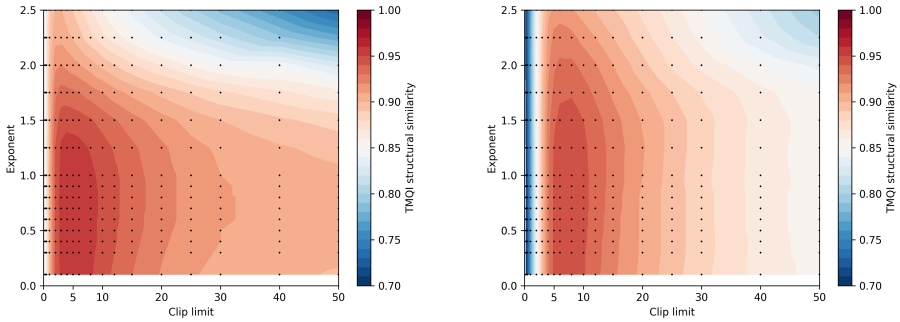
Equation (12) uses a Euclidian distance metric. Other distance metrics could be also used, e.g., Manhattan or maximum distances. Choosing a different distance did not visibly change the example images, but an image with a lot of directed structures might be sensitive to the distance metric selection. We always recommend the Euclidian metric because it is rotationally invariant.

The metric has to be invariant to translation, otherwise FFT could not be utilized for an efficient convolution calculation. While a position-dependent metric could be interesting, it would yield a much less efficient calculation scheme than the FFT-based convolution.

5.2. Locality and Clipping Limits

The two control parameters strongly affect the quality of the resulting images. However, finding the right balance can strongly depend on the visualization goal. Higher exponents mean stronger local contrast enhancement, but can also enhance noise. The contrast limit works against the distortions, but also limits the achievable feature amplification. Two image grids in Figures 7 and 8 demonstrate these issues.

The parameter sensitivity of the algorithm is an important question. The TMQI structural similarity score is depicted in Figure 10 as the function of the exponent and the clip limit. There is a central region where the TMQI score is relatively stable and any parameter pair from this range is a reasonably good choice, e.g., $a = 1.0$, $clip\_limit = 5$ seems to be a good starting point. As will be discussed in the Optimization section, the calculation can be sped up without sacrificing the image quality too much. With lower precision calculations, the parameter space is quickly discoverable.



(**a**) chest CT parameter sensitivity (**b**) head CT parameter sensitivity

Figure 10. Parameter sensitivity of the algorithm for (**a**) the chest CT, and (**b**) the head CT image.

5.3. Edge Enhancement and Auto-Leveling

As was mentioned in Equation (7), base-detail layer separation is a frequent approach. Edges could be separated from the image, and added back to the compressed image. However, this technique is applicable to all tone mapping approaches. Such edge enhancement was not applied for the reference operators, and, for the sake of fair comparison, the proposed method does not contain such edge enhancement either. Similarly, auto-leveling, a method which slightly stretches image contrast, was not applied in the traditional TMOs or in our proposed method.

5.4. Asymptotic Complexity

Asymptotic complexity is dominated by the calculation of the weighted histogram array. Calculating weights directly between pixel pairs has a complexity of $\mathcal{O}(n^4)$.

Creating the indicator array is proportional to the number of pixels: $\mathcal{O}(n^2)$. Fast Fourier Transform of a $n \cdot n$ array requires two one-dimensional FFTs for every line (n), and the cost of one line's transform is $\mathcal{O}(n \log n)$. The number of 2D arrays is proportional to the number of different pixel-value levels (m). Taking all of these into account, the asymptotic complexity is:

$$\mathcal{O}(2m \cdot n^2 \log n). \tag{18}$$

As long as $2m < n^2 / \log n$, this is a faster approach than weights taken directly from pixel pairs. This condition is easily met for traditional photographs which usually have $n > 1000$ pixels for image width and height while the number of discrete levels is usually between 256 and 1024. In some medical applications, the image resolution is lower, while the number of levels is higher. For instance, computed tomography usually has images with 512×512 pixels with 4096 different quantization levels.

Color images are usually processed in an 8- or 10-bit discretized brightness, lightness or value channel, but conversion from another color space with a large dynamic range might yield a huge number of levels, e.g., directly processing RAW data from digital cameras. A huge number of discrete levels might render the approach non-beneficial, or would require approximations, as is explained later in Section 6.

5.5. Memory Consumption

Similar to the $\mathcal{O}$ notation, we use $\mathcal{S}$ for space complexity.

Convolution-based calculation produces the whole histogram array at once before it could be used for further calculation. This means $\mathcal{S}(m \cdot n^2)$ space complexity.

However, local histograms could be calculated directly using the pairwise weight calculation. This approach only needs to store one local histogram ($\mathcal{S}(m)$) the input image and the output image ($\mathcal{S}(n^2)$).

This means that there is a trade-off between space complexity and time complexity. The convolution-based algorithm has better asymptotic time complexity but an order of magnitude worse space complexity than the direct pairwise method.

6. Optimization

There are two easy ways to optimize performance. First, the number of shades can be effectively reduced. Locally, the only factor taken into account for a single pixel is the weighted number of pixels which have lower or higher values than the current given pixel. The number of shades could be reduced significantly without a noticeable difference in the output, if the values are binned with error diffusion, also known as dithering. When the local histograms are ready, the final intensities should be calculated using linear interpolation between the discretization levels. Usual histogram processing methods are not sensitive to a large amount of gray levels, and this dithered downsampling of the histogram has not been used before, to the best of our knowledge. Dithering is also used after the tone mapping in order to decorrelate the discretization error. Both dithering steps use the Floyd–Steinberg algorithm [48].

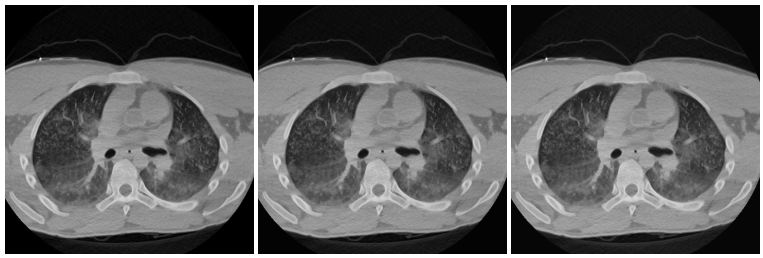
Another optimization possibility is to downsample the image during the histogram calculation, and perform the tone mapping using the downsampled histogram. Technically, this means a histogram column gets a contribution from several image pixels. No pixel data is lost during this downsampling, but the spatial resolution of the histogram array is decreased. This interpolation technique is mentioned in [11]. These histograms can be bilinearly interpolated for any interior point when it is required for the tone mapping.

Both spatial downsampling and dithering build on the fact that important features in the images usually have a larger area than a few pixels, and the local histograms do not change too fast; therefore, the 3D histogram can be approximated well with one which has reduced resolution along all axes. As long as these assumptions are valid, the tone mapping can approximately linearly speed up with downsampling in terms of the number of pixels and/or shades. Memory consumption scales down in a similar manner. The effect of the techniques can be seen in Figures 11 and 12 for dithering and downscaling, respectively. Note that the downscaling is meant for each axis. For instance, a 64× downsample means only eight samples along each axis, and 64 samples in total for a CT image with 512×512 pixels.

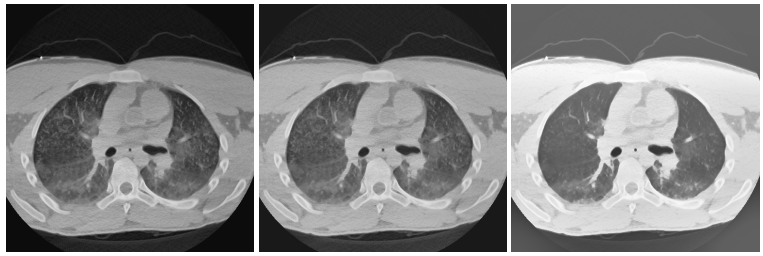
In both cases, the linear interpolation makes the approximation robust, which to a certain extent can compensate for the downsampling error, both in spatial and in color space. In exchange for precision loss, the computation time could be greatly reduced, as is shown in Figure 13. The execution time (t) is approximately linear in the number of pixels and in the discretizaton levels (D), not counting a small constant overhead (O). The number of pixels is inversely proportional to the square of the downscaling factor(x):

$$t \approx O + c\frac{D}{x^2}, \tag{19}$$

where c is a constant. The same result could be derived from Equation (18), with the $\log n \approx$ constant approximation.



(**a**) 2048 levels, S = 0.957 (**b**) 512 levels, S = 0.956 (**c**) 256 levels, S = 0.956



(**d**) 128 levels, S = 0.952 (**e**) 64 levels, S = 0.954 (**f**) 16 levels, S = 0.869

Figure 11. Calculating the histograms using a decreasing number of discretization levels. While quality slightly degrades after a while, the linear interpolation and dithering make the algorithm robust. TMQI structural similarity slowly decreases as the approximation becomes coarser.

(**a**) original resolution, S = 0.957 (**b**) 2× downsampled, S = 0.957 (**c**) 4× downsampled, S = 0.957

(**d**) 8× downsampled, S = 0.957 (**e**) 16× downsampled, S = 0.958 (**f**) 64× downsampled, S = 0.960

Figure 12. Calculating the histograms using spatial downsampling along each axis. Even significant downsampling does not cause very visible artefacts, which is also reflected in the TMQI score. However, local differences might appear, e.g., compare the middle region of the left lung in (**a**) and (**f**).

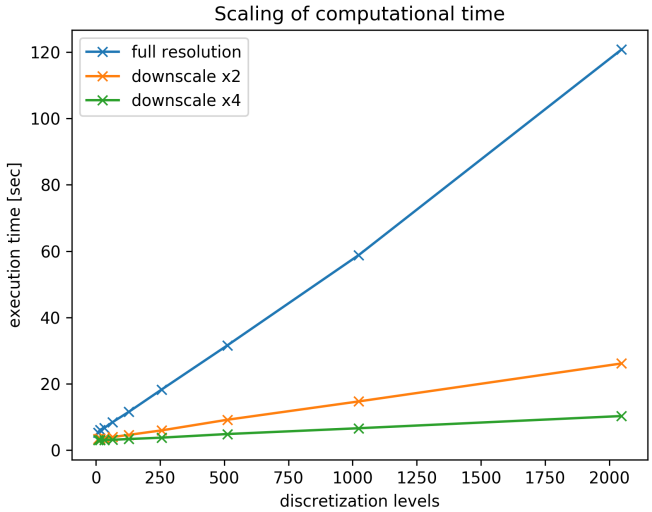


Figure 13. Approximate execution time scales with the number of pixels and the number of discretization levels plus a constant overhead because of data pre- and post-processing.

7. Conclusions

Our proposed method yields good local contrast for CT images while maintaining a similar image structure to the reference CT image. This could contribute to improving the visualization of pathologies. The proposed method performs well in terms of structural similarity compared to popular tone mapping algorithms. The computation cost can be effectively reduced with approximations.

Supplementary Materials: The following are available at http://www.mdpi.com/1999-4893/11/8/111/, algorithm source code.

Author Contributions: Methodology, D.V.; Software, D.V.; Supervision, A.C.T.M., A.S.-P., D.W. and M.P.; Visualization, D.V.; Writing—Original draft, D.V.; Writing—Review & editing, D.V., A.C.T.M., A.S.-P., D.W. and M.P.

Funding: This research was been funded by the Research Council of Norway through project no. 221073 'HyPerCept—Colour and quality in higher dimensions'.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barrett, J.F.; Keat, N. Artifacts in CT: recognition and avoidance. *RadioGraphics* **2004**, *24*, 1679–1691, doi:10.1148/rg.246045065. [CrossRef] [PubMed]
2. John, A.; Huda, W.; Scalzetti, E.M.; Ogden, K.M.; Roskopf, M.L. Performance of a single lookup table (LUT) for displaying chest CT images1. *Acad. Radiol.* **2004**, *11*, 609–616, doi:10.1016/j.acra.2004.01.002. [CrossRef] [PubMed]
3. Fayad, L.M.; Jin, Y.; Laine, A.F.; Berkmen, Y.M.; Pearson, G.D.; Freedman, B.; Van Heertum, R. Chest CT window settings with multiscale adaptive histogram equalization: Pilot study. *Radiology* **2002**, *223*, 845–852. [CrossRef] [PubMed]
4. Chang, A.E.; Matory, Y.L.; Dwyer, A.J.; Hill, S.C.; Girton, M.E.; Steinberg, S.M.; Knop, R.H.; Frank, J.A.; Hyams, D.; Doppman, J.L. Magnetic resonance imaging versus computed tomography in the evaluation of soft tissue tumors of the extremities. *Ann. Surg.* **1987**, *205*, 340–348. [CrossRef] [PubMed]
5. Seeram, E. *Computed Tomography-E-Book: Physical Principles, Clinical Applications, and Quality Control*; Elsevier Health Sciences: St. Louis, Missouri, USA, 2015.

6. Lehr, J.L.; Capek, P. Histogram equalization of CT images. *Radiology* **1985**, *154*, 163–169, doi:10.1148/radiology.154.1.3964935. [CrossRef] [PubMed]
7. Zuiderveld, K. *Graphics Gems IV*; Chapter Contrast Limited Adaptive Histogram Equalization; Academic Press Professional, Inc.: San Diego, CA, USA, 1994; pp. 474–485.
8. Cohen-Duwek, H.; Spitzer, H.; Weitzen, R.; Apter, S. A biologically-based algorithm for companding computerized tomography (CT) images. *Comput. Biol. Med.* **2011**, *41*, 367–379, doi:10.1016/j.compbiomed.2011.03.015. [CrossRef] [PubMed]
9. Acharya, T.; Ray, A.K. *Image Processing: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
10. Nikvand, N.; Yeganeh, H.; Wang, Z. Adaptive windowing for optimal visualization of medical images based on normalized information distance. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; doi:10.1109/icassp.2014.6853787.
11. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Gr. Image Process.* **1987**, *39*, 355–368, doi:10.1016/s0734-189x(87)80186-x. [CrossRef]
12. Kim, Y.T. Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Trans. Consum. Electron.* **1997**, *43*, 1–8, doi:10.1109/30.580378. [CrossRef]
13. Nikolova, M.; Wen, Y.W.; Chan, R. Exact Histogram Specification for Digital Images Using a Variational Approach. *J. Math. Imaging Vis.* **2012**, *46*, 309–325, doi:10.1007/s10851-012-0401-8. [CrossRef]
14. Stark, J. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **2000**, *9*, 889–896, doi:10.1109/83.841534. [CrossRef] [PubMed]
15. Pattanaik, S.N.; Ferwerda, J.A.; Fairchild, M.D.; Greenberg, D.P. A multiscale model of adaptation and spatial vision for realistic image display. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '98, Orlando, FL, USA, 19–24 July 1998; ACM Press: New York, NY, USA, 1998.
16. Farbman, Z.; Fattal, R.; Lischinski, D.; Szeliski, R. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM SIGGRAPH 2008 Papers on—SIGGRAPH '08*; ACM Press: New York, NY, USA, 2008.
17. Zhang, Z.; Su, Z. Tone mapping via edge-preserving total variation model. In Proceedings of the 2012 5th International Congress on Image and Signal Processing, Chongqing, China, 16–18 October 2012; doi:10.1109/cisp.2012.6469700. [CrossRef]
18. Tan, L.; Liu, X.; Xue, K. A Retinex-Based Local Tone Mapping Algorithm Using L 0 Smoothing Filter. In *Communications in Computer and Information Science*; Springer: Berlin, Germany, 2014; pp. 40–47.
19. Durand, F.; Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Gr.* **2002**, *21*, doi:10.1145/566654.566574. [CrossRef]
20. Reinhard, E.; Stark, M.; Shirley, P.; Ferwerda, J. Photographic tone reproduction for digital images. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '02, San Antonio, TX, USA, 21–26 July 2002; ACM Press: New York, NY, USA, 2002.
21. Fattal, R.; Lischinski, D.; Werman, M. Gradient domain high dynamic range compression. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '02, San Antonio, TX, USA, 21–26 July 2002; ACM Press: New York, NY, USA, 2002; doi:10.1145/566570.566573. [CrossRef]
22. Drago, F.; Myszkowski, K.; Annen, T.; Chiba, N. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Comput. Gr. Forum* **2003**, *22*, 419–426, doi:10.1111/1467-8659.00689. [CrossRef]
23. Reinhard, E.; Devlin, K. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. Vis. Comput. Gr.* **2005**, *11*, 13–24, doi:10.1109/TVCG.2005.9. [CrossRef] [PubMed]
24. Ferradans, S.; Bertalmio, M.; Provenzi, E.; Caselles, V. An Analysis of Visual Adaptation and Contrast Perception for Tone Mapping. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2002–2012, doi:10.1109/TPAMI.2011.46. [CrossRef] [PubMed]
25. Mantiuk, R.; Daly, S.; Kerofsky, L. Display adaptive tone mapping. In *ACM SIGGRAPH 2008 Papers on—SIGGRAPH '08*; ACM Press: New York, NY, USA, 2008.
26. Meylan, L.; Susstrunk, S. High dynamic range image rendering with a retinex-based adaptive filter. *IEEE Trans. Image Process.* **2006**, *15*, 2820–2830, doi:10.1109/tip.2006.877312. [CrossRef] [PubMed]

27. Fairchild, M.D.; Johnson, G.M. Meet iCAM: A next-generation color appearance model. In *Color and Imaging Conference*; Society for Imaging Science and Technology: Springfield, VA, USA, 2002; Number 1, pp. 33–38.
28. Reinhard, E.; Heidrich, W.; Debevec, P.; Pattanaik, S.; Ward, G.; Myszkowski, K. *High Dynamic Range Imaging*; Morgan Kaufman: San Francisco, CA, USA, 2006.
29. Mantiuk, R.K.; Karol, M.; Hans-Peter, S. High Dynamic Range Imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; pp. 1–42.
30. Banterle, F.; Artusi, A.; Debattista, K.; Chalmers, A. *Advanced High Dynamic Range Imaging: Theory and Practice*, 2nd ed.; AK Peters (CRC Press): Natick, MA, USA, 2017.
31. Kalender, W.A. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*; Publicis: Erlangen, Germany, 2011.
32. Barrow, H.; Tenenbaum, J. Recovering intrinsic scene characteristics. *Comput. Vis. Syst.* **1978**, *2*, 3–26.
33. Lepor, H. *Prostatic Diseases*; Number p. 966, v. 2000 in Prostatic Diseases; W.B. Saunders Company: Philadelphia, PA, USA, 2000.
34. Gross, B.H.; Kazerooni, E.A. *Cardiopulmonary Imaging*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2004.
35. Wright, F.W. *Radiology of the Chest and Related Conditions*; Taylor & Francis: London, UK, 2001.
36. Cooley, J.W.; Tukey, J.W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297, doi:10.1090/s0025-5718-1965-0178586-1. [CrossRef]
37. Blommaert, F.J.; Martens, J.B. An object-oriented model for brightness perception. *Spat. Vis.* **1990**, *5*, 15–41, doi:10.1163/156856890x00066. [CrossRef] [PubMed]
38. Mantiuk, R.; Myszkowski, K.; Seidel, H.P. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percep.* **2006**, *3*, 286–308, doi:10.1145/1166087.1166095. [CrossRef]
39. Yeganeh, H.; Wang, Z. Objective Quality Assessment of Tone-Mapped Images. *IEEE Trans. Image Process.* **2013**, *22*, 657–667, doi:10.1109/tip.2012.2221725. [CrossRef] [PubMed]
40. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Paramount, CA, USA, 2009.
41. Oliphant, T. *Guide to NumPy*; Continuum Press: Austin, TX, USA, 2015.
42. Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python, 2001. Available online: https://www.scipy.org (accessed on 12 June 2018).
43. Lam, S.K.; Pitrou, A.; Seibert, S. Numba. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC—LLVM '15*, Austin, TX, USA, 15–20 November 2015; ACM Press: New York, NY, USA, 2015.
44. Van der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T. scikit-image: Image processing in Python. *PeerJ* **2014**, *2*, e453, doi:10.7717/peerj.453. [CrossRef] [PubMed]
45. Van Rossum, G. PEP 8—Style Guide for Python Code. 2001. Available online: https://www.python.org/dev/peps/pep-0008 (accessed on 12 June 2018).
46. Rota, G.; Comida, F.; Anastasia, D. Luminance HDR. 2006–2017. Available online: chttp://qtpfsgui.sourceforge.net (accessed on 12 June 2018).
47. Eaton, J.W.; Bateman, D.; Hauberg, S.; Wehbring, R. GNU Octave Version 4.2.2 Manual: A High-Level Interactive Language for Numerical Computations. 2017. Available online: https://www.gnu.org/software/octave/doc/v4.2.2 (accessed on 12 June 2018).
48. Floyd, R.W.; Steinberg, L. An Adaptive Algorithm for Spatial Greyscale. *Proc. Soc. Inf. Disp.* **1976**, *17*, 75–77.

PAPER IV

algorithms

MDPI

Article

Image De-Quantization Using Plate Bending Model

David Völgyes [1,*], Anne Catrine Trægde Martinsen [2,3], Arne Stray-Pedersen [4,5], Dag Waaler [6] and Marius Pedersen [7]

1 Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway
2 Department of Physics, University of Oslo, 0316 Oslo, Norway; uxneti@ous-hf.no
3 Department of Diagnostic Physics, Oslo University Hospital, 0424 Oslo, Norway
4 Department of Forensic Sciences, Oslo University Hospital, 0424 Oslo, Norway; arne.stray-pedersen@medisin.uio.no
5 Institute of Clinical Medicine, University of Oslo, 0318 Oslo, Norway
6 Department of Health Sciences in Gjøvik, Norwegian University of Science and Technology, 2803 Gjøvik, Norway; dag.waaler@ntnu.no
7 Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway; marius.pedersen@ntnu.no
* Correspondence: david.volgyes@ntnu.no; Tel.: +47-4717-0724

Received: 21 June 2018; Accepted: 20 July 2018; Published: 24 July 2018

Abstract: Discretized image signals might have a lower dynamic range than the display. Because of this, false contours might appear when the image has the same pixel value for a larger region and the distance between pixel levels reaches the noticeable difference threshold. There have been several methods aimed at approximating the high bit depth of the original signal. Our method models a region with a bended plate model, which leads to the biharmonic equation. This method addresses several new aspects: the reconstruction of non-continuous regions when foreground objects split the area into separate regions; the incorporation of confidence about pixel levels, making the model tunable; and the method gives a physics-inspired way to handle local maximal/minimal regions. The solution of the biharmonic equation yields a smooth high-order signal approximation and handles the local maxima/minima problems.

Keywords: de-quantization; false contour removal; bit depth enhancement; biharmonic equation; partial differential equations

1. Introduction

Digital images are quantized signals, and quantization introduces quantization error. De-quantization approximates the original signal based on the digital samples. Most of the time, the intensities in the pixels of the digital images are represented with integer values in the range of $[0, 2^n - 1]$, where n is called the bit depth. Low bit depth images have fewer levels than the display which is used for presentation. De-quantization aims to decrease artifacts originating from the quantization process. Bit-depth extension or enhancement are also used to describe de-quantization.

One of the most disturbing artifacts of low bit depth images is the appearance of false contours. For instance, a sunset photo shows a large sky area, but the low number of available shades leads to large homogeneous regions, and a disturbing contour appears at the border between two regions (see Figure 1).

There are several ways to mitigate the problem. Dithering [1] was invented to de-correlate quantization error with the original signal. This is a very effective technique, and is still widely used.

Another class of low bit depth images were created using limited precision analog-to-digital converters and sensors, which has led to rounding or truncation. A large amount of old content exists

in such low bit depth format where a high quality source is not available, and the only way to improve quality is de-quantization and bit depth expansion.

Theoretically, both dithered and non-dithered images could be subjected to bit depth expansion, but the number of possible dithering methods is infinite, and every dithering method would require a dedicated algorithm. Most papers, including the present one, focus on non-dithered, truncated, or rounded images.

Figure 1. Example of non-continuous regions. The sky exhibits strong false contours, and the foreground ship partitions the sky into several regions. The photo was taken by the first author.

2. Related Work

Consider an image that is discretized using n bits. Any value between two discrete levels is rounded. The rounding strategy can be the nearest integer, or the nearest "smaller/larger than the signal" integer. In both cases, the original signal ($f_{original}$) is rounded to a discretized signal (g). The difference ($f_{original} - g$) is the lost information which is the target of the recovery operation . Using the $\mathcal{R}$ rounding operator and x, y spatial coordinates, the problem is as follows:

$$g(x,y) = \mathcal{R}\left\{f_{original}(x,y).\right\} \tag{1}$$

The lost signal must remain within certain limits. If the rounding operator is "rounding to the nearest integer", then the limit is ± 0.5 unit, but it could also be "rounding down" or other operators. The general formulation is:

$$\text{lower limit} \leq f_{original}(x,y) - g(x,y). \leq \text{upper limit} \tag{2}$$

What is the $f(x,y)$ approximation that makes the recovered signal the most plausible, most appealing?

$$f_{original}(x,y) \approx f(x,y). \tag{3}$$

Early strategies utilized blurring for false contour removal. For instance, Nishihara and Keda [2] used random blurring. Blurring and exploitation of the human visual system's properties remained a research interest, and led to low- and high-frequency separation techniques and adaptive cancellation, as reviewed by Daly and Feng [3]. The most challenging issue with this approach is the fact that

blurring might also affect real edges. The separation of false contours from real edges is a very difficult problem on its own.

The most recent approaches describe the false contour removal as a pure mathematical de-quantization challenge, and they do not aim to model the human visual system. Cheng's algorithm [4] uses flooding-based downward/upward distance maps to calculate step ratio and linearly interpolate. This method achieves very nice results for gradients, but fails to recover the local maximal/minimal (LMM) regions.

Wan et al. [5] developed a content-adaptive method using a virtual skeleton marking (VSM) model which transforms the 2D extrapolation problem into a 1D interpolation, but it is still based on linear interpolation. In a follow-up paper [6], a spatially varying sparsity-based de-quantization method was published, based on the observation that the second derivatives are positive/negative in the LMM regions and approximately zero in the other regions. The derived model is a Thikonov regularization-based anisotropic total variation minimization problem.

A similar approach was used by Ouyang [7], who developed a bounded interval regularization method, where the recovered signal must remain within limits, which is enforced with a penalty term outside limits. The algorithm in the allowed interval minimizes an energy function, which is quadratic in absolute value of the gradient ($|\nabla f|^2$). This formulation is equivalent to a bounded Laplace problem, where internal points (not exceeding the limits) satisfy the Laplace equation:

$$\Delta f(x, y) = 0. \tag{4}$$

This partial differential equation approach has several advantages, including the possibility of handling non-uniform quantization levels. However, it fails to reconstruct LMM regions, and perceived smoothness of the image might be low when the original signal's second derivative is significantly different from zero.

Wan et al. [8] developed an AC/DC separation model with maximum a posteriori AC estimation, where DC denotes the mean of the signal, and AC is the difference of the original signal and the DC component. They viewed the problem from a minimum mean square error perspective, and used a graph smoothness prior. The method works on blocks, where overlapping blocks are used to minimize artifacts at the edges.

3. Problem Statement

Partial differential equation-based signal de-quantization is a very appealing approach, but we identified the following areas where the previous approaches could be improved:

- non-continuous regions,
- incorporation of available additional information into the model,
- higher-order signal approximation for non-LMM regions.

Non-continuous regions appear when a foreground object splits the background into segments. Each segment belongs to the same underlying object (e.g., sky), and the false contours should be removed from all segments. The segments should match as if there was no foreground. See Figure 1 for an example.

The incorporation of available information means that sometimes additional information can be used to get a better approximation. Consider the following example: ski trails in white snow or a thin rope in a sunset have low contrast, and the difference between the foreground object and background might be just one discretization level. In these cases, it is very important to keep these edges, and removing them would degrade the impression of the images.

Finally, low-order signal approximation for medium intensities might cause a second-order false contour. By second-order, we mean that the intensity change is continuous, but the first derivative has a discontinuity and the observer might detect a break in the pace of lightness change.

4. Material and Methods

4.1. Plate Bending Model

The solution of the Laplace equation is similar to a stretched rubber plane, which has no stiffness. This is obviously a smooth solution, but it is also known that the solution always takes maximum value at the borders.

Reconstructing the LMM regions would be possible if the gradients from internal regions would be smoothly continued in the LMM regions, like a bended beam continues after the points were external forces bends the beam. This observation inspired a linear elasticity-based [9] approach:

- The signal is modeled as a thin plate.
- The maximum and minimum allowed values deform this plane.
- The stationary solution gives the approximation of the signal.

One simple model is the pure bending model, which can be described with the biharmonic equation:

$$\Delta^2 f(x,y) = \frac{\delta^4 f}{\delta x^4} + \frac{\delta^4 f}{\delta y^4} + 2\frac{\delta^2 f}{\delta x^2}\frac{\delta^2 f}{\delta y^2} = 0. \quad (5)$$

4.2. Boundary Conditions

The quantization error of the signal can be taken into account with an upper (U) and lower envelope (L). The original signal cannot be larger than the upper envelope, and cannot be smaller than the lower envelope:

$$L(x,y) \leq f(x,y) \leq U(x,y). \quad (6)$$

The distance between the envelopes does not have to be constant. If there is a priori knowledge about the signal, or the quantization levels are uneven, then appropriately chosen envelopes could take care of this problem. For instance, if the two envelopes touch each other, then it is equivalent with a Dirichlet boundary condition.

Numerical solution becomes much more tractable if mirrored periodic boundary conditions are assumed. This is a frequent assumption which does not limit the generality of the approach. Both "odd" and "even" mirroring could be used. In our examples, we used "odd" mirroring, which subtracts the reflected values two times from the edge value. Odd mirroring is depicted in Figure 2.

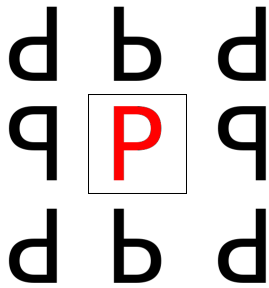


Figure 2. Two dimensional simplified example of mirrored periodic boundary conditions. The image is represented as a red P, and its edges are outlined with a black line. The reflected signals are in black.

Non-Continuous Regions

The envelopes should only be defined for the regions of interest. In irrelevant regions, the envelopes should take maximum/minimum allowed values. Using the plate analogy, supporting points (defined envelope values) can bend the plate. In irrelevant regions there is no external force;

the plate is continuous as the biharmonic equation governs it. We use a binary mask to define regions which should not be processed and where the envelope limits should be removed. After solving the the numerical problem, these regions are kept intact from the low-resolution input image.

4.3. Numerical Solution

The non-continuous regions and the additional information is taken into account through the upper–lower limits, and the higher-order smoothness is guaranteed by the biharmonic equation.

A straightforward solution can be implemented using finite differences schemes with a gradient descent method. However, a clipping step is required after every iteration step: if a value is lower than the lower envelope in the given point, or higher than the upper envelope, then approximation is set to be equal with the corresponding envelope. Laplace and biharmonic equations are part of the polyharmonic equations, and they can be described with similar equations where the difference is the order (p) of the Laplace-operator, which is 1 for Laplace and 2 for biharmonic operator. The gradient for time step is as follows:

$$\frac{\mathrm{d}f(x,y)}{\mathrm{d}t} = \Delta^p f(x,y). \tag{7}$$

While initially the solution function (f) does not satisfy the Laplace/biharmonic equations, evolving it with small time steps converges it to the solution. In this formulation, the solution of the problem is $f(x,y,t)$ in $t \to \inf$ limit.

During the discretization of the partial differential equations, we assume that the pixel sizes are equal in both dimensions, and the lowest-order central difference schemes are satisfactory for the numerical solution. With these assumptions, we use the following stencils approximating the Laplace and biharmonic operators:

$$\Delta \to \begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix} \tag{8}$$

$$\Delta^2 \to \begin{matrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & -8 & 2 & 0 \\ 1 & -8 & 20 & -8 & 1 \\ 0 & 2 & -8 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{matrix} \tag{9}$$

These stencils can be used everywhere in the image, including the edges, due to the periodic boundary conditions.

The convergence rate of this method is slow. Reducing execution time, the implementation uses a multi-grid approach where one level means a factor of 2 change in sampling points along an axis. First, both the envelopes and the initial image is downscaled. The image pixels take the average value of the higher-resolution pixels, the lower envelope takes the minimum of the higher-resolution envelope, and the upper envelope takes their maximum. This lower-resolution problem is solved first. When the approximation reaches convergence, a bicubic resampling is used to upscale the approximate solution, which serves as input for the next level. The pseudo-code of the iteration scheme is as follows:

```
1,   load initial image.

2,   if the smallest edge is longer than 31 pixels and divisible with two:

     2a,   downsample the problem with factor of 2:

           • lower-envelope:  minimum of grid points
           • upper-envelope:  maximum of grid points
           • image:  average of grid points
```

```
    2b,   solve the problem (jump to step 1)
    2c,   upscale the solution with a factor of two using bicubic interpolation
    2d,   use the upscaled image as initial image for the next step
3,  calculate gradient for the time step
4,  update the image
5,  clip the image if it exceeds limits (envelopes)
6,  continue with Step 3 if iteration number is less than specified
7,  return the image to higher level or as final solution
```

Alternatively, any other de-quantization method could be used as an initial input instead of Steps 2a–2d, and the algorithm should converge to the same solution.

5. Results

One-dimensional comparison can be seen in Figure 3 for demonstration purposes. The Laplace solution cuts the top and bottom of the dynamic range, while a higher-order solution allows overshoot and reconstructs values in the LMM regions.

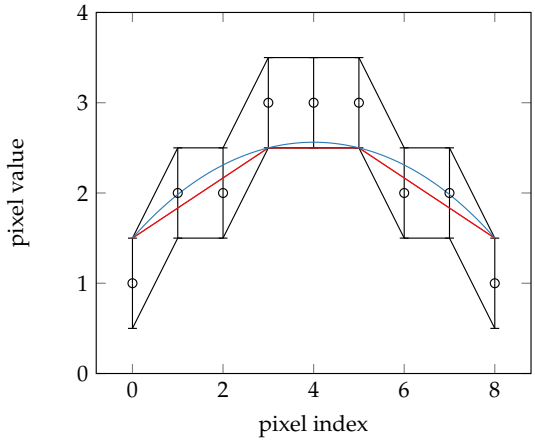


Figure 3. Discretized signal samples with upper and lower limits for the possible original function values. The minimized squared gradient magnitude (Laplace problem) assumption leads to a piecewise linear solution (red), while minimizing the bending energy or mean curvature leads to a higher-order smooth solution (blue).

Figure 4 shows two signals: a Gaussian and a Rosenbrock function [10]. The Gaussian does not have saddle points, while the Rosenbrock function is a frequently used test function in optimization due to its saddle-like shape with a relatively flat valley in the middle. The sub-figures show the high-bit-depth original, the reduced-bit-depth version, and the Laplace-reconstructed and biharmonic reconstructed solutions. The advantages of the higher-order approximation are obvious in the figure: the maximal region is smoothly recovered, and there is no second-order false contour.

Table 1 compares peak signal-to-noise ratio (PSNR) [11], normalized root mean square error (NRMSE), and structural similarity index (SSIM) [12] for a few selected (2, 3, 5) bit depths of Gaussian and Rosenbrock images extended with zero padding, Laplacian method, and with the proposed method.

Finally, an image of the night sky with the moon meant to demonstrate masking was used. When multiple objects were present in the scene, then the transitions led to overshoots. It is expected that the bended plate tries to keep its shape at the edges. However, if the image were segmented into

parts, then reconstruction might be applied partially, ignoring highly textured areas. Similarly, if the neighboring pixels differ in more than one discretization unit, then it can be assumed that they belong to different image components, and the assumption about signal smoothness does not hold. For an example, see the moon in Figure 5. If these regions (moon and sky) were not handled separately, then due to the stiffness of the plate in the model, the moon would generate a halo around itself. The synthetic examples were generated and the photo was taken by the first author.

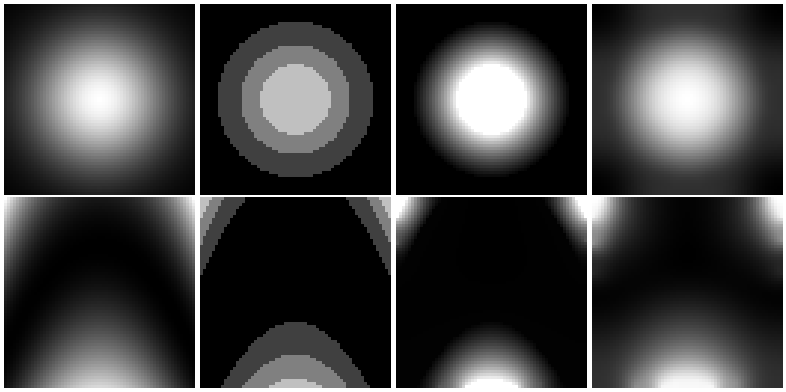


Figure 4. Top row: Gaussian function; Bottom row: Rosenbrock function. From left to right: 8-bit reference image, downscaled 2-bit image, Laplace reconstruction, and biharmonic reconstruction.

Table 1. Reconstruction of Gaussian and Rosenbrock functions evaluated with peak signal to noise ratio (PSNR), normalized root mean square error (NRMSE), and structural similarity index (SSIM) metrics. While at high bit depths both Laplace and biharmonic produced good results, at lower bit depth the biharmonic model had a significant advantage in all metrics.

Function	Bit Depth	Method	NRMSE	PSNR	SSIM
Gaussian	2	zero pad	0.3252	17.0239	0.4174
Gaussian	2	Laplace	0.3362	16.7344	0.4661
Gaussian	2	biharmonic	0.0713	30.2044	0.9462
Gaussian	3	zero pad	0.1618	23.0882	0.7585
Gaussian	3	Laplace	0.1535	23.5467	0.8187
Gaussian	3	biharmonic	0.1169	25.9098	0.8877
Gaussian	5	zero pad	0.0383	35.6112	0.9831
Gaussian	5	Laplace	0.0308	37.4985	0.9918
Gaussian	5	biharmonic	0.0371	35.8821	0.9885
Rosenbrock	2	zero pad	0.4011	19.3146	0.3118
Rosenbrock	2	Laplace	0.4457	18.3993	0.3855
Rosenbrock	2	biharmonic	0.2315	24.0895	0.8097
Rosenbrock	3	zero pad	0.2194	24.5547	0.5317
Rosenbrock	3	Laplace	0.2646	22.9290	0.4988
Rosenbrock	3	biharmonic	0.1520	27.7415	0.8251
Rosenbrock	5	zero pad	0.0561	36.4029	0.8999
Rosenbrock	5	Laplace	0.0721	34.2228	0.8555
Rosenbrock	5	biharmonic	0.0681	34.7163	0.9319

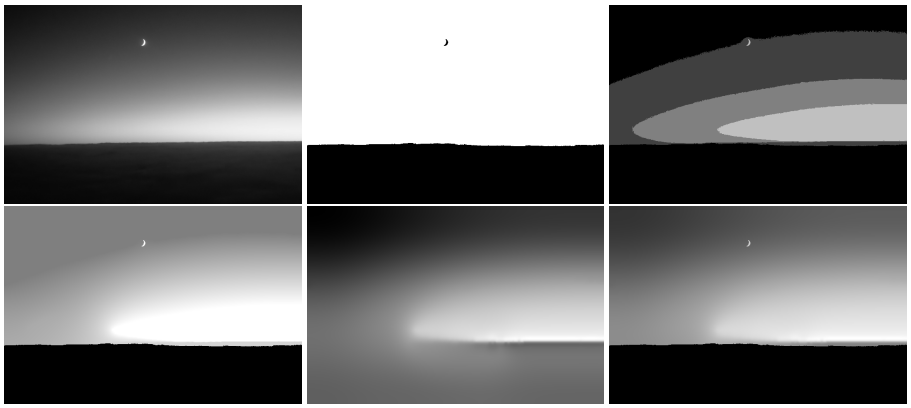


Figure 5. Reconstruction of night sky. Top row from left to right: original 8-bit image, mask for region of interest, 2-bit quantization. Bottom row from left to right: Laplace reconstruction with mask, biharmonic reconstruction solving inside the mask only, biharmonic reconstruction where the masked values were added back.

6. Discussion

One advantage of the biharmonic model is that it yields smooth surfaces. On the other hand, the bending model has an issue: it tries to continue the surface. This is very useful for large, related, but disconnected regions, but it leads to inferior image quality at the border of two different regions (e.g., hair and skin). This is easily understandable from the plate analogy. This is the trade-off for the handling of unconnected regions and higher-order approximation.

This could be mitigated with carefully chosen image masks where large flat regions are marked for reconstruction but textured regions should be left intact. For instance, the neighboring regions cannot have more than 1 unit difference, otherwise the difference does not originate from quantization error.

Note that limited applicability is a general phenomenon. The binary image is an extreme form of low-bit-depth images. The Laplace algorithm, Cheng's algorithm, and the proposed algorithm would assign a single value to every pixel in a binary image. This happens because the extreme conditions violate (implicit) model assumptions.

6.1. Dither and Noise

Another general challenge for bit depth enhancement is noise and artifacts from dithering. This affects every method, but more advanced methods tend to use more information from the image. Unfortunately, this also means higher sensitivity in handling the noise.

In the beginning of the article it was assumed that the image was not subject to dithering. However, if it was, it could generate serious artifacts: close to an edge, a dithering algorithm would generate points from both shades. There is a function which fits nicely to this data: a flat surface with equal distance from both quantization levels. This would lead to an image where there is a flat region in the neighborhood of the edges, which is exactly the opposite of the original aim. Further away from the edges, a smooth transition is achieved.

This artifact could be mitigated with relaxed upper and lower envelopes. The original assumption was that the rounding/flooring of the signal is perfect. Depending on the gradients in the image and the dithering method, a locally increased range for the allowed values (decreased lower envelope, increased upper envelope) would allow the algorithm to effectively mitigate dithering artifacts. However, the cost of this flexibility might be some blurring in neighboring areas. Similarly, noise which exceeds a quantization level deforms the plate.

From an algorithmic point of view, masking and relaxing is the same numerical problem, with two exceptions: in masked out points, the upper and lower envelopes are set to dynamic range limits, and after the iterative solution, in these points the original low bit depth pixels are used.

Adding a data fidelity term to the original problem, when the sum of the bending energy and the error term should be minimized, it would transform the problem into a denoising problem. This seems to be appealing, and modeling the signal as with a bended plate could be beneficial in some applications, however it does not suit the de-quantization problem. The reason lies in the fact that minimizing the error term would flatten out the reconstructed signal further from the edges. Technically, if the fidelity term is too small, it has no effect, and if it has significant effect, it would force the reconstructed signal to closely follow the flat regions.

Considering these arguments, our recommendation is to locally increase the distance between the envelopes, or remove the problematic pixels with noise filters or anomaly detection, as discussed in the next section.

6.2. Cross-Terms

There are previous higher-order polynomial approximations for this problem, but to our knowledge, the axes are handled individually. For instance, a fourth-order generalization of the Laplace equation without cross-terms could be an "almost biharmonic" equation where the cross-terms are downweighted, and while they give some contribution, their effect is much more subtle.

Not connecting the two axis has both advantages and disadvantages. In case of noise, it will not propagate into the other dimension, and will not generate wave-like patterns, but also smoothness will be limited. There are at least two ways to find balance between the approaches.

First, a lower-order stretching could be applied. In this approach, not only the plate bending energy should be minimized, which leads to the biharmonic equation, but an extra energy term which is proportional with the surface area. Applying this second energy term would lead to the Laplace equation. However, the sum of the two energies lead to non-trivial solutions, because the surface minimization tends to generate sharp edges at the border LMM regions. The weighting between the two energy terms could be chosen based on the application.

Second, anomalous points could be detected. The main idea is as follows: If the reconstructed function has extrema at a given point, and this point also lies at the border of two regions in the low-bit-depth image, then this originates from noise. The reason is simple. If a large smooth region has extrema, it is unlikely that it will be exactly on the edge of the borders. Therefore, this point should be masked. However, if the point was in fact a real extrema, removing it would not significantly change the image, because the rest of the points should still produce continuous smooth surface, and this process would regenerate the extrema, even without the control point. This process can be iteratively performed until all the anomalous points are taken care of. Actually, this approach can applied axis-by-axis, which can take care of possible ringing at edges. The main disadvantage of this approach is its time-consuming nature due to the iterations.

Third, alternative elasticity models could also be used. Referring to the literature, solving the biharmonic equation minimizes the mean curvature (B) [13] of a surface. Using principal curvatures (κ_1, κ_2), the mean curvature is as follows:

$$B = (\kappa_1 + \kappa_2)^2. \tag{10}$$

An alternative approach could be minimizing the total curvature (T) [13]:

$$T = \kappa_1^2 + \kappa_2^2. \tag{11}$$

This is equivalent to minimizing the thin plate energy functional. The biharmonic model is permissive towards saddle points, while the thin plate model prefers surfaces which can be approximated locally with a spherical surface.

The benefits of having higher-order plate-like approximation, having more traditional spline approximation, or a combined Laplace–biharmonic model depends on the content—especially on the structuredness, noise, and presence of saddle-like areas. From an application point of view, if similar high-resolution content exists, a simulated degradation–reconstruction process could be used to select the best solution from the aforementioned options in order to maximize a chosen metric, or to maximize observers' opinion score.

6.3. Mask Generation

Mask generation is a segmentation problem that is a big field on its own. Depending on the input image, even a simple strategy could work. False contours are only perceivable if the affected area is large enough, which means that the signal changes slowly locally. In this case, the quantized signal can change only one discretization level in a single step. If it changes more, then the edge belongs to an object border. However, if the difference between separate objects is only one discretization level, then this approach fails, and more sophisticated algorithms would be required (e.g., binary images). The mask in the example is hand-generated.

6.4. Advantages and Limitations

As far as we know, other approaches have not tried to recover signals in non-continuous regions. Our masking approach has the potential to handle such situations, but it requires a mask. Identifying which sub-regions should be handled together is a segmentation problem, and it is outside of the scope of this paper.

Does a single-pixel LMM region carry real information, or is it noise? The Laplace model effectively removes LMM regions (see the examples in [7]), and higher-order approximations (spline, biharmonic) try to take these points into account. If the pixel contains noise, these higher-order methods, including the proposed method, reconstruct a false local peak. On the other hand, if the data do not originate from noise, these methods better reconstruct the signal in the LMM region. In this case, the proposed model handles saddle points well, and yields good results for very low bit depths (2–3 bits).

The biharmonic equation is based on a simple linear elasticity model [9], where the bending energy should be minimal. This model assumes constant stiffness. It is easy to generalize this model to more general elastic models. However, the biharmonic model is high enough order for achieving smooth results but simple enough in terms of parameters, while applying statistical models [8] requires assumption about the signal distribution, and using neural networks [14] requires representative training data.

Handling non-equidistant discretization levels is trivial in the proposed model—only the upper and lower envelopes are needed to set appropriately.

In term of computational time, the proposed method is slower than many competing methods, because large flat areas need significant time to converge, and the convergence is also affected by the cross-terms. The Laplace method converges faster, and the Laplace operator can be more efficiently calculated than the biharmonic. However, elastic problems have been studied for a long time, and optimized solvers were developed to handle them [15], which means that even more generalized models would be tractable.

While the proposed method requires both masking and long computation until it achieves convergence, it processes the whole image. Many algorithms use block-based processing [8], which requires overlapping blocks in order to avoid artifacts, or try to detect the false contours, and operate around the contours [2,14]. Using blocks and/or the neighborhood of the false contours becomes a limitation at extremely low bit depths (2–3 bits), where processing blocks are smaller than the distance between the contours.

7. Implementation

Our algorithm is implemented in Python3.6 [16] using Numpy 1.13.3 [17], Scipy 1.0.1 [18], and Numba 0.38.0 [19]. The source code follows PEP8 [20] recommendations, and it is part of the supplemental material. All of the custom calculation codes and example images are available as a supplement to the paper.

8. Conclusions and Future Directions

Our proposed method based on the bending of deformable plates gives an intuitive way to handle non-continuous regions, and allows the incorporation of extra information into the model while ensuring a high-order signal approximation. To our knowledge, this is the first algorithm with such properties for the de-quantization problem.

Several questions remain open for future research. Mask generation requires the separation of foreground–background objects. It is a research area on its own, and low-bit-depth input makes it even more challenging.

Using biharmonic, Laplace, or other differential equations assumes an energy-like functional to be minimized. These are derived from physical assumptions such as minimal surface or minimal bending energy. The biharmonic equation is one of the simplest models in elasticity, with location-independent "stiffness". Total curvature minimization was mentioned as an alternative approach. More advanced physical models [9] (e.g., location-dependent elastic properties) could take further parameters into account. In general, elastic problems are most frequently solved with finite element methods. Solving these advanced models would require more computational power, and choosing the local parameters might be challenging.

If more advanced models are not derived from physical analogies, then machine learning could be used to learn signal properties. Normal images and their simulated down-quantized version could be used as training samples, similar to the super-resolution problem where this approach is already used [21], or to detect false contours [14] for later processing.

Finally, temporal consistency and high-throughput processing become interesting research questions when de-quantization is applied to video sources [22,23].

Author Contributions: Conceptualization, D.V.; Methodology, D.V.; Supervision, A.C.T.M., A.S.-P., D.W. and M.P.; Writing—original draft, David Völgyes; Writing—review & editing, D.V., A.C.T.M., A.S.-P., D.W. and M.P.

Funding: This research has been funded by the Research Council of Norway through project No. 221073 "HyPerCept—Colour and quality in higher dimensions".

Acknowledgments: The authors would like to thank Faouzi Alaya Cheikh and Ivar Farup for their feedbacks and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schuchman, L. Dither Signals and Their Effect on Quantization Noise. *IEEE Trans. Commun.* **1964**, *12*, 162–165. [CrossRef]
2. Nishihara, S.; Ikeda, K. False-contour removal by random blurring. *Comput. Gr. Image Process.* **1982**, *20*, 391–397. [CrossRef]
3. Scott J. Daly, X.F. Decontouring: Prevention and removal of false contour artifacts. In *Human Vision and Electronic Imaging IX*; Rogowitz, B.E., Pappas, T.N., Eds.; SPIE: San Jose, CA, USA, 2004; Volume 5292, pp. 1–20.
4. Cheng, C.H.; Au, O.C.; Liu, C.H.; Yip, K.Y. Bit-depth expansion by contour region reconstruction. In Proceedings of the IEEE International Symposium on Circuits and Systems, Taiwan, 24–27 May 2009; pp. 944–947.

5. Wan, P.; Au, O.C.; Tang, K.; Guo, Y.; Fang, L. From 2D Extrapolation to 1D Interpolation: Content Adaptive Image Bit-Depth Expansion. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, Melbourne, Australia, 9–13 July 2012; pp. 170–175.
6. Wan, P.; Au, O.C.; Tang, K.; Guo, Y. Image de-quantization via spatially varying sparsity prior. In Proceedings of the 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 953–956.
7. Ouyang, T.; Tumblin, J. *Removing Quantization Artifacts in Color Images Using Bounded Interval Regularization*; Northwestern University: Evanston, IL, USA, 2006.
8. Wan, P.; Cheung, G.; Florencio, D.; Zhang, C.; Au, O.C. Image Bit-Depth Enhancement via Maximum A Posteriori Estimation of AC Signal. *IEEE Trans. Image Process.* **2016**, *25*, 2896–2909. [CrossRef] [PubMed]
9. Gould, P.L. *Introduction to Linear Elasticity*; Springer: New York, NY, USA, 2013.
10. Rosenbrock, H.H. An Automatic Method for Finding the Greatest or Least Value of a Function. *Comput. J.* **1960**, *3*, 175–184. [CrossRef]
11. Huynh-thu, Q.; Ghanbari, M. Scope of validity of PSNR in image. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]
12. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
13. Greiner, G. Variational design and fairing of spline surfaces. In *Computer Graphics Forum*; Blackwell Science Ltd.: Edinburgh, UK, 1994; Volume 13, pp. 143–154.
14. Park, M.H.; Lee, J.; Park, R.H.; Kim, J.S. False contour reduction using neural networks and adaptive bi-directional smoothing. *IEEE Trans. Consum. Electron.* **2010**, *56*, 870–878. [CrossRef]
15. Logg, A.; Mardal, K.A.; Wells, G.N. *Automated Solution of Differential Equations by the Finite Element Method*; Springer: New York, NY, USA, 2012.
16. Van Rossum, G.; Drake, F.L. *Python 3: Reference Manual*; SohoBooks: New York, NY, USA, 2009.
17. Oliphant, T. *Guide to NumPy*; Continuum Press: Austin, Texas, USA, 2015.
18. Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python, 2001. Available online: https://scipy.org (accessed on 12 June 2018).
19. Lam, S.K.; Pitrou, A.; Seibert, S. Numba. In Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC-LLVM'15, Austin, TX, USA, 15–20 November 2015.
20. Van Rossum, G. PEP 8—Style Guide for Python Code. 2001. Available online: https://www.python.org/dev/peps/pep-0008 (accessed on 12 June 2018).
21. Romano, Y.; Isidoro, J.; Milanfar, P. RAISR: Rapid and Accurate Image Super Resolution. *IEEE Trans. Comput. Imaging* **2017**, *3*, 110–125. [CrossRef]
22. Ahn, W.; Kim, J.S. Flat-Region Detection and False Contour Removal in the Digital TV Display. In Proceedings of the IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6 July 2005.
23. Jin, X.; Goto, S.; Ngan, K.N. Composite Model-Based DC Dithering for Suppressing Contour Artifacts in Decompressed Video. *IEEE Trans. Image Process.* **2011**, *20*, 2110–2121. [CrossRef] [PubMed]

Complete list of publications

[1] D. Völgyes et al. “How Different Iterative and Filtered Back Projection Kernels Affect Computed Tomography Numbers and Low Contrast Detectability”. In: *Journal of Computer Assisted Tomography* 00.00 (2016), p. 1. ISSN: 0363-8715. DOI: 10.1097/RCT.0000000000000491.

[2] D. Völgyes et al. “Applicability of a clinical cardiac CT protocol in post mortem studies”. In: *Journal of Forensic Radiology and Imaging* 12 (2018), pp. 25–30. ISSN: 2212-4780. DOI: 10.1016/j.jofri.2018.01.003.

[3] D. Völgyes et al. “A Weighted Histogram-Based Tone Mapping Algorithm for CT Images”. In: *Algorithms* 11.8 (July 2018), p. 111. ISSN: 1999-4893. DOI: 10.3390/a11080111.

[4] D. Völgyes et al. “Image De-Quantization Using Plate Bending Model”. In: *Algorithms* 11.8 (July 2018), p. 110. ISSN: 1999-4893. DOI: 10.3390/a11080110.

[5] H. K. Andersen, D. Völgyes and A. C. T. Martinsen. “Image quality with iterative reconstruction techniques in CT of the lungs-A phantom study”. In: *European Journal of Radiology Open* 5 (2018), pp. 35–40. ISSN: 2352-0477. DOI: 10.1016/j.ejro.2018.02.002.

[6] D. Völgyes et al. *Mixing CT images*. The 7th National PhD Conference in Medical Imaging, Oslo, Norway. Nov. 2015.

[7] H. K. Andersen et al. “Comparison of different iterative and FBP reconstruction techniques with respect to image quality in chest CT examinations”. In: *ECR 2016 Book of Abstracts - B - Scientific Sessions and Clinical Trials in Radiology*. Vol. 7. 1. Mar. 2016, S185–S185. DOI: 10.1007/s13244-016-0475-8.

[8] D. Völgyes. *PerceptionMD*. Zenodo. July 2018. DOI: 10.5281/zenodo.1324409.

[9] D. Völgyes. *DCM2HDR: DICOM to HDR image conversion*. Zenodo. June 2018. DOI: 10.5281/zenodo.1290825.

[10] D. Völgyes. *Zenodo-get: a downloader for Zenodo records*. Zenodo. June 2018. DOI: 10.5281/zenodo.1261813.

[11] D. Völgyes, I. Voldsbekk and M. Strømstad. *Example MRI scans from sleep deprivation experiment*. July 2018. DOI: 10.5281/zenodo.1299153.

[12] D. Völgyes. *HDRIHaven: Public domain high dynamic range image dataset*. Zenodo. June 2018. DOI: 10.5281/zenodo.1285800.

[13] D. Völgyes. *Dual energy CT scan of ordinary objects*. Zenodo. June 2018. DOI: 10.5281/zenodo.1253035.

[14] M. Z. Pajak et al. "NEMA NU4-2008 Performance Evaluation of Albira: A Two-Ring Small-Animal PET System Using Continuous LYSO Crystals". In: *Open Medicine Journal* 3.1 (Apr. 2016), pp. 12–26. ISSN: 1874-2203. DOI: 10.2174/1874220301603010012.

[15] K.-E. Hines et al. "COMPET: High resolution high sensitivity MRI compatible pre-clinical PET scanner". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 732 (2013). Vienna Conference on Instrumentation 2013, pp. 581–585. ISSN: 0168-9002. DOI: 10.1016/j.nima.2013.05.148.

[16] M. Rissi et al. "Characterization of a high resolution and high sensitivity preclinical PET scanner with 3D event reconstruction". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 695 (2012). New Developments in Photodetection NDIP11, pp. 135–138. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.10.042.

[17] M. Rissi et al. "High resolution and high sensitivity PET imaging with COMPET". In: *Radiotherapy and Oncology* 102 (Mar. 2012), S131–S132. ISSN: 0167-8140. DOI: 10.1016/S0167-8140(12)70222-0.

[18] K. Hines et al. "COMPET a high-resolution, high-sensitivity pre-clinical PET scanner". In: *Proceedings of the International Workshop on New Photon-detectors (PhotoDet2012). June 13-15, 2012. LAL Orsay, France.* Vol. 1. 2012, p. 19.

[19] K. Nagy et al. “Presentation of Test Results of an Integrated Small Animal Sequential PET/MRI System”. In: *Europen Journal of Nuclear Medicine and Molecular Imaging*. Vol. 38. Springer 233 Spring ST, New York, NY 10013 USA. 2011, S289–S290.

[20] M. Magdics et al. “Performance evaluation of scatter modeling of the GPU-based “Tera-Tomo” 3D PET reconstruction”. In: *2011 IEEE Nuclear Science Symposium Conference Record*. Oct. 2011, pp. 4086–4088. DOI: 10.1109/NSSMIC.2011.6153777.

[21] *Detector modeling techniques for pre-clinical 3D PET reconstruction on the GPU*. Fully 3D Concerence. July 2011, pp. 375–378.

[22] *TeraTomo project: a fully 3D GPU-based reconstruction code for exploiting the imaging capability of the NanoPET/CT system*. World Molecular Imaging Congress. Kyoto, Japan, Sept. 2010.

[23] G. Hesz et al. “Timing calibration method for NanoPET/CT system”. In: *IEEE Nuclear Science Symposuim Medical Imaging Conference*. IEEE. Oct. 2010, pp. 2848–2850. ISBN: 978-1-4244-9105-6. DOI: 10.1109/NSSMIC.2010.5874314.

[24] P. Major et al. “Local energy scale map for NanoPET/CT system”. In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. Oct. 2009, pp. 3177–3180. DOI: 10.1109/NSSMIC.2009.5401698.

[25] E. Hild et al. “Characterisation of silica nanoparticulate layers with scanning-angle reflectometry”. In: *From Colloids to Nanotechnology*. Heidelberg, Germany: Springer Berlin Heidelberg, 2004, pp. 61–67. ISBN: 978-3-540-45119-8. DOI: 10.1007/978-3-540-45119-8_11.