Eindhoven University of Technology

MASTER

A system for computed rotational angiography

Habets, D.F.

Award date:
2000

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

TECHNISCHE UNIVERSITEIT EINDHOVEN

Department of Mathematics and Computing Science

MASTER'S THESIS

A system for Computed Rotational Angiography

by

D.F. Habets

Supervisors: D.W. Holdsworth Ph.D., dr. ir. J.J. van Wijk
Advisors: ir. N.H.L. Kuijpers, dr.ir. J.B.O.S. Martens, dr.ir. H.M.M. van de Wetering

December 2000

Abstract

Computed Rotational Angiography (CRA) is an x-ray technique that provides high-quality three-dimensional anatomical information for neuro-interventional procedures. CRA requires specific re-projection techniques that deliver high-quality interactive visualizations of CRA data.

This thesis describes how reconstruction algorithms back-project two-dimensional images to form a new three-dimensional volume. These CRA volumes are re-projected using two re-projection algorithms which together deliver the image quality and performance desired. The technical image quality of these re-projection algorithms are evaluated against Digital Subtraction Angiographic (DSA) images using a measure called: two-alternative forced choice (2AFC). To support proper registration of the two-dimensional DSA images to the three-dimensional CRA volumes a registration algorithm has been designed.

Keywords: computed tomography, volume rendering, technical image quality, 2D-3D registration, angiography, aneurysms, 2AFC.

Acknowledgements

I would like to thank Dr. David Holdsworth, who gave me the opportunity to work on this project and learn about the various aspects of medical imaging. Under his supervision I came to appreciate the specific requirements that medical imaging puts on Computer Graphics.

I would also like to thank Dr. Jack van Wijk, who helped organize all the information and knowledge that I gained while working on this thesis. Without his help the thesis would not have been on the right track.

I would also like to thank Mr. Michael Thornton and Dr. Ramesh Galigekere for their assistance and discussion about image construction and computing in general. Many thanks are also extended to Mr. Chris Norley, who helped me find the right pieces of data whenever I needed them.

I also would like to thank Dr. Brian Chapman for assistance with the two-alternative forced choice image quality measurement and for supplying the necessary software.

For making my time in Canada and at the Robarts Research Institute memorable:
Jeremy Gill, Floyd Harriot, Derek Hyde, Richard Goodine, Ali Sodagar and Bryon Sol.

Last but certainly not least I would to thank my parents who have made it possible for me to pursue my dreams and me supported throughout all of my decision.

Contents

List of Figures

List of Tables

List of abbreviations

2AFC	two-alternative forced-choice
2D	two-dimensional
3D	three-dimensional
AVM	arterio-venous malformation
CBP	convolution back-projection
CRA	computed rotational angiography
CT	computed tomography
DRR	digitally reconstructed radiograph
FOV	field of view
MRA	magnetic resonance angiography
RAD	rapid application development
ROC	receiver operator characteristic
SNR	signal-to-noise ratio
VTK	the visualisation toolkit
XRII	X-ray image intensifier

List of definitions

Angiogram An angiogram is a series of X-ray visualizations of the heart and blood vessels. A radiopaque substance, that is, a material that does not allow the passage of X rays through it, is injected into a vein or artery, and X-ray pictures are then taken in rapid succession. The series of pictures reveals the size and shape of veins or arteries in organs and tissues. An angiogram is used as a diagnostic tool with certain diseases.

Digital Subtraction Angiography DSA is a procedure in which a computerized technique is used to observe arterial blood circulation. In DSA, X-ray images of arterial blood vessels are taken both before and after the introduction of a radiopaque dye. These images are converted into digital data; and a computer "subtracts" one set of data from another with the resulting image showing only the dye and, therefore, the shape of the artery through which the blood is flowing.

Endoscopy Endoscopy is an examination of the interior organs and body cavities using a light and pliable tube (endoscope). Endoscopes make use of fiber optics, fine glass tubes that conduct light. Endoscopy can be done either via a natural body opening or a surgical incision.

Fluoroscope A fluoroscope is a machine that aids in diagnosis. It projects continuously X-ray pictures of what is happening inside a person's body onto a screen. The screen is activated when X rays strike it. By using a fluoroscope, a physician can study the movements of internal organs and the passage of contrast materials through those organs.

Third generation scanner A third-generation scanner has the x-ray source (the x-ray tube) as its apex, and this configuration is referred to as a source fan. Third-generation scanners require excellent detector calibration and stability.

Fourth generation scanner A fourth-generation scanner has a single detector as the apex of the fan beam, a so-called detector fan. Fourth-generation scanners require excellent x-ray tube output stability.

Sinogram Data structure in which the raw CT data projections are stored one after another as the projection angle is incremented. If the imaged object is a small point, then this object will trace out a sine wave on the sinogram when viewed as a 2D image.

Source fan see third generation scanner

1 Introduction

This chapter introduces the basic concepts of and motivation for advanced three-dimensional (3D) imaging for cerebrovascular diseases. The main two clinical problems that computed rotational angiography (CRA) is intended for, will be described briefly, followed by an overview of the currently used imaging techniques. Next CRA is introduced and its benefits for these clinical problems are discussed. Finally the research goals and approach taken in this thesis are described and an overview of this thesis is given. A list of definitions of medical terms is given as an appendix.

1.1 Motivation: 3D imaging during neurovascular interventional therapy

The brain, being the 'controller' behind the human body, has a complex network of vessels that are vital for the correct function of the human brain and therefore the human body. Sometimes this complex network of vessels is diseased, these so-called cerebrovascular diseases can influence daily life as well as cause life threatening situations. The standard approach to diagnose these cerebrovascular diseases is by imaging of the brain.

Recent techniques for neurovascular interventional therapy for the treatment of various cerebrovascular diseases have been made possible through the use of image guidance provided by real-time x-ray fluoroscopy and Digital Subtraction Angiography (DSA). Two of the most significant clinical applications are:

1. embolization for intracranial (or cerebral) aneurysms, and
2. embolization of cerebral arterio-venous malformations (AVMs).

In both applications an improved imaging technique may provide benefits for diagnosis, treatment planning and therapy. Currently however, no conventional medical imaging system provides a combination of real-time performance and quantitative, high-contrast 3D anatomical and functional information during neuro-interventional procedures. This thesis focuses on improved imaging techniques that will aid in the first two stages: diagnosis and treatment planning.

1.1.1 Intracranial aneurysms

An intracranial aneurysm is a cerebrovascular disease that manifests itself as a pouching or ballooning of the vessel wall. Aneurysms are commonly found in the brain along the course of an artery, although they could develop within any area of the human body. Intracranial aneurysms occur in at least 2% of the population and aneurysm rupture occurs in over 30,000 cases each year in North American alone [3, 4]. Treatment options are:

1. clipping of the intracranial aneurysm, and

2. coil embolizations using small platinum coils (FDA approved September 1995), see figure 1.

Angiographic imaging is used in all stages of treatment for intracranial aneurysms. High-resolution DSA images are acquired at various view angles and are used to determine the most appropriate therapy: surgical clipping or endovascular occlusion (in the form of coil embolization). This choice is influenced by the size and shape of the aneurysm and aneurysmal neck which are determined from multiple DSA images. During endovascular therapy, fluoroscopic images are used for catheter guidance and DSA images are used to assess aneurysm filling.

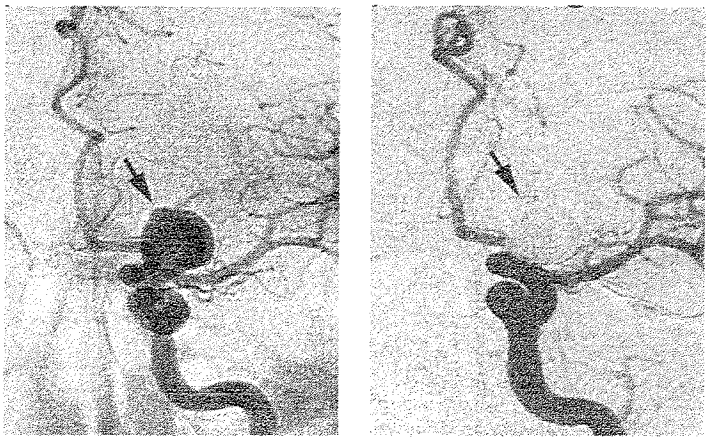

Figure 1: a) a DSA image of a cerebral aneurysm, before embolization with platinum coils. b) Same aneurysm after platinum coil embolization. The coils in figure (b) have been subtracted out.

Risks to the patient during endovascular therapy are:

1. stroke, due to thrombus production on the coils and/or catheter;
2. systemic reaction to the contrast material; and
3. deterministic and stochastic effects of ionizing radiation.

1.1.2 Arteriovenous malformations

AVMs are a congenital disorder of the blood vessels in the brain where an abnormal connection between arteries and veins exist. It has been estimated that over 400,000 individuals in North America have an intracranial AVM, affecting mostly individuals between 20 and 40 years of age. These lesions are usually detected in patients as the result of a seizure or hemorrhage. Treatment options are:

1. surgical excision;

2. radio-surgical treatment; and
3. endovascular embolization using a liquid or particulate embolic agent (e.g. cyanoacrylate glue), see figure 2.

Treatment planning for surgery, radio-surgery or embolization depends on detailed knowledge of the 3D location and volume of the AVM. Quantification of the AVM volume is made from DSA images. As with aneurysm embolization, real-time fluoroscopic guidance is essential during endovascular intervention and DSA images are used to assess treatment success.

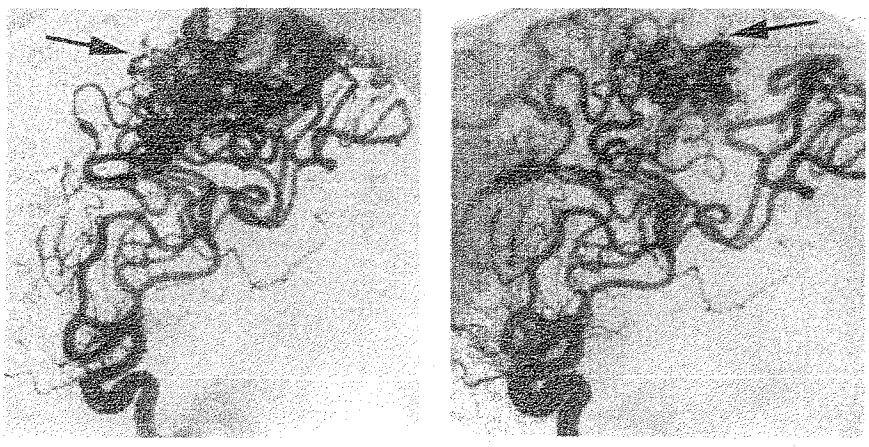

Figure 2: a) Angiogram of pre-embolized AVM, b) Angiogram of post-embolized AVM showing reduction in nidus size.

Risks to the patient during AVM therapy are:

1. all the risks of catheterization describe above; and
2. incorrect catheter placement and cyanoacrylate glue delivery.

1.2 Imaging techniques

Because of the importance of imaging in medicine, various imaging techniques have been introduced: DSA, MRA, spiral CT and CRA. Only CRA is able to provide a combination of real-time and high resolution 3D anatomical information during an intra-arterial neuro-interventional procedure. An overview of the various imaging techniques and their physical imaging mechanism is given in table 1.

1.2.1 Digital Subtraction Angiography

DSA is the current gold standard for cerebrovascular disease imaging because of its ability to produce high resolution 2D images and the facilitation of providing immediate subtracted images and road mapping capabilities. Its limitations however are:

	x-ray	magnetic resonance
2D	DSA	-
3D	CRA, spiral CT	MRA

Table 1: An overview of imaging techniques and their physical imaging mechanism

1. the inability to provide high resolution 3D images during neuro-interventional procedures; and
2. the exposure of the patient to ionizing radiation doses.

1.2.2 Magnetic Resonance Angiography

Magnetic resonance angiography (MRA) is commonly used during diagnosis of intracranial vascular disease and recent advances in contrast-enhanced MRA increase the likelihood that MRA will replace intra-arterial DSA in many cases. However, MRA has certain major limitations which must be overcome before MRA can be used to guide interventional therapy. These limitations are:

1. signal loss due to slow flow or disturbed flow (saturation);
2. real-time imaging capabilities are still at the developmental stage, interventional systems are not widely available;
3. the contrast resolution is less than that obtained using DSA; and
4. signal-to-noise ratio (SNR) and spatial resolution are less than obtained using DSA.

1.2.3 Spiral Computed Tomography

The most recent spiral (or helical) CT scanners acquire contiguous images during patient translation following intravenous injection of contrast agent, from which a 3D volume can be derived. Spiral CT scanners have been used during diagnosis and treatment planning, but for use during interventional therapy several limitations have to be overcome. These limitations are:

1. the CT gantry reduces access to the patient;
2. the spiral CT scanner as yet has no capability for fluoroscopic imaging (unless combined with a conventional DSA);
3. up to a 30 second breath hold may be required for contrast imaging;
4. spatial resolution is not isotropic; and
5. small-vessel resolution is limited by partial volume artifacts.

1.2.4 Computed Rotational Angiography

The limitations of the previous clinical imaging systems must be overcome before the full potential of intra-vascular neuro-radiological procedures can be realized. CRA has been developed to accomplish this goal.

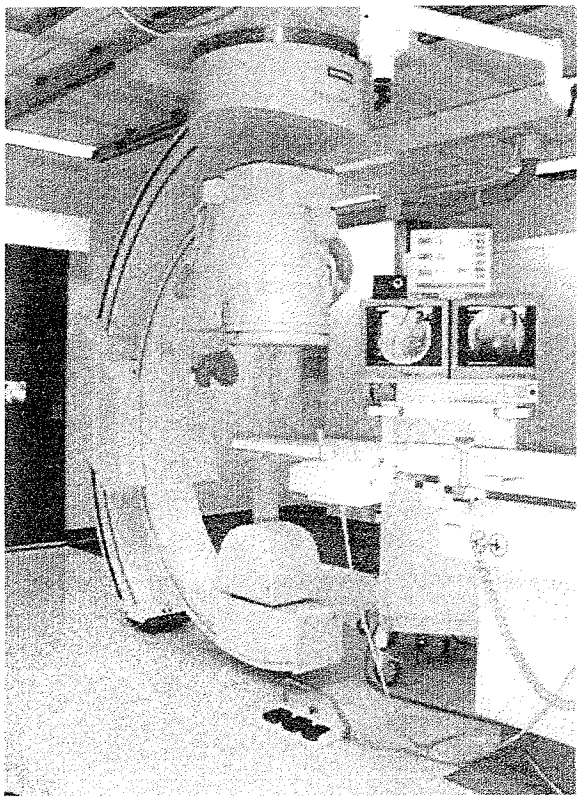

Figure 3: The CRA system (Siemens Multistar) is compromised of a C-arm mounted x-ray image intensifier for producing high contrast, high resolution 3D images during neuro-interventional procedures.

Unlike competing systems, the CRA system is based on a modified DSA system, see figure 3. This DSA system is used to collect x-ray projections around a patient during an intra-arterial contrast injection. DSA uses these x-ray projections from many angles (200^o at 45^o /s acquired in only 4.4 s) around the patient for 3D CT reconstruction [1, 2]. This results in a volume CT reconstruction from 2D projected data of contrast enhanced vessels.

1.3 Research goals

Using CRA it is possible to generate 3D volumes of an object of interest, which in this case is the brain vasculature. To present these volumes software is needed

that correctly produces an interactive 3D visualization. This leads to the first goal:

- To develop software for interactive 3D visualization of CRA data.

Since the target audience of this software are radiologists, the main concern is to present the 3D volumetric data in a meaningful way. Radiologists are used to working with DSA images while diagnosing various cerebrovascular diseases, therefore it is useful to generate simulated DSA images from these 3D CRA volumes, such simulated DSA images will be called Digitally Reconstructed Radiographs (DRR). Hence, the next goal is:

- To produce diagnostic-quality DRRs at any angle.

For these DRR to be useful in a clinical situation, they have to be compared with the current gold-standard DSA images. This leads to the final goal, which is:

- to compare DSA and DRR images with a user image quality study.

These three main goals lead to various sub-goals which are described in detail in the following chapters.

1.4 Outline

This section describes the following chapters and how they are related.

Chapter 2: Computed Rotational Angiography

This chapter gives an introduction to CRA reconstruction, which is based on a re-projection of 2D projected images to 3D volumes using CT. It will introduce the noise and errors inherent to x-ray imaging and give the Feldkamp algorithm that is used to perform the CT computations. Possible future enhancements are also covered.

Chapter 3: Visualization

This chapter both covers the DRR construction and the interactive 3D visualization of the CRA data. Several compromises had to be made, specifically compromises between speed and image quality. Several future enhancements are also covered.

Chapter 4: Image quality comparison

This chapter describes the image quality comparison between DRR and DSA images. This comparison is performed using a user image quality study known as two-alternative forced-choice (2AFC). 2AFC is introduced in this chapter and the involved procedures are explained. To minimize any misregistration influences on the study, a 2D-3D image registration is used to obtain equivalent DRRs matching with DSAs. Use of the 2AFC methodology in a small patient study is also discussed.

Chapter 5: Conclusion and future work

In this chapter the results obtained, limitations and future improvements are discussed. Attention is given to future applications of the CRA technique and possible necessary changes are covered.

2 Computed Rotational Angiography

Computed Rotational Angiography is a 3D x-ray imaging system that can produce high-resolution angiographic images of the cerebral vasculature. To fully understand the nature of these images it is necessary to first introduce some basic parameters affecting x-ray image quality. This is then followed by an overview of the CRA volume reconstruction procedure, which is based on a modified Feldkamp cone-beam algorithm. Finally some results of these reconstructions are shown and current limitations are discussed. The inverse of CRA construction is CRA visualization, therefore a thorough understanding of the CRA construction will help in understanding the following chapters.

2.1 Overview of x-ray imaging

The x-ray images obtained with the C-arm, are the data used to construct the CRA volume, therefore the imaging parameters affecting the x-ray image quality, will also affect the CRA volume quality. The scanner that is used, is of the so-called "third generation" type with a fan angle of 18°, see figure 4. A third-generation scanner has the x-ray source (the x-ray tube) as its apex, and this configuration is referred to as a source fan.

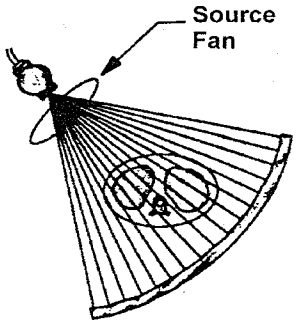

Figure 4: A 3rd generation scanner.

Several parameters, such as the x-ray generator type or scatter reduction used, exist that affect the x-ray image quality, but these are beyond the scope of this thesis, see table 2.

Of all these parameters the following two are of importance for CRA construction:

1. x-ray exposure; and
2. the focal spot selection.

The x-ray exposure influences the image density, the image contrast and the patient exposure. Image density is defined as the blackening of the x-ray image

and is influenced by three factors: the x-ray tube current and exposure time, and the x-ray tube voltage. An increase in tube current causes a proportional increase in the number of x-rays reaching both patient and image detector, and therefore influences the image density. The tube voltage has a more dramatic effect on image density as is stated by the 10-kVp rule[1] while increasing patient exposure.

Tube voltage selection also influences image contrast, which is defined as the difference in radiographic density of adjacent anatomic structures. Changing the x-ray tube voltage will directly change the x-ray beam energy, which influences the penetrability of the x-rays. A higher tube voltage will result in a reduction of contrast compared with low kilovoltage techniques, but when the tube current is adjusted such that the same amount of radiation reaches the image detector, then the high kilovoltage technique uses a much lower radiation exposure, due to the fact that x-ray beam is more penetrating and less absorption occurs in the patient.

Patient x-ray exposure is thus dependent on the exposure time, the tube current and the tube voltage. One goal of CRA is to minimize patient exposure and contrast fluid usage.

The focal spot size primarily influences the amount of geometric unsharpness, but also affects motion blur, due to the limiting effects it has on tube current and exposure time. An increase in magnification or focal spot size will result in an increase in image blur. Motion blur is affected as follows: the larger a focal spot the more motion blur will be suppressed. Therefore the focal spot size will be determined weighting the influence of motion blur and image blur.

Motion blur has to be considered in CRA, because bone tissue is removed by volumetric subtraction of two CRA volumes, one constructed using x-ray images with no contrast fluid in the vessels, the other using x-ray images with contrast fluid. A large amount of motion blur results in a poor subtraction.

Factors affecting x-ray image quality
x-ray exposure
focal spot size
detector size and spacing
pixel matrix
patient motion
scanner motion
x-ray generator type
x-ray scatter rejection method used

Table 2: Overview of parameters affecting image quality.

In summary, all the various x-ray parameters directly affect the x-ray image quality and indirectly the quality of the CRA volume. These x-ray parameters are inter-related and affect the x-ray image quality in various ways. An optimal set of values for these parameters is determined based on various issues, such as the application of the x-ray images, the object being imaged, etc.

[1] An increase of 10 kVp is equivalent to doubling the milliampere-seconds.

2.2 CRA volume construction

Angiographic imaging is important in all stages of the treatment of intracranial aneurysms and AVMs. Computed rotational angiography (CRA) has overcome the limitations of current systems and is able to provide 3D image guidance during interventional therapy, while reducing patient X-ray dose, contrast usage and providing qualitative and quantitative 3D capabilities.

This section introduces CT reconstruction and how it is used in CRA. CRA reconstruction requires some additional corrections beyond standard CT reconstruction, which are given. The main artifacts that may occur in the CRA (and in CT in general) are given as an appendix. Finally some results are presented.

2.2.1 CT reconstruction: convolution back projection

The problem of CT reconstruction is defined as follows: construct an unknown two-dimensional image (or function) given the ray sums (or line integrals) at a finite number of positions and angles, see figure 5.

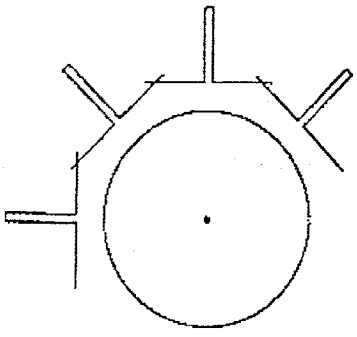

Figure 5: A point object and some of its projections.

In radiography this unknown function is the attenuation coefficients over a cross section. The ray sums are derived by making x-ray transmission measurements and then taking their logs. To give an exhaustive and complete description of CT reconstruction would be beyond the scope of this thesis, but some references are [7, 8, 9, 10, 11].

Convolution back projection is one of many reconstruction methods available. To describe this method, the principle is used that if a good reconstruction system can be made for an image consisting of a single point, then it is also possible to construct a reconstruction of any image, since an image is a collection of points.

Now consider figure 5 having a unity valued pixel in the center with all others zero. All projections are the same for all directions. Back projection simply distributes each projection measurement equally along the entire line, resulting in a blurred reconstruction of the point, as shown in figure 6. The reconstructed image of the points has its maximum at the center where the rays converge, while decreasing with increasing radius.

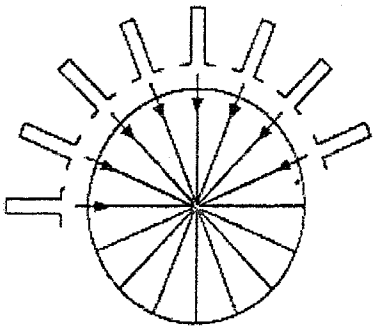

Figure 6: Back projection of a point object.

To prevent this blurring, each projection is modified prior to back projection so that the blurring will be canceled and a sharp point will remain. The modified projected point for removing blur is shown in figure 7. The negative lobes in the modified projected point cancel, when back projected, the blur and results in the desired reconstructed point.

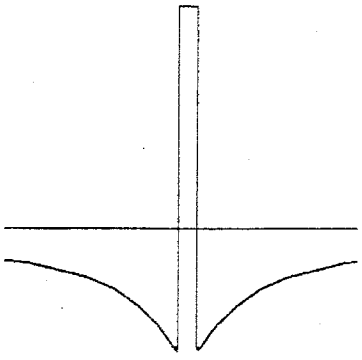

Figure 7: A modified projected point for removing blur.

Now that a correct reconstruction can be made for a single point, it is also possible to reconstruct any image. Each projection is decomposed into an array of points and these are convolved with the waveform in figure 7. These filtered, or convolved, projections are now used for back projection.

Mathematically convolution back projection is described as follows (also see figure 8):

$f(r, \varphi) = \int_0^{2\pi} \int_{-\infty}^{\infty} p(\theta, t)\, h(r\, \cos(\theta - \varphi) - t)\, dt\, d\theta$
where
(r, φ) denotes a point in the target image in polar coordinates;
$p(\theta, t)$ is the raw projection data as a function of the detector linear position t and the X-ray source rotation angle θ; and
$h(t)$ is the convolution kernel.

Several choices for the filter $h(t)$ can be made. An ideal filter is a ramp in the

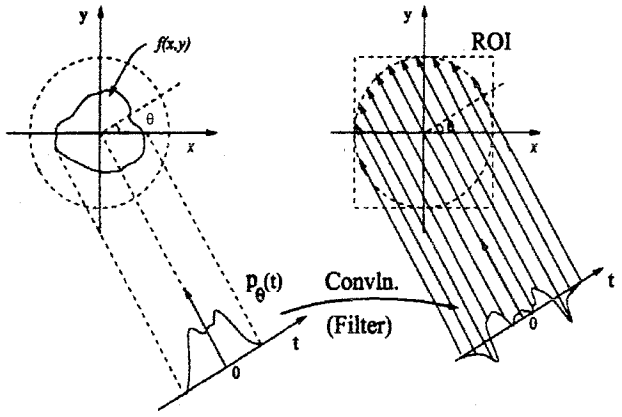

Figure 8: parallel-beam CT geometry

frequency domain, i.e. the filter value is the absolute value of the frequency[7]. A well-known filter is the Shepp-Logan filter, which is used in the current implementation of CRA reconstruction. The Shepp-Logan filter is defined as follows:

$C_{SHLO}(t) = \frac{2}{d}|\sin(\pi\, t\, d)|(\frac{\sin(\pi\, t\, d)}{\pi\, t\, d})^2$
where
d is the pixel spacing in the projection image.

The convolution kernel is now constructed from taking the Fourier transform of these filters. For a detailed description of this and various other convolution kernels, please refer to [7, 13, 14].

2.2.2 CT reconstruction: fan-beam data

The convolution back projection as described in the previous section uses a parallel geometry. More commonly used however is the fan-beam geometry, see figure 9. The fan-beam back projection is derived from the parallel-beam geometry with a simple coordinate transformation using the Jacobian.

A surface integral, I, of $f(x, y)$ is defined by $I = \int\int f(x, y)\, dx\, dy$. This integral can be transformed from the (x, y) coordinate system to another system (α, β) with the use of the Jacobian J to give:

$I = \int\int f(x(\alpha, \beta), y(\alpha, \beta))\, J\, d\alpha\, d\beta,$
where
$J = \begin{vmatrix} \frac{\delta x}{\delta \alpha} & \frac{\delta x}{\delta \beta} \\ \frac{\delta y}{\delta \alpha} & \frac{\delta y}{\delta \beta} \end{vmatrix}$
where $x = x(\alpha, \beta)$ and $y = y(\alpha, \beta)$

Fan-beam CT for a stationary axis of rotation The goal is to transform the back projection integral from parallel-beam coordinates (t, θ) to the fan-beam coordinates (s, β) for a flat detector. The coordinate transform is given by:

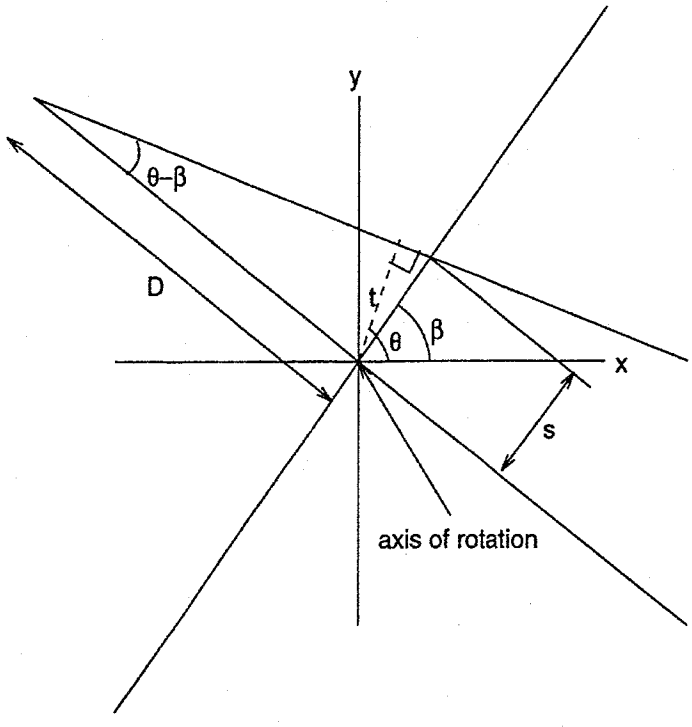

Figure 9: Fan-beam CT geometry with a flat detector and a stationary axis of rotation.

$t(s, \beta) = sq$
$\theta(s, \beta) = \beta + \tan^{-1}(s/D)$
where
$q = \frac{D}{\sqrt{s^2+D^2}}$ and
D is the distance of the X-ray source to the origin on the reconstruction system.

Using the Jacobian the reconstruction for fan-beam geometry with a flat detector becomes:

$$J(s, \beta) = \begin{vmatrix} \frac{\delta t}{\delta s} & \frac{\delta t}{\delta \beta} \\ \frac{\delta \theta}{\delta s} & \frac{\delta \theta}{\delta \beta} \end{vmatrix} = q^3$$

Then the new back projection equation becomes

$f(r, \varphi) =$
$\int_0^{2\pi} \int_{-\infty}^{\infty} p(\theta(s, \beta), t(s, \beta))\, q^3\, h(r\, \cos(\beta + \tan^{-1}(s/D) - \varphi))\, ds\, d\beta$

which (after some calculation) is simplified to

$f(r, \varphi) = \int_0^{2\pi} \frac{1}{U^2} \int_{-\infty}^{\infty} p(\theta(s, \beta), t(s, \beta))\, q\, h(s' - s)\, ds\, d\beta$
with $s' = \frac{D\, r\, \cos(\beta-\varphi)}{r\, \sin(\beta-\varphi)+D}$ and $U = \frac{r\, \sin(\beta-\varphi)+D}{D}$

Hence, there are three steps involved in the reconstruction process for fan-beam geometry with a flat detector:

1. weight the raw projection data $p(\theta(s, \beta), t(s, \beta))$ by the factor q;
2. convolve the weighted projection data with the kernel $h(s)$; and
3. back project the convolved projections while applying the weighting factor $1/U^2$.

Fan-beam CT for a non-stationary axis of rotation The back projection integral given above is for a stationary axis of rotation, however in reality this is too simplistic, therefore it is expanded to a back projection integral for a non-stationary axis of rotation.

The goal is once again to transform the back projection integral from parallel-beam coordinates (t, θ) to the fan-beam coordinates (s, β) for a flat detector, however now the axis of rotation is non-stationary. The coordinate transform is now given by:

$t(s, \beta) = (s - \tau)\, q$
$\theta(s, \beta) = \beta + \tan^{-1}(s/D)$
where
$q = \frac{D}{\sqrt{s^2+D^2}}$,
D is the distance of the X-ray source to the origin on the reconstruction system and
τ is the perpendicular shift between the rotation axis and the mid-line of the fan beam.

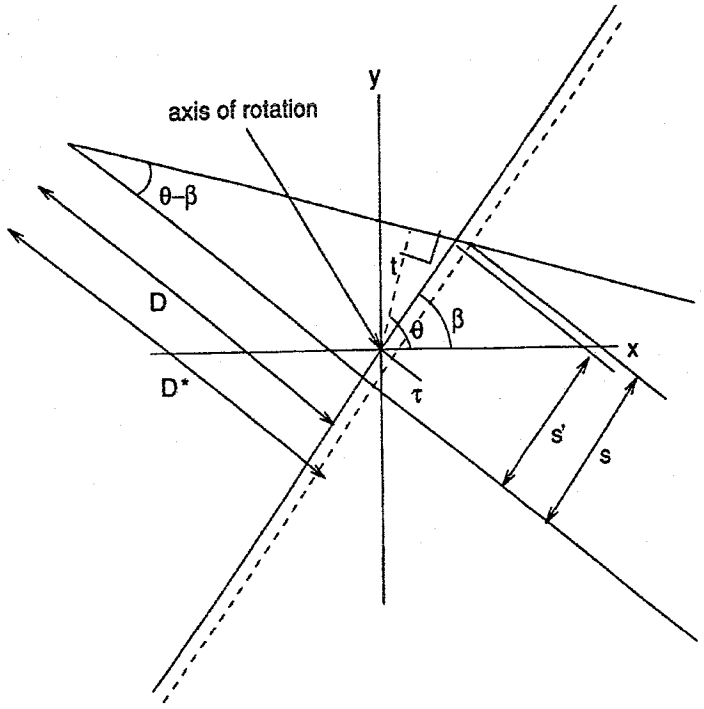

Figure 10: Fan-beam CT geometry with a flat detector and a non-stationary axis of rotation.

Using the Jacobian the reconstruction for fan-beam geometry with a flat detector and a non-stationary axis becomes:

$$J(s,\beta) = \begin{vmatrix} \frac{\delta t}{\delta s} & \frac{\delta t}{\delta \beta} \\ \frac{\delta \theta}{\delta s} & \frac{\delta \theta}{\delta \beta} \end{vmatrix} = (D^2 + \tau\, s + D\Delta(\tau)) \frac{D}{(D^2+s^2)^{3/2}}$$

where
$\Delta(\tau)$ is the partial derivative of τ with respect to β.

Then the new back projection equation (after some calculation) becomes

$$f(r,\varphi) = \int_0^{2\pi} \frac{1}{U^2} \int_{-\infty}^{\infty} p(\theta(s,\beta), t(s,\beta)) \frac{D+\tau\, s/D+\Delta(\tau)}{\sqrt{D^2+s^2}} h(s'-s)\, ds\, d\beta$$

with
$s' = \frac{D\, r\, \cos(\beta-\varphi)+\tau\, D}{r\, \sin(\beta-\varphi)+D}$ and
$U = \frac{r\, \sin(\beta-\varphi)+D}{D}$

Hence, there are three steps involved in the reconstruction process for fan-beam geometry with a flat detector and a non-stationary axis of rotation:

1. weight the raw projection data $p(\theta(s,\beta), t(s,\beta))$ by the factor $\frac{D+\tau\, s/D+\Delta(\tau)}{\sqrt{D^2+s^2}}$;
2. convolve the weighted projection data with the kernel $h(s)$; and
3. back project the convolved projections while applying the weighting factor $1/U^2$.

2.2.3 CT reconstruction: cone-beam data

The back projection algorithms described above are for 2D geometries, but the scanning geometry of data acquired with the C-arm system is 3D in nature. The Feldkamp algorithm is an extension of the fan-beam back projection algorithm to an ideal Feldkamp geometry which approximates the scanning trajectory of the C-arm system, see figure 11.

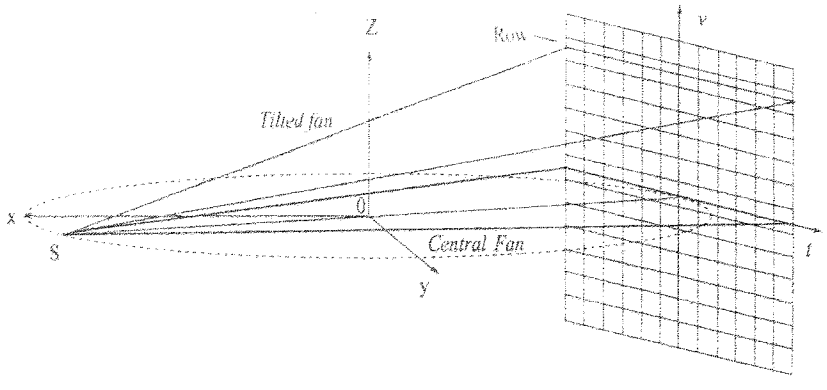

Figure 11: The Feldkamp geometry.

The data acquired using the C-arm system is a cone-beam of a 3D object, but this cone-beam consists out of a set of tiled fan-beams and a central fan which has an ideal reconstruction. The goals in now to use the 2D fan-beam algorithm, as given earlier, for a 3D cone beam coordinate system. This is done by using the perspective projection p of the voxel v, from S onto the detector at angle (β).

$f(r, \varphi, z) = \int_0^{2\pi} \frac{1}{U^2} \int_{-\infty}^{\infty} p(\theta(s, \beta), v, t(s, \beta))\, q\, h(s' - s)\, ds\, d\beta$
with
$\theta(s, \beta)$ and $t(s, \beta)$ defined as before,
$v = (\frac{D}{r \sin(\beta-\varphi)+D})\, z$,
$U = \frac{r \sin(\beta-\varphi)+D}{D}$,
$s' = \frac{D\, r \cos(\beta-\varphi)}{r \sin(\beta-\varphi)+D}$,
$q = \frac{D}{\sqrt{s^2+v^2+D^2}}$ and
D is the distance of the X-ray source to the origin on the reconstruction system.

Hence, there are three steps involved in the reconstruction process for fan-beam geometry with a flat detector:

1. weight the raw projection data $p(\theta(s, \beta), v, t(s, \beta))$ by the factor q;
2. convolve the weighted projection data with the kernel $h(s)$; and
3. back project the convolved projections while applying the weighting factor $1/U^2$.

The Feldkamp algorithm is an approximate cone-beam algorithm as opposed to the exact cone-beam reconstruction algorithms given in [19, 20]. The three advantages of approximate cone-beam reconstruction are:

1. incomplete scanning loci can be used;
2. the computational efficiency is high; and
3. artifacts may be less with the approximate reconstruction.

The exact cone-beam reconstruction algorithms have to satisfy the completeness condition, which states that there exist at least a source position on any plane intersecting an object to be reconstructed. This condition may not be satisfied in many cases of X-ray CT (i.e. CRA).

The approximate reconstruction uses typically less raw data than an exact reconstruction. Also the computational structure is straightforward and allows for a parallel reconstruction of ROI.

Image noise, blurring and ringing artifacts may be less with approximate reconstruction. For instance the direct Fourier method in [19] resulted in more ringing than the Feldkamp method. It appears that ringing is inherent to all exact cone-beam reconstruction formulas that take the second derivative of data.

2.2.4 CRA volume reconstruction

CRA volume reconstruction uses Feldkamp's algorithm to reconstruct the CRA volume, however several additional steps are performed, which are listed here.

Prior to reconstruction of the image data several corrections are applied to the projection data. The first correction is an angle-dependent 2D polynomial distortion correction to account for the effect of the earth's magnetic field on the electronic optics of the XRII [2]. The second correction is a correction of the non-ideal, but reproducible, motion of the gantry [16]. And finally all images are corrected for variations in pixel sensitivity before reconstruction.

Reconstructions are then carried out using the modified Feldkamp cone-beam algorithm to perform convolution-back-projection over (π + fan angle), resulting in reconstructed volumes with an isotropic voxel spacing of 0.56 or 0.3mm depending on the field of view (FOV) used during acquisition (40 or 20 cm).

2.3 Results

Using the techniques given in the previous sections a CRA system has been build using a prototype C-arm DSA system (Multistar, Siemens Medical Systems, Erlangen) at the J.P. Robarts Research Institute. This equipment and custom-made software allows for rapid cone-beam reconstruction of volumes as large as 400^3 voxels and in vitro and in vivo model has been used for investigations of CRA performance. Figure 12 shows several slices through the CRA data showing good spatial resolution and the validity of the correction algorithms. Note that all of the original x-ray projections required to produce these slices were acquired in just 4.4s.

Figure 12: 3 example results: an axial CRA slice, a sagittal CRA slice and the same CRA slice with window-level suppressing background tissue. All examples have had the bone tissue subtracted.

3 Visualization

As discussed in the previous chapters, 3D image reconstruction is increasingly being used in medical applications for diagnosis, therapy planning, image-guidance during intervention and patient examination. This chapter describes a system developed for visualizing the CRA data. First background information and a characterization of the possible re-projections is given, followed by a section on interactive visualization and a section on diagnostic quality visualization. Finally some results are presented and possible future improvements are discussed.

3.1 CRA re-projection

Using the techniques described in the previous chapter it is possible to back-project the various DSA images such that a CRA volume of the desired object, in this case the cerebral vasculature, is constructed. This 3D structure needs to be re-projected as a 2D image such that a radiologists can use these 2D projections for e.g. diagnosis and treatment planning. The benefit of using the CRA volume and re-projections thereof to generate new 2D images is threefold:

1. a patient does not need to be subjected to additional radiation and contrast fluid;
2. projections that are difficult to construct using standard DSA are easily created; and
3. animations or interaction reveal the 3D structure of the vasculature.

3.1.1 CRA data

Before a description is given of the chosen visualization approach, it is important to first understand the type of volume that is used and create a feeling for the need of (among other things) perspective distortion in these re-projections.

As stated before the object being imaged is the cerebral vasculature and this imaging is done with x-rays while injecting contrast fluid. This contrast fluid ensures that the x-rays through the vessels will be highly attenuated and therefore have a high intensity in the resulting image. However, background tissue (e.g. bone tissue) is also present in the volume and therefore this volume needs to undergo a preprocessing step to remove unwanted signals, i.e. the background tissue. Several approaches for this preprocessing step are available, from which for CRA usually volumetric subtraction is chosen. A mask (non-contrasted) CRA volume is obtained during the same session as the contrast-enhanced data followed by a voxel-by-voxel subtraction of spatially-registered volumes, which effectively removes the background tissue. The technique is sensitive to patient motion, however the Contrast-to-Noise-Ratio (CNR) is much greater than when using grey-scale segmentation to separate vessel, soft-tissue and bone tissue. This subtracted CRA volume is then used to visualize the cerebral vasculature.

The current gold standard in imaging of the cerebral vasculature is DSA. These DSA images have perspective distortions which add important depth-cues for a better understanding of the 3D structure. It is important to realize that this perspective distortion is dependent on the acquisition geometry and that this same geometry must be used for creating the desired re-projections of the CRA volume.

3.1.2 Re-projection types

Volume rendering in general is concerned with the re-projection of 3D volume to a 2D image. This re-projection involves compositing several voxels into one pixel such that a meaningful representation of the data is constructed. The compositing function determines in what way these voxels are combined to obtain a pixel and determine the final appearance of a re-projected volume. Various compositing choices are possible, but for the intended application two are of importance: maximum intensity projection and radiographic (or x-ray) projection.

Maximum Intensity Projection Maximum Intensity Projection (MIP) is a common way to visualize angiographic volume data. This technique is fairly forgiving when it comes to noisy data and produces images that provide an intuitive understanding of the underlying data. MIP is the least computationally intensive compositing function of the two discussed in this section since only the maximum of a set of voxels is desired. One problem however is that it is not possible to tell from a still image where this maximum value occurred along a ray. A common solution is to present a movie, where each frame is a MIP image rotated around a selected axis, resulting in a better understanding of the three-dimensional structure.

Intuitively a MIP re-projection of the CRA data is an image with for each pixel p, the maximum intensity (or density, located at sample point s) $\mu(s)$ coming from ray direction r, in other words:

$$I_{MIP}(p, r) = max(\mu(s)) \text{ where } 0 \leq s \leq length(r)$$

Therefore a ray does not have to be completely traversed as soon as a sample point s reaches full intensity.

Also a modified version of the MIP re-projection has been proposed by various authors (e.g. [17]) adds a depth cue to the image in order to improve the 3D impression of the image. This depth cue MIP (dMIP) uses an intensity ramp that dims parts of the volume further away from the image plane.

Intuitively a dMIP re-projection of the CRA data is an image with for each pixel p, the maximum intensity (or density, located at sample point s) $f_{depth}(\mu(s), s)$ coming from ray direction r, in other words:

$I_{dMIP}(p, r) = max(f_{depth}(\mu(s), s))$
where
$0 \leq s \leq length(r)$ and
$f_{depth}(\mu(s), s)$ is the intensity mapping

The compositing function now favors points closer to the view point. As a results, a back-to-front re-projection will have to traverse the whole ray, but a front-to-back re-projection can stop traversing the ray as soon as an intensity mapped sample point s reaches full intensity.

Digitally Reconstructed Radiograph As stated earlier a Digitally Reconstructed Radiograph (DRR) is a perspective radiographic re-projection of the CRA data. Such a DRR can be made for arbitrary values for the following 5 intrinsic parameters:

- the pixel size (width, height);
- the source-detector distance; and
- the image center (x, y),

and for the following 6 extrinsic parameters:

- the x, y and z shift; and
- the alpha, beta and gamma rotation.

Intuitively DRRs are re-projections of CRA data with for each pixel p on the image plane, the attenuation line integral accumulated along a ray direction r, in other words:

$$I_{SUM}(x, r) = \int_0^{length(r)} \mu(s)\, ds$$

This effectively simulates the behavior of x-rays going through an object. The resulting DRRs (after image processing) can be used in Algebraic Reconstruction Techniques (ART) to improve the quality of the CRA data, see [18, 24, 31].

3.1.3 Fundamental re-projection algorithms

The various types of re-projections have been defined above, but no details were given how the various voxels are sampled. In practice there are three main approaches to this sampling problem: ray-, voxel- and texture-driven.

Ray-driven (or image-order) re-projection determines the value of each pixel in the re-projection by sending a ray through the pixel into the volume according to geometric parameters. This ray is then sampled and the resulting values are combined using one of the compositing functions. It is important to realize that the ray only in rare cases passes through voxel centers, therefore interpolation is a requirement. In the following section three common interpolation algorithms will be given. The ray-casting approach clearly is computationally intensive since each sample requires interpolation, many samples are used to determine the value of one pixel and many voxels are used multiple times.

Voxel-driven (or object-order) re-projection processes samples in the volume based on the organization of the voxels in the dataset and the specified geometry.

A voxel is projected onto the view plane and using the compositing function the pixel is updated. This approach to re-projection is fast and efficient, however it is more prone to image artifacts than ray-driven approaches due to the discrete selection of the projected image pixel. Splatting is also an example of a voxel-driven technique, see [32].

Texture-driven re-projection consists out of two steps, first the volume is re-sampled using some form of interpolation, then slice-by-slice each voxel slice is combined with the current image in the frame buffer. The main advantage of this approach is that part of the re-projection can be done in hardware, however this also brings about several limitations. Texture-driven re-projections tend to have more artifacts than ray-driven approaches due to limited precision of the frame buffer for storing partial results at each pixel during blending. Also, only a few compositing functions are supported by hardware.

Interpolation Interpolation is a fundamental concept of re-projection and (in part) it determines the resulting quality of the re-projection. Various interpolation methods exist, but only three will be discussed in this section: nearest neighbor, linear and cubic interpolation.

Nearest neighbor interpolation is performed by selecting the nearest voxel and using this voxel as the sampled point. Nearest neighbor interpolation is the simplest and fastest interpolation method, but can suffer from severe artifacts, such as shifting of images which results in a blocky appearance of the final re-projection. Shifting can occur up to one-half of a pixel and this prohibits proper image registration.

Linear interpolation (or tri-linear for the 3D case) is one of the most popular techniques. It offers a better quality at the cost of more computation. The interpolation formula is:

$$f(d) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} d + f(x_0)$$

where

d is the distance from the location at which to interpolate to the first sample x_0

Linear interpolation is easily extended to multiple dimensions and results in better quality re-projections, however some artifacts still occur due to the fact that it is not a radial interpolation.

Cubic convolution results in an even better re-projection quality, however it is far more computationally involved than linear interpolation. The cubic convolution function's weights are:

$$w(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1 & 0 \leq |x| \leq 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 \leq |x| \leq 2 \\ 0 & 2 \leq |x| \end{cases}$$

The value a allows for control of the shape of the kernel and allows for sharpening. The usual range of values for a is between -3 and 0, where -3 results in the strongest sharpening, in most cases -0.5 is the best value to use. Cubic convolution results in the best re-projection quality, but requires the costly computation of the kernel coefficients.

For an overview of the exact number of calculations required for the various interpolation methods on a three-dimensional signal, see table 3.

	Nearest neighbor	Linear	Cubic convolution
Multiply	0	7	52
Addition/Subtraction	3	14	39

Table 3: Computational cost of various interpolation methods

3.1.4 Commercially available solutions

Various solutions are currently commercially available for performing these re-projections (or volume rendering). However these solutions have severe limitations which make them inappropriate for the goals set for CRA and DRR.

Of the currently available software packages, none was able to properly visualize the CRA data because of several limitations in their design. These limitations are :

1. most packages produce parallel re-projections, since these are less computationally involved than perspective re-projections;
2. most packages will only generate MIP re-projections, but are not capable of generating DRRs; and
3. several packages introduced image blurring due to poor interpolation.

Perspective re-projection is essential for DRRs to compete with DSAs, which by physical construction contain perspective distortions that give clues to the volumetric structure being imaged. MIPs, which for reasons given in the previous section are less desired than radiographic re-projections, will make a final comparison against DSA images impossible.

The only hardware solution currently commercially available, the VolPro card, had also several limitations:

1. limited scaleabilty (no more than 128mb voxel memory);
2. only parallel re-projections; and
3. no proper support for Linux.

For these reasons it was decided not to use existing solutions, but to design and implement algorithms that are capable of producing the required DRRs and MIPs. The algorithms have been developed entirely in software (C++ using VTK), rather than with accelerated hardware, to ensure compatibility with future microprocessor architectures. This approach will also give complete control over the quantitative aspects of images during re-projection, and allow the porting of the algorithms to the fastest available CPU in future clinical implementations. The set of algorithms used are described in the next section. Currently the algorithms are available on a Dec Alpha 600MHz processor with the Linux operating system.

3.1.5 System overview

Interaction requires fast re-projection, but generally has less strict requirements for image quality. Diagnostic quality re-projection values image quality and has less strict requirements for re-projection time. Therefore a difference is made between re-projection for interaction and diagnostic quality re-projection. As follows from the previous section, ray-driven projection algorithms result in the highest re-projection quality, therefore this is the choice of algorithm to produce the diagnostic quality re-projections. Voxel-driven algorithms are much faster, but their image quality is less and therefore these type of algorithms are perfectly suited for interaction quality re-projections. Texture-driven algorithms are not used, because of the limitations previously described.

Interaction with the CRA volume consists out of rotating the volume, clipping the volume and changing the view angle of the geometry. Also the opacity transfer function, which maps voxel scalars to opacities, may be altered, however this operation is usually done only once. The user determines the opacity function such that only the vasculature is visible. The color transfer function, which maps voxel scalars to an intensity, is fixed to a gray-scale ramp.

Diagnostic quality re-projections are created after the user has selected an appropriate view by changing the rotation, clipping and view angle of the volume. This re-projection may therefore take seconds as opposed to the interactive re-projection with must be as fast as possible. This diagnostic quality re-projection may be window-leveled as to optimize contrast.

The approach chosen here is the use of progressive refinement. The re-projection software progressively renders the image in a higher quality depending on the requested frame rate. If a high frame rate is requested, which is the case during interaction, then the fastest algorithm is used, however in periods of inactivity the higher quality rendering will occur and seamlessly replace the lower quality re-projection.

3.2 Interactive visualization

Interactive visualization requires low rendering times, but image quality is a lesser concern. As stated before the possible interactions (listed by importance) are:

- volume rotations;
- volume clipping;
- view angle changes;
- opacity transfer function changes; and
- source-detector distance changes.

The type of algorithms used for re-projection during interaction are object order algorithms that use spatial coherence to compress the volume and to allow for

skipping of transparent voxels. Object-order algorithms are inherently more efficient than image-order algorithms, because the volumetric data is accessed in storage order as opposed to an arbitrary view-dependent direction. Real-life data inherently contains a degree of coherence, therefore using this coherence is an efficient way to gain re-projection speed while preserving image quality [34].

The opacity transfer function determines the opacity of each voxel, therefore the compression is dependent on this function. This results that two object-order algorithms are used for the interactive visualization. One algorithm uses a pre-computed Run-Length-Encoded (RLE) ([39]) representation of the volume, which assumes a fixed opacity transfer function. The other algorithm uses an un-encoded volume, a min-max octree and a summed-table for rendering while changing the opacity transfer function. The min-max octree is used by a background process to create a new RLE representation followed by a seamless switch between the two object-order algorithms.

3.2.1 The shear-warp factorization

The algorithms used for interactive rendering are based on a factorization of the viewing transformation such that the projection from 3D volume to 2D image are simplified, as described in [28]. This factorization involves using an an efficient mapping from the object coordinate system to an intermediate coordinate system, which allows for an efficient 2D re-projection. This intermediate coordinate system is the sheared object space and is defined as:

- The coordinate system where, by construction, all viewing rays are parallel to the third coordinate axis.

This definition implies that for a perspective re-projection each slice must be scaled as well as translated to obtain parallel viewing rays.

An overview of the object-order perspective re-projection algorithm using the shear-warp factorization is as follows:

1. The volume is sliced such that the slices are most perpendicular to viewing direction.
2. Each slice is translated, scaled and then re-sampled using one of the previously defined interpolation functions.
3. The re-sampled slices are then traversed in a front-to-back order using one of the previously defined compositing functions. This results in a 2D intermediate image in sheared object space.
4. The 2D intermediate image is now transformed to image space by warping.

The perspective shear-warp factorization The perspective shear-warp factorization is given here, for the parallel shear-warp factorization see [28]. The perspective shear warp algorithm was first described by Lacroute however it has been adapted here to fit problem specific issues. The algorithm has been

adapted such that it allows for a faster MIP, dMIP and DRR generation by using a more efficient representation of the volume. This same classification is also used for a novel ray-driven algorithm which is described in the section on diagnostic quality visualization.

The perspective shear-warp factorization assumes that a four-by-four viewing transformation M_{view} and the volume dimensions (N_x, N_y, N_z) are known. The viewing transformation matrix M_{view} transform homogeneous points from object space into image space. First a definition of the various coordinate systems is necessary: object coordinates, standard object coordinates, sheared object coordinates and image coordinates. These definitions follow the system described by Lacroute.

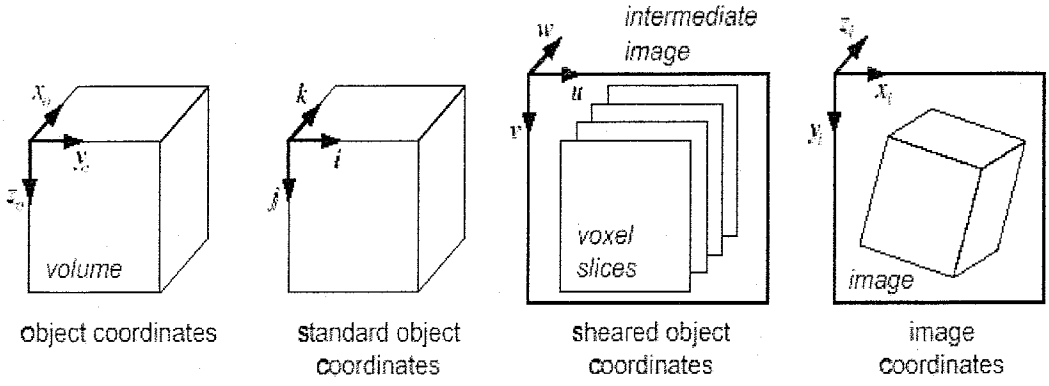

Figure 13: The four coordinate systems

The *object coordinate system* is the coordinate system of the volumetric data. The origin is located at one corner and the unit distance along each axis equals the length of one voxel along that axis. These axes are labelled x_o, y_o and z_o.

The *standard object coordinate* system is constructed by permutating the axes of the object coordinate system such that the principal viewing axis becomes the third coordinate axis. The principal viewing axis is the object-space axis that is most parallel to the viewing direction. These axes are labelled i, j, and k, where k is the principal viewing axis.

The *sheared object coordinate system* is constructed by applying the shear-warp matrix to the standard object coordinate system. The sheared object coordinate system is also the coordinate system of the intermediate image. The origin is located at the upper-left corner of the intermediate image. In the sheared object space all viewing rays are parallel to the third coordinate axis. The axes are labeled u, v and w.

The *image coordinate system* is the coordinate system of the final image. The final image is constructed by warping the sheared-object coordinates into image coordinates. The origin is located at the upper left corner of the image. The axes are labelled x_i, y_i and z_i.

The goal is now to factor an arbitrary perspective viewing transformation M_{view} such that:

$$M_{view} = M_{warp} \cdot M_{shear}$$

The M_{shear} contains both a translation as well as a scaling factor such that each viewing ray becomes perpendicular to the slices. M_{warp} is a perspective transformation matrix.

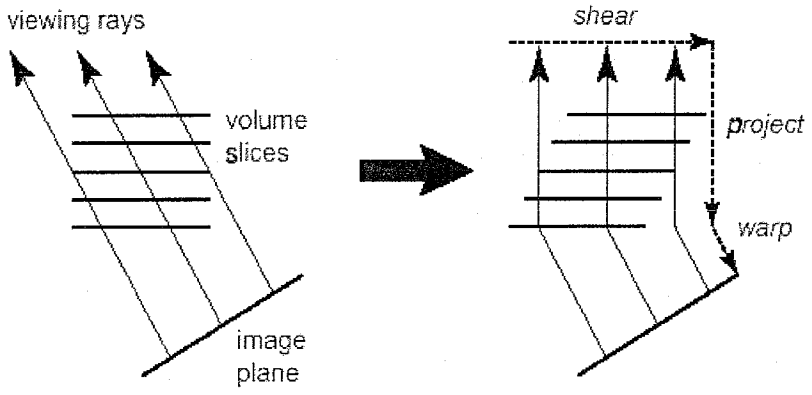

Figure 14: The shear-warp factorization

The principal viewing axis needs to be determined to transform the object coordinate system to the standard object coordinate system. However, because of the divergent viewing rays, it may not always be possible to determine one consistent principal axis. The solution to this problem, which mainly occurs with large view angles, is to subdivide the volume into at most six sub-volumes. The volume is subdivided through the eye-point along planes oriented at 45 degree angles relative to the slices. Each sub-volume is then rendered separately and tiled together to form the complete image. A few voxel overlap in each sub volume is required to ensure a smooth transition between the tiled images.

The principal viewing axis is found by forming vectors from the eye-point to each corner of the bounding volume and to find the principal viewing direction for each of corner. Let $\overrightarrow{p_o}$, a point in homogeneous coordinates, be a corner of the bounding volume in object space, then the vector in object coordinates from the eye-point to this corner is $\overrightarrow{p_o} - \overrightarrow{e_o}$. The principle viewing axis now corresponds to the largest absolute vector component of $\overrightarrow{p_o} - \overrightarrow{e_o}$ or in other words:

$$\max(|\overrightarrow{p_{o,x}} - \overrightarrow{e_{o,x}}|, |\overrightarrow{p_{o,y}} - \overrightarrow{e_{o,y}}|, |\overrightarrow{p_{o,z}} - \overrightarrow{e_{o,z}}|).$$

This is repeated for each corner of the bounding volume. If not all corners have the same principle viewing axis, then this volume has to be subdivided as described before.

The transformation from object coordinates to standard coordinates now follows by permutating the volume using one of the following three permutation matrices P:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ if the principle viewing axis is the x_o axis

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ if the principle viewing axis is the y_o o axis

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ if the principle viewing axis is the z_o axis

However if the eye-point is located in the $k = 0$ plane, then an additional transformation T_{shift} is required that translates the origin of the volume. The transformation matrix M'_{view} from standard object space into image space is now defined as:

$$M'_{view} = M_{view} P^{-1} T^{-1}_{shift}.$$

This transformation matrix M'_{view} is now factorized into a shear component M'_{shear} and a component M'_{warp}. The M'_{shear} component satisfies the following condition: *after the shear and scale transformation the viewing direction must be perpendicular to the* (i, j) *plane.*

The M'_{shear} component shears the volume such that the ray from the eye-point through the origin of standard object space becomes perpendicular to the slices. The location of the eye in standard object space is: $\vec{e_s} = T_{shift} P \cdot \vec{e_o}$.

The required shear transformation is:

$$M_{sh} = \begin{bmatrix} 1 & 0 & -\frac{e_i}{e_k} & 0 \\ 0 & 1 & -\frac{e_j}{e_k} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ where $e_i = e_{s,i}/e_{s,w}$, $e_j = e_{s,j}/e_{s,w}$ and $e_k = e_{s,k}/e_{s,w}$ are the homogenized eye coordinates.

Now the requirement for T_{shift} becomes clear since it guarantees that e_k is non-zero.

The scale component of M_{shear} scales the slices such that the sheared viewing rays becomes parallel. The choice is made that the voxel slice in $k = 0$ is scaled by a factor of one. Now consider a viewing ray through point $\vec{p}$ in the $k = 0$ plane and point $\vec{q}$ in some other plane. The perspective transformation must scale the voxel slice containing $\vec{q}$ by $\frac{p_i}{q_i}$ in the i direction and $\frac{p_j}{q_j}$ in the j direction. Using the similar triangles it follows that, $\frac{p_i}{q_i} = \frac{1}{1 - q_k/e_k}$. The value for $\frac{p_j}{q_j}$ equals the same expression. Thus the required perspective transformation is:

$$M_s = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{e_k} & 1 \end{bmatrix}$$

The complete perspective shear factor is the product of M_s and M_{sh} and after substituting the homogeneous coordinates for the eye-point is:

$$M'_{shear} = M_s \cdot M_{sh} = \begin{bmatrix} 1 & 0 & -\frac{e_{o,i}}{e_{o,k}} & 0 \\ 0 & 1 & -\frac{e_{o,j}}{e_{o,k}} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{e_{o,w}}{e_{o,k}} & 1 \end{bmatrix}$$

The warp transformation has to compensate for the shear transformation, i.e. $M'_{view} = M'_{warp} \cdot M'_{shear}$. Therefore the warp transformation equals the permuted view transformation multiplied by the inverse of the shear transformation:

$$M'_{warp} = M'_{view} \cdot M'^{-1}_{shear} = M'_{view} \cdot \begin{bmatrix} 1 & 0 & \frac{e_{o,i}}{e_{o,k}} & 0 \\ 0 & 1 & \frac{e_{o,j}}{e_{o,k}} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{e_{o,w}}{e_{o,k}} & 1 \end{bmatrix}$$

The shear matrix just derived transforms standard object space into a deformed space. And now a transformation from this deformed space to the intermediate image coordinate system is defined. As before a uniform scale is chosen such that the front-most slice is scaled by a factor of one. Then a translation to position the origin of the intermediate image coordinate system at the top-left corner of the intermediate image is performed.

The shear matrix M'_{shear} in the previous section guarantees that the voxel slice located in the $k = 0$ plane has a factor of one. However this is not necessarily the front voxel slice, depending on the stacking order of the voxel slices in the intermediate image. The sign of the largest component of $\overrightarrow{p_o} - \overrightarrow{e_o}$ determines the staking order: if positive then $k = 0$ is the front slice, otherwise the slice in the $k = k_{max}$ plane is the front slice.

If the front slice is in the $k = 0$ plane then the scale factor for the front slice after applying M'_{shear} is already one, so no scaling is required. However if the front slice is in the $k = k_{max}$ plane, then M'_{shear} results in a scale factor of:

$$f = \frac{1}{1 - \frac{e_{o,w}}{e_{o,z}} k_{max}}$$

This scaling must be inverted by the transformation to intermediate image coordinates. The required scaling matrix is now:

$$S = \begin{bmatrix} f' & 0 & 0 & 0 \\ 0 & f' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ where}$$

$$f' = \begin{cases} 1 & if\ slice\ k = 0\ is\ in\ front \\ 1 - \frac{e_{o,w}}{e_{o,z}} & otherwise \end{cases}$$

To choose the translation, the corners of the front and rear voxel slices are transformed by $S \cdot M_{shear}$. From the transformed coordinates the bounding box of the volume after transformation is found and the translation $T(t_i, t_j, 0)$ that positions the upper-left corner of the bounding box at the origin is chosen.

Combing all matrix factors described above, the final shear-warp factorization for perspective projection is:

$$M_{shear} = T \cdot S = \begin{bmatrix} 1 & 0 & \frac{e_{o,x}}{e_{o,z}} & 0 \\ 0 & 1 & \frac{e_{o,y}}{e_{o,z}} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{e_{o,w}}{e_{o,z}} & 1 \end{bmatrix}$$

The warp factor is:

$$M_{warp} = M'_{view} \cdot M^{-1}_{shear} \cdot S^{-1} \cdot T^{-1}$$

The third and fourth row of M_{warp} are not used because only a 2D warp is required and the intermediate image is located in the $k = 0$ plane and the z_i coordinate is not used.

Summarizing the complete perspective factorization:

$$M_{view} = M_{warp} \cdot M_{shear} \cdot T_{shift} \cdot P$$

is computed as follows:

1. find the eye-point in object space;
2. find the principal viewing axis and the correct permutation matrix P, if required split the volume;
3. determine T_{shift} and compute the permuted view matrix from M_{view} and P;
4. compute the shear and perspective scale coefficients from the eye-point;
5. compute the scale and translation that complete the transformation at intermediate image coordinates;
6. compute the M_{shear} and M_{warp} matrices.

For the perspective shear-warp factorization there are 2 main properties:

1. Scanlines of pixels in the intermediate image are parallel to the voxels in the volume data; and
2. All voxels in a given slice have the same scale factor.

3.2.2 Perspective re-projection for classified data

The previous section described the mathematical basis of the shear-warp factorization. The efficiency of the re-projection algorithm is based on first the efficient projection from object space to intermediate image space, secondly on the (relatively) computationally inexpensive two-dimensional warp from intermediate image space to image space. This section will describe for pre-classified data: the implementation details of the data representation, the encoding that uses data coherency to skip voxels, how clipping is performed and implementation details for the various compositing functions.

Volume data representation One of the properties of the shear-warp factorization is that voxel scanlines are parallel with pixel scanlines in the intermediate image. This property implies that the re-projected data structure and the volume can be traversed simultaneously in scanline order. Therefore the coherence will be represented using a scanline-based run-length encoding (RLE). The coherence is based on the voxel intensity and the opacity transfer function which maps a voxel intensity (or scalar) to an opacity value. For CRA data a lot of voxels will receive a low opacity value implying that these voxels are classified as not being part of the structure of interest. Recall the formation of the DSA images in which the vessels are represented with a high intensity. A proper classification will therefore map these values to a high opacity, while background tissue will be mapped to a low opacity. The operator is able to select a threshold at which a voxel has a negligible opacity and therefore this voxel will not contribute significantly to the final appearance of the image. Although no exact numbers are available for the DSA volumes, in similar work it has been found that typically less then 12% of the voxels are classified [34].

The voxel classification is used to RLE the volume data such that voxels of no importance can be skipped efficiently. Previous work uses a binary decision: a voxel is either transparent or non-transparent. Transparent voxels are not stored and non-transparent voxel are then stored continuously. However the algorithm constructed here does not store non-transparent voxels with identical intensity separately (e.g. [28]), but packs these voxels together. This choice requires that any data used for e.g. shading has to be stored in a separate structure, because even though voxels might have the same intensity they can have different values used for e.g. shading such as surface normal. However, the three main visualizations of the CRA data do not use this shading but are purely based on the voxel intensity. Although no visualization has been implemented that uses these additional voxel attributes, it is expected that the amount of delay due to the use of separate structures is minimal if both data representations can be kept in memory simultaneously.

The RLE of the volume data occurs as a pre-processing step and encodes the volume using four data structures of which for the current visualization options only three need to be kept in memory at one time. The first structure is an array of run lengths, the second structure contains the essential voxel data for each non-transparent voxel, the third structure allows for slice-based access to the RLE volume data and the fourth structure contains the shading data for each voxel (non-transparent or transparent).

The run length array contiguously stores the length of each run. A run can either represent a empty run of transparent voxels or a run of non-transparent voxels. The second structure contains the compressed voxel data of non-transparent voxels storing the voxel intensity in two bytes and the voxel opacity in two bytes. Of the 16 bits used to represent the voxel intensity only 12 bit are used to store the intensity, while the 4 most significant bits are used to store a "repeat count" for the voxel intensity. A "repeat count" of zero represents a run of one voxel, a count of one represents two voxels, etc. This repeat count is used to save some calculations because for each slice holds that the scale factor is the same for each voxel, therefore if a run of voxels in a slice has equal intensity, then some calculations can be re-used. The third structure stores pointers into the first and second data structure. One pointer points to the first run of a slice,

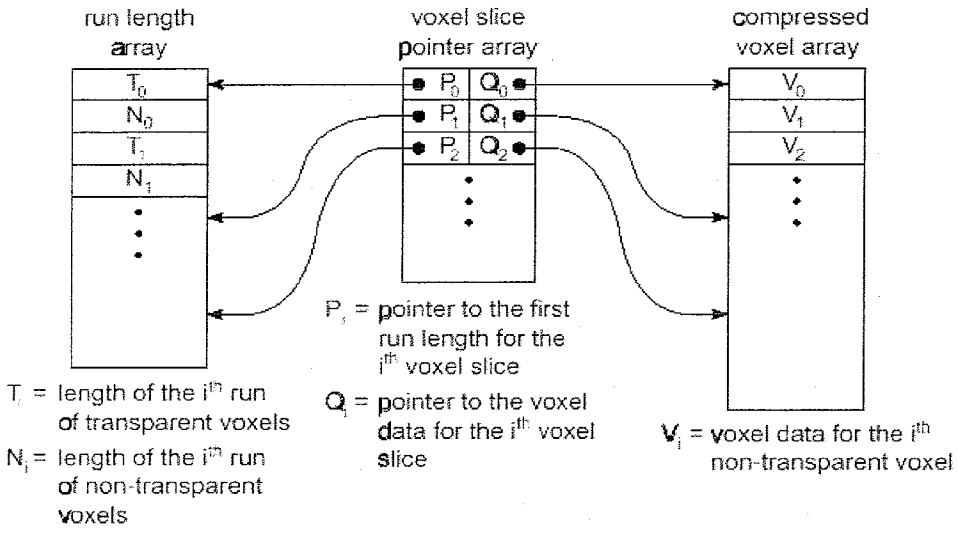

Figure 15: The structures used to represent the volume

the second pointer points to the first non-transparent voxel of that slice. This structure enables the algorithm to access the RLE data slice-by-slice. The fourth structure (not shown) contains for each voxel (non-transparent or transparent) the pre-computed surface normal using two bytes and the pre-computed gradient using two bytes. The first three structures are pre-computed for each major axis, while the fourth structure is only computed once and is independent of the voxel classification and view point.

Volume data re-projection The convolution is performed in a different way than previously described in the literature. Each slice must be translated and scaled when it is transformed into sheared-object space and for this a box reconstruction filter and a box low-pass filter are used. To resample at least two voxel scanlines are necessary and instead of traversing and decoding the voxel scanlines for each image scanline, a splatting-type convolution is used. Lacroute uses a gather-type convolution, which requires that each voxel scanline is traversed at least twice, and the reasoning behind this is that no partially computed voxels need to be stored. However for re-projecting large volumes in perspective geometries, it is more efficient to splat each voxel's partial result to a temporary buffer represented with a 2D array of 32 bit than to decode volume scanlines multiple times. The maximum number of voxel scanlines that splat onto an image scanline is constant within each slice.

The intermediate image is represented with a run-length encoded data structure that differentiates between runs of maximum intensity and non-maximum intensity pixels. This RLE is not pre-computed, but modified during rendering and is used to skip pixels that have maximum intensity. This RLE is only performed for MIP and dMIP. DRR require that each voxel's contribution is added

to the image and therefore early ray termination is not possible and the RLE of the intermediate image is not necessary.

Once the compositing phase has finished, then the intermediate image is warped to the final image. This image warp is similar to the image warp presented in [28].

In short: the algorithm traverses front-to-back, slice by slice, scanline by scanline as follows:

```
for each voxel slice s do
{
   for each voxel scanline l do
   {
      traverse and decode the voxel scanline l;
      for each non-transparent voxel v, splat its contribution into
         the temporary buffer
   }
   for each scanline i in the intermediate image
   {
      traverse and decode the intermediate image scanline i;
      convolute the results the corresponding temporary
         buffer scanline j with i
   }
}
```

Clipping the volume data Clipping the volume consists out of removing unwanted information from the volume to provide cut-away views or to focus on a particular part of the data set. The CRA data is initially clipped to a sphere with a radius equal to the half the size of the volume. This clipping is optional, but data outside of this spherical volume is artificial since the Feldkamp algorithm only correctly reconstructs within this sub-volume. The voxel outside of this region receive an intensity value of zero. The user can also define other clipping planes which place bounds on the loop that iterates over the voxels in each voxel scanline. Voxels outside of these user-defined clipping planes are not altered.

User-defined clipping planes are heavily used in CRA volumes during interaction. The ROI, the cerebral aneurysm, is usually much smaller than the volume size and for determining the desired orientation the information outside of this ROI is of lesser concern. Common ROI for cerebral data range from 200x200x200 voxels to 100x100x100 voxels.

Implementation requirements for various compositing functions As stated before a DRR does not allow for early ray termination since all voxel intensities need to be summed. MIP re-projections only determine the maximum intensity along each ray, therefore early ray termination is performed when an intermediate pixel has reached the maximum intensity.

dMIP requires the distance from the voxel to the view point. Instead of calculating this distance for each voxel a lookup table is used. The distance from a voxel to the view point is broken down into two components:

1. the distance d_{vi} from the voxel to the intermediate image; and
2. the distance d_{if} from the intermediate image to the final image.

The distance d_{vi} is calculated by using the sample spacing along a viewing ray. The sample spacing along each viewing ray Δx is constant and is stored for each intermediate image pixel in a 2D array. For each voxel the distance d_{vi} is now calculated as follows:

$$d_{vi} = k \cdot \Delta x$$
where k is the voxel slice number and
Δx is the distance between the slices along a viewing ray in image space.

The distance d_{if} from the intermediate image to the final image is calculated for each intermediate image pixel and applied to the pixel intensity before the image warp is performed. This factorization of the distance allows for an efficient dMIP re-projection.

3.2.3 Perspective re-projection for unclassified data

The opacity transfer function is selected such that only the vasculature in the CRA volume is visible, all other tissue will have a low opacity. However for most volumes a correct opacity transfer function is not immediately known, therefore a fast classification algorithm is necessary allowing the user to manipulate this transfer function. The re-projection of unclassified volumes uses an un-encoded volume and the voxel traversal is slice-by-slice, scanline-by-scanline, however the classification of each voxel is performed during rendering.

The fast classification algorithm uses a min-max octree to classify the CRA volume. The min-max octree, which is independent of the opacity transfer function and the view point, is constructed as a preprocessing step when the CRA volume is loaded into memory.

An octree recursively subdivides the volume as a tree. Each node of the octree has eight child nodes and each represents a cubic region of space. The min-max octree stores for each node the minimum and maximum voxel intensity in the cubic region covered by this node. Next to the minimum and maximum voxel intensity, each node contains a flag indicating whether the cubic region contains only transparent, non-transparent or both transparent and non-transparent voxels. Also, each node contains a pointer to its eight child nodes.

The min-max tree is constructed bottom to top. First the base of the octree is computed by looping over the voxels in the volume to compute the minimum and maximum voxel intensities for a pre-defined maximum detailed octree node. Then the next octree level is computed by combining the results of eight higher-detail octree nodes into one parent node. This process continues until the root node has been constructed. The resulting min-max octree is now pruned by combing child nodes which have a min-max difference that falls below a specified threshold. This user defined threshold trades off level-of-detail with octree traversal time.

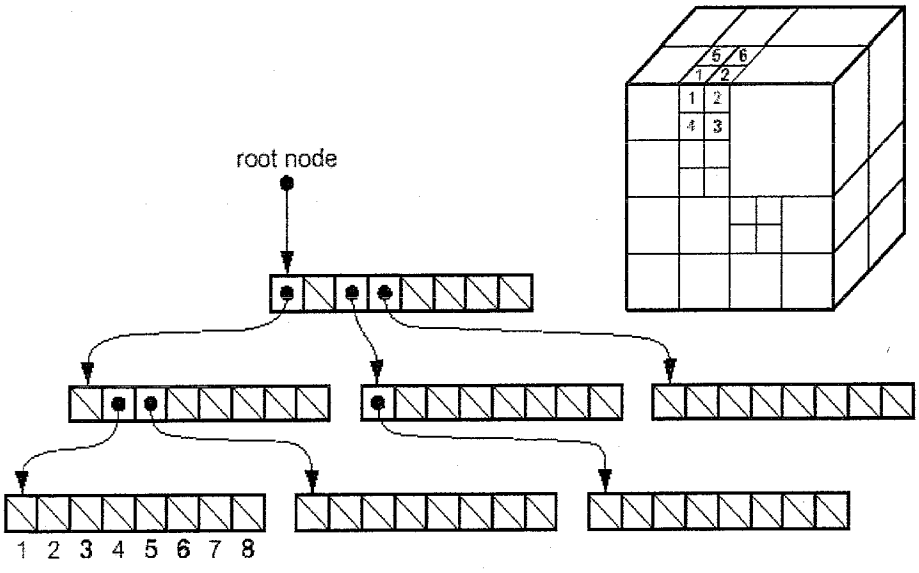

Figure 16: The min-max octree

The second structure used by the fast classification algorithm is a summed opacity table. This table is dependent on the opacity transfer function, but can be computed very quickly since in the current implementation only the voxel intensity is used to map to a voxel transparency. The summed opacity table is defined as follows:

$$t[i] = (\Sigma j : 0 \leq j \leq i : f(j))$$

where

i is the voxel intensity (scalar value); and

$f(i)$ is the voxel intensity mapped to its opacity value.

This summed opacity table is now used to classify the voxels represented by an octree node as follows:

$$if \begin{cases} t[max] - t[min] < threshold & \textit{the node is transparent} \\ t[max] - t[min] > threshold & \textit{the node is not transparent} \end{cases}$$

where

(max, min) are the minimum and maximum voxel intensities in a node.

For each scanline, the octree and summed opacity table are used to determine which part of the scanline is non-transparent. If a scanline is classified as non-transparent, then it is divided and the appropriate child nodes in the octree are classified. This process continues until the child node with the highest detail is still not transparent, then each voxels is classified and rendered. If the scanline (or a sub-division thereof) classified as transparent, then this scanline is skipped as in the previous algorithm. Opaque regions in the intermediate image are also

skipped as before. This approach results that non-transparent voxels are never missed, but some transparent voxels may be rendered.

3.2.4 Optimizations for interaction

Several standard optimizations for volume rendering are present in the interactive quality re-projection algorithm. The algorithm described above allows for early-ray termination and efficient transparent voxel skipping using the RLE based on the opacity transfer function. The fast classification algorithm allows for efficient rendering of unclassified data.

The algorithm has been implemented using the Visualization ToolKit (VTK) and uses a progressive refinement feature present in this toolkit. The toolkit allows for several levels of detail that progressively improve upon the quality of the rendered image. The decision was made to represent two levels for the pre-classified rendering algorithm, one level using a sub-sampled CRA volume, the other using the CRA volume as read from the file.

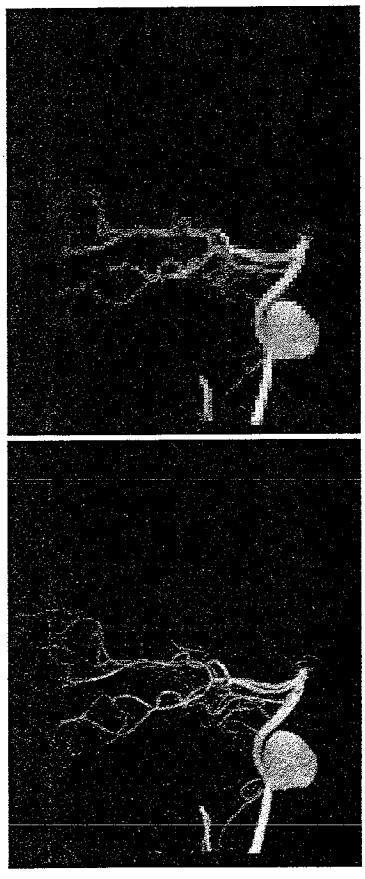

Figure 17: The lowest interactive sampling level and the highest interactive sampling level.

The second optimization deals with the fact that an unclassified volume can be rendered using the fast classification algorithm, however this algorithm is slower than the re-projection algorithm using the pre-classified data. Therefore over several renderings the CRA volume is RLE using the specified opacity transfer function. This encoding of the volume is done in the background, but also uses partial results obtained while rendering using the fast classification algorithm.

3.3 Diagnostic quality visualization

Diagnostic quality visualization requires a high quality image, but rendering time is a lesser concern. The diagnostic quality rendering is used after the user has selected a proper opacity transfer function, volume rotation, volume clipping planes and source-detector distance. The user may change the rotation, clipping planes and source-detector distance, but the opacity transfer function is fixed.

3.3.1 Perspective re-projection

The perspective diagnostic quality re-projection uses perspective ray casting. Rays, in a perspective geometry, do not maintain a continuous and uniform sampling rate of the underlying volume data (i.e. the rays diverge). Common solutions to this problem are to under-sample or to over-sample the volume. Under-sampling occurs when the rays are cast such that the desired sampling rate is achieved at the front of the volume close to the eye-point (something similar to this is done in the shear-warp algorithm presented in the previous section). Over-sampling occurs when the rays are cast such that the desired sampling rate is achieved at the back of the volume. Under-sampling is not acceptable for the required image quality, because under sampling will result in aliasing in the rear of the volume. Over-sampling is not an option either because of the size of the volume data.

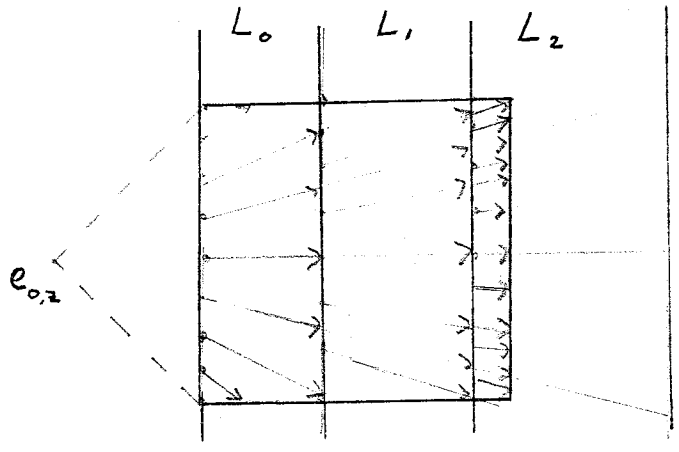

Figure 18: A volume divided into sections

A solution to the problem of diverging rays and non-continuous sampling is to use an adaptive sampling scheme which ensures that the sampling rate stays

close to the density of the underlying volume data. Therefore an adaptive perspective front-to-back ray casting algorithm which divides the view frustum into regions based on exponentionally increasing distances from the eye-point was designed. Continuous rays are then cast front-to-back through the volume and rays are split as soon as a region boundary has been reached. This splitting of the rays ensures that the volume is regularly sampled and solves the diverging ray problem.

The variations of the algorithm presented here were designed, of which the first one uses a breadth-first front-to-back slice-order processing, and the volume is projected to an intermediate image which is most perpendicular to the view direction. This intermediate image is similar to the intermediate image of the shear-warp factorization and is warped onto the final image plane as is done in the shear-warp algorithm. The second variation of the algorithm does not use the breadth-first front-to-back slice-order processing, but instead ray casts through the volume using tri-linear interpolation.

As mentioned, the first variation uses breadth-first slice-order processing along one of the three major axes. Therefore for each major axis a volume needs to be present in memory. The volume representation used for these three volumes is the RLE representation of the shear-warp algorithm and the major viewing axis is determined as described for the shear-warp algorithm. The breadth-first approach allows the use of RLE structure and each temporary results per section is stored in a 2D array.

From this point on it is assumed that the volume is being projected along slices perpendicular to the Z-axis. The algorithm determines the distance k along the Z-axis from the viewpoint to the front of the volume and creates the first region to consist of k Z-slices. Each successive region is twice as deep as the previous. In general the depth of a region is determined by:

region i is $k \cdot 2^i$ slices thick
where
k is the distance from the viewpoint to the front of the volume

Defining the regions in this way effectively ensures that any two perspective rays sent through any region are twice as far apart at the rear as they are at the front of this region. In other words, the density at the front of a region is twice as high as at the back of the same region. The front-to-back processing uses a bilinear interpolation to generate new rays which start between the two rays.

In short the first variation of the algorithm for slice-by-slice access is:

```
given the z-position of the eye in object coordinates: e_{o,z}
construct the exponential regions, front-to-back: L_0, L_1, ..., L_N
for region L_0
{
   initialize the viewing rays
   for each slice s going front to back
   {
      decode the region slice by slice, sample the ray segments
      composite into the buffer b_0
   }
```

```
}
for each region r = L_1 to L_N
{
    generate new intermediate rays by interpolation
    for each slice s going front to back
    {
        decode the region slice by slice, sample the ray segments
        composite into the buffer b_r
    }
}
for each buffer i = 0 to N
{
    resample the buffer b_i to the resolution of the intermediate image.
    composite the buffer b_i onto the intermediate image
}
warp the intermediate image to the final image plane
```

The second variation works in a similar way however it is does not traverse slice by slice. The second variation uses the same idea of regions and it processes each ray segment breadth-first. This is done to allow for better interpolation which preserves small vessel detail. However the ability for better interpolation comes at the cost of losing the cache coherent data access.

3.4 Future improvements

Improvements to the algorithms can be improvements to the re-projection image quality and improvements to the re-projection speed. The algorithms as described above allow for interactive manipulation and re-projection of the volumetric CRA data. However both algorithms can be greatly improved upon by re-projection the data in a distributed manner.

The interactive quality re-projection algorithm can be distributed over multiple systems by having each node calculate part of the intermediate image. Each node should obtain a sub-volume and re-project this sub-volume to an intermediate image. All these intermediate images should then be 'blended' together into the intermediate image which is then warped to the final image. Possible problems however are that the intermediate images calculated by the various nodes should keep additional information such as the opacity of the pixels to allow for proper blending. Also each sub-volume should have a few voxels overlap and the contributions of these voxels need to be averaged in the intermediate image. The diagnostic quality re-projection algorithm can be distributed over multiple machines using a similar technique.

Improvements regarding the image quality will have to be looked for in better various issues (both involving the re-projection and the actual volume), such as:

- improved interpolation algorithms;
- improved classification preserving fine vessel details; and
- higher resolution volumes.

3.5 Results

The results shown here were obtained by implementing the algorithms described above using C++ and the Visualization ToolKit (VTK). The re-projection time listed here is an average of the rendering time for several volumes. All volumes used were real-life data, cropped to a ROI and pre-classified unless otherwise mentioned. All algorithms used early-ray termination for the MIP re-projection. As can be seen from the table, the algorithms described in this chapter above are faster than a standard ray caster using tri-linear interpolation supplied with VTK. The main reason for the difference in speed is because of the RLE, the regular sampling and the cache coherency.

re-projection algorithm	*MIP*	*DRR*
interactive, pre-classified data	0.12-0.13 sec	0.30-0.34 sec
interactive, un-classified data	0.16-0.19 sec	0.31-0.35 sec
diagnostic, bi-linear interpolation	12-15 sec	17-21 sec
diagnostic, tri-linear interpolation	22-25 sec	30-35 sec
standard VTK ray caster using oversampling	35-40 sec	50-58 sec

Table 4: The various re-projection algorithms and their performance

Figure 19: 1) DSA projection 2) Corresponding MIP re-projection

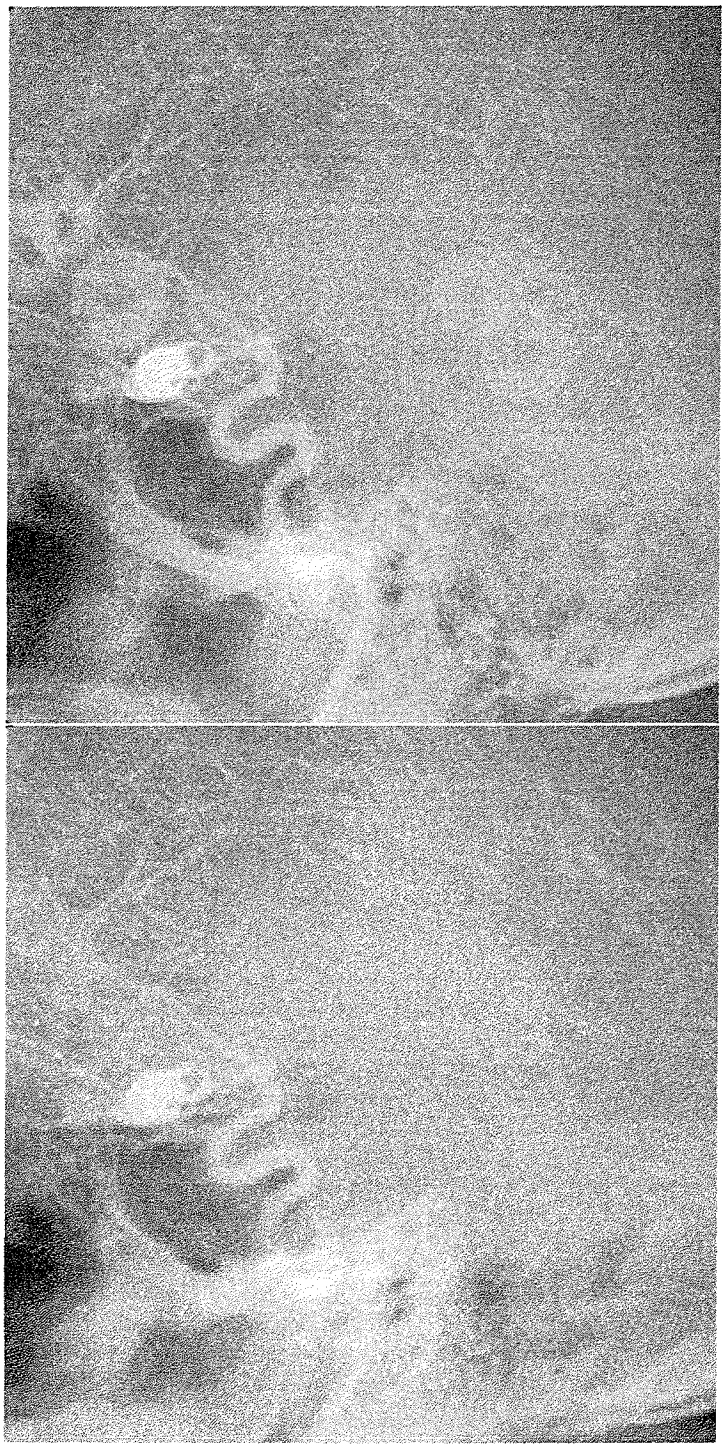

Figure 20: 1) DSA projection 2) Corresponding voxel-driven DRR re-projection

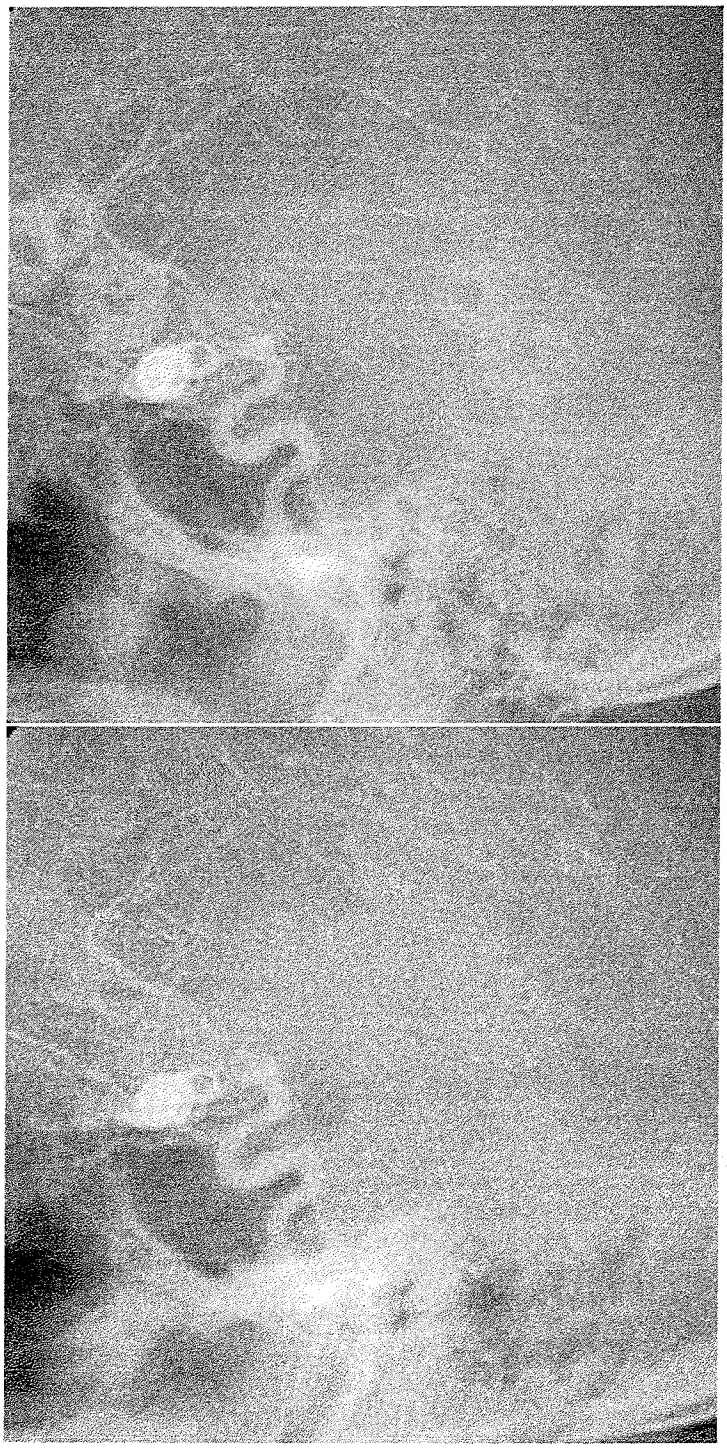

Figure 21: 1) DSA projection 2) Corresponding ray-driven DRR re-projection

4 Image quality comparison

This chapter describes a comparison of Computed Rotational Angiography (CRA) with Digital Subtraction Angiography (DSA), the gold standard, using a two alternative forced-choice (2AFC) methodology. 2AFC is a technical image quality measurement for intracranial CRA. First the 2AFC methodology is explained, followed by a 2D-3D registration algorithm used to validate 2AFC results. Finally an overview is given on a patient study performed using the 2AFC methodology and the results are presented.

4.1 Two-alternative forced choice

The goal of the technical image efficacy measurements (i.e. technical image quality) is to evaluate the effect of changes in imaging technology on the imaging performance. There are two problems to this task: first, CRA is still actively being developed and therefore any measure of technical image efficacy must be fast and economical; second, the value of images as determined by technical image efficacy should be closely related to the value of images at the clinical level.

Image efficacy in clinical radiology is usually defined in terms of how well an observer performs in extracting desired information from the image, [26]. Observer performance is quantified using signal detection experiments such as receiver operator characteristic (ROC) or multiple alternative forced-choice (MAFC) experiments. The presence or absence of this desired information is determined by reference to a suitable gold standard imaging technique (i.e. DSA for CRA). Unfortunately these observer performance experiments are expensive, time-consuming and only evaluate the specific task at hand. An acceptable measurement of technical image efficacy must however be closely related to this information extraction process.

Another measure used in clinical radiology is the signal-difference-to-noise-ratio (SDNR) or contrast-to-noise ratio (CNR). This measure is used because of its simplicity: detecting a signal known exactly on a uniform background with superimposed noise. However for medical images it is not clear what the importance of observed SDNR changes is to the observer tasks.

Chapman argued, for Magnetic Resonance Angiography (MRA), that two alternative forced-choice provides an upper bound for vascular pathology detection using a given imaging modality, [27]. 2AFC cannot discriminate between images with an equal vessel segment detection rate but unequal pathology detection rate, but this will not affect the intended application.

Signal detection theory describes the statistics of detecting a signal embedded in a background. Two types of stimuli are distinguished. A 'noise-only' stimulus, which consists of carrier stimuli on which a signal could be superimposed. A 'signal-plus-noise' stimulus, which consists of a stimuli that has a signal superimposed on the background carrier stimuli. The absence or presence of a signal in a stimulus is determined by reference to a gold standard.

In a 2AFC experiment the observer is presented with two stimuli, a 'signal-plus-noise' and a 'noise-only' image, between which a choice has to be made

by referencing a given signal (which was obtained from the gold standard). A signal here is defined as a specific part of a vessel that has to be detected. The figure-of-merit for a 2AFC test, which is the proportion of signal-plus-noise stimuli correctly identified, is defined as follows:

$$P_c = \int_{-\infty}^{\infty} p(x|signal - plus - noise)(\int_{-\infty}^{x} p(x'|noise - only))dx$$

under the assumption that the stimuli evoke a response x describing the likelihood of a signal being present ($p(x)$) and also assuming that the observer always picks the stimuli with the highest evoked response. Therefore the higher the P_c value for an image, the higher the assumed technical efficacy. This P_c value in a 2AFC experiment is equivalent to the area under the ROC curve.

The P_c variable is bounded between zero and one, and it's common to transform this bounded variable into an unbounded variable d', known as the detectability index. Under the assumption that the signal-plus noise and noise-only response distributions are independent Gaussians with equal variances (σ^2) separated by a distance μ, then d' is defined as

$$d' = \frac{\mu}{\sigma}.$$

In other words d' measures the separation between noise-only and signal-plus-noise response distributions, thus d' and P_c are related by

$d' = \sqrt{2}z(P_c)$ where
$z(P_c)$ is the normal deviate.

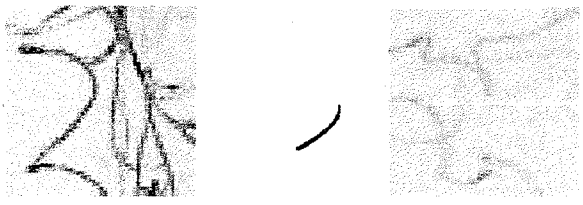

Figure 20: 2AFC example

The 2AFC measurement used for CRA use DSA images as the gold standard for comparison with CRA. The 2AFC measurements use various signal-plus-noise ROIs within one CRA image and for each signal-plus-noise ROI the arterial signal is defined by tracing a line through the vessel in the corresponding DSA image to capture its shape and the width of the vessel is estimated using X-ray pixels. The noise-only ROIs are defined by selecting other ROIs in the CRA image without associating them to a specific arterial signal. This results in three images: one containing the signal and noise, the second containing the traced vessel and the last one containing only noise. These three images are then to be presented to an operator who has to choose which of the two CRA images contains the signal presented by the traced vessel.

The reason why 2AFC is used for CRA image efficacy measurement, is that it is a direct measurement of vessel visualization and this is one, if not the most, important aspects of intercranial CRA. A drawback however is the fact that the experiments are labor and resource-intensive. Reducing mis-registration

between the CRA and DSA images is also of great importance and this will be solved by registering the DSA images to the CRA volume using a registration algorithm described in the following section.

4.2 2D-3D Image registration

The previously described technical image quality measurement, two-alternative forced choice, has been applied to operator determined projections of the CRA data. This small clinical study is discussed in the next section. However, to minimize mis-registration effects a user-assisted registration algorithm has also been designed which registers the DSA image to the CRA volume re-projections (e.g. the DRR image). DSA to CRA registration involves determining the 11 parameters involved in DRR image construction such that a similarity measure is maximized.

Several approaches to 2D-3D registration can be found in the literature ([21, 22]) and these can be categorized into two main categories:

1. feature-based methods; and
2. intensity-based methods.

Feature-based methods use segmented images. The benefit of this is that a reduced amount of data is used which enables the algorithms to run fast. However, the segmentation stage is hard to automate and segmentation errors lead to errors in the registration. The vast amount of data required in an image quality study using 2AFC prohibits the use of feature-based methods.

Intensity-based methods compare voxel and pixel values directly using measures based on image statistics. Although these algorithms are slower they require almost no segmentation and can easily be automated. Automation is the main reason for choosing this type of registration.

The DSA and DRR images are compared using a similarity measure and a search is performed to maximize the similarity measure. However there are some differences between the DSA and DRR images, which can be categorized in:

1. differences due to changes in the imaged object itself; and
2. differences in image formation.

Changes in the object can be interventional instruments, tissue deformation, coils, etc. Differences in the image formation are, among other:

- DSA images are directly influenced by the x-ray parameters as given in Chapter 2, but DRR images are not;
- the use of (possibly) different x-ray energies;
- different resolutions of each image; and

- distortions in the DSA image caused by non-linearities in the optics, which are not present in the DRR image.

Various similarity measures are available, but it is known from the literature that a similarity measure based on gradient differences will be accurate and robust for the task at hand [22]. Initially the DRR and DSA image are used to create four gradient images: a horizontal and vertical gradient image of the DRR and the DSA images. These gradient images are constructed by convolution both images with a Sobel filter. The gradient images are then subtracted from each other using a suitable intensity scaling s to give the vertical and horizontal gradient difference:

$$I_{\Delta V}(i,j) = \frac{dI_{DSA}}{di} - s\frac{dI_{DRR}}{di} \text{ and}$$
$$I_{\Delta H}(i,j) = \frac{dI_{DSA}}{dj} - s\frac{dI_{DRR}}{dj}$$

The similarity measurement is then defined as:

$$S = \sum_{i,j} \frac{A_v}{A_v + (I_{\Delta V}(i,j))^2} + \sum_{i,j} \frac{A_h}{A_h + (I_{\Delta H}(i,j))^2} \text{ where}$$
A_v and A_h are the variance of the respective gradient DSA images.

The 2D-3D registration algorithm uses this measure S to match the DSA to the CRA image for a user-specified error ϵ. The algorithm uses a specific ROI in the DSA image to match to the CRA volume. This ROI is used for two reason:

1. it speeds up the matching since less data needs to be processed; and
2. the matching will be more accurate for the ROI as when matching would be done for the whole image.

The algorithm currently uses as the search space the three rotations and three translations of the CRA volume. The other five parameters involved in DRR generation: the pixel size, the source-detector distance and image center are fixed. The algorithm uses a user supplied approximation of the correct 6 parameters as a starting point and then alters each value one at a time as follows:

1. initially S is calculated for the approximation $(T_0, T_1, T_2, R_0, R_1, R_2)$;
2. then each parameter is altered one at a time $T_0 \pm w_{T_0}\delta$ and a new DRR image is constructed, resulting in 12 new DRR images,
   where w_{T_0} is a user-supplied weight factor which determines the parameter specific rate of change and δ is the overall rate of change;
3. the similarity measurement S is then calculated for all these twelve images and the values that result in a better similarity are changed proportional to the amount of improvement; and
4. steps two and three are repeated until all six parameters are at an optimum. At this point δ is reduced and the process restarts at step 1 using the new approximation $(T_0', T_1', T_2', R_0', R_1', R_2')$ as long as the amount of mis-registration is above ϵ and δ is above a user specified threshold.

The weights ($w_{T_0}, w_{T_1}, w_{T_2}, w_{R_0}, w_{R_1}, w_{R_2}$) are user supplied but usually kept at unity.

The current algorithm does not use a multi resolution approach, but it is expected that using a coarser resolution for the initial steps and progressively refining the resolution while δ becomes smaller, will reduce processing time and help avoid local minima. Processing time is currently limited by minimizing the search space to a few degrees around the user determined approximation and a few pixels for each translation. This ensures that the matching will perform within minutes several minutes.

4.3 Results

CRA reconstructions and MIP CRA re-projections were used to generate images which were compared to DSA images using a 2AFC approach in a small patient study. The hypothesis is that these re-projections will provide more accurate and precise measurements of vessel geometry and visibility than DSA alone, [23], and that also provide for better contrast and increased spatial resolution.

Study: details

The volumetric data used in this study was acquired using the prototype C-arm system and techniques shown in chapter 2. Calibration and correction of the geometric distortions in the XRII and techniques to alleviate view-aliasing artifacts were applied, [15, 16, 17, 18, 24]. This resulted in CRA volumes with 400^3 voxels and an isotropic voxel spacing of 0.4mm with an acquisition time of $4.4s$ at $45°s^{-1}$.

The patient group consisted out of 14 ambulatory outpatients from the London Health Sciences center. All patients were imaged using the approved protocols and all patients provided informed consent. Both CRA and DSA exams were performed in the same session. The mean age of the patient group, consisting out of six males and eight females, was 57 years. The study examined specifically:

- 17 vessels with 18 aneurysms;
- 67 angiographic views;
- 7 aneurysms coil follow-ups;
- 4 aneurysms clip follow-ups;
- 7 aneurysms pre-treatment.

The observers consisted out of four neuro-radiologists and six scientists. 164 vessel segments were examined using the 2AFC technique described earlier.

Study: results

The 2AFC technique covered 164 vessel segments resulting in a total of 1640 comparisons. The signal selection was performed by tracing the vessel segment in the DSA images and a set of pairs of correct CRA re-projections and random alternate image were presented to the observer as is shown in figure 20.

This resulted in the following data:

Observer	# Incorrect signal selections	P_c
1	17	0.8963
2	10	0.9390
3	19	0.8841
4	20	0.8780
5	17	0.8963
6	10	0.9390
7	28	0.8293
8	16	0.9024
9	19	0.8841
10	10	0.9390

with a mean of 0.8988 and a standard deviation of 0.0343.

This compares well with another 3D imaging technique, Magnetic Resonance Angiography (MRA) which in a similar test resulted in a P_c being between 0.8 and 0.9.

As discussed in the first chapter, one of the goals of CRA is to reduce patient risk and therefore one has to consider the amount of X-ray dose and contrast fluid usage. X-ray radiation has a theoretical risk of 0.02% of generating a fatal cancer due to interventional therapy [40]. Adverse effects (e.g. nausea, minor allergic reactions) limit the amount of contrast agent that can be introduced, therefore limiting the number of vessel images that can be acquired.

Therefore the amount of x-ray dose and contrast used was measured and from this data, the two graphs 21 and 22 have been constructed.

This shows that there is no significant difference in patient dose between CRA and DSA. However the amount of contrast used for CRA construction is far less than the amount used for single-plane DSA images, and therefore CRA also reduces cost and patient risk.

Study: conclusion

The 2AFC test results show that the CRA re-projections are equivalent to DSA images, but have a higher resolution. The CRA volume construction requires the total same amount of radiation, but less contrast and therefore has a lower cost. CRA also provides tomographic source images for precise vessel measurement.

The data used in this preliminary study, was non-optimal in the sense that the re-projections were user-selected MIP re-projections. The MIP algorithm

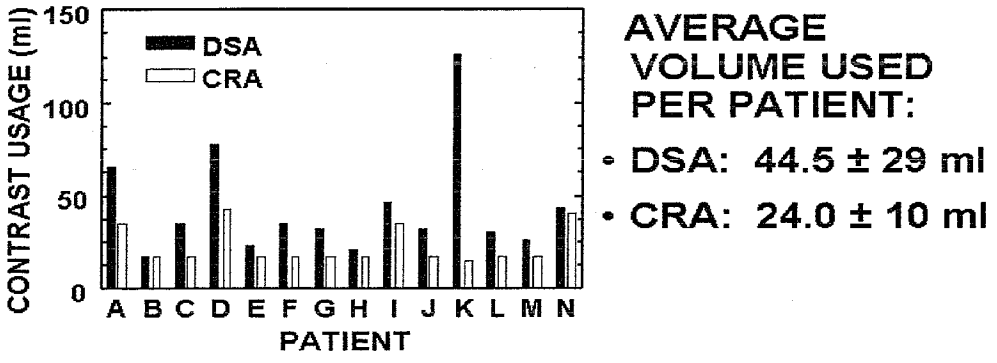

Figure 21: The x-ray dose for the DSA and CRA procedures

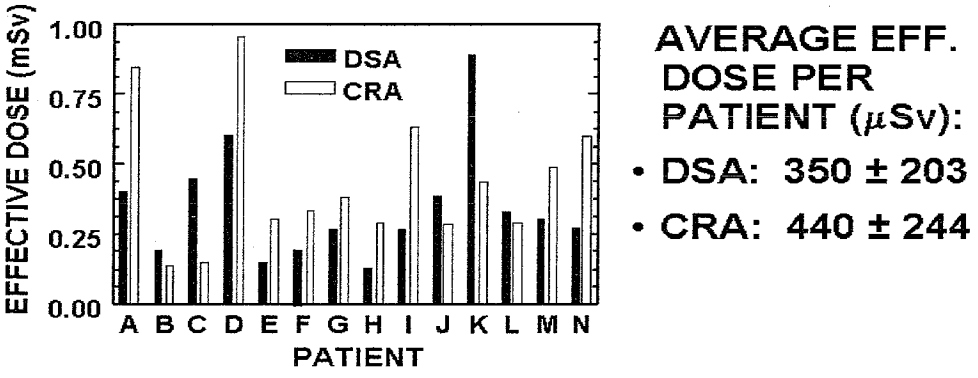

Figure 22: The contrast usage for the DSA and CRA procedures

does result in some decrease in small vessel detail, though this loss is well understood in the literature, [25]. Future tests will be performed using DRR re-projection which mimic the physical image acquisition process better and also the user-selection (of angle, zoom factor, etc.) will be improved on by using the automated 2D-3D registration algorithm described earlier.

5 Conclusion and future work

The re-projection algorithms presented in this thesis allow for the interactive manipulation of CRA volume data and the construction of diagnostic quality DRR. The technical image quality of these re-projections has been verified using a 2AFC measurement. The CRA technique is still being worked on and new applications are being developed, but the goals set for this thesis have been met.

Future work will mainly consist out of :

1. improvements of the current algorithms; and
2. the search for solutions of new problems posed by novel applications for CRA.

Great improvements to the speed of both the CRA construction and CRA re-projection can be made by distributing the task at hand over multiple machines. CRA construction allows for this and experiments distributing this task have already been performed and the preliminary results are very promising. The CRA re-projection algorithms can, by design, be easily distributed over multiple machines. Also work will be performed implementing more compositing operations resulting in additional ways of viewing the CRA data.

A new application currently being investigated is the use of CRA during interventional procedures. This poses new requirements on the CRA re-projection, which will have to work in real-time. Research will be done in achieving this real-time behavior by distributing the re-projection task. Another problem that will be solved in the near future is surgical tool tracking in which CRA images, real time x-ray images are blended together and surgical tools are graphically represented with their relative position in the CRA volume. This will allow the surgeon to work with greater precision and gain more inside in the three-dimensional aspects of the vasculature.

Appendix A - Artifacts in CT reconstruction

An artifact is defined as a difference between the reconstructed and true attenuation coefficients. CRA, being based on 3D-CT techniques, is also susceptible to some of these artifacts. Therefore it is necessary to characterize these artifacts and understand their origin because some of these artifacts can be compensated for. The most important sources of artifacts are:

- geometric errors;
- data acquisition errors; and
- attenuation measurement errors.

The possible artifacts that occur will be characterized by example. The images used were obtained using the CRA system and are of an anesthetized 25 kg pig, in which the carotid artery is injected with contrast agent. Figures 1, 2 and 3 illustrate the CRA reconstruction procedure, which was previously defined[12].

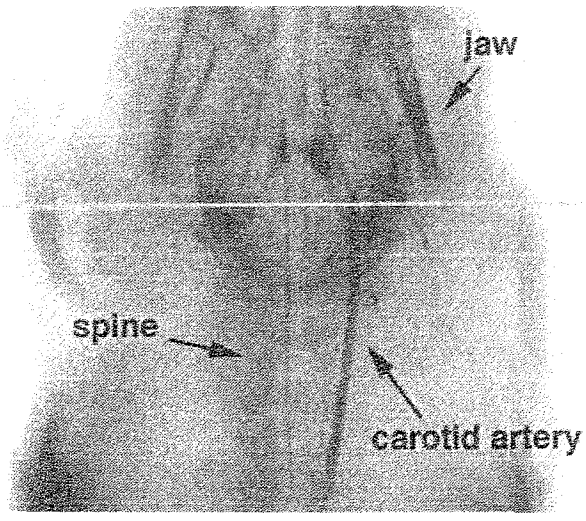

Figure 1: Radiograph showing the location of the subsequent reconstruction.

Geometric errors involve mis-interpretation or inconsistency in the geometric configuration of the acquisition system. Examples are patient motion and mechanical errors such as reconstruction with an incorrect center of rotation. Figures 4, where artifacts are visible around the carotid artery, illustrate the severity of a small error in the center of rotation. This explains the requirement for a high geometrical precision.

Data acquisition errors relate to incomplete or incorrect acquired x-ray projection data. Artifacts caused by these acquisition errors are: view aliasing, incomplete angular views, missing projections, and truncated projections.

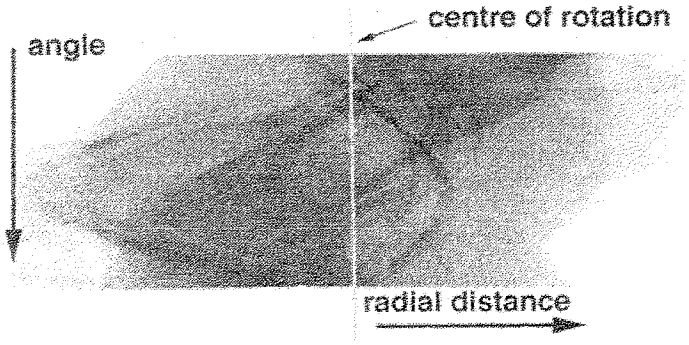

Figure 2: Sinogram through the slice over 200^o.

Figure 3: CT reconstruction using correct reconstruction data.

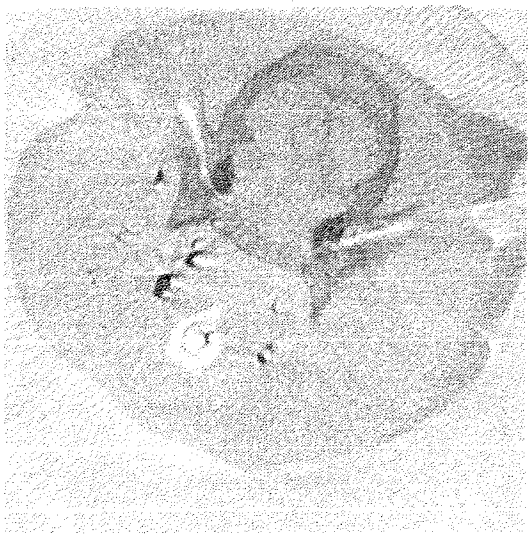

Figure 4: An incorrect reconstruction with center of rotation set to 200 pixels.

View-aliasing caused by insufficient angular sampling causes "spoke" artifacts. These artifacts tend to be worse near the periphery of the image. The images used here were reconstructed using only 130 views and therefore suffer from view-aliasing.

Incomplete angular views are a result from not acquiring projection data during a complete rotation, which results in an incorrect x-ray attenuation map. As stated previously, a correct reconstruction can be obtained by rotating over $\pi + fan\,angle$. Figure 5 illustrates the error caused by incomplete angular views.

Missing projection views cause streak-type artifacts as illustrated in figure 6. In a fourth generation CT scanner a missing view can for instance be caused by a bad detector.

Truncated projection views is a rare artifact, but occurs when data is not collected over the entire FOV of the object, resulting in a truncated x-ray projection, see figure 7.

Attenuation measurement errors caused by detector miscalibration is one of the most significant errors in CT. Miscalibrated detectors in third generation CT scanners result in a ring artifact, see figure 8. Even errors less than 1% can be significant in clinical systems.

Although care is taken to prevent these artifacts from occurring, understanding their causes is important for generating correct visualizations of the volumetric data and for understanding the limitations that have to be dealt with. Especially in applications where CRA data is combined with other imaging modalities, for instance X-ray data, is it important to understand that there can be a small difference between these images due to limitations of their re-

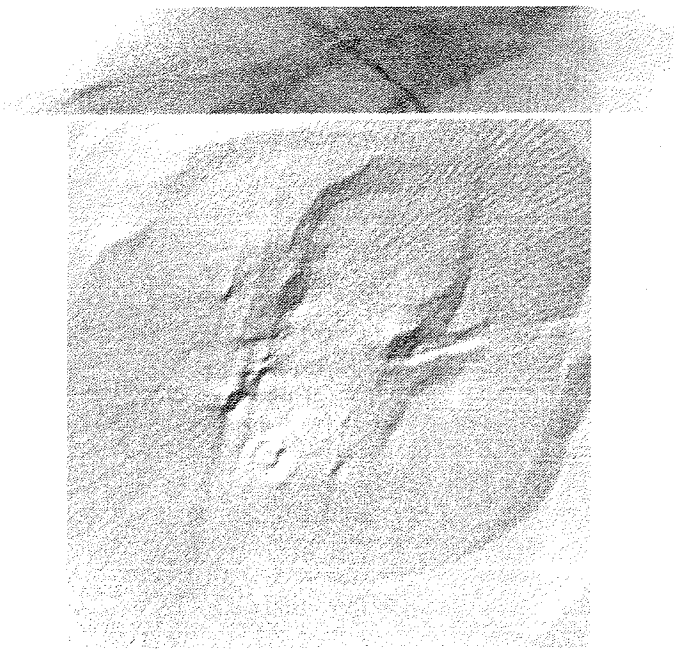

Figure 5: 1) A sinogram truncated after 100^o. 2) Image reconstructed using the incomplete sinogram.

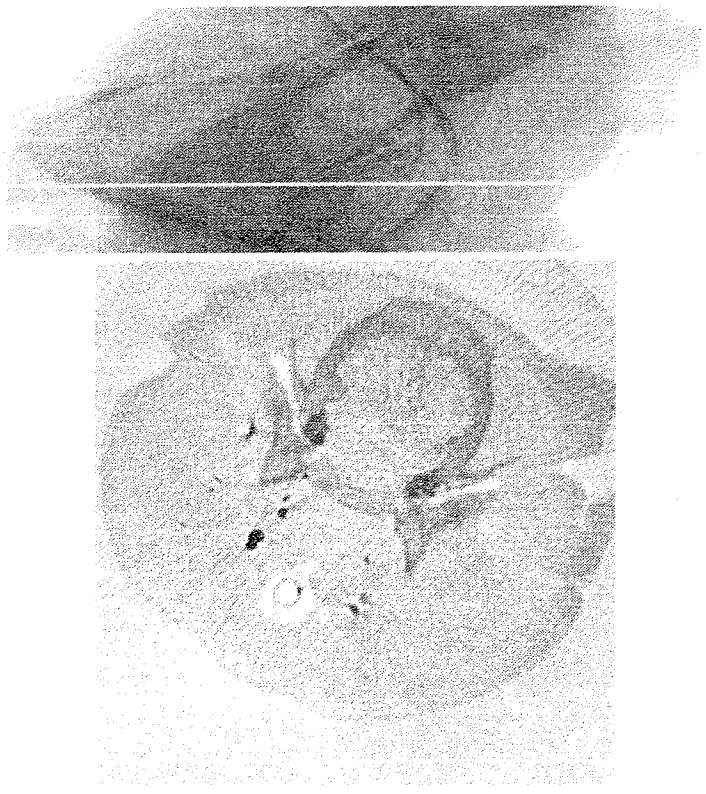

Figure 6: 1) Sinogram with a missing projection. 2) Streak artifact visible near the carotid artery.

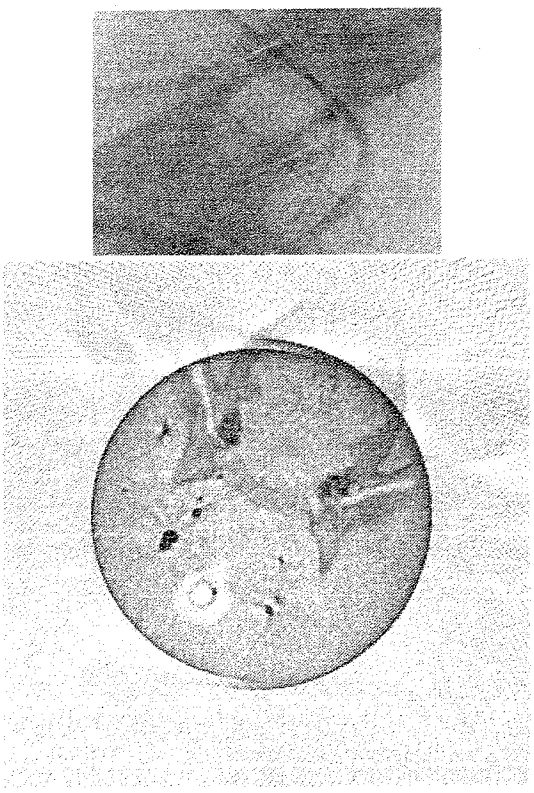

Figure 7: 1) A truncated sinogram. 2) Reconstruction using the truncated sinograms, note that the data outside the circle is artificial.

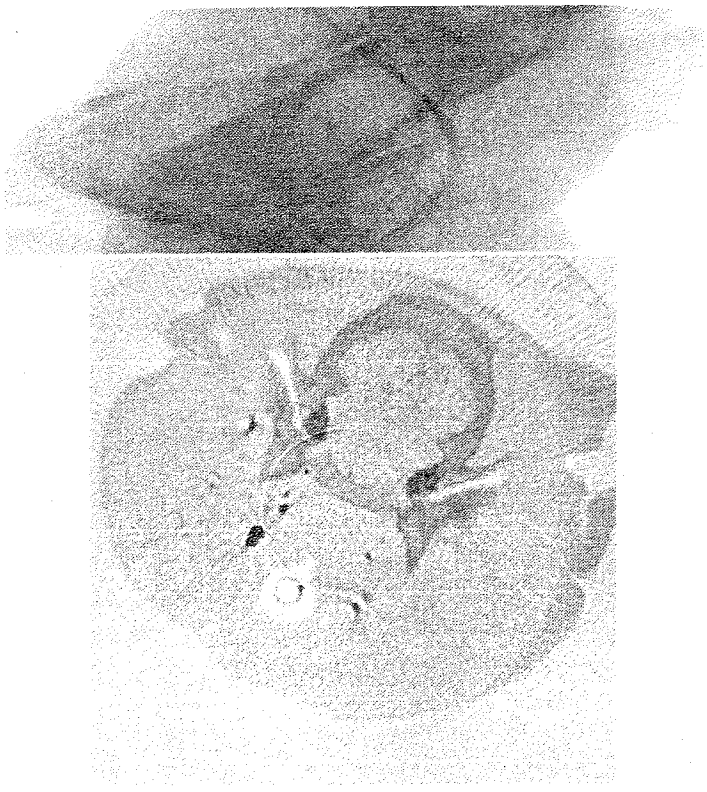

Figure 8: 1) sinogram with detector sensitivity reduced to 90%. 2) Reconstruction of the data showing a ring artifact.

spective imaging techniques.

References

[1] R. Fahrig, A.J. Fox, S. Lownie and D.W. Holdsworth, "Use of a C-arm system to generate true three-dimensional computed rotational angiograms: preliminary in vitro and in vivo results", ANJR 18, 1507-1514 (1997).

[2] R. Fahrig, M. Moreau and D.W. Holdsworth, "Three-dimensional computed tomographic reconstruction using a C-arm mounted XRII: correction of image intensifier distortion", Med. Phys. 24, 1097-1106 (1997).

[3] The International Study of Unruptured Intracranial Aneurysms Investigators, "Unruptured intracranial aneurysms - risk of rupture and risks of surgical intervention", New Eng. J. of Med. 399, 1725-1733 (1998).

[4] G.J. Rinkel, M. Djibuti and J. van Gijn, "Prevalence and risk of rupture of intracranial aneurysms: a systematic review", Stroke 29, 251-256 (1998).

[5] G. Guglielmi and F. Vinuela, "Intracranial aneurysms. Guglielmi electrothrombotic coils", Neurosurgery clinics of North America 5, 427-435 (1994).

[6] B. Schueler, "The AAPM/RSNA Physics tutorial for residents", Imaging and Therapeutic Technology, volume 18 number 3, 729-744 (1998).

[7] A. Macovski, "Basic concepts of reconstruction algorithms", Principles of reconstruction algorithms, Chapter 110, 3877-3887.

[8] R.A. Brooks, G. Di Chiro, "Principles of computer assisted tomography (CAT) in radiographic and radio-isotopic imaging", Phys. Med. Biol., vol 21, 689-732, 1976.

[9] R.H. Huesman, G.T. Gullberg, W.L. Greenberg, T.F. Budinger, "Donner algorithm for reconstruction tomography", Lawrence Berkeley Laboratory, University of California, Berkely, 1977.

[10] C.R. Crawford, G.T. Gullberg, B.M.W. Tsui, "Reconstruction for fan beam with an angular-dependent displaced center-of-rotation", Med. Phys, vol 15, 67-71, 1988.

[11] J.A. Concepcion, J.D. Carpinelli, G. Kuo-Petravic, S. Reisman, "CT fan beam reconstruction with a non-stationary axis of rotation", IEEE Trans. on Med. Imaging, vol 11, 111-116, 1992.

[12] D. Holdsworth, "Artifacts in computed tomography", Medical Biophysics 515a/475a course material for Medical Imaging, 1998.

[13] G.N. Ramachandran, A.V. Laksminarayanan, Proc. Natl. Acad. Sci. U.S., 68, 2236-2240 (1971).

[14] L.A. Shepp, B.F. Logan, IEEE Trans. Nucl. Sci, NS-21, 21-43 (1974).

[15] L.A. Feldkamp, L.C. Davis, J.W. Kress, "Practical cone-beam algorithm", J. Opt. Soc. Am. A, Vol. 1, No. 6, 1984.

[16] R. Fahrig, A. Fox, D Holdsworth, "Characterization of a C-arm mounted XRII for 3D image reconstruction during interventional neuroradiology", Proc SPIE Medical Imaging 2708, 351-360 (1997).

[17] W. Heidrich, M. McCool, J. Stevens, "Interactive maximum projection volume rendering", 11-18, IEEE (1995).

[18] G.H. Glover, N.J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstruction", Med. Phys., 799-807 (1981).

[19] C. Axelsson, P.E. Danielsson, "Three-dimensional reconstruction from cobe-beam data in $O(n^3 logn)$ time", Phys. Med. Biol., 39: 477-491, 1994.

[20] H. Kodo, F. Noo, M. Defrise, "Cone-beam filtered back-projection algorithm for truncated helical data", Phys. Med. Biol., 43: 2885-2909, 1998.

[21] J. Weese, T.M. Buzug, C. Lorenz, and C. Fassnacht, "An approach to 2D/3D registration of a vertebra in 2D X-ray fluoroscopies with 3D CT images", Proc. CVRMed/MRCAS, 119-128 (1997).

[22] G. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill and D.J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration", IEEE Trans. on Med. Imaging, vol 17, 586-595 (1998).

[23] D.W. Holdsworth, A.J. Fox, S.P. Lownie, "Image-guided neurovascular therapy using 3D computed rotational angiography", Medical Research Council of Canada research proposal (1998).

[24] R.R. Galigekere, K. Wisent, D.W. Holdsworth, "Techniques to alleviate the effects of view aliasing in computed tomography", Medical Physics (1998).

[25] M. Prokop, H.O. shin, A. Schanz, C.M. Schaefer-Prokop, "Use of maximum intensity projections in CT angiography: a basic review", radiographics 17, 433-451 (1997).

[26] B.E. Chapman, A.R. Sanderson, K.C. Goodrich, A.L. Alexander, D.D. Blatter, D.L. Parker, "Observer performance methodologies for evaluating blood vessel visibility in MR angiograms using accurate geometric registration to high resolution x-ray angiograms", MRM 37, 519-529 (1997).

[27] B.E. Chapman, K.C. Goodrich, A.L. Alexander, D.D. Blatter, D.L. Parker, "Evaluation of measures of technical image quality for intracranial magnetic resonance angiography", Computers and Biomedical Research 32, 530-556 (1999).

[28] P. Lacroute, M. Levoy, "Fast volume rendering using a shear-warp factorization of the viewing transformation", ACM SIGGRAPH 94, 451-457 (1994).

[29] H. Pfister, J. Hardenbergh, J. Knittel, H. Lauer, L. Seiler, "The VolumePro real-time ray-casting system",

[30] L.G. Brown, "A survey of image registration techniques", ACM Computing surveys 24, 326-276 (1992).

[31] P.M. Joseph, C. Ruth, "A method for simultaneous correction of spectrum hardening artifacts in CT images containing both bone and iodine", Med. Phys. 24, 1629-1634 (1997).

[32] K. Mueller, R. Yagel, "Fast perspective volume rendering with splatting by utilizing a ray-driven approach", IEEE 1996 Symposium on volume visualization , 65-72 (1996).

[33] B. Lichtenbelt, R. Crane, S. Naqvi, "Introduction to volume rendering", Prentice Hall, NJ: 1998.

[34] M.F. Santarelli, V. Positano, L. Landini, A. Benassi, "Volume rendering in medicine: the role of image coherence", Computers in cardiology 24, 323-326 (1997).

[35] M. Brady, K. Jung, H. Nyugen, T. Nyugen, "Two-phase perspective ray-casting for interactive volume navigation", IEEE Visualization 97, 183-189 (1997).

[36] J. Freund, K. Sloan, "Accelerated volume rendering using homogenous region encoding", IEEE, 191-195 (1997).

[37] D.L. Wilson, D.D. Royston, J.A. Noble, J.V. Byrne, "Determining x-ray projections for coil treatments of intracranial aneurysms", IEEE Trans. on med. imaging 18, 973-980 (1999).

[38] W. Schroeder, K. Martin, B. Lorensen, "The visualization toolkit", 2nd ed., Prentice Hall (1998).

[39] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes, "Computer Graphics: principle and practice", 2nd ed. in C, Addison-Welsey (1997).

[40] P. Bergeron, R. Carrier, D. Roy, N. Blais, J. Raymond, "Radiation to patients in neurointerventional procedures", AJNR 15, 1809-1812 (1994).