UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

JERÔNIMO GUSTAVO GRANDI

Multidimensional Similarity Search for 2D-3D Medical Data Correlation and Fusion

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Prof. Dr. Anderson Maciel
Advisor

Porto Alegre, September 2014

CIP – CATALOGING-IN-PUBLICATION

Jerônimo Gustavo Grandi,

Multidimensional Similarity Search for 2D-3D Medical Data Correlation and Fusion /

Jerônimo Gustavo Grandi. – Porto Alegre: PPGC da UFRGS, 2014.

88 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2014. Advisor: Anderson Maciel.

1. 2D-3D Image fusion. 2. Medical images. 3. Volumetric visualization . 4. Within-modality registration. I. Maciel, Anderson. II. Título.

Multidimensional Similarity Search for 2D-3D Medical Data Correlation and Fusion

ABSTRACT

Images of the inner anatomy are essential for clinical practice. To establish a correlation between them is an important procedure for diagnosis and treatment. In this thesis, we propose an approach to correlate within-modality 2D and 3D data from ordinary acquisition protocols based solely on the pixel/voxel information. The work was divided into two development phases. First, we explored the similarity problem between medical images using the perspective of image quality assessment. It led to the development of a 2-step technique that settles the compromise between processing speed and precision of two known approaches. We evaluated the quality and applicability of the 2-step and, in the second phase, we extended the method to use similarity analysis to, given an arbitrary slice image (2D), find the location of this slice within the volume data (3D). The solution minimizes the virtually infinite number of possible cross section orientations and uses optimizations to reduce the computational workload and output accurate results. The matching is displayed in a volumetric three-dimensional visualization fusing the 3D with the 2D. An experimental analysis demonstrated that despite the computational complexity of the algorithm, the use of severe data sampling allows achieving a great compromise between performance and accuracy even when performed with low gradient intensity datasets.

Keywords: 2D-3D image fusion, medical images, volumetric visualization, within-modality registration.

Busca de similaridade para correlação e fusão de imagens médicas multidimensionais

RESUMO

Imagens da anatomia interna são essenciais para as práticas médicas. Estabelecer correlação entre elas, é um importante procedimento para diagnóstico e tratamento. Nessa dissertação, é proposta uma abordagem para correlacionar dados multidimensionais de mesma modalidade de aquisição baseando-se somente nas informações de intensidade de pixels e voxels. O trabalho foi dividido em duas fases de implementação. Na primeira, foi explorado o problema de similaridade entre imagens médicas usando a perspectiva de análise de qualidade de imagem. Isso levou ao desenvolvimento de uma técnica de dois passos que estabelece um equilíbrio entre a velocidade de processamento e precisão de duas abordagens conhecidas. Avaliou-se a qualidade e aplicabilidade do algoritmo e, na segunda fase, o método foi estendido para analisar similaridade e encontrar a localização de uma imagem arbitrária (2D) em um volume (3D). A solução minimiza o número virtualmente infinito de possíveis orientações transversais e usa otimizações para reduzir a carga de trabalho e entregar resultados precisos. Uma visualização tridimensional volumétrica funde o volume (3D) com a imagem (2D) estabelecendo uma correspondência entre os dados. Uma análise experimental demonstrou que, apesar da complexidade computacional do algoritmo, o uso de amostragem, tanto na imagem quanto no volume, permite alcançar um bom equilíbrio entre desempenho e precisão, mesmo quando realizada com conjuntos de dados de baixa intensidade de gradiente.

Palavras-chave: Fusão de imagens 2D-3D, Imagens médicas, Visualização volumétrica, Regristro de dados de mesma modalidade.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
CBE	Cell Broadband Engine
CC	Correlation Coefficient
CPU	Central Processing Unit
CT	Computed Tomography
CTA	Computed Tomography Angiography
CUDA	Compute Unified Device Architecture
DICOM	Digital Imaging and Communications in Medicine
DSA	Digital Subtraction Angiography
fMRI	Functional Magnetic Resonance Imaging
GPU	Graphics Processing Unit
HVS	Human Vision System
IQA	Image Quality Assessment
MI	Mutual Information
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean squared error
MSSIM	Mean Structural Similarity
NCC	Normalized Cross-Correlation
NMI	Normalized Mutual Information
NMR	Nuclear Magnetic Ressonance
OpenCV	Open Source Computer Vision Library
OpenGL	Open Graphics Library
OpenMP	Open Multi-Processing
PCA	Principal Component Analysis
PET	Positron Emission Tomography

PS3	PlayStation 3
PSNR	Peak Signal-to-Noise Ratio
RAM	Random Access Memory
RAW	Raw image file
SAD	Sum of Absolute Differences
SIFT	Scale-Invariant Feature Transform
SPECT	Single Proton Emission Computed Tomography
SSD	Sum of Squared Differences
SSIM	Structural Similarity
US	Ultrasound

CONTENTS

1 INTRODUCTION

Images of the inner anatomy are essential for clinical practice as they support diagnosis, surgery planning, lesions characterization and physiological studies to treat human diseases. Several image acquisition modalities are available. They emphasize different internal body structures (anatomical or functional) and present varied number of dimensions. Some of them are captured and analyzed in real-time, others through batch processes for posterior analysis. Computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, x-ray and laparoscopic images are the most common. Today, all types of medical images are available in digital representations. This allows computational systems to perform medical data analysis and image processing. Virtual three-dimensional representations of datasets from MRI and CT, segmentation and classification of organs functional segments, similarity analysis and image registration are some examples.

To establish a correlation between medical images from multiple sources is an important procedure for diagnosis and treatment. While within-modality image correlation allows following disease evolution, cross-modality correlation permits to combine a perception of the different structures visible in either modality. Fusion of computed tomography (CT) and ultrasound (US), for example, can improve the diagnostic value to an extent beyond the sum of the individual modalities (WEIN et al., 2008). Despite its importance, this correlation procedure is most often performed manually in clinical practice. The physician's confidence in a lesion identification and classification with multimodal acquisition sources is still a challenging clinical procedure (WOOD; CHUNG; ALLEGRA, 2000). Results are then quite dependent of the physician's skills to produce a mental image of the data correlation.

It is expected that fast and accurate automatic medical image correlation of within-modality and cross-modality can lead to an improved diagnosis and treatment. Image features captured in different timestamps or with different imaging devices would be highlighted in a fused single view, increasing accuracy in diseases localization and precise apparatus placement in lesion ablation.

The registration of medical images is widely explored (MAINTZ; VIERGEVER, 1998; KHALIFA et al., 2011; MANI; RIVAZHAGAN, 2013). Solutions for a variety of images correlations have been proposed. They approach registration between pre-operative data from same or different dimensionalities and modalities, such as 2D-3D: X-ray CT and MRI (TOMAEVIC; LIKAR; PERNUš, 2002); 3D-3D: whitin-modality MRI (BEN SASSI et al., 2008), CT-MRI (DEAN et al., 2012). These works focus on registration between images in the three main anatomical axes, axial, coronal and sagittal.

On the other hand, in registration between pre- and intra-operative modalities, such as US-CT (WEIN et al., 2008; CROCETTI et al., 2008), US-MRI (KING et al., 2010; NIKAS et al., 2003; LINDSETH et al., 2003) the captured data are often misaligned to the main

axes. To overcome the user dependent acquisition, previous solutions track the position and orientation of the US imaging transducer and patient's body to match the modalities coordinates system.

1.1 Goal

The ideal system would be able to perform real-time fusion of distinct modalities, even when the data is freely captured disrespecting the anatomical axes. This leads to a variety of subproblems, such as, to establish a matching criteria, defining the geometrical transformations to align the distinct data, implementing algorithm optimizations and finding the best fitting plane in the three-dimensional data given the 2D arbitrary image. The challenges involved in each sub-problem are complex research subjects and involve the computer science and medicine fields.

In this thesis, we design an approach to correlate within-modality 2D and 3D data from ordinary acquisition protocols based solely on the pixel/voxel information. Our method uses similarity analysis to, given an arbitrary slice image (2D), find the location of this slice within the volume data (3D). Our solution minimizes the virtually infinite number of possible cross section orientations and uses optimizations to reduce the computational workload and output accurate results.

Volumetric rendering techniques are applied for visualization. The volumetric medical image is cross-sectioned to highlight the 2D image location.

An extensive evaluation was performed to assess the precision, accuracy and efficiency of the proposed technique.

1.2 Contributions

Following the technical contributions of this thesis are summarized:

- A 2-step algorithm for multidimesional medical images matching;
- A multidimensional similarity search for 2D-3D data correlation and fusion;
- Three sets of experiments performed to evaluate the quality and applicability of our methods;
- A visualization tool that integrates real-time volumetric rendering of dataset cross-sectioned with arbitrary plane image.

1.3 Outline of this thesis

In the next chapter, we introduce the fundamental concepts for the understanding of this thesis. The related works are presented in Chapter 3. In the next two chapters we deal with the investigation, implementation and evaluation of two medical image matching techniques. They are the result of a spiral methodology, where, in the first research cycle (Chapter 4), we have explored the similarity between images using the perspective of image quality assessment for image similarity. It led to the development of a two-step technique for within-modality medical image matching. We evaluate its quality and applicability and, in the second cycle (Chapter 5), we extend the method to match arbitrary plane images within volumetric data. In the end of this cycle, we have evaluated the technique with controlled data and in Chapter 6 we validate our prototype in a medical context. The conclusion and next research cycles are in Chapter 7.

2 BACKGROUND

This chapter introduces the fundamental concepts developed throughout this thesis. They are: medical imaging (Sec. 2.1), medical image registration (Sec. 2.2), volumetric visualization of medical data (Sec. 2.3) and image similarity assessment (Sec. 2.4).

2.1 Medical Imaging

Medical imaging aims to produce images from the human body internals. It is important for the understanding of the human anatomy, function and physiological behavior of organs and tissues. Thus, it is an essential tool for diagnosis as well as monitoring and evaluation the treatment of an illness.

Medical imaging modalities can be classified into anatomical (or structural) and functional (or metabolic) (DHAWAN, 2011). The anatomical imaging modality can be characterized depending on its ability to discriminate different body structures such as bone, soft tissue and body fluids. The main anatomical imaging modalities include X-ray imaging (X-ray radiograph, X-ray computer tomoghaphy (CT)), ultrasound (US), and magnetic ressonance imaging (MRI). The functional imaging modality is characterized depending on its ability to discriminate different levels of metabolism caused by specific biochemical activity. Major functional imaging modalities include functional MRI (fMRI), single proton emission computed tomography (SPECT), positron emission tomography (PET) and fluoroscopy imaging.

Only the anatomical modalities are covered in this work and detailed in the next subsections.

2.1.1 Radiograph

The radiograph imaging (Fig. 2.1a) records a two-dimesional projection of a three-dimensional (3-D) human body anatomical structure. The radiograph images are composed in gray gradient where the tissue density is mapped to an intensity value. This imaging modality is effective to accuse fractures, tumors, pneumonia, among other information being less used to visualize soft tissues.

The X-ray is situated between the ultraviolet and gamma rays in the elelectromagnetic spectrum. Because of their wavelength, X-rays have high energy providing excellent capability of straight penetration and transmission in human body. There are two main wavelength ranges in the X-ray spectrum, the called soft X-rays, identified as having wavelengths from 10nm to 0.1nm and the hard X-rays with wavelenghts shorter than 0.1nm down to 0.001nm.

For medical applications, it is important to select X-rays with wavelengths that provide

linear attenuation coefficients for human body. The intensity attenuation is determined at each scan location by measuring the difference between the source and the detector due to the mass attenuation coefficients of body's physiological structures. The soft X-ray range is quite reasonable to discriminate bones, soft tissues and air. For this reason are the most used for medical diagnostic purposes.

Wavelengths shorter than those used for diagnosis provide much higher photon energy and therefore less attenuation. This hard X-ray is commonly applied in therapeutic procedures. These high energy particles create molecule ruptures that affects cancerous cells as they generally have poorer repair capacities than healthy tissues.

The rays emitted by the X-ray may have a harmful effect on the human body depending on the time of exposure to its radiation.

2.1.2 Computed Tomography

Conventional radiography creates a 2-D projection image of a 3-D object. While radiographs may be adequate for many diagnostic applications, it does not provide 3D quantitative and qualitative information about the anatomical structures and associated pathology necessary for diagnostic and treatment of a number of diseases. For example, the volume of an organ and the tumor size and its shape are important features for diagnostic and surgery planning purposes (DHAWAN, 2011). The basic principle of X-ray CT is the same as that of X-ray digital radiography. But, in the CT imaging, X-rays beams are emitted from different angles through an X-ray tube that rotates around the axis of the pacient's body. Mathematical reconstruction techniques creates cross-sectional slice images of X-Ray attenuation inside the human body (RUSS, 2011). A stack of axial slices are obtained from few millimeters between them, comprising a three-dimensional dataset, as shown in Figure 2.1b. Both for X-Ray and CT, the vasculature can be emphasized by injecting contrast agent, a fluid with high attenuation. The overall imaging is then denoted Angiography or Computed Tomography Angiography (CTA).

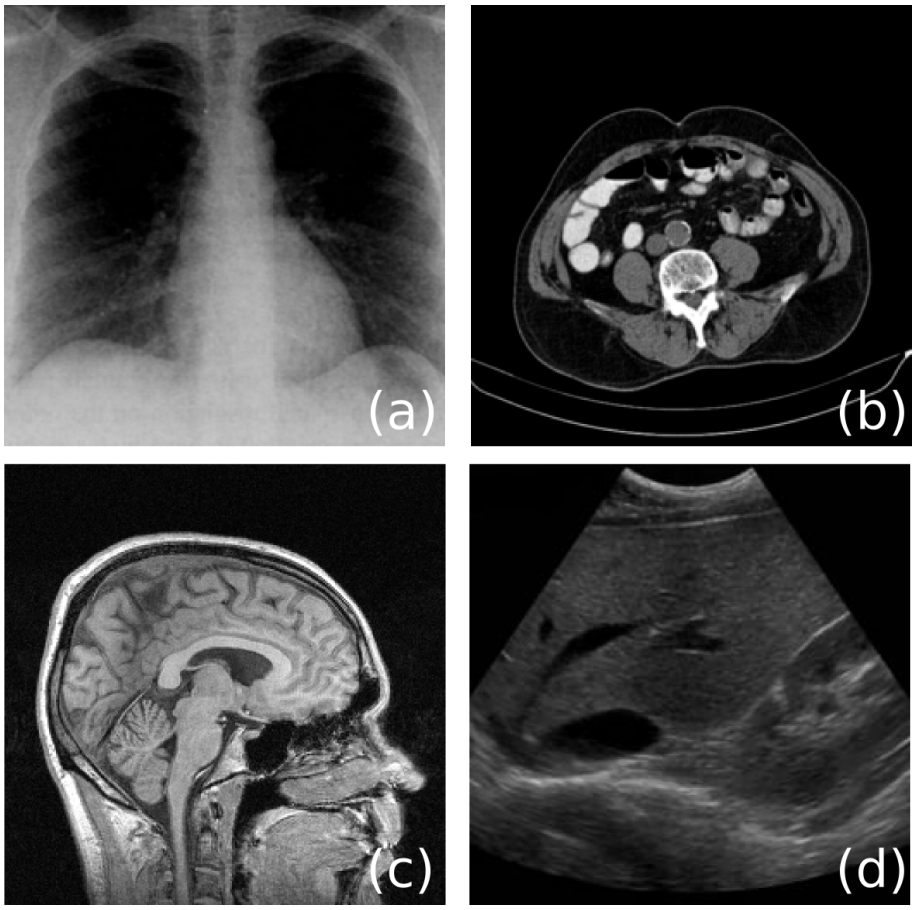
2.1.3 Magnetic Resonance

Magnetic resonance (MRI), like the CT, is a tomographic imaging method that produces 3D images of the human body (DHAWAN, 2011). But, unlike CT it is not based on the transmission of external radiation for imaging, and is therefore considered a non-invasive modality. MRI uses nuclear magnetic ressonance (NMR) property of selected nuclei. Among the nuclei that generate signals of MRI, the most important is the hydrogen nucleus. The hydrogen protons are readily available in water and other fluids in the body and vary in density within a tissue if its chemical composition changes, providing the basis for better imaging. Morover, they are stable and very sensitive to magnetic resonance core in tissues. Because of the underlying physics of MRI acquisiton, it provides high-resolution images with superior soft-tissue contrast when compared to X-ray CT. An MRI image example is shown in Figure 2.1c.

2.1.4 Ultrasound

Medical ultrasound imaging (Fig. 2.1d) is a non invasive technique that uses sound pulses in order to visualize biological tissues. Different from CT and MRI modalities, the ultrasound allows real-time imaging with user interaction. An ultrasound transducer device is positioned at the patients skin and transmits sound pulses through the patient's body. The sound pulses are reflected at the boundary between two different density tissues. The

Figure 2.1 - Visual representation of the main medical imaging acquisiton sources. (a) Radiograph, (b) Computed tomography, (c) Magnetic ressonance (d) Ultrasound.



Source: Compiled by author.

higher the difference the stronger the resulting echoes are. The echoes are detected and captured by the transducer and sent to a computer for image processing and visualization. The depths of the boundaries are estimated based on the time interval between sending the pulse and receiving the echoes, which is also known as the pulse-echo principle (GILL, 2012). The received signal is post processed and the resulting echoes are displayed using the depth information and a brightness value according to the signals strength of the received echoes. More details on image formation in real-time ultrasound can be found in HEDRICK; HYKES (1995) and (DHAWAN, 2011).

2.1.5 Medical Images Classification

We summarize this section with a classification of the medical imaging methods discussed, seen in Table 2.1. We highlighted the main modalities' characteristics regarding the focus of this thesis.

2.2 Medical Image Registration

Image registration (JV HAJNAL, 2001) also known as image matching or alignment, tries to establish spatial correspondence between bidimensional and/or tridimensional images. Several similar scenes can be captured and registered. The captured data may

Table 2.1 - Classification of the medical imaging methods.

Modality	Radiograph	CT	MRI	US
Acquisition	X-ray		magnetic ressonance	sound pulse-echo
Structures	hard tissues		soft tissues	hard and soft tissues
Operative phase	pre-operative			intra-operative
Invasive	yes		no	

Source: Compiled by author.

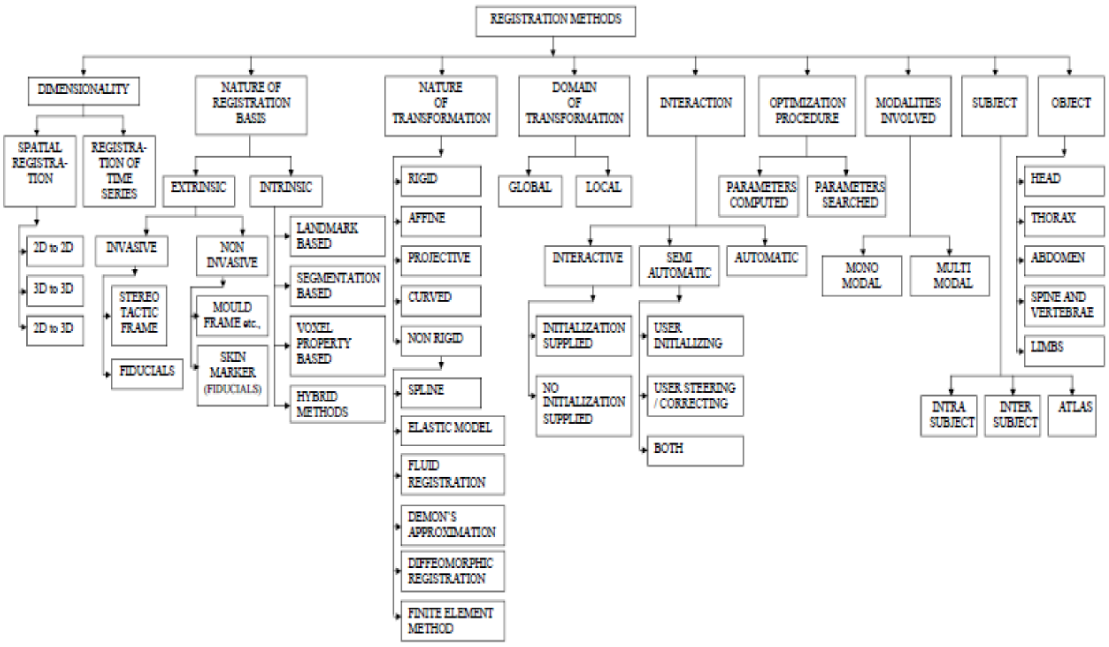
be acquired from different sensors (image modalities), different time instants (temporal/functional data) or from different viewpoints.

Since information captured from distinct imaging modalities is usually complementary, means to proper integrate this separate images are needful. The main step in the integration process is to spatially align the modalities involved. This step, plays an important hole in medical image analysis, both in research and medical applications.

Medical image registration can be applied during diagnosis, surgery planning, guiding treatment, and monitoring disease progression. It has fundamentally many uses in medicine such as multi-modality fusion, image-guided surgery, monitoring changes in size, shape, or image intensity of tissues over time, image segmentation, functional brain mapping and many others.

All current image registration methods can be classified based on the criteria formulated by Maintz et al (MAINTZ; VIERGEVER, 1998) and Mani et al (MANI; RIVAZHAGAN, 2013). The diagram in Figure 2.2 shows all criteria and in the following subsections they are explained.

Figure 2.2 - Medical image registration classification.



Source: Diagram extracted from MANI; RIVAZHAGAN (2013).

2.2.1 Dimensionality

The registration process involves the calculation of transformations between coordinate system. There are three main dimensionalities involved, 2D-2D, 2D-3D and 3D-3D. In addition, images can be captured at time intervals creating a time series.

- 2D-2D: The simplest dimensionality registration. It uses only rotations and translations to image alignment.

- 2D-3D: Establish correspondence between 3D volumes and projected images, such as, x-ray. The main application for this type of registration is in image guided surgery.

- 3D-3D: Registration of volumetric data provide spatial relationship between the internal organs. The input images can be from the same or distinct modalities. The scanner devices should be calibrated to match the voxels size.

- Temporal series: The medical images are captured over time. They are used to monitor the progress of bodily disorders and diseases, for example, to monitor the tumor growth.

2.2.2 Modalities Involved

The correlation between medical images can be classified in within-modality and cross-modality. The clinically established modalities are computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and single-photon emission computed tomography (SPECT) (TURGEON et al., 2005; MATTES et al., 2003; DU et al., 2006; WONG; BISHOP, 2007).

- Within-modality: The same imaging device is used to capture the same type of information. This method has application in growth monitoring, verify results of surgical interventions (kidney removal) and others.

- Cross-modality: different imaging devices are used to capture information from the same anatomy. Multi-modal registration is more complex due to distinct nature of the signals. The applications of these methods are more diverse, but predominately used during the diagnostic phase.

2.2.3 Nature of registration basis

Can be divided into intrinsic, methods based only on information of the captured image, or extrinsic, use external objects inserted in the imaged space during the capture procedure.

- Intrinsic: The main approaches are based on the voxels intensity and on the image characteristics (landmarks) (CHUI; RANGARAJAN, 2003).

- Extrinsic: Fiducial marks are inserted in the imaged area or attached to the patient body. The markers must be visible by all modalities. Detection algorithms locate patterns and calculate the image transformations needed for the correct alignment. This registration method can be easily automated and is computationally efficient.

2.2.4 Nature of the transformation

The movements or distortions required to proper align the images. The main transformations are rigid, affine and non-rigid.

- Rigid Transformation: only rotations and translations are applied. Examples are the non-deformable register of body structures, such as bones. The transformations are applied globally. Rigid transformations are popular because they have a good registration approximation and because they are fast computed (few parameters have to be taken into account during the calculations).

- Affine transformation: Preserves the parallelism between lines. This transformation allows more freedom when compared to rigid transformation by adding scale and shear in each dimension.

- Non-rigid Transformation: The human body, due to its composition, has many soft tissues, e.g. organs and skin structures. Therefore, a rigid or affine transformation does not always provide a good registration approximation. Thus, the registration using non-rigid transformations is ideal to establish better correlation between soft tissue structures. This transformations may be non-linear, relies on complex equations and remains a research problem due to the high degree of freedom in the deformation process. Several algorithms have emerged for non-rigid registration. All techniques require high computational power, which is the biggest problem for most medical applications. The main techniques used in non-rigid transformations are: splines, elastic models and finite elements.

2.2.5 Domain of transformation

The transformation can be global if it is applied to the entire data and local if it is applied to a data's portion.

2.2.6 Optimization procedure

Optimization procedures are used to fast and reliably determine the optimum transformation parameters for the image alignment. The transformation parameters are computed in an explicit way from the available data, or by finding an optimum of some cost function defined. The cost function determines the similarity between the two images given a certain transformation. The functions are less complex in mono modal registrations since there is a linear relationship between the source and the target image and the similarity metric is straightforward (MANI; RIVAZHAGAN, 2013).

For the non-rigid registration, it is hard to choose/build the optimizer because more parameters are generally required to describe it.

2.2.7 Subject

Subject refers to the patients, whose images are to be registered. The images to be registered may be captured from the same patient or from different patients. Based on the subject, registration algorithms can be classified into intra-subject, captured from a single patient, inter-subject, captured from different patients and atlas, acquired from a single patient and registred with a image constructed from many subjects.

2.2.8 Object

The object in medical image registration is the part of the anatomy involved. The main anatomic structures imaged are: head (brain, retina, dental), thorax (breast, cardiac), abdomen (liver, kidney, prostate), spine and vertebrae and limbs.

2.2.9 Important classifications for this work

This work is specially focusing on:

- Dimensionality (Subsection 2.2.1): we are dealing with 2D and 3D images correlation;
- Modalities Involved (Subsection 2.2.2): within-modalities;
- Nature of registration basis (Subsection 2.2.3): pixel/voxel-based intrinsic methods;

2.3 Volumetric Visualization of Medical Data

Both clinical diagnostics and registration need means to visualize the informations contained within a volume. In earlier methods, the visualization was done by extracting surfaces from the volumetric information and transform them into graphical primitives that could be rendered with traditional techniques. This approach is referred to as indirect volume rendering (MEIBNER et al., 2000). However, such transforming approach is susceptible to loose volume information. In this way, direct volume rendering techniques (MAX, 1995) have been proposed, that construct and render the volumetric dataset directly, without changes is the representation. The pioneer methods, were volume ray-casting (LEVOY, 1988) and volume splatting (WESTOVER, 1990). But, due to the high computational demands of this methods, they were not capable to build 3D volumetric renderization in real-time. Figure 2.3 shows the visual differences of indirect and direct methods.

Today, the advance of graphics hardware and their programmability capabilities allowed texture based direct volume rendering with high quality and better performance (ENGEL; KRAUS; ERTL, 2001; ENGEL et al., 2004). Some examples of medical volumetric rendering are shown in Figure 2.4.

The volumetric rendering techniques are applied in many scientific fields, such as biology, archeology, geology, material science, computational science and engineering; and entertainment like games and movies (HADWIGER et al., 2006). Medical imaging was one of the early fields that adopted volume rendering. In medical imaging, 3D data is typically acquired by some kind of scanning device (Section 2.1). Stacks of two-dimensional images from CT and MRI are the main data sources for these applications. Both modalities generate an image series that represents cross-sections through the patient's anatomy. The unitary volumetric value is called voxel.

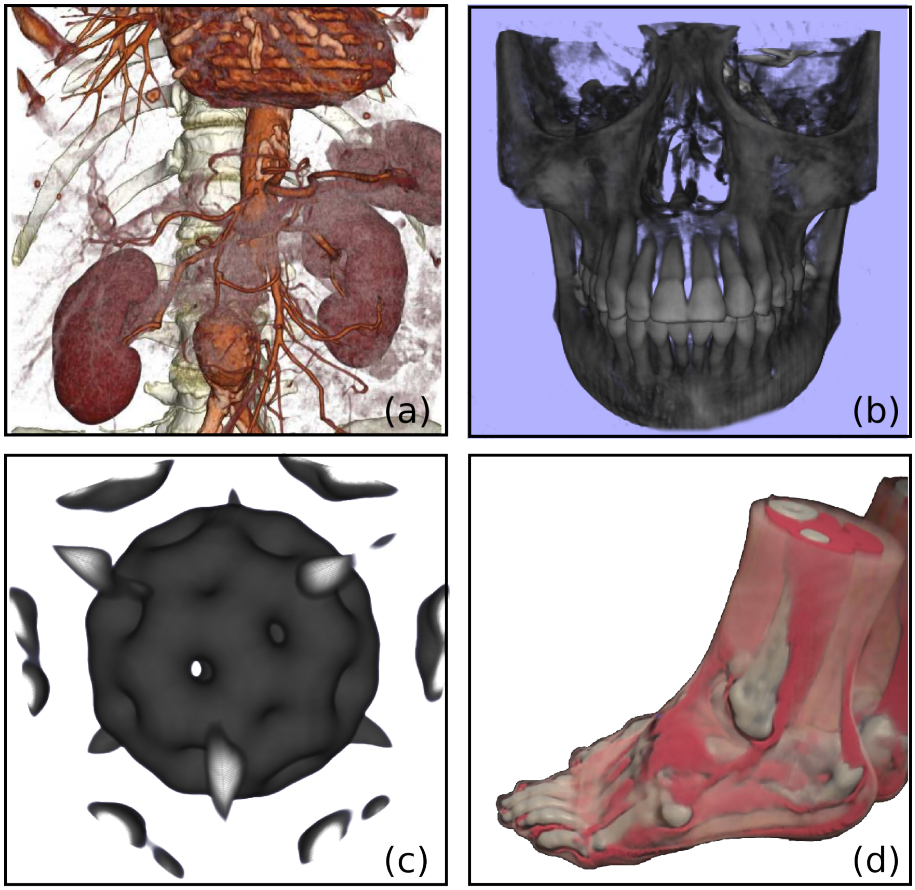
Although medical professionals are able to perform diagnoses based on the raw data, three-dimensional visualization is proving to be useful especially in complex organs like the liver and its vascular system, diagnosis, treatment planning, surgical simulation, medical training and intra-operative procedures (TOMBROPOULOS, 1999). The three-dimensional visualization methods provides very precise anatomical details and it allows a faster learning curve for medicine students and also a reduction of medical errors caused by misinterpretation of two-dimensional data that may result in complications to the patient. This occurs especially with less experienced physicians (MARESCAUX et al., 1998).

Figure 2.3 - Comparison between indirect and direct. (a) Traditional surface representation of a volumetric dataset. (b) Direct volume rendering of the same dataset.

(a) (b)

Source: Image taken from (BRUCKNER, 2008).

Figure 2.4 - Volumetric rendering examples. (a) high-quality volumetric rendering of an abdomen CT scan, (b) skull CT, (c) simulation of the electron density of a Buckminster-Fullerene molecule and (d) feet of the Visible Female CT dataset.



Source: (a)Image taken from (BRUCKNER, 2008), (b) and (c) Compiled by author and (d)Image taken from (HADWIGER et al., 2006)

2.4 Image Similarity Measures

Similarity functions quantify signal correspondences between a target and a reference image. They should have global maximum when the evaluated images are identical or are correctly aligned.

In the remaining of this section, we describe the most commonly used measures: sum of squared differences, structural similarity, normalized cross-correlation, mutual information and scale-invariant feature transform.

2.4.1 Mean Squared Error

Mean squared error (MSE), can be computed by averaging the squared intensity of the pixel differences in two images, normally one is the reference image and the other is the distorted one (to be evaluated). The result is a quantitative score that describes the degree of similarity or the level of distortion between them (WANG; BOVIK, 2009). The MSE between two images is given by:

$$MSE = \frac{1}{k}\sum_{i=1}^{k}(x_i - y_i)^2, \tag{2.1}$$

where x and y are the two images data, k is the number of signal samples (pixels) and x_i and y_i are the values of the ith samples in x and y, respectively. The difference between the original and distorted images is given by $e_i = x_i - y_i$. Similarity values close to zero represent better image match whereas $MSE = 0$ means that a perfect match was found.

According to WANG; BOVIK (2009), MSE has many positive features, here we highlight two of them:

- It is easy to compute and memoryless. The complexity is of only one multiplication and two additions per sample.
- It is a convention. It has been extensively used for optimization and assessment in signal compression, restoration, denoising, reconstruction, and classification. Moreover, throughout the literature, it provides a convenient and extensive standard against which new algorithms have most often been compared.

The main disadvantage of MSE is the restrictive assumption that the image intensities are identical. Therefore non-matching structures entering one of the images easily cause larger increases of the error value than the shift to a correct alignment. In that sense, its robustness with respect to outliers and noise is generally bad. This measure is mostly useful for registering images of the same modality.

2.4.1.1 Mean of Absolute Differences

The MSE measure is very sensitive to a small number of voxels that have very large intensity differences between images A and B. This might arise, for example, if contrast material is injected into the patient between the acquisition of images A and B or if the images are acquired during an intervention and instruments are in different positions relative to the subject in the two acquisitions.

One common modification to reduce the sensitivity with respect to outliers is to use the Mean of Absolute Differences (MAD).

$$MAD = \frac{1}{k}\sum_{i=1}^{k} \mid (x_i - y_i) \mid, \tag{2.2}$$

2.4.1.2 *Peak Signal-to-Noise Ratio*

In the literature of image processing, MSE is often converted into a peak signal-to-noise ratio (PSNR) measure:

$$PSNR = 10log_{10}\frac{L^2}{MSE}, \tag{2.3}$$

where L is the dynamic range of allowable image pixel intensities. For images that have allocations of 8 bits/pixel of gray-scale, $L = 2^8 - 1 = 255$. The PSNR is useful when comparing images with different dynamic range, otherwise it does not add new information relative to the MSE.

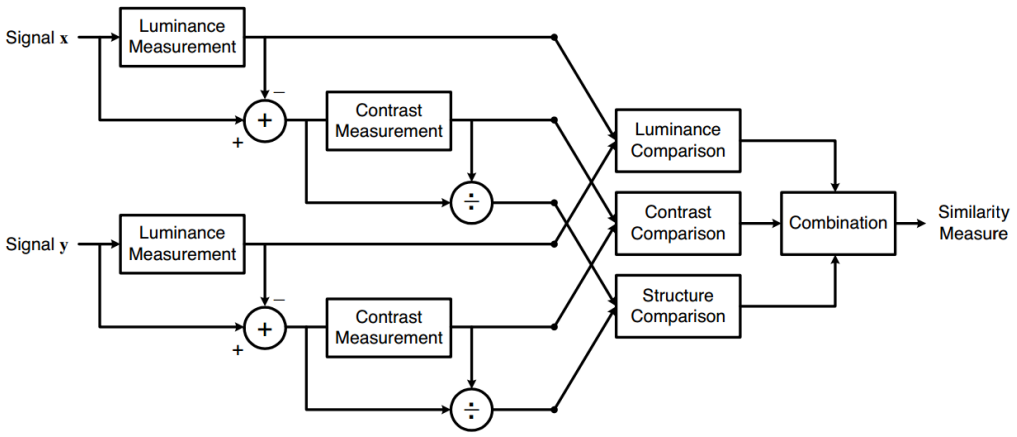
In summary, MSE deals with images in the pixel domain and it is not similar to the way that human visual system perceives images. Moreover, psychophysical and physiological features of the HVS are not accounted for by the MSE.

2.4.2 Structural Similarity Approach

The motivation of the structural similarity (SSIM) approach (WANG et al., 2004) is to find a more direct way to compare a reference and a distorted image. The pixels of natural images have strong dependencies that carry important information about the structure of objects in the scene, especially when pixels are spatially close. In this way, Wang et al. (WANG, 2001; WANG et al., 2002) proposed a new framework to assess image quality, based on the hypothesis that the human visual system is highly adapted to extract structural information from images seen.

The measure of structural information change can provide a good approximation to perceived image distortions. The system diagram of the proposed quality assessment system is shown in Fig. 2.5 where x and y are two nonnegative image signals. If we consider one of the signals to have perfect quality, then the similarity measure can serve as a quantitative measurement of the quality of the second signal. The task is separated into three comparisons: luminance, contrast and structure.

Figure 2.5 - Diagram of the structural similarity (SSIM) meassurement system.



Source: Image extracted from (WANG et al., 2004).

All three comparisons are mixed into one equation, formulated as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{2.4}$$

where μ_x, σ_x and μ_y, σ_y denote the mean intensity and standard deviation of images x and y, respectively, while σ_{xy} denote their cross correlation. C_1 and C_2 are small constant values used to avoid instability problems when the denominator is too close to zero. The SSIM mathematical formulation is detailed in WANG; BOVIK (2006).

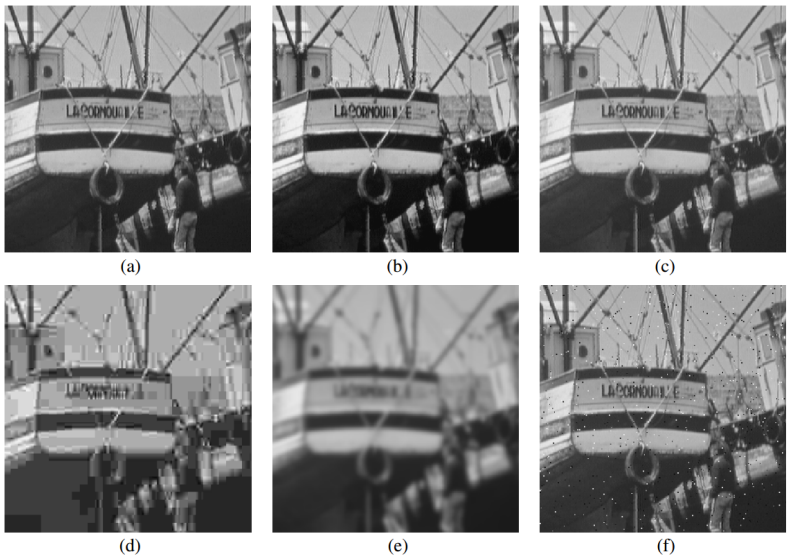
The mean SSIM (eq. 2.5) is used to assess the overall quality of an entire image.

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j), \tag{2.5}$$

where X and Y are the reference and the distorted images, x_j and y_j are the image at the jth local window and M is the number of local image windows.

The major difference between error sensitivity metric and the structural similarity approach is that the first estimates perceived errors to quantify image distortion while the second considers image distortion as perceived changes in structural information variation. Figure 2.6 shows an example extracted from WANG et al. (2004) that illustrates the differences between the approaches. As the MSE does not takes into account the image structural formation, it cannot predict witch image has the best quality, or is the less degraded. The SSIM has more consistent results.

Figure 2.6 - Quality comparison between MSE and MSSIM. Images with different types of distortions, all with MSE = 210. (a) Original image (8 bits/pixel; cropped from 512x512 to 256x256 for visibility). (b) Contrast-stretched image, MSSIM = 0:9168. (c) Mean-shifted image, MSSIM = 0:9900. (d) JPEG compressed image, MSSIM = 0:6949. (e) Blurred image, MSSIM = 0:7052. (f) Salt-pepper impulsive noise contaminated image, MSSIM = 0:7748.



Source: Images and data extracted from WANG et al. (2004).

2.4.3 Normalized Cross-Correlation

Normalized Cross-Correlation (NCC), also denoted as correlation coefficient (CC) is a very common technique in statistics, signal processing and many other fields. It assesses the amount of linear relationship between the intensity values in the images by computing the average product of its demeaned values, divided by their standard deviation. It is a

slightly less strict assumption than the mean-squared error. The correlation coefficient is presented in equation 2.6.

$$NCC = \frac{\sum_k (i_k - \bar{i})(j_k - \bar{j})}{\sqrt{\sum_k (i_k - \bar{i})^2 \sum_k (j_k - \bar{j})^2}}, \tag{2.6}$$

where $\bar{i}$ and $\bar{j}$ are the mean voxel values and have to be calculated previously. The CC is a normalized version of the cross correlation measure C (Equation 2.7)

$$C = \frac{1}{N} \sum_k i_k . j_k, \tag{2.7}$$

Unlike MSE, the NCC measure is independent of linear changes. If the intensities i and j are completely independent, the value of NCC will be around zero. If the linear mapping is correct for all intensities, its value will be 1 or -1 (depending if the linear mapping is positive or negative).

2.4.4 Mutual Information

Image registration can be thought as a process to maximize the amount of shared information in two images. The most commonly used measure of information in image and signal processing is the Shannon-Wiener entropy measure (Equation 2.8).

$$H = -\sum_i p_i \log_2 p_i, \tag{2.8}$$

where p_i are the marginal probability distribution function of image. Entropy will have a maximum value if all symbols have equal probability of occurring and have a minimum value of zero if the probability of one symbol occurring is one, and the probability of all the others occurring is zero.

2.4.4.1 Joint Entropy

Joint entropy measures (eq. 2.9) the amount of information that the combined images have. If A and B are totally unrelated, then the joint entropy will be the sum of the entropies of the individual images. The more similar the images are, the lower the joint entropy compared to the sum of the individual entropies.

$$H(A, B) \leq H(A) + H(B), \tag{2.9}$$

2.4.4.2 Conditional Entropy

The conditional probability $p(b \mid a)$ is the probability that B will take the value b given that A has the value a. The conditional entropy is therefore the average of the entropy of B for each value of A, weighted according to the probability of getting that value of A.

$$H(B \mid A) = H(A, B) - H(A) = -\sum_{i=1}^{n} \sum_{i=1}^{m} p_{ij} \log_2 p_{i|j}, \tag{2.10}$$

2.4.4.3 Mutual Information

Mutual information (MI) (MAES et al., 1997) is the most successful and commonly used universal and highly accurate similarity measure for aligning multimodal images and

2-D/3-D rigid and nonrigid registration. The MI equation is described below:

$$MI(A,B) = H(A) - H(B \mid A) = H(B) - H(A \mid B) = \sum_{i=1}^{n}\sum_{i=1}^{m} p_{ij} \log_2 \frac{p_{ij}}{p_i q_j} \quad (2.11)$$

where $H(.)$ is the Shannon's entropy (eq. 2.8) of the signals, $H(.,.)$ (eq. 2.9) is their joint entropy, and $H(. \mid .)$ (eq. 2.10) is the conditional entropy.

2.4.5 Scale-Invariant Feature Transform (SIFT)

Another approach that can be used to detect similarity between images is by feature matching. These techniques compute abstractions of image information and make local decisions in every image point where there are "interesting" points of a given type. The scale invariant feature transform (SIFT) (LOWE, 1999, 2004) is a robust image feature matching algorithm. It aims to solve many of the practical problems in interest points extraction and their use for matching images. One of the most important properties of the SIFT is that it is invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. The interest points are localized in both the spatial and frequency domains, reducing the probability of disruption by occlusion, clutter, or noise. The features are highly distinctive, which allows it to be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition. However, interest points extraction is costly even with a cascade filtering approach, in which the more costly operations are applied only at locations that pass an initial test. Preliminary tests with our set of medical images revealed that the SIFT can be used with our data, but it is not suitable for our problem. The CT datasets have very similar images among their neighbors. This causes the algorithm to detect the same features in more than one image, confusing the detection. The algorithm was compared with our solution and the results are exposed in Section 4.2.

The problem with feature-based techniques to matching medical images was discussed in the work by McLauglin et al. (MCLAUGLIN et al., 2001). The authors compared the intensity-based gradient difference similarity presented in (PENNEY et al., 2001) and a feature-based technique presented in (KITA; WILSON; NOBLE, 1998) in registration on DSA images of a neurovascular phantom. In this case, the intensity-based registration algorithm was found to give more accurate registrations, 1.2mm error, against 2.3mm error of feature-based section.

3 RELATED WORK IN MEDICAL IMAGE CORRELATION AND REGISTRATION

The work developed in this thesis uses similarity metrics to match an input image with volumetric data. We are interested in match multidimensional medical images. General purpose methods to match and assess similarity between images is widely explored. The survey conduced by ZITOVá; FLUSSER (2003) presents and discusses several works that match many distinct images sources with a variety of methods. Limiting the research field, a comprehensive analysis of medical image match and registration were done in the surveys elaborated by MAINTZ; VIERGEVER (1998), KHALIFA et al. (2011) and MANI; RIVAZHAGAN (2013).

Although there are several classifications for works in the area, we have divided the related works into three groups: pre- and intra-operative correlation, intrinsic 2D/3D correlation and feature matching. The methods are detailed in the following sections. The works were arranged in those groups according with their main contribution. Some of them could be classified in more than one group.

3.1 Extrinsic Pre- and intra-operative correlation

One branch of related works include those that focus on 2D-3D image correlation of pre- and intra-operative procedures. The most common, but not limited to, correlation is performed between US-CT and MRI data. WEIN et al. (2008) developed a technique to automatically register US images with CT. The US transducer position is tracked with magnetic sensors and a 3D sweep acquisition is performed. To accomplish a successful match against the modalities, they simulate medical ultrasound effects from CT data and elaborate a robust similarity function. The registration was tested in 25 patients with indeterminate lesions in liver and kidney with average registration error of 8.1mm in relation to the ground truth, which is high for precise interventional navigation. And, despite the user dependent US acquisition, they only present results matching the informations in axial image plane.

In the same direction, KING et al. (2010) explored the registration of segmented preoperative MRI images to intraoperative 3D US data. The technique incorporates knowledge of the physics of US acquisition into the registration process. It increases the similarity measure when the alignment is correct. They also track the US transducer position to match the modalities coordinate system. Beyond the rigid registration, the authors developed a registration correction for respiratory motion. The technique is validated with six subjects and two patients and they achieve accuracy of 2.1 to 4.4mm.

The objective of the work developed by LINDSETH et al. (2003), is to integrate

intraoperative 3D ultrasound and preoperative MR images in one 3D scene. The setup focus on visualizing brain shift during the ultrasound-based neuronavigation. Commercial optical positioning tracker captures the US freehand movement. The integration of pre- and intraoperative data into a common coordinate system is achieved using skin-fiducials. Furthermore, they developed a multimodal volume visualizer that supports both orthogonal and oblique slicing relative to the volume axis where the 2D US images are displayed along with 3D MR volumetric rendering.

NIKAS et al. (2003) co-register intraoperative ultrasonography with MRI using an optical 3D tracker. The tracker records the patient's head position and the US transducer position to define the rigid transformations between them. In the same direction, Crocetti et al. (CROCETTI et al., 2008) use a commercial real-time CT-US fusion system in targeting radio-frequency (RF) ablation of a liver target. They register the real organ with the CT images using fiducial markers to calibrate the coordinate systems. The position of the US transducer is electromagnetic tracked. A monitor shows the intervention with the biopsy needle in the US images and the CT data in real-time in a side-by-side visualization.

The works presented in this section have solid contributions to improve the medical analysis and treatment. However, all techniques need means to track the environment, the patient and the imaging apparatus to transform the distinct acquisition modalities into the same coordinate system. Moreover, some of them do not benefit from the free acquisition capabilities of the US modality, fusing the data only in the axial anatomical plane. On of them does not fuse the modalities into a single visualization. The long term goal of our project is to use only pixel/voxel intensity information to correlate and fuse user dependent intra-operative acquisition data with pre-operative CT and MRI dataset eliminating the need of external tracker.

3.2 Intrinsic 3D/2D registration methods

Intrinsic image matching depends on the relationship between images intensities. It is based solely on the pixel/voxel intensities or with the help of landmarks placed in the imaged scene. This work assumes that the same anatomical structures have similar intensity values. Thus, the sum of squared (SSD), absolute differences (SAD) or peak signal-noise ratio (PSNR) can be used as matching criteria. The choice between them depends on the assumption regarding the noise that corrupts the image intensities and the image dynamic range. In the case that a linear relation is assumed between the signal intensities, the optimal criterion should be the cross-correlation and correlation coefficient (AVANTS et al., 2008), (KIM; FESSLER, 2004).

The use of methods previous created to assess image quality was introduced by XIAO; ZHENG (2011). They developed a novel image fusion scheme by integrating the structural information of the image using the SSIM technique and directive contrast information for line, edges and contours extraction. The contrast information is defined as the ratio of the local maxima of the detail components in a Gaussian window and the mean of the corresponding approximate components. They perform a qualitative and quantitative experiment comparing their technique against 5 other image fusion approaches by fusing an MRI slice with a CT slice. Their technique exacted the largest similarity features, this information is claimed to be a good quantitative metric.

The use of mutual information as matching criterion to the problem of medical image registration was first proposed by MAES et al. (1997). The method presented applies MI to measure the statistical dependence between the image intensities, which is assumed to be

maximal if the images are geometrically aligned. In maximization of MI, no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion was validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution.

PENNEY et al. (1998) present a extensive comparison of six intensity similarity techniques to register a CT scan into fluoroscopy images. Fiducial markers were used for ground truth. The authors report that pattern intensity and gradient difference measures were able to register accurately and robustly even when soft tissue structures and interventional instruments were present. Mutual information (MI) was the worst technique with their setup. However, WEST et al. (1997) achieve a great success in the registration of 3-D to 3-D images (single or multi modality) using MI.

Based on the analysis in PENNEY et al. (1998), PENNEY et al. (2001) developed an automated intensity-based algorithm for updating the 3D position of an interventional instrument using a single plane angiogram registered to a 3D volume. They align preoperative CT and intraoperative fluoroscopy using two similarity measures: pattern intensity and gradient difference. Comparative experiments obtained surface-target error of the order of 1-2mm. This work was expanded to register 3D cerebral magnetic resonance with 2D X-ray angiograms (HIPWELL et al., 2003) and matching 3D X-ray digital subtraction angiography images (BYRNE et al., 2004).

ROHLFING; MAURER JR. (2002) developed a image similarity method based on probabilities rather than intensities or geometric features for registration of 3D MR images and X-ray projection images. The method successfully identifies the correct position of an MR image with respect to a set of orthogonal digitally reconstructed radiographs computed from a co-registered CT image.

SCHARFE; PIELOT; SCHREIBER (2010) addressed functional and structural data registration. They register 2D PET scans with MRI. They propose an implementation of the normalized mutual information (NMI) similarity criteria on a cell broadband engine (CBE), which allows multi-core processing in a PS3 game console. However, the images are captured only in the axial plane, reducing the problem to 2D-2D slice matching. Other assumptions are made in relation to the anatomy as the brain shape is constrained by the skull.

3.3 Chapter summary

To the maximum of our knowledge, based on the research done, we could not find works that address the similarity between images not aligned to the main axis planes using pixel/voxel information. The research was done in many fields that deal with volumetric data. We believe that our work has great potential contributions both in general purpose areas that generate analyses of volumetric information and in the medical field.

4 TWO-STEP ALGORITHM FOR MULTIMODAL MEDICAL IMAGES MATCHING

In this chapter, we explored the problem of similarity between images using the perspective of image quality assessment (IQA) (WANG; BOVIK, 2006). Regardless that techniques in this field have focused on evaluating the loss of quality in images and videos after its compression and transmission, the general goal of image quality assessment methods is to be flexible enough to be used in a variety of different applications. With this in mind, we explored IQA techniques to assess similarity between medical images. Some of these techniques are based on simple and easy to compute mathematical approaches, as the mean squared error an its variation peak signal-to-noise ratio. Other more complex methods use common features or make assumptions about the human vision system, i.e. structural similarity algorithms.

The main contribution of this chapter is an image similarity analysis system to match same or different protocols of two and three-dimensional images from within-subject data. We also present a set of experiments to assess the precision an efficiency of the method to produce correct matches between real-time and previously stored data. In Section 4.1 we describe how we analyzed and combined existing techniques, leading to the development of a two-step algorithm that settles the compromise between processing speed and precision of two known approaches. Another contribution is a set of experiments we performed to evaluate the quality and applicability of our methods (Section 4.2).

The discussion about the results, limitations, insights and future improvements in our approach are raised in Section 4.3.

4.1 Two-step Algorithm Development

This section exposes the medical data characteristics and the overall design of our medical image similarity strategy based on a two-step algorithm using PSNR and SSIM.

4.1.1 Medical data characteristics

The data we are dealing with are axial slice images from the thorax to the pelvis of a male human body acquired with a computed tomography (CT) system. In total, four datasets from the same individual were converted and prepared for use in the testing phase. Our datasets were acquired from the same individual in one single exam session but using four different protocols. The difference that distinguishes the datasets is the perfusion phases of a radio-contrast agent injected in the patient. The phases are shown in Fig. 4.1 and are described below:

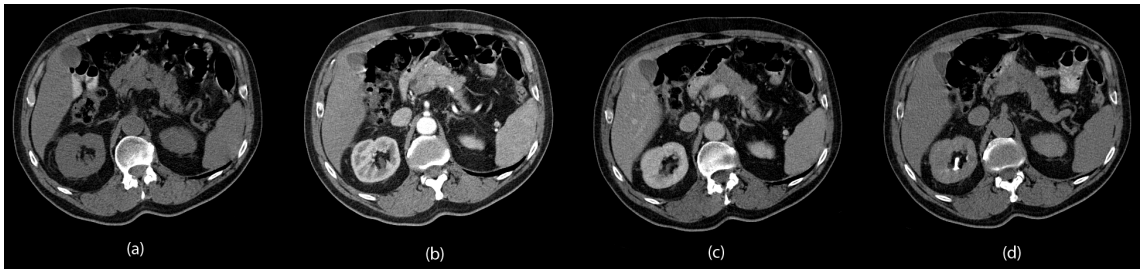
- No contrast phase: before contrast injection;

- Arterial phase: performed approximately 30 seconds after the contrast injection;
- Venous phase: performed 70-90 seconds post contrast;
- Late phase: performed 5-10 minutes post contrast.

CT image slices are stored in datasets, in DICOM format, a standard for handling, storing, printing, and transmitting information in medical imaging. The DICOM format has lots of metadata, from individuals information to the type of imaging device used. We discarded them, storing only the raw images.

The resulting images have 8- or 16- bit stored in one color channel. The pixel data represents luminance values. The datasets are composed of 140 images (slices) with 512x512 pixels resolution. The datasets were also scaled down by factors of two and four. This gives us three resolutions to be experimentally evaluated: 512x512 (native), 256x256 and 128x128. The inter slice distance is very small ($2.5mm$). Thus, in Fig. 4.4 we show six neighbor slices from the dataset with arterial contrast phase. As observed, the images change slightly from one to another.

Figure 4.1 - Image samples from four datasets with distinct acqusition protocols. (a) without contrast, (b) arterial phase contrast, (c) venous phase contrast, (d) late phase contrast.



Source: Compiled by author.

4.1.2 Our two-step image matching algorithm

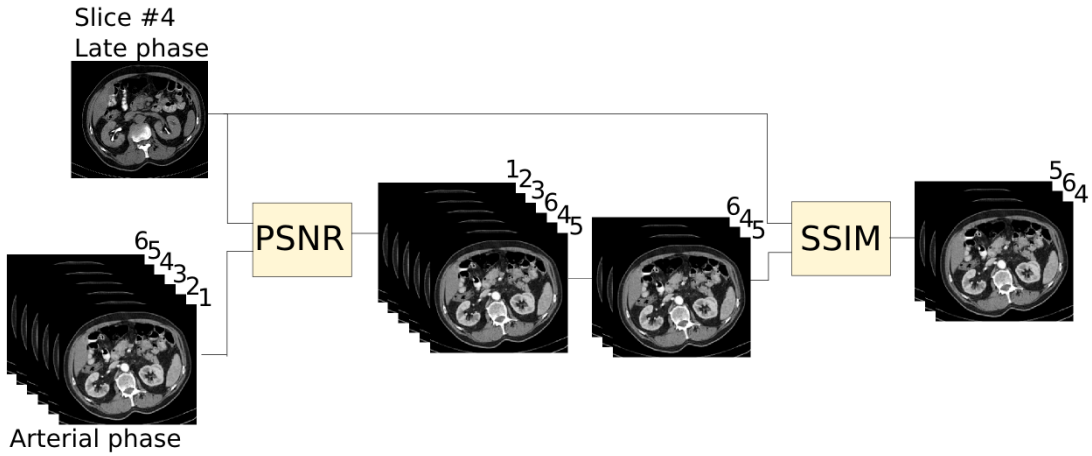
With the above image characteristics, one possible experimental question is: given a slice in the late phase (Fig. 4.1d), which slice in the arterial phase (Fig. 4.1b) is the most similar to the given late slice? This question can be answered by a human specialist that observes the reference image and compares it with all images in the dataset. This analysis is rather effective, but its efficiency worsens as the dataset grows with more images, the process is tedious and boring, and the quality of the results depends on the user experience and attention. Computer methods exist that can solve this problem with greater efficiency and reliability, as described in the related works, Chapter 3. The inputs for any of these methods are: one reference image and one complete dataset (stack of images). The goal is to find a slice in the complete dataset that is the most similar to the reference input slice. This must be done without requiring any additional information. The reference image may be in different radiocontrast phase than the stack of images.

In a quest for a computer automated solution for this problem, we first implemented the PSNR and SSIM algorithms (Subsections 2.4.1.2 and 2.4.2) in C++. Preliminary tests reveled that PSNR is fast but is imprecise in some cases. On the other hand, SSIM is slower but obtains more accurate results in average. We then developed an algorithm that

settles the compromise between processing speed and accuracy of the two approaches. Comparative experiments and results are discussed later in Section 4.2.

Our algorithm has two steps. In a first step, PSNR calculation is performed to quickly create a list of slices that best match with the reference image. In the second step, SSIM is computed only for the top ranked slices that resulted from PSNR and reorders them in the ranking according to their best matches. The algorithm pipeline is shown in Figure 4.2

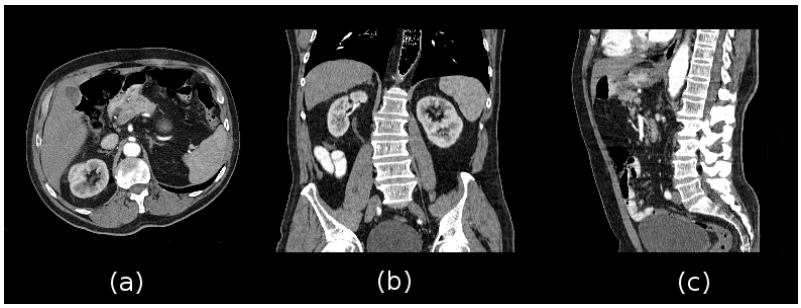
Figure 4.2 - The 2-step algorithm pipeline. The input data are a 2D image (in this case, the fourth slice of a CT in the axial view in late radio-contrast phase) and a 3D dataset (CT arterial radio-contrast phase). First, the 2D slice is compared with all dataset images using PSNR. The 3D dataset is ranked from the most to the less similar images. Only the n top ranked slices are then compared with the 2D image using the SSIM. After that, new similarity values are assigned to the top ranked slices and they are reordered again in decreasing order.



Source: Compiled by author.

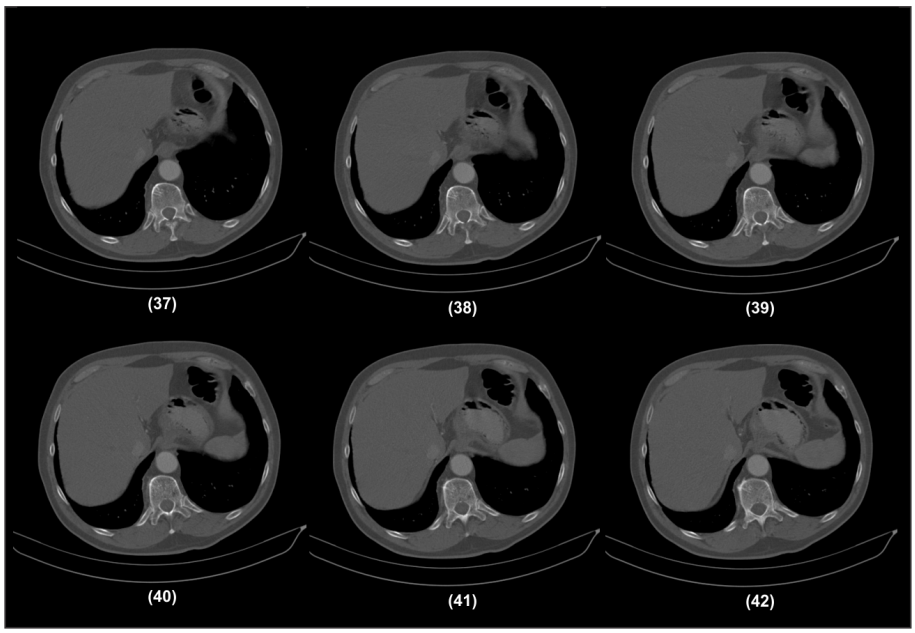
The computer vision and image processing library OpenCV gave the necessary support with the structures that facilitate the data storage and manipulation. It also provides easy to use methods to build the algorithms both in CPU and GPU. The system is capable to load CT and MRI datasets and display the images in the three main viewing axes: axial, coronal and sagittal (Fig. 4.3).

Figure 4.3 - Three main anatomical views. (a) Axial, (b) Coronal, (c) Sagittal.



Source: Compiled by author.

Figure 4.4 - Dataset slices sequence with arterial contrast phase. The images were taken from the middle of the dataset. The numbers represents the slice number inside the dataset. The images has slightly changes from one to another.



Source: Compiled by author.

4.2 Image Matching Evaluation

We set up a test protocol to assess and compare the performance and effectiveness between our 2-step approach, PSNR and SSIM algorithms to predict similarity. In this section we detail the metrics used. Our concern is that the algorithms run faster and deliver the best possible results in the detection of similarity between images. Thus, we evaluate both their precision and accuracy. Accuracy stands for nearness of the results average to the ideal value, while precision is how close the results are to each other. In this thesis, accuracy and precision can be referred as quality as well. Since all algorithms are accurate enough (the results are always scattered around the optimum), we are more concerned with measuring and evaluating the precision achieved by our technique compared to others.

The relevant data are logged in runtime and saved automatically for analysis. The machine used during the tests is a Core i7-2600 running at 3.4GHz and a GeForce 560Ti with 384 CUDA cores running at 880MHz.

4.2.1 Experimental protocol

The long term goal of this project is to find a section plane in a 3D dataset that corresponds to reference planar image obtained in real time from another modality. However, to push the limits of the system and to assess its performance in this situation, we decided, for testing purposes, to replace the reference planar image with a reference dataset. In this case, the comparison is performed between all the slices of one dataset with the other one.

All datasets were previously visually aligned. This was necessary to establish a pattern that we can trust as a correct matching score. Although both datasets were obtained from the same subject and in a single test session, the stacks of images were slightly shifted, then requiring this preliminary alignment. The difference in DICOM was small (about 3-5 slices), but this offset could affect the confidence in the results. After correction, we can

rely that, for example, the correct match for slice 42 in the first dataset is the slice 42 in the second.

At every dataset comparison, we store the most similar slice (highest score) and its n neighbors sorted in decreasing order (most similar to less similar). This information is useful when the algorithm outputs a different slice as the most similar. It allows to calculate the distance (delta) to the exact matching. Moreover, it lets us know if the correct match neighbors are within the rank. This information may also be an accuracy indicative. The number of ranking positions, n, directly affects the second step of the 2-STEP algorithm. It means that more comparisons need to be made with the slower SSIM algorithm.

4.2.2 Test variables

The independet variables used in our test are:

- Dataset resolutions: 512x512x140, 256x256x140, 128x128x140;
- Orthogonal orientation planes: Axial (XY), Coronal (XZ) and Sagittal (YZ);
- Algorithms: 2-STEP, PSNR, SSIM;
- Datasets radiocontrast phases: arterial, venous and late;

The dependent variables colletected are:

- Runtime: in seconds;
- Correct match distance: in slices;
- Ranking: image slice precision in the maximum distance of ten slices;

4.2.3 Precision evaluation

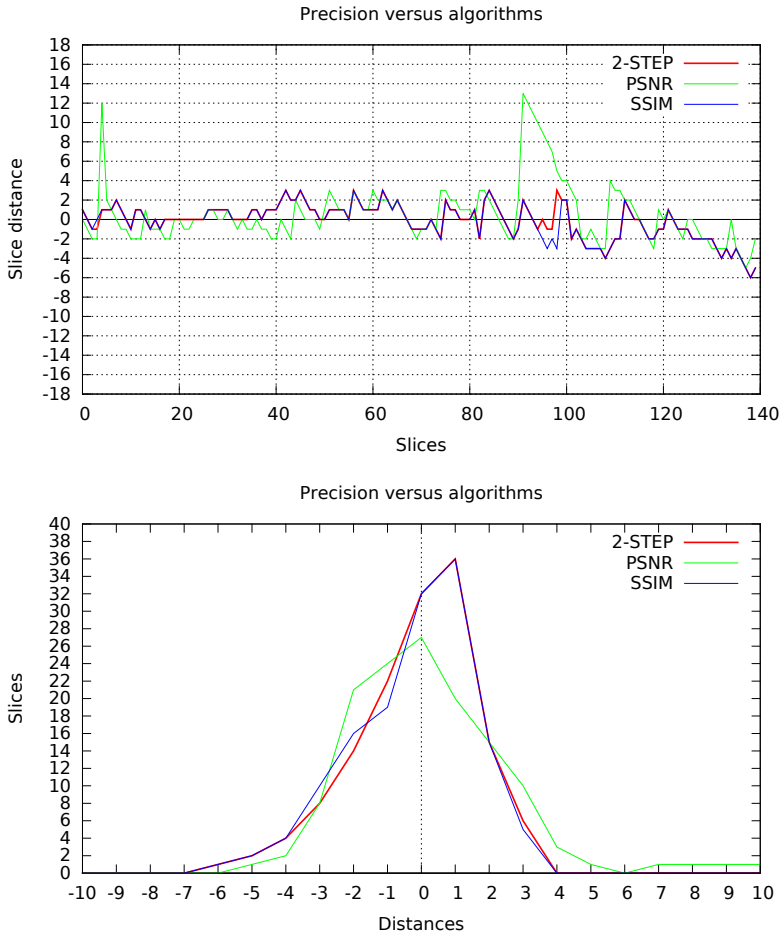
The hypothesis for precision evaluation is that PSNR will be less precise against the other two algorithms, but faster. If the results of PSNR falls into the ten most similar slices, the 2-STEP can achieve precision very similar to SSIM with lower runtime. If not, the precision of 2-STEP is compromised.

To evaluate the hypothesis, in this first test, we arbitrated the comparison of the venous phase dataset against the late phase. Both with a resolution set to 256x256x140 in the axial orientation plane. The graph in 4.5(a) shows the results. As we can observe, the PSNR has some instabilities to predict the correct match (the red spikes). The 2-STEP is more stable, as all correct matches are in PSNR rank results. This way, the second step of the algorithm solves the instability. The SSIM results were very similar to the 2-STEP, as expected.

Another view of the results is shown in graph 4.5(b). This graph highlights how many slices are closest to the correct match. The zero in the center of the x axis is the best result. -1 (and subsequently) means that the slice was matched with the previous slice in the other dataset. The opposite occur with +1 and subsequent distances. The red spikes seen in the previous graph were cut out because they are out of the range (-10 to 10).

The graph in 4.6(a), shows the instability problem hypothesized. The PSNR has a high instability in the beginning of the dataset comparison (slices 1-30). In this case, the set of best ranked slices does not contain the correct match. Consequently, the second step of the 2-STEP algorithm cannot predict the best match successfully, even though the results are slightly better in the instability zone. This comparison was made between venous and late phases datasets at 256x256x140 resolution.

Figure 4.5 - Precision evaluation. Comparison of venous and late datasets with 256x256x140 resolution in the axial orientation plane. The upper chart shows the precision to predict the best match with 2-STEP and PSNR algorithms (SSIM hidden as it gives the same results as 2-STEP). The green peaks show some PSNR instabilities. The lower chart, shows the closest slices to the correct match, the zero in x axis.

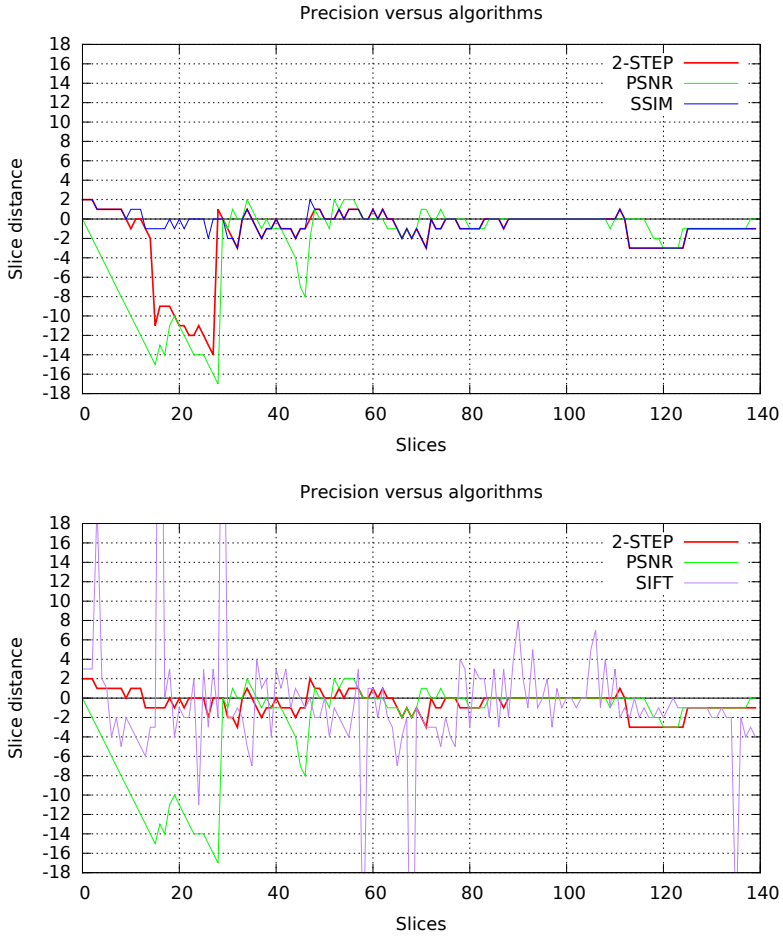


Source: Compiled by author.

The solution for the instability problem was to increase the number of entries in the slice ranking to increase the chances for the correct match to be inside (graph 4.6(b)). It increased the 2-STEP computation time as it has to perform more comparisons. The ranking was increased to 20 entries instead of the default 10, the runtime increased from 6.9 to 12.89 seconds (87%). Another hypothesis raised from the literature review is that SIFT would perform slower and less accurately than our solution. The results with the SIFT algorithm are also displayed in the graph 4.6(b). Notice that the algorithm is very unstable, it outputs a number of sharp spikes that diverge from the optimal result.

The total runtime for this test is shown in the chart of Fig. 4.7. PSNR is significantly fast against 2-STEP and SSIM, but lacks on robustness as discussed above. The interesting conclusion here is that 2-STEP performs much faster than SSIM with the same level of precision. SIFT (ommited in the graph) was the slowest, concluding the calculations on CPU in 264.4 seconds.

Figure 4.6 - Instability correction of the 2-STEP algorithm by widening the ranking selection. In the left graph (a), the PSNR instability affects the 2-STEP algorithm due the fact that the optimal slices are not in the first step ranking. After the increase to 20 ranking entries the instability is solved (b). SSIM results are not presented here as it has the exact same plot as 2-STEP. SIFT presents even worse spikes (peaks ommitted in the graph as they are higher than 20 slices away).



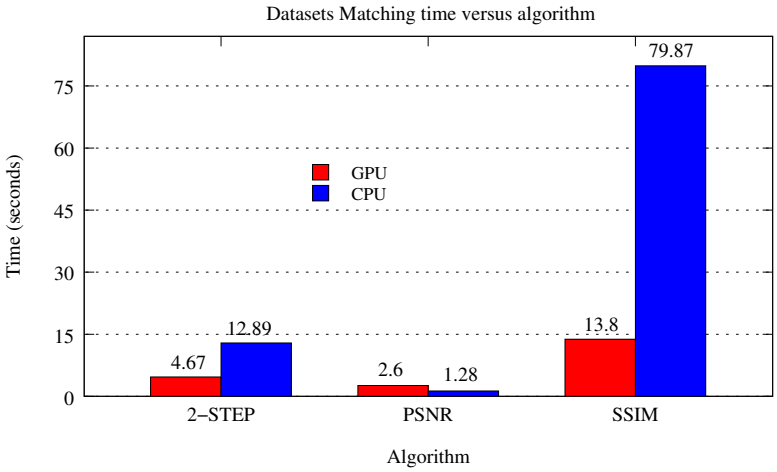
Source: Compiled by author.

4.2.4 Dataset resolutions evaluation

Another hypothesis is that there will be little difference in the precision results with lower datasets resolution. On the other hand, the runtime will drop quadratically as the resolutions decrease.

This test compares the runtime and similarity precision obtained with the 2-STEP algorithm for of all datasets combinations for each resolution. In this test the orientation plane was fixed in axial view. The ranking set used for the 2-STEP is the larger one (20 comparisons in the second step). The results are exposed in the graph of Fig. 4.8 and the runtime comparison in Fig. 4.9. Precision and time results displayed correspond to the results from all dataset combinations (3 combinations of 2-by-2). Notice that, while the runtime greatly drops, the loss of precision is small when lowering the resolution, confirming the hypothesis. The resolution 256x256 best balances perforance and precision. In very low resolution, the GPU implementation has no advantages against the one in CPU. Only with higher resolutions the parallelization makes difference.

Figure 4.7 - The algorithms runtime for the tests 4.6. The 2-STEP algorithm was evaluated with 20 slices ranking. In the GPU bar, both steps were executed in the GPU processor. PSNR does not take advantages of the GPU parallelization.



Source: Compiled by author.

4.2.5 Evaluation of orientation planes

We hypothesize that the 2-STEP algorithm will perform with similar precision in the three orthogonal cutting planes evaluated. The independent variables are: sum of the three combinations of dataset comparison, with 256x256x140 resolution running the 2-STEP algorithm. Results are presented In the graph of Fig. 4.10.

Unlike our hypothesis, the datasets in axial orientation presented a lower precision rate, represented by the red line in the chart. One cause for this behavior is that in this orientation plan, most of the times, large areas of the image contain no information (black parts seen in Figs. 4.3(a) and 4.4). Thus, with less information, the algorithm can not make the best analysis, leading to results comparatively worse than the other axes. This difference in information population can be seen in Fig. 4.3.

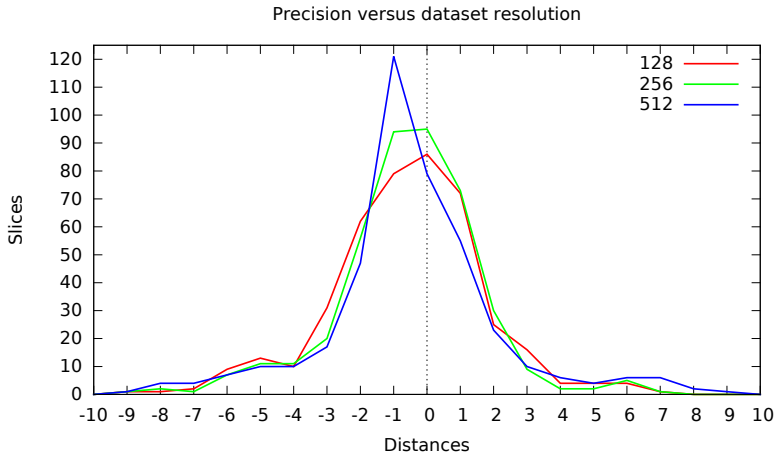
4.3 Chapter summary

The two-step technique developed and evaluated integrates two similarity methods into a technique that match within-modalities with distinct acquisition protocols. An experimental evaluation demonstrated that it is faster than SSIM and as much more accurate and precise than PSNR.

Our experiments compare the two-step algorithm with other representative algorithms found in the literature. The fastest algorithm tested, PSNR, generates unstable results, with some errors peaks. The SSIM is the opposite. It produces more precise matching, but is more computationally demanding. Higher processing times make it unaffordable for a robust solution in our application area. Our solution is an algorithm that solves the problem in two steps. It first ranks the most similar images using PSNR and then applies SSIM only to a set of top ranked images, thus saving computation power.

The tests showed that a vital point to obtain good accuracy with our method is the number of elements in the set of slices stored in the ranking created after calculating the first step. For most datasets tested, ten slices stored are sufficient, but in some situations the number of comparisons is not enough for a perfect match with SSIM in the second step.

Figure 4.8 - Precision achieved in three distinct resolutions. The results of all dataset combinations were merged for each resolution. The orientation plane is in axial view.



Source: Compiled by author.

For these odd cases, results without error peaks were achieved by doubling the ranking storage. This consequently increases the algorithm runtime. In the future a solution with adaptive ranking comparisons could better balance performance and accuracy.

Interestingly, one hypothesis not confirmed is that all orientation planes would have similar precision. The results reveled that in the axial view the algorithm results in a higher number of incorrect matches with the datasets tested. It was not worrying, however, as the results are in the interval. We believe that this difference is caused by the lack of information in black regions of the image that disturbs the similarity check. They are actually large areas that look similar in all images. Further investigation should be carried out to segment and remove these areas or use partial images as input.

One challenging issue was how to define what is acceptable error. Making a visual comparison of the CT slices, it is difficult to notice significant changes between neighboring images. In the evaluation, we have considered as acceptable the results sitting within the distance of 5 slices from the optimal. Remember that our dataset has 2.5mm between slices.

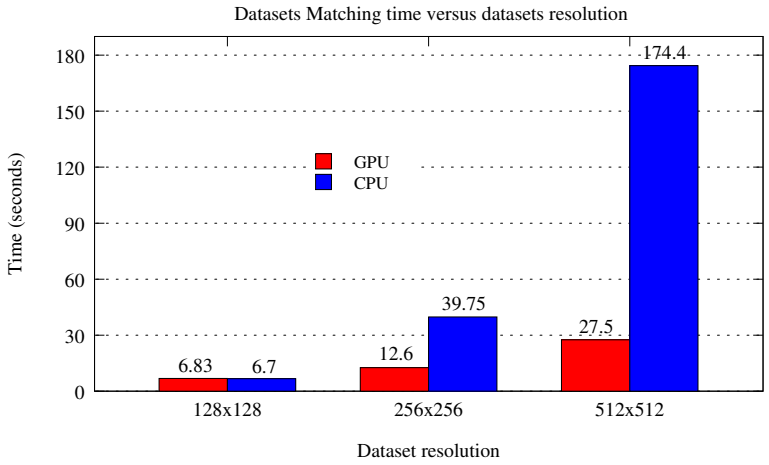
The algorithms presented here are highly parallelizable. In the pursuit for the best possible performance, they have also been implemented to take advantage of graphics cards parallel processing power. As expected, the test results showed significant reduction of the computation time for the proposed technique with the same results quality.

We also tried to increase performance by resampling the images to lower resolutions. Datasets resolutions lower than native only slightly altered the precision of the results and highly increased performance. Among the tested resolutions, the 256x256x140 has shown to be a good compromise between correct matches and execution time.

In the next chapter we extend the search to non-orthogonal axes, allowing slices at arbitrary angle as input.

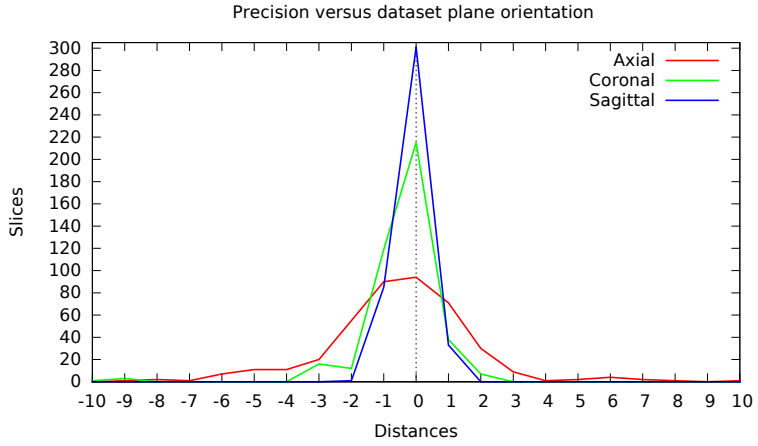
A possible secondary application envisioned for the technique developed is to align different datasets. Before the tests reported, we performed preliminary experiments. With these experiments we noticed that our datasets were not perfectly aligned and required registration. They basically had different starting and ending slice locations for each of the different radiocontrast phases. A difference around 3 to 5 slices. Such difference was detected when our first plots showed a peak about 3 unities away from zero. A visual

Figure 4.9 - The sum of all dataset combinations runtime for each dataset resolution from Fig. 4.8. In CPU the time grows exponentially with the resolution, but in GPU it only doubles.



Source: Compiled by author.

Figure 4.10 - Sum of all dataset combinations in the three orientation planes. The coronal and sagittal planes have stable results, while the axial plane is unstable in some regions.



Source: Compiled by author.

inspection of the slices has proven later that they were actually shifted. The method might then be promising also for fast dataset registration.

5 ARBITRARY IMAGE-PLANE SEARCH AND VISUALIZATION FOR MULTIDIMENSIONAL MEDICAL IMAGE CORRELATION

The work presented in Chapter 4 and most related works in medical images registration (Chapter 3) address similarity between images aligned on one of the three planes: axial, coronal and sagittal. However, in daily practice, the acquisition tend to be outside the main planes. This occurs especially when image capture is user dependent, for example, during ultrasonic imaging.

In this chapter, we propose an approach to correlate 2D and 3D data from ordinary acquisition protocols based solely on the pixel/voxel information. Our method uses local similarity analysis to, given an arbitrary slice image (2D), find the location of this slice within the volume data (3D). The approach aims at establishing a correlation between medical images that can be arbitrarily oriented and that only partially cover a section plane of the volume.

One can imagine a brute force solution that compares the image with all possible total and partial cross sections of the volume using some similarity criteria. The most similar section will correspond to the planar image location within the volume. As the computational cost of such exhaustive search is prohibitive, we propose here a smarter approach that minimizes the virtually infinite number of possible cross section orientations.

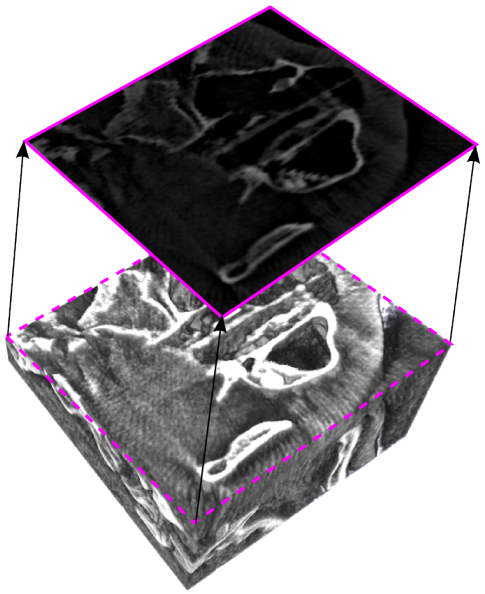
In the following sections, we present all steps and methods of our approach.

5.1 Inputs and Arbitrary Plane Extraction

Two elements compose the input data for our algorithm: a 3D volume dataset; a planar image of any size generated as a subset of the volume aforementioned.

Given an input 3D volume, the planar image can be generated, for testing purposes, as a cross-section of the volume (Fig. 5.1). The volume voxels are first mapped in the interval -1 to 1 in x,y and z axis. We then generate the arbitrary image-plane selecting three non collinear points, p_1, p_2 and p_3. The normal vector $\vec{n}$ is calculated as $\vec{n} = \hat{v}_1(p_2, p_1) \times \hat{v}_2(p_3, p_1)$. To complete the plane equation, the distance from the coordinate system origin, d coefficient is calculated as $d = -(\vec{n} \cdot p_1)$. A 2D nearest-neighbor sampling is then applied to generate the final image. The plane coordinate system has the same spacing as the original coordinate system. In this conditions, the nearest-neighbor sampling guarantees that the sampled image-plane exactly reproduces the original data.

Figure 5.1 - Visual representation of the arbitrary oriented image sampling. The image (solid pink contour) is sampled from the original volume (dashed line) in a plane arbitrarily oriented and positioned using nearest-neighbor.



Source: Compiled by author.

5.2 Image-Volume Search and Similarity Criteria

The algorithm (Fig. 5.2) works locally evaluating similarity between subimages and subvolumes sampled from the original input data (Sec. 5.1). A size s defines the resolution of the subimage ($SI_{s\times s}$) and subvolume ($SV_{s\times s\times s}$).

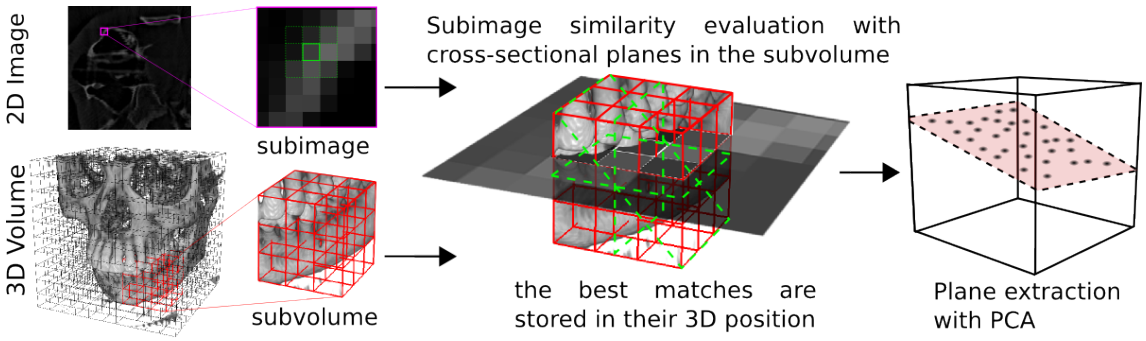
The similarity of each SI is evaluated against nine oriented planes of the SV (three cardinal planes plus six intermediary planes, see Figure 5.3 for a visual representation). The similarity is calculated by averaging the squared intensity of the pixels differences (SSD), explained in Subsection 2.4.1. Each of the nine planes are stored if their computed similarity is below a threshold (>85% of similarity). Finally, only the most similar plane is picked and its voxel position, similarity value and orientation are stored in a best match list L^k for later use (Section 5.3). The pseudo-code of the image-volume search can be seen in Algorithm 1.

Since the evaluation is done within-modality and only in a small area of the original image, the SSD is the method that delivers the lowest execution time with good match precision (see the evaluation made in Section 4.2).

At this point, we have only addressed the similarity comparison problem. The approach described in this section uses local similarity evaluation to reduce the amount of plane comparisons. Figure 5.4 shows the three-dimensional localization of all calculated similarity points. Even though we can clearly see where the plane searched may be located, we need means to explicitly calculate it. Another problem, present in the four images in the first line, is the sparse black voxels. These artifacts may potentially influence on the plane fitting calculations.

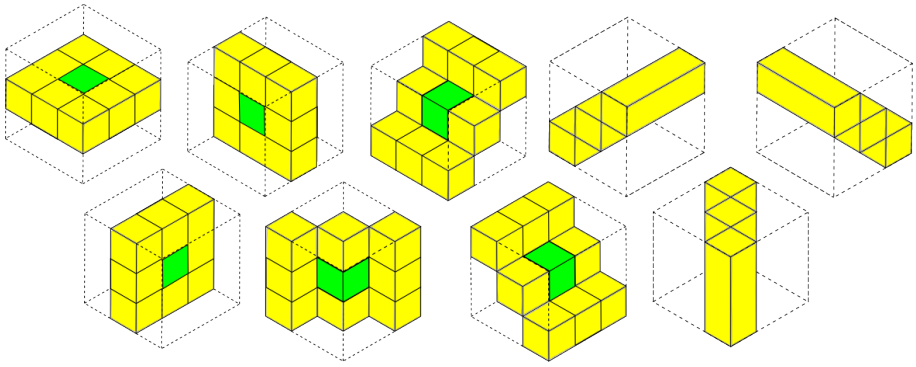
These regions have pixel/voxel intensity near zero, both in the image and volume. The similarity for some volume planes fell inside the threshold interval and they were inserted in the L^k list. To avoid that, two strategies were carried out. We analyze the L^k list to find

Figure 5.2 - For each pixel and voxel a square subimage and subvolume with the neighbouring data is taken. Nine image-planes with different angles are generated from the subvolume and their similarity evaluated against the subimage. For each voxel, the most similar subplane among the 9 orientations with every subimage are stored. After all comparisons, PCA is computed to extract the best fitting plane from the voxels set.



Source: Compiled by author.

Figure 5.3 - The nine image-planes sampled from the subvolume. The subimage similarity is evaluated against all image-planes. Only the most similar is stored. The green cube is the voxel analyzed and the yellow cubes are its neighbors.



Source: Compiled by author.

most recurrent plane orientation excluding all other planes. The Empty-space skipping is the second strategy and it is discussed in Section 5.4.1.

5.3 Plane Fitting

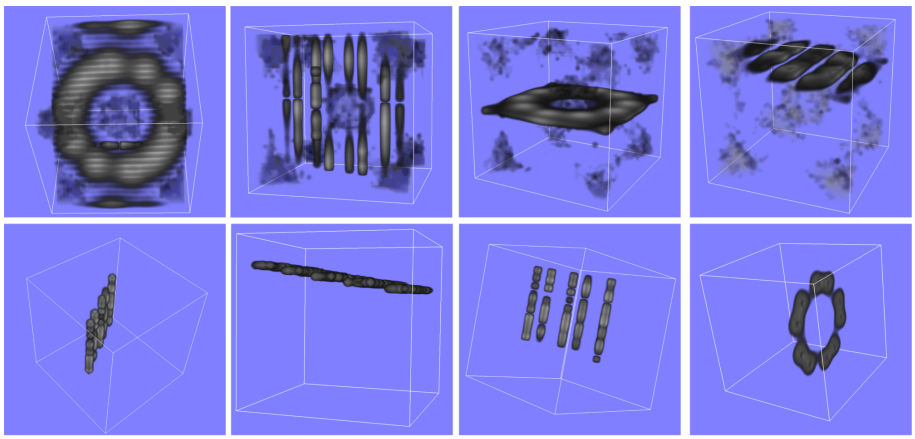
After the similarity calculations we need to extract the best plane that fits all voxels stored in the list L^k. The method is based on the first order 3D plane fitting proposed by BERKMANN; CAELLI (1994). The problem of determining the plane equation of L^k is a least-square plane fitting estimation. The plane is represented as a point x and a normal vector $\vec{n}$. The centroid of L^k is defined as: $\bar{p} = \frac{1}{k} . \sum_{i=1}^{k} p_i$.

The solution for $\vec{n}$ is given by analyzing the principal components of L^k. Eigenvalues and eigenvectors of the covariance matrix C, expressed in eq. A.4, are then calculated.

$$C = \frac{1}{k} \sum_{i=1}^{k} (p_i - \bar{p}).(p_i - \bar{p})^T, C.\vec{v_j} = \lambda_j.\vec{v_j}, j \in \{0, 1, 2\} \tag{5.1}$$

Matrix C is symmetric and positive semi-definite and its eigenvalues (λ_j) are real numbers. The eigenvector $\vec{v_j}$ with the smaller eigenvalue is therefore the approximation

Figure 5.4 - Three-dimensional localization of all calculated similarity points. The top images shows the artifacts present in may add. The were removed in bottom images using an analysis of the most recurrent plane orientation.



Source: Compiled by author.

of the normal $\vec{n} = (n_x, n_y, n_z)$ or $-\vec{n}$. The d component of the plane equation is the dot product of $\vec{n}$ and the $\bar{p}$ of L^k.

5.4 Algorithm Optimization

Two simple optimization strategies were applied in the algorithm to avoid unnecessary calculations: empty-space skipping and image downscaling/downsampling.

5.4.1 Skipping Empty-Space

Empty-space skipping is useful when a dataset contains large portions of completely dark space. The imaging apparatus, such as CT and MR, capture and store all information inside its view area, as shown in Figure 5.5. The data captured are composed by intensity values of the patient's body tissues and the surrounding body areas (it includes the table under the patient present in the bottom of the figure). The surrounding areas represented as very low image intensities are not important for diagnosis. The "black" image portions may also introduce false positive information, confusing the plane fitting calculation. This occurs mainly in our algorithm because we analyze small samples of the original image. In order not to waste time on features that do not contain relevant information, empty-space skipping can be employed to save computation power and avoid the false positive problem. In this way, the subimages and subvolumes with more than 80% of very low intensity values, less than 20 in an 8-bits (0-255) image, are skipped from the similarity calculation process. On the other hand, with less comparisons, datasets with high amount of low intensity areas may have inferior plane fitting accuracy.

A similar idea of the empty-space skipping is used in volumetric rendering techniques to skip dark volume regions, thus, reducing memory access and save computation power (HADWIGER et al., 2006).

We perform the calculations to quantify the amount of data that would be skipped using this strategy. The datasets and their comparisons performed are presented in Table 5.1. The results were calculated with 128x128x128 dataset resolution minus one pixel in each border (to avoid the composition of the subimages with information outside the

Algorithm 1 - Arbitrary plane search pseudo-code.

Require: Img {2D image}
Require: Vol {3D volume}
Require: $size$ {image and volume size in each dimension}

{We exclude one pixel in each image and volume border }
for $pixel = 0$ ... $pixel < size \times size$ **do**
  $Sub\_Img \leftarrow pixel$ and its 8 neighbors

  for $voxel = 0$ ... $voxel < size \times size \times size$ **do**
    $Sub\_Vol \leftarrow voxel$ and its 26 neighbor

    {The nine volume planes Figure 5.3}
    for 0 ... 9 **do**
      take a subimage Sub_VolImg from Sub_Vol with size equal to Sub_Img
      $similarity\_value \leftarrow$ SSD(Sub_Img,Sub_VolImg)

      if $similarity\_value$ < previous Sub_VolImg and $similarity\_value$ < theashold **then**
        $most\_similar\_candidate \leftarrow voxel\_position$ and $similarity\_value$ and $plane\_orientation$ {Temporary structure}
      end if
    end for

    $L^k \leftarrow most\_similar\_candidate$

  end for

end for

return L^k {List with the most similar voxels found.}

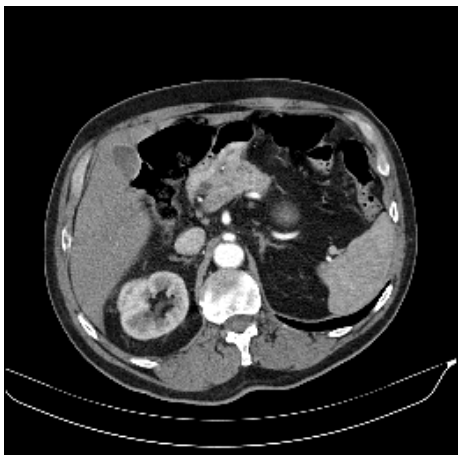
volume) for nine 3x3 subimages orientations. It results in 126x126x126 resolution times 9 subimages, totaling 18,003,384 maximum comparisons.

These sets of data are presented in Figure 5.8, explained in the next sections and they are used in the experimental evaluations. Note that the table shows only the comparison reduction for datasets. The image comparison is suppressed as the amount of black regions of the input image varies with its angle and position. To know exactly the total comparisons decrease, it is necessary to sum both dataset and image reductions.

5.4.2 Image/Volume Downscaling and Downsample

The performance of our technique is bounded by the image and volume resolutions. To overcome this problem, the second strategy adopted is the image/volume downscaling and downsample (ZITOVá; FLUSSER, 2003). Both are basic optimization strategy widely adopted in image processing that reduce the number of image/volume fragments tested. This optimization is handy to our technique since we are not interested in visual results (i.e. display the downscaled data).

Figure 5.5 - A CT slice from a patient's thorax. The surrounding areas (black image portions) are composed of unimportant information. This parts are likely to be skipped. The lines in the inferior represent the CT table where the patient is on top of.



Source: Compiled by author.

Table 5.1 - Number of comparisons performed for each dataset. The resolution was set to 126x126x126 (volume borders excluded) times 9 subimages. The Random dataset performs the comparisons for all voxels. On the other hand, only 12.1% of the voxels are compared in Engine.

Full resolution		
Dataset	# Comp.	%
Random	18,003,384	100%
Bucky	8,015,703	44.5%
Engine	2,179,032	12.1%
Skull	6,092,194	33.8%
Thorax	8,719,107	49.37%

Source: Compiled by author.

In the in image/volume downscale we reduce the data resolution before load it in the computer's memory.On the other hand, in the downsample case, the full resolution data is uploaded on memory. A downsample factor is stipulated and the algorithm performs the calculations only in the sampled pixels/voxels. Since we have the full resolution information, our image-volume search algorithm (Sec. 5.2) still uses neighbor pixel/voxels from the original resolution to build the subimage and the subvolume.

The scaling and sampling factors adopted are 2x and 4x. For example, a 128x128x128 resolution with downscaling of 2x becomes 64x64x64 and with 4x becomes 32x32x32. The Table 5.2 shows the decrease of comparisons when the downsample factor increase. Observe that the empty-space skipping optimization has less contribution with higher downsample factors are applied.

5.5 Visualization

The technique developed works together with systems for volumetric visualization. The objective is to slice the volumetric data in a cross-section where the plane calculated in Section 5.3 passes through.

Table 5.2 - Number of comparisons performed for each 128x128x128 dataset with downsample factors of 2x and 4x. The values of the second column of each downsample factor are: the amount of the original data that are processed with the downsample factor plus empty-space skipping skipping; the values in brackets represents the contribution of the skipping optimization in the downsampled volumes.

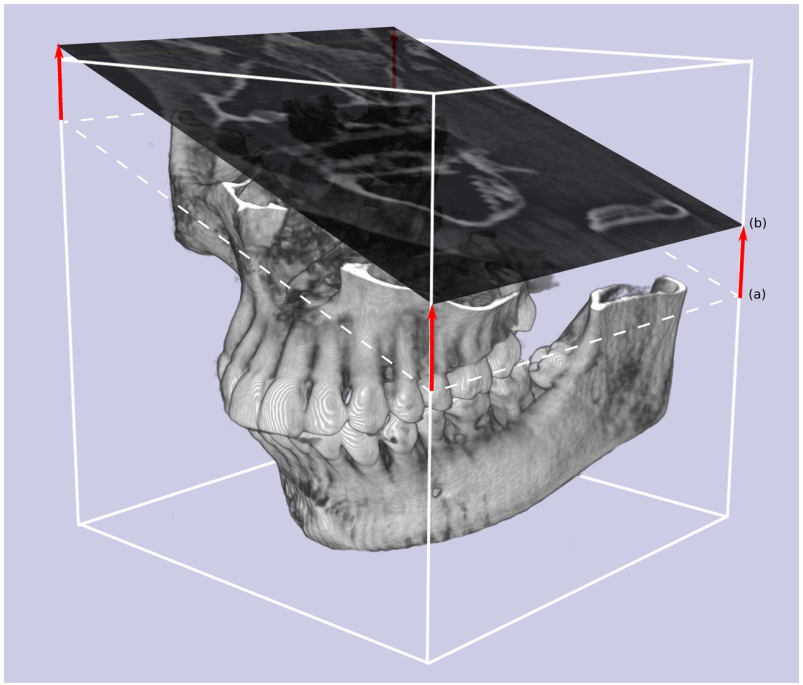
	2x Downsample		4x Downsample	
Dataset	Comp.	%	Comp.	%
Random	9,001,692	50% (0%)	4,500,846	25% (0%)
Bucky	7,753,518	43.1% (13.8%)	4,334,386	24.1% (3.6%)
Engine	7,026,922	39.1% (21.9%)	4,239,345	23.5% (5.8%)
Skull	7,510,705	41.7% (16.5%)	4,304,642	23.9% (4.3%)
Thorax	7,841,485	43.5% (12.8%)	4,344,748	24.1% (3.4%)

Source: Compiled by author.

We modified an implementation of a volumetric render system developed by ENGEL et al. (2004). The system provides real-time rendering achieved by GPU implementation of direct texture rendering (discussed in Section 2.3). Great visual results are obtained by the use of trilinear interpolation of the 3D texture mapping.

We have implemented two main modifications in the volumetric render algorithm.

Figure 5.6 - Skull cross-sectioned based on an input 2D image. Plane position and orientation is calculated and the result visualized in 3D. The actual plane location (a) has been shifted to (b) for illustration purposes.



Source: Compiled by author.

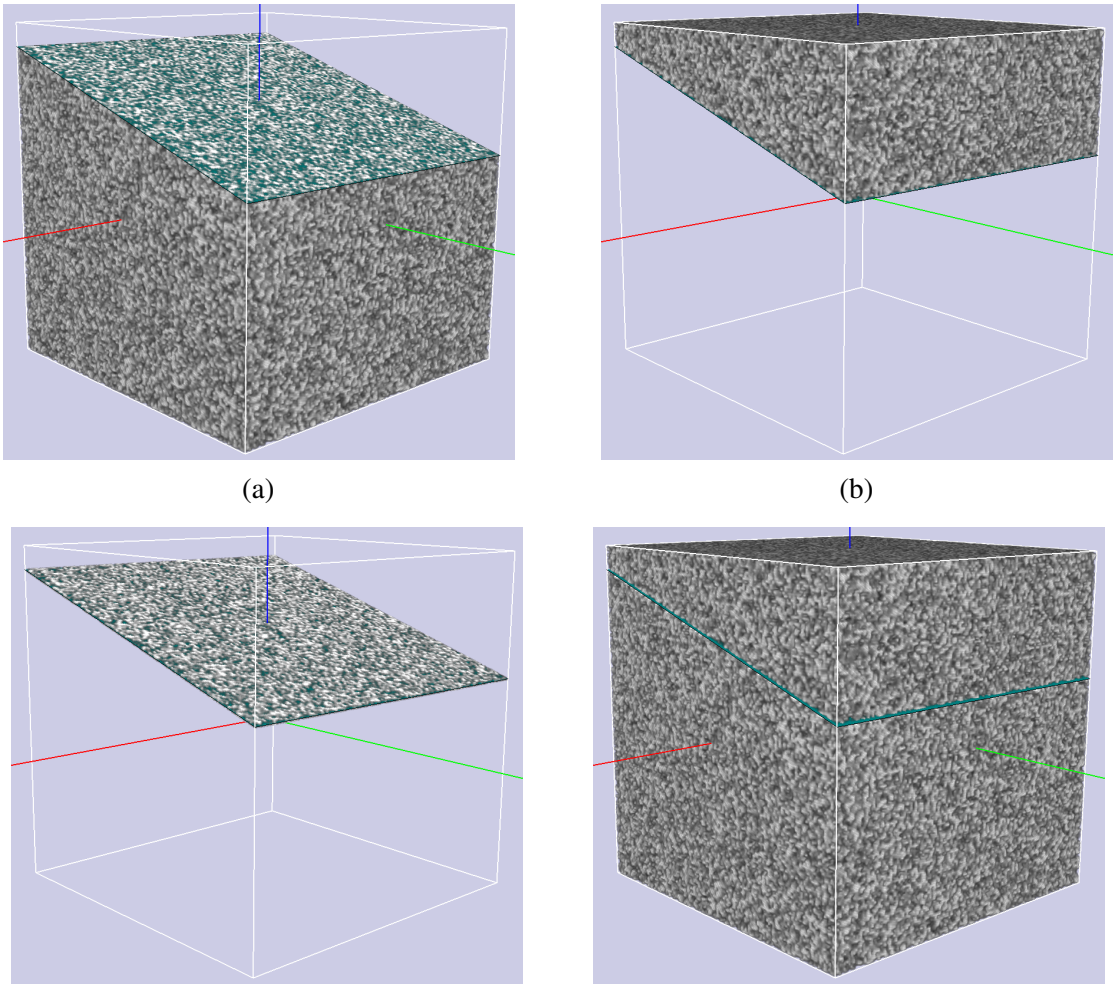
First, the loader function was replaced to deal with the type of data in our datasets. Basically, our 3D data is stored in a RAW file that contains, for each fragment, 16-bits or 8-bits intensity values. We had to parse the RAW information to correctly fit in a 3D OpenGL texture.

Second, we modified the system to display a plane and the volume cut based on the plane information. Figure 5.7 shows the result.

The plane equation is uploaded to the fragment shader. We implemented four possible volume cuts by calculating the dot product of the volume voxels V_k and the plane equation pe. To discard the volume voxels that are above the plane we test if $V_k \cdot pe > 0$ (Fig.5.7a). To discard the voxels below the plane we test if $V_k \cdot pe < 0$ (Fig.5.7b). To show only the voxels that are on the plane we discard the upper and the lower voxels using the dot product changing the zero to > 0.2 and < -0.2 (Fig.5.7c). This makes the slice thicker. The opposed result is shown in Figure 5.7d. The four visualization types can be changed in execution time.

Then, in Figure 5.6, the original input image is displayed at the location found in the volumetric data, fusing the 3D with the 2D image. Transparency and other rendering effects can also be applied to highlight features of interest and apply the best information visualization techniques available. Ideally, a real time implementation of the algorithm would allow to control the input image interactively and display an animated visual feedback.

Figure 5.7 - The four possible volume representations for the same plane.



(a) (b) (c) (d)

Source: Compiled by author.

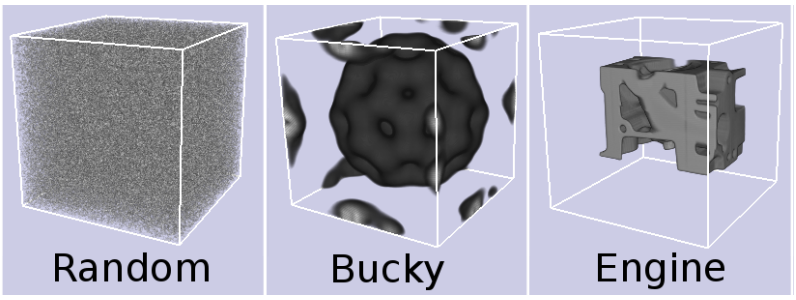
5.6 Evaluation and results with controlled data

We set up a test protocol to assess and compare the performance and effectiveness of our technique among different datasets, resolutions and downsample factors. In this section we detail the metrics used. Our concern is that the algorithm run faster and deliver the best possible results in terms of finding the best fitting plane.

5.6.1 Experimental Setup

The datasets evaluated are: Bucky, Engine and Random. They are general purpose datasets (see Figure 5.8). Random is composed with randomized intensity values and is hardly optimized due the lack of black regions (Tables 5.1 and 5.2). Bucky is a simulation of the electron density of a Buckminster-Fullerene molecule and Engine is a scan of two cylinders of an engine block. Both, Bucky and Engine can be obtained online (FAU-3D, 2014). All datasets are stored in one color channel of 8-bit (0-255) precision. Each voxel represents luminance values. The dataset resolutions and downsample factors used in all tests are listed below:

Figure 5.8 - Volumetric representation of the experimental datasets.



Source: Compiled by author.

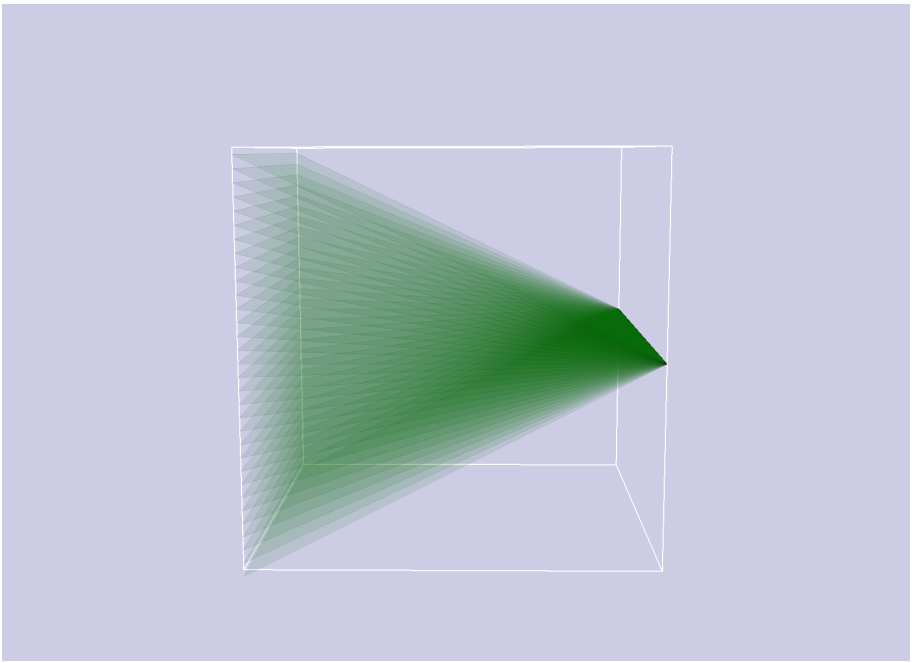
- **Resolutions**: 128x128x128 (original), 64x64x64 and 32x32x32;
- **Downsampling factors**: 1x (original), 2x and 4x for Image XY, Volume XY and Volume Z.

We defined a notation for quick reference: when we say Bucky 128_142, it means that the input image has 128x128 pixels and the volume is downscaled from 128x128x128 to 128x32x64 in runtime. Note that all original datasets are 128x128x128 except Bucky that is 32x32x32 and was upscaled to equalize the comparisons.

The goal of our algorithm is to find the best fitting plane given a 2D slice as input. However, for testing purposes, we created sets of 32 planes and average the results. The set represents a sweep with one fixed edge and angles increasing by 2° successively. The Figure 5.9 shows the planes sweep. We repeat the execution for the hole plane set. The planes are the same for all tests to allow comparison among datasets. Despite that the algorithm was designed to work with partial images, for testing purposes, all tested planes fit inside the volumetric cube and are composed with only one rotation axis .

We store the plane equations of all input 2D images and use them as the ground truth for precision evaluation. The error between the input p and the result q planes is measured with two variables: the mean of the euclidean distance between the correspondent vertex

Figure 5.9 - One example of a plane sweep set.



Source: Compiled by author.

in each plane (Eq. 5.2) and angle between p and q normals (Eq. 5.3). The runtime for each execution is measured in seconds.

$$d(p,q) = \frac{1}{N}\sum_{i=1}^{N}(p_i - q_i) \tag{5.2}$$

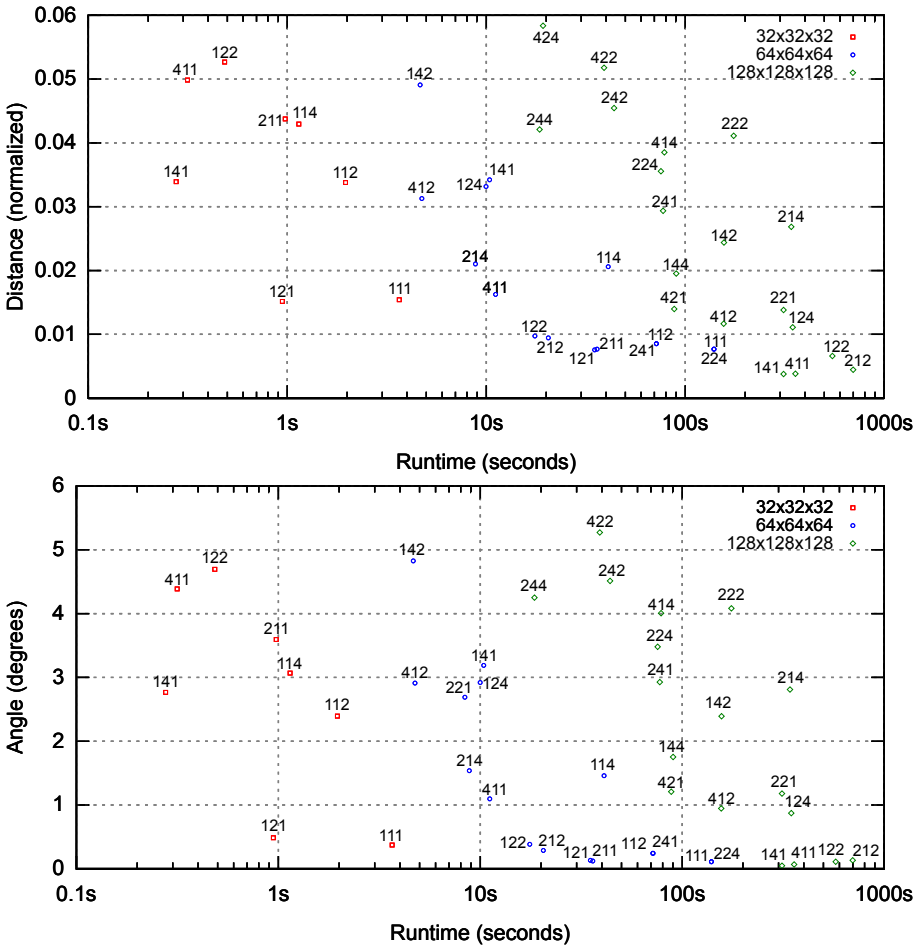
$$cos\alpha = \frac{\hat{n_p} \cdot \hat{n_q}}{|\hat{n_p}|\,|\hat{n_q}|} \tag{5.3}$$

All tests were conducted using an Intel Core i7-2600 running at 3.4GHz with 8GB of RAM. The algorithm was developed in C++, compiled with GCC 4.8.2 in Arch Linux kernel 3.12 box. Eight parallel threads are used with openMP API.

5.6.2 Accuracy-Performance compromise evaluation

Due to the high number of data dimensions it evaluates, our algorithm has a computational complexity of $O(n^7)$. Nevertheless, we hypothesize that it can obtain consistent results working on a very reduced set of samples from the input data. Moreover, we believe that there is an optimal combination of resolution and image scaling factor that is sufficiently accurate and performs in low runtime. As some data are more homogeneous than others, accuracy and runtime can also be dependent on the nature of the dataset. The experiment described below should test some of these hypotheses. To eliminate uncertainties due to heterogeneous datasets, the first test was performed only with the Random dataset. The test aims to find out the best compromise between quality and runtime of the all downscale and downsample combinations. The charts in Figure 5.10 show the plots of all resolution and downsample combinations that fit in the interval. We noticed that distance and angle errors are correlated, allowing us to evaluate only one of them in next experiments and suggesting a possible unified metric to be investigated in future works. The 128 resolution without downscale is very accurate (0.17° error), but performs

Figure 5.10 - Top: Error in terms of distance vs. runtime in log scale. Bottom: Error in terms of angle vs. runtime in log scale.



Source: Compiled by author.

in inviable time (1200s, out of the chart range). Smaller resolutions and coarser samplings, however, still provide similar high precision with times up to 3 degrees of magnitude smaller. As examples, locate points 32_121, 64_122, 128_141 in the chart. For a typical 200x200x200mm dataset, the normalized distance error of 0.49 obtained for 32_121 is equivalent to 3.2mm, and it took less than 1 second to compute.

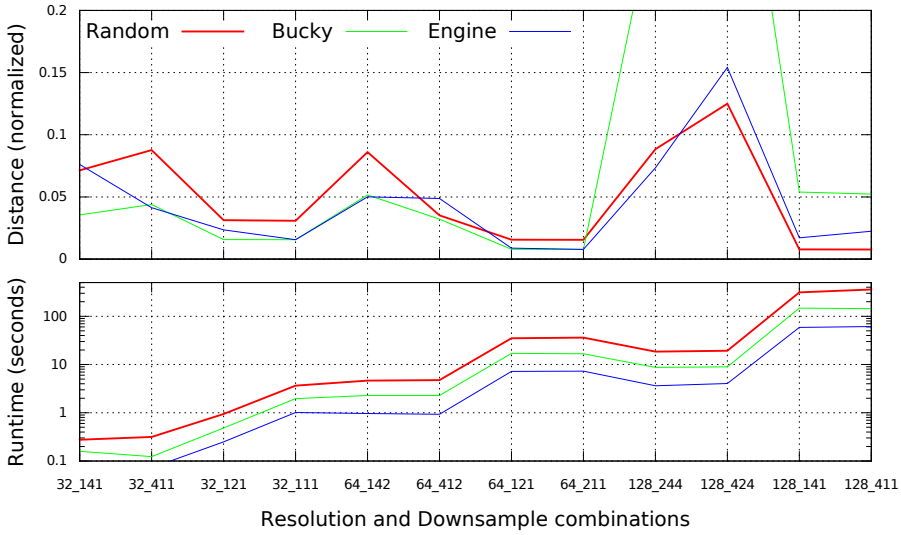
Next, to find out how dataset differences interfere with precision, we selected the two most accurate and the two fastest downscale combinations of each resolution for further analysis. These combinations were evaluated for all datasets and results are shown in Fig. 5.11. The combinations are sorted from the lowest to the highest runtime in each resolution. As expected, Random is the slowest dataset. Plane distances (errors) are consistent between datasets, except in part for Bucky that presented a higher error in 128_424. We believe this is due to the lack of information as it was originally upscaled.

The four setup combinations that equilibrate the performance and quality are 32_121 and 32_111 (prioritizing runtime); and 64_121 and 64_211 (prioritizing quality).

5.6.3 Sweep Planes Accuracy evaluation

In the charts plotted above we used the mean of the sweep planes. To assess instabilities in all angles, here we plotted all planes errors for the four selected resolutions and

Figure 5.11 - Runtime impact of different datasets in 12 selected resolution and sampling combinations. Error and runtime are consistent among datasets, with the Skull being the fastest. Notice that an excellent time-precision compromise is found for combination 32_121.



Source: Compiled by author.

downsample scale combinations in the previous assessment. The charts Figure 5.12 show the results for 32x32x32 resolution and charts Figure 6.2 show the results for 64x64x64 resolution.

As observed, for these optimal cases the results are consistently stable. The maximum distance error of 0.52 (Random dataset) is equivalent to 5.3mm. Instabilities appear only with severe downsamples. Chart Figure 5.14 shows that the planes with higher angles are more prone to errors for severely downsampled data.

5.6.4 Runtime versus optimizations evaluation

In chart Figure 5.15 we show the impact of the optimization approaches adopted. To assess the performance gain with the optimization strategies, we compare the algorithm execution time in different cases. We use only the downscale and downsample selected in Subsection 5.6.2. Random 64_111 is our ground truth as it is never optimized. Engine 64_111 is only optimized with the empty-space skipping (Subsection. 5.4.1), with higher angles the portion of the input image with black pixels increase. The correspondent images extracted with $30°$, $15°$ and $3°$ are in Figure 5.16. Note that the chart's y axis is in logarithm scale.

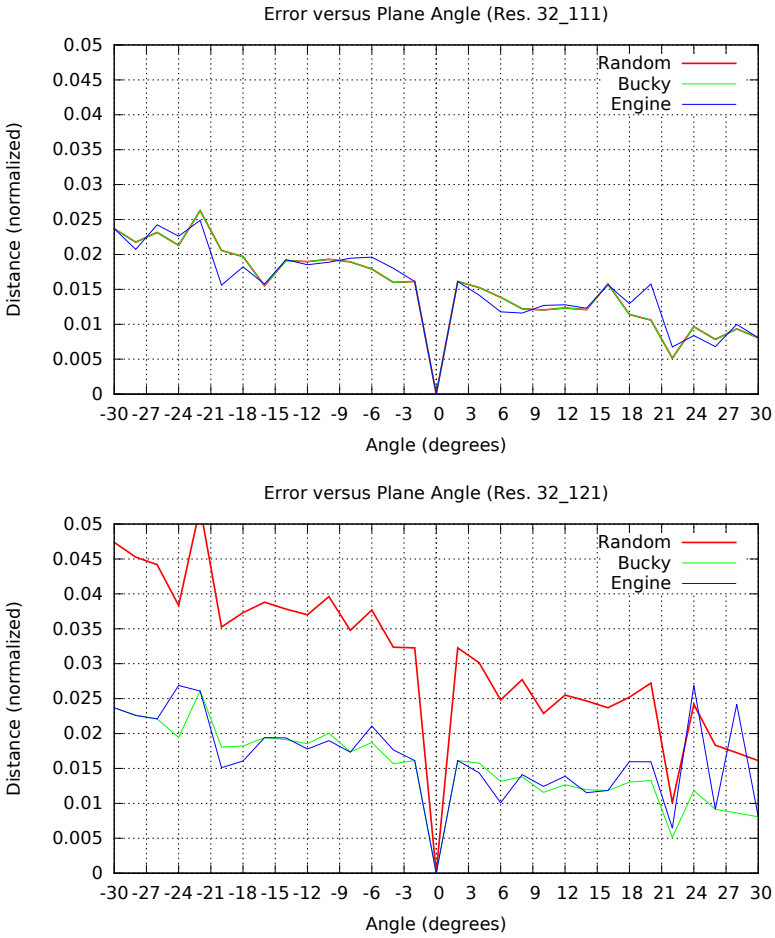
5.7 Chapter Summary

We have proposed a simple algorithm for fast and accurate search of an arbitrary cross-sectional plane image within volumetric datasets. An experimental analysis demonstrated that despite the computational complexity of the algorithm, there are combinations of downsample and downscale data that allows achieving a great compromise between performance and accuracy even when performed with low gradient intensity datasets. We came up with four optimal setups as result of the evaluation performed. They are:

- Prioritizing runtime:
  - 32_111
  - 32_121
- Prioritizing precision:
  - 64_121
  - 64_211

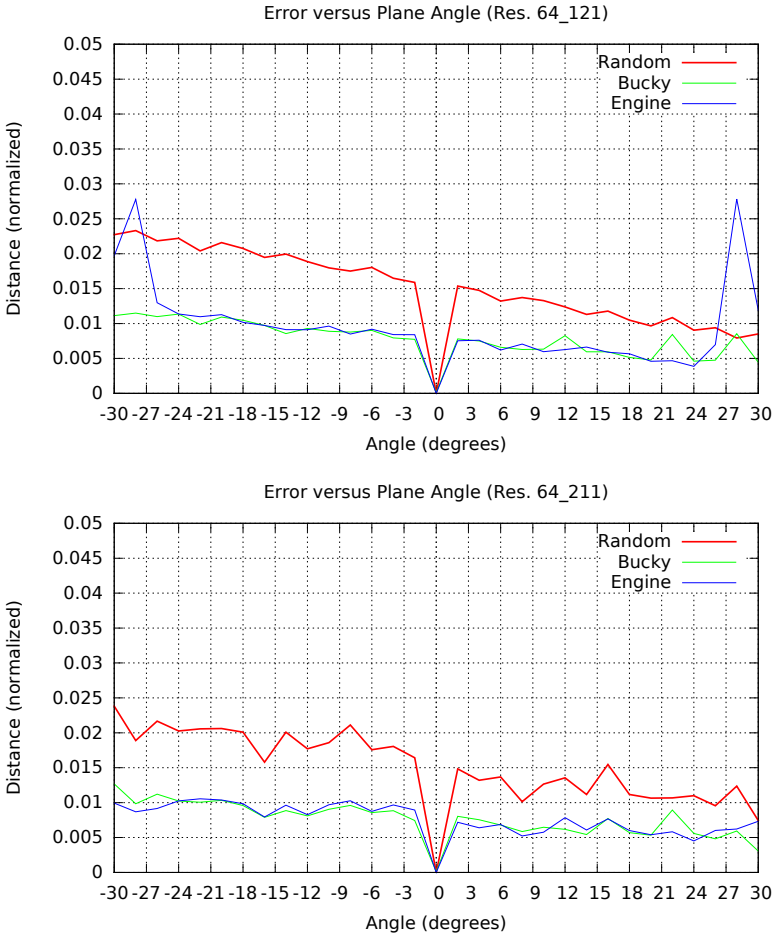
In the next chapter we extend the algorithms evaluation to medical data to assess their ability in patient data correlation and visualization.

Figure 5.12 - Sweep planes precision evaluation of three ordinary datasets with 32_111 and 32_121. (a) Shows the plane fitting precision with downscale resolution of 4x without downsample. (b) Precision with same downscale but now with downsample of 2x in volume slice. Note that Random has the worst precision but it still in the acceptable error.



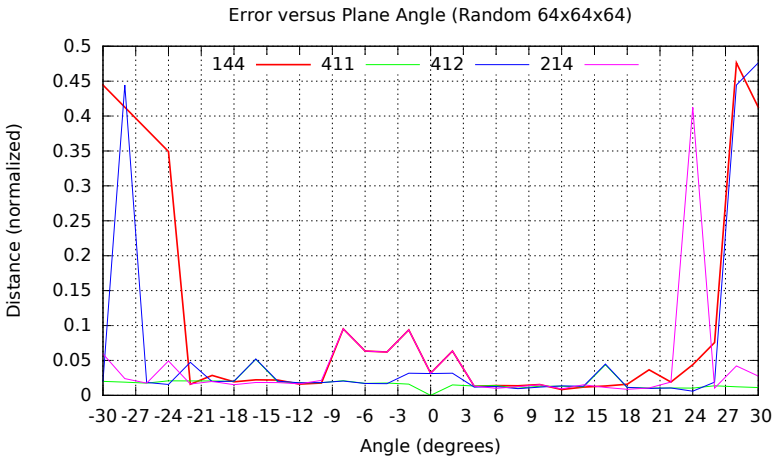
Source: Compiled by author.

Figure 5.13 - Sweep planes precision evaluation of three ordinary datasets with 64_121 and 64_211. (a) Shows the plane fitting precision with image downscale resolution of 2x and downsample of 2x in volume slices. (b) precision with same downscale but now with downsample of 2x in image slice. Random still has the worst precision but is in the acceptable error.



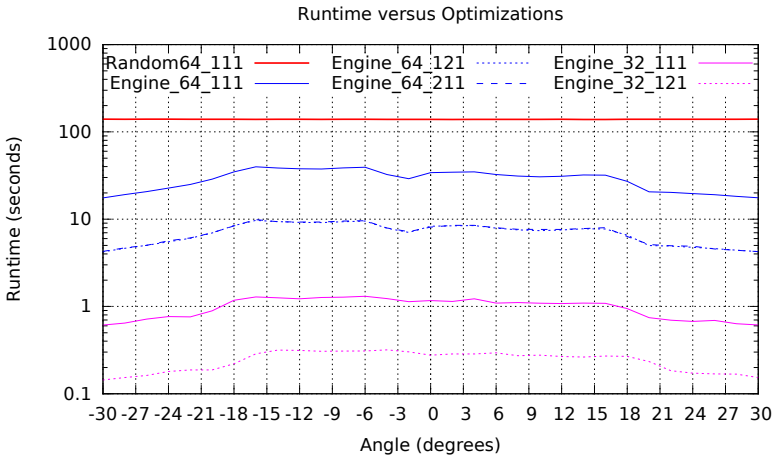
Source: Compiled by author.

Figure 5.14 - Precision versus sweep planes with severe downsampling . Instabilities occurs in higher angles with severe image and volume downsacle and downsample of 2x and 4x.



Source: Compiled by author.

Figure 5.15 - Performance with agorithm optimization.



Source: Compiled by author.

Figure 5.16 - Three image samples taken in different plane angles.

(a) 30° = 14,830 black pixels (b) 15° = 12,389 black pixels (c) 0° = 13,556 black pixels

Source: Compiled by author.

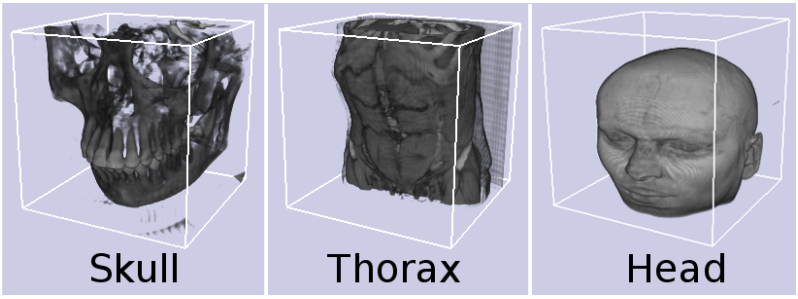
6 APPLICATION OF ARBITRARY PLANE SEARCH AND VISUALIZATION FOR IMAGE CORRELATION WITH MEDICAL DATA

The deep evaluation performed in chapter 5.6, has defined the optimal parameters to achieve the ideal balance between performance and precision. In this chapter, we use the same evaluation protocol to assess medical data.

6.1 Medical Data Characteristics

Three datasets, two from conventional CT (Skull and Thorax) and one from MRI (Head) were evaluated. Their 3D render can be seen in Figure 6.1. Skull covers the mandible and the maxillary and Thorax comprehends the base of the lungs to the kidneys. Skull and Head can be obtained in (FAU-3D, 2014). The Thorax datasets was stored in DICOM format. We exclude all DICOM's meta-data saving only the information regarding to the image. The datasets were originally in different resolutions. Head and Skull were in 256x256x256 with 8-bits and Thorax was in 16-bits precision with 512x512x512 resolution. It was necessary to bring the data to equivalent evaluation conditions of the previous assessment, so all datasets were downscaled to 128x128x128 resolution with 8-bits precision.

Figure 6.1 - Medical datasets evaluated in the experiments.



Source: Compiled by author.

The resolutions and downsample combinations are shown below:

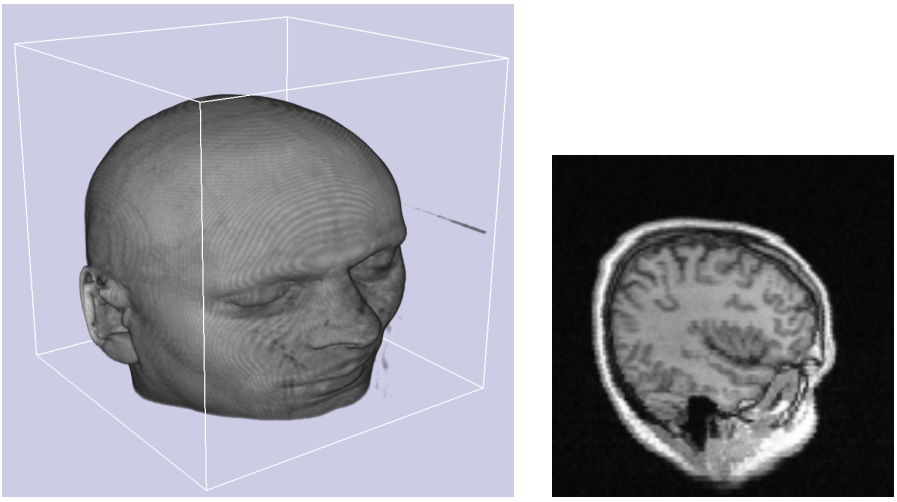
- **32_111:** Data downscaled to 32 pixels of resolution per dimension without downsample;
- **32_121:** Same downscale as the above item, volume slices downsampled 2x;
- **64_121:** Data downscaled to 32 pixels of resolution per dimension, volume slices downsampled 2x;
- **64_121:** Same downscale as the above item, image downsampled 2x;

6.2 Plane Search and Fitting Scenario

In this section, we present a detailed analysis and the visual results of one plane search and fitting.

The algorithm inputs for this scenario are: the MRI Head dataset (Figure 6.2a) and an arbitrary image previous taken from the Head volume, Figure 6.2b. In this scenario, we are evaluating the image and dataset with 64 pixels/voxels resolution per dimension. Note that it is hard to guess the image (Fig. 6.2b) position and orientation that fits inside the volumetric data.

Figure 6.2 - The evaluation input. (a) MRI Head dataset. (b) An arbitrary orientated image slice taken from Head.



(a) (b)

Source: Compiled by author.

6.2.1 Optimization data

In Table 6.1, we show the data amount prone to empty-space leaping optimization. The information present in this example is represented in resolution 64 per dimension. Approximately 48% of the input data are skipped by the similarity calculations. The "#Comp". column represents the maximum comparisons with 62x62x62 volume resolution plus 9 planes and 62x62 image resolution (1 voxel in each border is ignored). In the "Black" column are the number of voxels and pixels skipped. It results in 1,906,782,363 comparisons and represents only 23.12% of the entire pixel/voxel evaluation.

Table 6.1 - Data amount prone to empty-space leaping optimization. Volume = 43.38% voxels skipped. Image = 45.48% pixels skipped.

Head	#Comp.	Black	%
Volume	2,144,952	1,023,501	47.71%
Image	3,844	1,863	48.46%

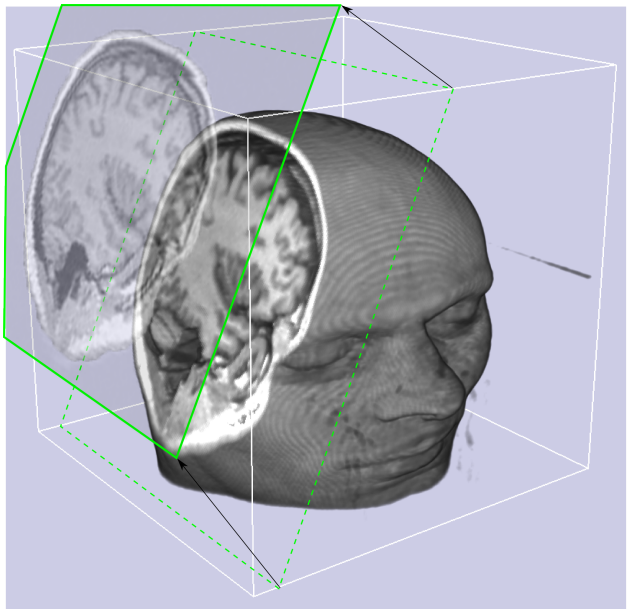
Source: Compiled by author.

Regarding to the downsample factors, the 121 and 211 downsample combinations are equivalent in the amount of data they evaluate.

6.2.2 Plane fitting

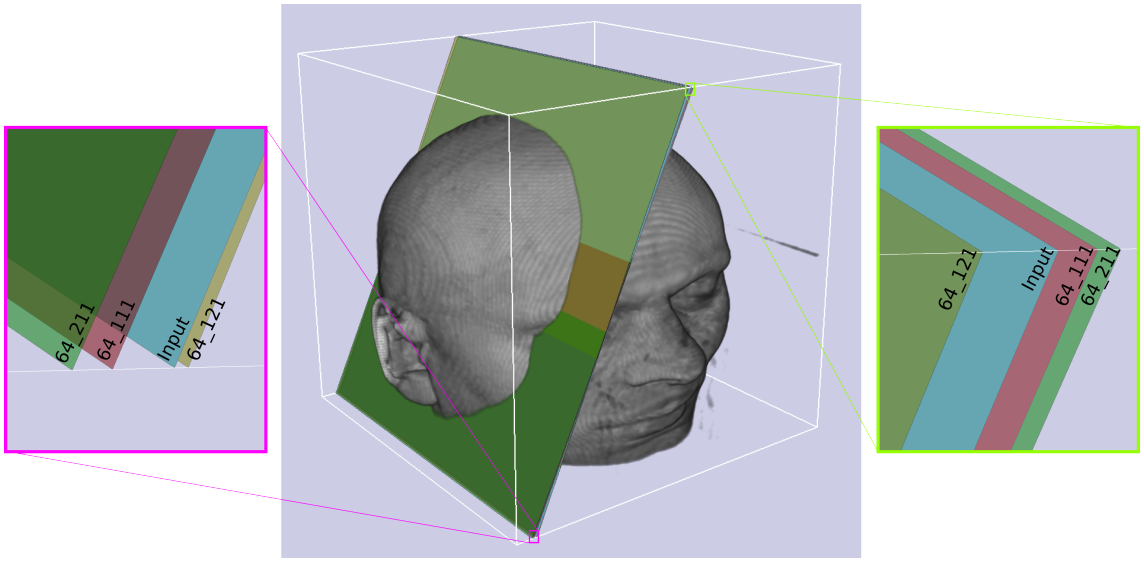
We performed the similarity evaluation for the combinations of downsample with fixed 64 pixels resolution per dimension. The Figure 6.3 shows the match of the input image slice with the cross sectioned volume. Figure 6.4 visually shows the planes precision error when compared with the input image plane. The maximum error obtained was $\approx 2.5mm$. Finally, Figure 6.5 shows the other possible visualizations.

Figure 6.3 - Volume cross sectioned (dashed plane) where the input image was found. The input image (green plane) was fused in the same visualization with transparency to show the similarity between the image plane found and the input image.



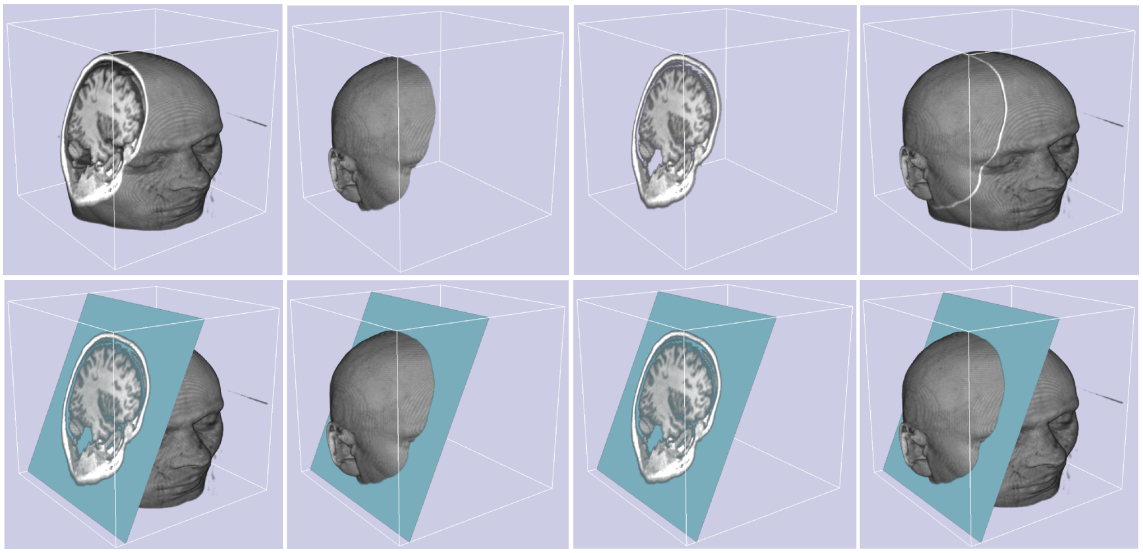
Source: Compiled by author.

Figure 6.4 - Plane fitting errors with different resolution and downsample combinations. The worst case, 64_121(green plane) was fitted with 190 similarity points and has 0.011 normalized distance and $0.33°$. It is equivalent to ≈ 2.5mm error.



Source: Compiled by author.

Figure 6.5 - Other possible visualizations of the cross sectioned volume.

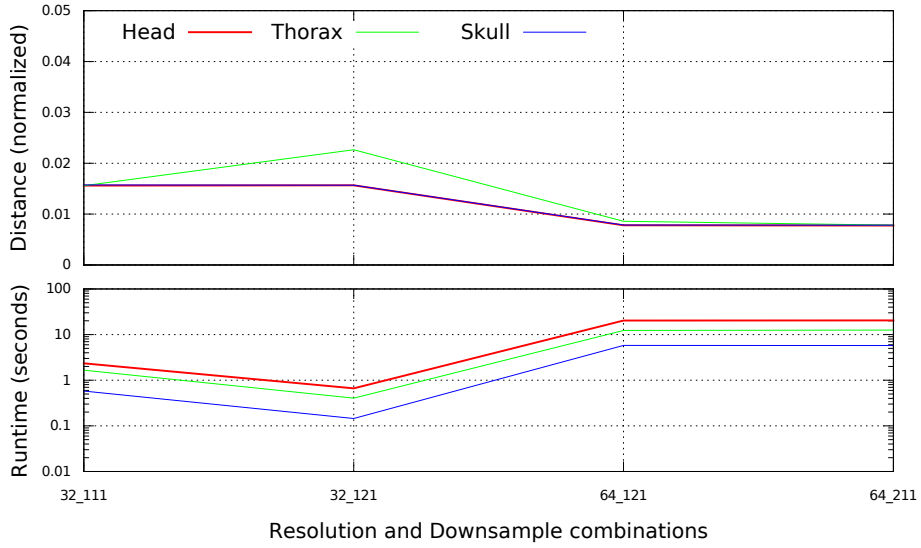


Source: Compiled by author.

6.3 Assessment Results

We evaluated the precision and performance of the three datasets. As expected (see Fig. 6.6), the results achieved correspond to the previous evaluation with controlled data. The optimizations have heavy impact on runtime and it still maintain great precision.

Figure 6.6 - Precision and runtime evaluation. Head and Skull achieve the same precision even with distinct number of comparisons performed.



Source: Compiled by author.

6.4 Chapter summary

In this chapter, we have evaluated our technique with medical images. The technique is robust in the search and match process. The precision was not affected by the complex input data. It delivers high precision and acceptable processing time. The technique has potential to be applied in other scientific fields that deal with volumetric informations.

7 DISCUSSION AND CONCLUSION

In this thesis we have investigated means to establish correlation and fusion of within-modality 2D and 3D medical images based on only pixel/voxel intensities. It was divided into two development cycles.

First, we approach the problem of matching 2D and 3D data aligned to anatomical planes. We optimize the correlation search integrating two similarity measures into one technique. It ranks all data in a coarse but fast computation and in a second step, a more refined but computational demanding similarity measure evaluates only the best classified data rearranging to the matching results achieved. This 2-step approach proved to have a good compromise between computational coast and effectiveness to match intra-subject medical data even with distinct acquisition protocols, especially when parallelized with GPU threads.

In the second development cycle, we use the knowledge acquired in the previous investigation and expand the similarity search to arbitrary aligned image planes. Due to the virtually infinite comparisons possible when using the previous technique applied to this problem, we propose an approach to reduce the comparisons set and keep high quality matches. The technique works locally evaluating similarity in small image and volume subsamples. It generates a three-dimensional point cloud with only the highest similarities found. By performing principal component analysis, we extract the best fitting plane that comprehends all the cloud points. We minimized the expected computational cost working on a very reduced set of samples, with image and volume downscale and downsample factors combined.

The design of this algorithm is the main contribution of this work. Other contribution of this thesis is the set of experiments accomplished with experimental and medical data. They reveled that there are downscaling and downsampling optimal combinations that can be applied to achieve a balance between precision and performance. The plane fitting is good enough to correctly match dozens of arbitrary oriented planes with precision higher than those found in literature.

The long-therm goal of the project started here is to address real-time correlation and fusion of pre-operative and user dependent intra-operative medical image modalities. The development cycles performed in this thesis have accomplished the foundations of the long-term goal. Despite that, they individually have their own contributions and possible real applications, both in medical and general context. In the medical context, the 2-step technique could be applied to perform fast CT dataset alignment of different acquisition protocols. In a broader approach, as the similarity measures are vastly adopted in IQA field, it could be proposed to assess loss of quality in images and videos after its compression and transmission and match ordinary images in databases. Expanding the possible applications of our arbitrary plane search, the technique could be applied to automatically align many

intersecting US frames with different orientation in the same 3D reference frame. The resulting data could then be visualized in 3D as an incomplete volume.

The work developed in this thesis has potential applications in the medical field. We believe that in the future, our technique would be integrated in medical context contributing in the analysis and operative procedures with a seamless multi-modal visualization. Nevertheless, due its general formulation, the technique can be extended and modified to benefit other fields and other data structures.

7.1 Limitations and Future work

The techniques developed through this thesis are applied in the context of within-modality correlation. Despite their usefulness in this context, a generalization of the algorithms to deal with cross-modal is planned for future works. Based on the related works, we consider two possible paths. First, change the similarity criteria to an information theory approach, such as mutual information. Second, simulating one modality into another, as performed by WEIN et al. (2008) allowing us to still use pixel/voxel intensity approaches.

Even though the evaluations have shown that the techniques perform precise matching in acceptable runtime using two simple optimizations, they are far from the desired real-time execution. To achieve higher performance, more sophisticate optimizations have to be addressed in the future. Some strategies are: adaptive multi resolution, such as Powell method, gradient descent or quasi-Newton (KLEIN; STARING; PLUIM, 2007) and KD-tree space subdivision. Furthermore, the technique can be boosted up with massive GPU parallelization.

We plan to explore in depth a possible error metric to assess similarity between arbitrary oriented planes unifying to their position and orientation error in space.

Other limitation not addressed in this work, is that the input 2D image have to be in the same scale as the input volume and cannot be rotated.

REFERENCES

AVANTS, B. B. et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. **Medical Image Analysis**, [S.l.: s.n.], v.12, n.1, 2008, p.26–41.

BEN SASSI, O. et al. MR Image Monomodal Registration Using Structure Similarity Index. In: IMAGE PROCESSING THEORY, TOOLS AND APPLICATIONS, 2008. IPTA 2008. FIRST WORKSHOPS ON. **Proceedings...** [S.l.: s.n.], 2008. p.1–5.

BERKMANN, J.; CAELLI, T. Computation of surface geometry and segmentation using covariance techniques. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.: s.n.], v.16, n.11, 1994, p.1114–1116.

BRUCKNER, S. **Efficient Volume Visualization of Large Medical Datasets - Concepts and Algorithms**. Saarbrücken, Germany, Germany: VDM Verlag, 2008.

BYRNE, J. V. et al. Assessment of a technique for 2D–3D registration of cerebral intra-arterial angiography. **The British Journal of Radiology**, [S.l.: s.n.], v.77, 2004, p.123–128.

CHUI, H.; RANGARAJAN, A. A New Point Matching Algorithm for Non-rigid Registration. **Comput. Vis. Image Underst.**, New York, NY, USA, v.89, n.2-3, Feb. 2003, p.114–141.

CROCETTI, L. et al. Targeting liver lesions for radiofrequency ablation: an experimental feasibility study using a ct-us fusion imaging system. **Investigative radiology**, [S.l.: s.n.], v.43, n.1, Jan. 2008, p.33–9.

DEAN, C. et al. An evaluation of four CT-MRI co-registration techniques for radiotherapy treatment planning of prone rectal cancer patients. **British Journal of Radiology**, [S.l.: s.n.], v.85, n.1009, 2012, p.61 – 68.

DHAWAN, A. P. **Medical image analysis**. 2nd.ed. Oxford: Wiley-Blackwell, 2011. (IEEE Press series on biomedical engineering). Previous ed.: New York; Chichester: Wiley, 2003.

DU, J. et al. Intensity-based robust similarity for multimodal image registration. **Int. J. Comput. Math.**, [S.l.: s.n.], v.83, n.1, 2006, p.49–57.

ENGEL, K. et al. Real-time Volume Graphics. In: ACM SIGGRAPH 2004 COURSE NOTES, New York, NY, USA. **Proceedings...** ACM, 2004. (SIGGRAPH '04).

ENGEL, K.; KRAUS, M.; ERTL, T. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. In: HWWS '01: PROCEEDINGS OF THE ACM SIGGRAPH/EUROGRAPHICS WORKSHOP ON GRAPHICS HARDWARE, New York, NY, USA. **Proceedings...** ACM, 2001. p.9–16.

FAU-3D. The Volume Library. In: VOLUMETRIC DATASETS. **Proceedings...** Online at: http://lgdv.cs.fau.de/External/vollib/, 2014.

GILL, R. **The physics and technology of diagnostic ultrasound** : a practitioner's guide. Abbotsford, N.S.W: High Frequency Publishing, 2012.

HADWIGER, M. et al. **Real-time Volume Graphics**. Natick, MA, USA: A. K. Peters, Ltd., 2006.

HEDRICK, W. R.; HYKES, D. L. Image formation in real-time ultrasound. In: JOURNAL OF DIAGNOSTIC MEDICAL SONOGRAPHY 11. **Proceedings...** [S.l.: s.n.], 1995. p.246–251.

HIPWELL, J. et al. Intensity-based 2-D - 3-D registration of cerebral angiograms. **Medical Imaging, IEEE Transactions on**, [S.l.: s.n.], v.22, n.11, 2003, p.1417–1426.

HAJNAL, J. V.; HAWKES, D. J.; HILL, D. L. G. (Ed.). **Medical Image Registration**. [S.l.: s.n.], 2001.

KHALIFA, F. et al. State-of-the-Art Medical Image Registration Methodologies: a survey. In: EL-BAZ, A. S. et al. (Ed.). **Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies**. [S.l.]: Springer US, 2011. p.235–280.

KIM, J.; FESSLER, J. Intensity-based image registration using robust correlation coefficients. **Medical Imaging, IEEE Transactions on**, [S.l.: s.n.], v.23, n.11, Nov 2004, p.1430–1444.

KING, A. et al. Registering Preprocedure Volumetric Images With Intraprocedure 3-D Ultrasound Using an Ultrasound Imaging Model. **Medical Imaging, IEEE Transactions on**, [S.l.: s.n.], v.29, n.3, March 2010, p.924–937.

KITA, Y.; WILSON, D.; NOBLE, J. Real-time registration of 3D cerebral vessels to X-ray angiograms. In: WELLS, W.; COLCHESTER, A.; DELP, S. (Ed.). **Medical Image Computing and Computer-Assisted Intervention — MICCAI'98**. [S.l.]: Springer Berlin Heidelberg, v.1496, 1998. p.1125–1133. (Lecture Notes in Computer Science, v.1496).

KLEIN, S.; STARING, M.; PLUIM, J. Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines. **Image Processing, IEEE Transactions on**, [S.l.: s.n.], v.16, n.12, 2007, p.2879–2890.

LEVOY, M. Display of surfaces from volume data. In: IEEE COMPUTER GRAPHICS AND APPLICATIONS. **Proceedings...** [S.l.: s.n.], 1988. p.v.8,p 29–37.

LINDSETH, F. et al. Multimodal Image Fusion in Ultrasound-Based Neuronavigation: improving overview and interpretation by integrating preoperative mri with intraoperative 3d ultrasound. **Computer Aided Surgery**, [S.l.: s.n.], v.8, n.2, 2003, p.49–69.

LOWE, D. Object recognition from local scale-invariant features. In: COMPUTER VISION, 1999. THE PROCEEDINGS OF THE SEVENTH IEEE INTERNATIONAL CONFERENCE ON. **Proceedings...** [S.l.: s.n.], 1999. v.2, p.1150 –1157 vol.2.

LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. **Int. J. Comput. Vision**, Hingham, MA, USA, v.60, n.2, Nov. 2004, p.91–110.

MAES, F. et al. Multimodality image registration by maximization of mutual information. **Medical Imaging, IEEE Transactions on**, [S.l.: s.n.], v.16, n.2, 1997, p.187–198.

MAINTZ, J.; VIERGEVER, M. A. A survey of medical image registration. **Medical Image Analysis**, [S.l.: s.n.], v.2, n.1, 1998, p.1 – 36.

MANI, V.; RIVAZHAGAN, D. Survey of Medical Image Registration. **Journal of Biomedical Engineering and Technology**, [S.l.: s.n.], v.1, n.2, 2013, p.8–25.

MARESCAUX, J. et al. Virtual reality applied to hepatic surgery simulation: the next revolution. In: SURGERY. **Annals...** [S.l.: s.n.], 1998. p.228:627–634.

MATTES, D. et al. PET-CT Image Registration in the Chest Using Free-form Deformations. **IEEE Trans. Med. Imaging**, [S.l.: s.n.], 2003, p.120–128.

MAX, N. Optical models for direct volume rendering. **Visualization and Computer Graphics, IEEE Transactions on**, [S.l.: s.n.], v.1, n.2, 1995, p.99–108.

MCLAUGLIN, R. A. et al. Intensity-based Registration versus Feature-based registration for neurointerventions. In: MEDICAL IMAGE UNDERSTANDING AND ANALYSIS (MIUA). **Proceedings...** [S.l.: s.n.], 2001. p.69–72.

MEIBNER, M. et al. A practical evaluation of popular volume rendering algorithms. In: VVS '00: PROCEEDINGS OF THE 2000 IEEE SYMPOSIUM ON VOLUME VISUALIZATION, New York, NY, USA. **Proceedings...** [S.l.: s.n.], 2000. p.81–90.

NIKAS, D. C. et al. Coregistered intraoperative ultrasonography in resection of malignant glioma. **Neurosurgical Focus**, [S.l.: s.n.], v.14, n.2, 2003, p.1–5.

PENNEY, G. P. et al. A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration. **IEEE Trans. Med. Imaging**, [S.l.: s.n.], v.17, n.4, 1998, p.586–595.

PENNEY, G. P. et al. Validation of a two- to three-dimensional registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images. **Medical Physics**, [S.l.: s.n.], v.28, n.6, 2001, p.1024–1032.

ROHLFING, T.; MAURER JR., C. R. A Novel Image Similarity Measure for Registration of 3-D MR Images and X-Ray Projection Images. In: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION-PART II, 5., London, UK, UK. **Proceedings...** Springer-Verlag, 2002. p.469–476. (MICCAI '02).

RUSS, J. **The image processing handbook**. Boca Raton: [s.n.], 2011.

SCHARFE, M.; PIELOT, R.; SCHREIBER, F. Fast multi-core based multimodal registration of 2D cross-sections and 3D datasets. **BMC Bioinformatics**, [S.l.: s.n.], v.11, n.1, 2010, p.1–11.

TOMAEVIC, D.; LIKAR, B.; PERNUš, F. “Gold Standard” 2D/3D Registration of X-Ray to CT and MR Images. In: DOHI, T.; KIKINIS, R. (Ed.). **Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002**. [S.l.]: Springer Berlin Heidelberg, v.2489, 2002. p.461–468. (Lecture Notes in Computer Science, v.2489).

TOMBROPOULOS, R. Z. Techniques and applications for tree-dimensional visualization in medicine. In: BIOENGINEERING CONFERENCE. **Proceedings...** [S.l.: s.n.], 1999.

TURGEON, G.-A. et al. 2D-3D registration of coronary angiograms for cardiac procedure planning and guidance. **Medical Physics**, [S.l.: s.n.], v.32, n.12, 2005, p.3737–3749.

WANG, Z. **Rate scalable foveated image and video communications**. [S.l.]: University of Texas at Austin, 2001.

WANG, Z.; BOVIK, A. Mean squared error: love it or leave it? a new look at signal fidelity measures. **Signal Processing Magazine, IEEE**, [S.l.: s.n.], v.26, n.1, jan. 2009, p.98 –117.

WANG, Z.; BOVIK, A. C. **Modern Image Quality Assessment**. [S.l.]: Morgan & Claypool Publishers, 2006. (Synthesis Lectures on Image, Video, and Multimedia Processing).

WANG, Z. et al. Why Is Image Quality Assessment So Difficult? In: PROC. IEEE INT. CONF. ACOUST., SPEECH, AND SIGNAL PROCESSING. **Proceedings...** [S.l.: s.n.], 2002. p.3313–3316.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. **Image Processing, IEEE Transactions on**, [S.l.: s.n.], v.13, n.4, april 2004, p.600 –612.

WEIN, W. et al. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. **Medical Image Analysis**, [S.l.: s.n.], v.12, n.5, 2008, p.577 – 585. <ce:title>Special issue on the 10th international conference on medical imaging and computer assisted intervention - {MICCAI} 2007.

WEST, J. et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. **JOURNAL OF COMPUTER ASSISTED TOMOGRAPHY**, [S.l.: s.n.], v.21, 1997, p.554–566.

WESTOVER, L. Footprint Evaluation for Volume Rendering. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 17., New York, NY, USA. **Proceedings...** ACM, 1990. p.367–376. (SIGGRAPH ’90).

WONG, A.; BISHOP, W. Indirect Knowledge-Based Approach to Non-Rigid Multi-Modal Registration of Medical Images. In: ELECTRICAL AND COMPUTER ENGINEERING, 2007. CCECE 2007. CANADIAN CONFERENCE ON. **Proceedings...** [S.l.: s.n.], 2007. p.1175–1178.

WOOD, T.; CHUNG, M.; ALLEGRA, D. Radiofrequency ablation of 231 unresectable hepatic tumors: indications, limitations, and complications. **Annals of Surgical ...**, [S.l.: s.n.], v.7, n.8, Sept. 2000, p.593–600.

XIAO, Z.-S.; ZHENG, C. xun. A novel image fusion scheme by integrating local image structure and directive contrast. In: INFORMATION FUSION (FUSION), 2011 PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON. **Proceedings...** [S.l.: s.n.], 2011. p.1–6.

ZITOVá, B.; FLUSSER, J. Image registration methods: a survey. **Image and Vision Computing**, [S.l.: s.n.], v.21, n.11, 2003, p.977 – 1000.

APPENDIX A BUSCA MULTIDIMENSIONAL DE SIMILARIDADE PARA CORRELAÇÃO E FUSÃO 2D-3D DE DADOS MÉDICOS

Resumo da Dissertação em Português

A.1 Introdução

Imagens da anatomia interna humana são essenciais para as práticas clínicas, elas auxiliam diagnósticos, planejamento de cirurgias, caracterização de lesões e estudos fisiológicos para tratamento de doenças humanas. Diversas modalidades de aquisição de imagens estão disponíveis. Cada uma delas enfatizam diferentes estruturas internas (anatômicas e funcionais) e apresentam diferentes dimensionalidades. Algumas delas são capturadas e analisadas em tempo real, e outras capturadas e armazenadas para uma análise posterior. Tomografia computadorizada (TC), ressonância magnética (RM), ultrassom (US), raios-X e laparoscopia são as mais comuns. Hoje, todas as modalidades de imagens médicas estão disponíveis em formato digital, isso permite que sistemas computacionais realizem análises e processamentos automáticas dos dados médicos. Representações tridimensionais virtuais dos datasets de TC e RM, segmentação e classificação de segmentos funcionais de órgãos, análise de similaridade e registro de imagens de diferentes modalidades são alguns exemplos.

Estabelecer correlações entre imagens médicas de múltiplas fontes de aquisição é procedimento importante para diagnósticos e tratamentos. Enquanto correlação entre imagens de mesma modalidade permitem acompanhar a evolução e o crescimento de uma doença, correlação entre imagens de diferentes modalidades permite combinar a percepção da mesma estrutura vista por cada modalidade. Fusão entre imagens de TC e US, por exemplo, pode melhorar significativamente a validade de um diagnóstico quando comparada com a análise separada das modalidades (WEIN et al., 2008). Apesar da sua importância, esses procedimentos de correlação são frequentemente realizados manualmente nas práticas clínicas. A confiança do médico na identificação da lesão é ainda um desafio durante os procedimentos clínicos (WOOD; CHUNG; ALLEGRA, 2000). Os resultados são, então, muito dependentes das habilidades do médico para produzir uma imagem mental da correlação dos dados. Espera-se que uma correlação automática rápida e precisa de imagens médicas de mesma e diferentes modalidades possa levar a um melhor diagnóstico e tratamento. Características das imagens capturadas em diferentes instantes de tempo ou

com diferentes dispositivos seriam enfatizadas em uma única visualização, aumentando a precisão na localização de doenças e posicionamento preciso de aparelhos de tratamento. O registro de imagens médicas é amplamente explorado (MAINTZ; VIERGEVER, 1998; KHALIFA et al., 2011; MANI; RIVAZHAGAN, 2013). Soluções para uma variedade de combinações de correlações de imagens já foram propostas. Elas abordam registro entre dados pré-operatórios de mesma ou diferente dimensionalidade e modalidades, como por exemplo, 2D-3D: raios-X com TC e RM (TOMAEVIC; LIKAR; PERNUš, 2002); 3D-3D: RM-RM (BEN SASSI et al., 2008) e TC-RM (DEAN et al., 2012). Esses trabalhos focam no registro em um dos três eixos anatômicos principais, axial, coronal e sagital.

Por outro lado, no registro entre modalidades pré- e intra-operatórias, como US-TC (WEIN et al., 2008; CROCETTI et al., 2008), US-RM (KING et al., 2010; NIKAS et al., 2003; LINDSETH et al., 2003) os dados capturados estão normalmente desalinhados aos eixos principais. Para resolver o problema da aquisição dependente do posicionamento do aparelho de aquisição, as soluções anteriores rastreiam a posição e orientação do transdutor do ultrassom e do corpo do paciente para registrar as modalidades no mesmo sistema de coordenadas.

A.1.1 Objetivos

Nesta dissertação, projetamos uma abordagem para correlacionamento de dados 2D e 3D de uma mesma modalidade de aquisição baseada somente nas informações dos pixels e voxels. Nossa abordagem usa análise de similaridade para, dada uma imagem (2D) arbitrária, achar a localização dessa imagem dentro de um volume (3D). Nossa solução minimiza os infinitos cortes transversais possíveis e utiliza otimizações para diminuir a carga computacional para exibir resultados precisos. Técnicas de rendering volumétrico são aplicadas para visualização. Os volumes médicos são cortados transversalmente para enfatizar a imagem 2D localizada. Uma avaliação extensa foi realizada para avaliar a precisão e a eficiência da técnica proposta.

A.1.2 Contribuições

A seguir, as contribuições técnicas deste trabalho são resumidas.

- Um algoritmo de dois passos para correlacionamento multidimensional de imagens médicas;
- Uma nova abordagem para busca de similaridade para correlação e fusão de dados multidimensional.
- Três conjuntos de experimentos realizados para avaliar a qualidade dos resultados e aplicabilidade do nosso método;
- Uma ferramenta de visualização que integra em tempo real rendering volumétrico de datasets com cortados transversalmente em planos arbitrários de acordo com uma imagem bidimensional de entrada.

A.1.3 Organização da dissertação

Após a apresentação dos trabalhos relacionados (capítulo 3) e embasamento técnico (capítulo 2), os dois próximos capítulos tratam da investigação, implementação e avaliação de duas técnicas para correlação de imagens médicas: Uma técnica baseada em dois passos para correlação de imagens médicas de mesma modalidade alinhadas a um dos

eixos ortogonais (capítulo 4) e outra técnica para correlação de imagens arbitrariamente orientadas com datasets (capítulo 5). As discussões, concluões e trabalhos futuros estão no capítulo 7.

A.2 Trabalhos Relacionados

Vários trabalhos anteriores abordam o problema de correlação/registro de imagens. Esta seção apresenta trabalhos que realizam correlação de imagens 2D-3D e discute apenas os trabalhos mais intimamente relacionados com o foco desta dissertação. Para uma análise mais abrangente dos trabalhos relacionados, as pesquisas por MAINTZ; VIERGEVER (1998), KHALIFA et al. (2011) e MANI; RIVAZHAGAN (2013) apresentam vários trabalhos focados em registro de imagens médicas.

WEIN et al. (2008) desenvolveu uma técnica para registrar automaticamente Imagens dos US com TC. A posição do transdutor US é rastreada com sensores magnéticos e a aquisição das imagens é feita através de uma varredura 3D realizada pelo transdutor. Para cumprir o registro entre as modalidades, eles simularam efeitos de ultrassom nas imagens de TC e desenvolveram uma função para o cálculo da similaridade entre o ultrassom e as imagens simuladas. O registro foi testado em 25 pacientes com erro de registro médio de 8,1 milímetros em relação aos dados de referencia, que é muito alto para uma intervenção médica precisa.

SCHARFE; PIELOT; SCHREIBER (2010) abordou o registro funcional e estrutural de imagens 2D PET com ressonância magnética. Eles propõem uma implementação da técnica de comparação de similaridade utilizando informação mútua normalizada (NMI) na Cell Broadband Engine (CBE) do console PlayStation 3, que permite o processamento multi-core. No entanto, as imagens são processadas apenas no plano axial, reduzindo o problema para registro de imagens 2D-2D.

NIKAS et al. (2003) correlacionam ultrassonografia com RM usando um dispositivo de rastreamento óptico 3D. O rastreador óptico captura a posição da cabeça do paciente e do transdutor US para definir as transformações rígidas entre eles. Na mesma direção, CROCETTI et al. (2008) usa um sistema comercial que registra imagens de TC e US em tempo real para segmentação e posterior tratamento por radiofrequência (RF) de tumores no fígado. Eles registram o órgão real com as imagens tomográficas utilizando marcadores fiduciais para calibrar o sistema de coordenadas do aparelho enquanto a posição do transdutor eletromagnético é rastreada.

Embora esses trabalhos confirmem o interesse médico para fundir diferentes fontes de dados, eles ou dependem de marcas fiduciais invasivas ou sistemas de rastreamento para registro. O nosso método, ao invés disso, usar apenas as informações dos pixels e voxels.

A.3 Embasamento Técnico

Nesta seção são detalhadas três abordagens para análise de similaridade entre imagens, erro médio quadrático (MSE), peak signal-to-noise ratio (PSNR) e similaridade estrutural (SSIM) que são fundamentais para o entendimento desta dissertação.

A.3.1 Erro médio quadrático

Erro quadrático médio (MSE) é uma métrica simples e popular, que é calculada pela média quadrática da intensidade da diferença dos pixels de duas imagens, normalmente uma é a imagem de referência e o outro é uma distorcida. O resultado é uma pontuação

quantitativa que descreve o grau de semelhança ou o nível de distorção entre eles (WANG; BOVIK, 2009).

O MSE entre dois sinais é dado por:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2, \tag{A.1}$$

onde x e y são os dados de duas imagens, N é o número de amostras do sinal (quantidade de pixels). A diferença entre o sinal original e o distorcido é dada por: $e_i = x_i - y_i$

Na literatura de processamento de imagens, MSE pode ser convertido para o PSNR:

$$PSNR = 10log_{10}\frac{L^2}{MSE}, \tag{A.2}$$

onde L é o intervalo dinâmico permitido para os valores de intensidade de cada pixel. O PSNR é útil quando as imagens comparadas possuem intervalos dinâmicos diferentes, caso contrário ele não adiciona novas informações relativas ao MSE.

A.3.2 Similaridade Estrutural

A motivação para a abordagem de similaridade estrutural (SSIM) é encontrar uma maneira mais direta e precisa para comparar imagens. Os pixels das imagens possuem fortes dependências e eles carregam informações importantes sobre a estrutura dos objetos na cena, especialmente quando os pixels estão próximos espacialmente. Desta forma, WANG et al. (2004) propôs um novo framework para avaliação da qualidade de imagens, baseado na hipótese que o sistema visual humano é altamente adaptado para extrair informações estruturais das imagens vistas. A medida de informação estrutural pode fornecer uma boa aproximação para distorções de imagem percebidas pelos humanos. A tarefa é separada em três comparações: luminosidade, contraste e estrutura. Todos os três elas são misturadas em uma equação, formulada como:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{A.3}$$

onde μ_x, σ_x e μ_y, σ_y denotam a intensidade média e desvio padrão das imagens x e y respectivamente, enquanto σ_{xy} denota sua correlação cruzada. $C1$ e $C2$ são constantes pequenas utilizadas para evitar problemas de instabilidade quando o denominador está próximo de zero.

A.4 Algoritmo de Dois Passos para Correlacionamento de Imagens Médicas Multidimensionais

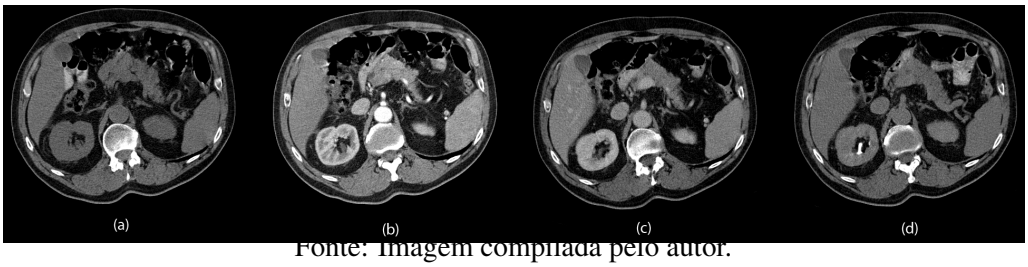
Nesta seção são descritos o projeto do algoritmo para correlacionamento de imagens médicas multidimensionais utilizando duas fases análise de similaridade.

A.4.1 Características dos Dados Médicos

Os dados que estamos tratando são datasets de fatias de imagens de tomografia computadorizada, capturadas no plano axial que compreendem a região inferior do tórax até a pélvis de um corpo homem adulto. No total, quatro datasets do mesmo indivíduo foram convertidos e preparados para o uso em nosso sistema. Os datasets foram capturados do em uma única sessão de exame utilizando quatro diferentes protocolos. A diferença que

distingue os datasets é a fase de perfusão de um agente de contraste radiológico injetado no paciente. As fases são mostradas na figura A.1. As imagens resultantes possuem 16 bits armazenados em um canal de cor que representa valores de luminância. Os datasets são compostos de 140 fatias com resolução de 512x512 pixels. Os conjuntos de dados também foram escalados para baixo por fatores de dois e quatro. Isso nos dá três resoluções para serem avaliadas nos experimentos: 512x512 (nativo), 256x256 e 128x128. A distância entre fatia é de 2,5 mm.

Figure A.1 - Imagens exemplos retirados dos quatro datasets. (a) antes do contraste. (b) fase arterial: 30 segundos depois da injeção de contraste, (c) fase venosa: 70-90 segundos após a injeção de contraste, (d) fase tardia: 5-10 minutos depois.



Fonte: Imagem compilada pelo autor.

A.4.2 Algoritmo de dois passos para correlação de imagens

Com as características apresentadas acima, uma possível questão a ser explorada é: dada uma fatia na fase tardia (Figura A.1(d).) qual fatia mais semelhante presente na fase arterial (Figura A.1(b))? Esta pergunta pode ser respondida por um ser humano especialista que observa a imagem de referência e a compara com todas as imagens do conjunto de dados. Esta análise é muito eficaz, mas sua eficiência piora à medida que o conjunto de dados cresce. Métodos computacionais podem resolver este problema com uma maior eficiência e confiabilidade. As entradas para qualquer uma destes métodos são: uma imagem de referência e um dataset. O objetivo é encontrar uma fatia do conjunto de dados completo que é a mais parecida com a fatia de entrada de referência sem o uso qualquer informação adicional.

Implementamos primeiro as técnicas PSNR e SSIM em C++. Testes preliminares revelaram que PSNR é rápido, mas é impreciso em muitos casos. Por outro lado, SSIM é mais lento, mas obtém resultados muito precisos em média. Nós, em seguida, desenvolvemos um novo algoritmo que estabelece um equilíbrio entre velocidade de processamento e precisão das duas abordagens. Os dois passos do algoritmo são: No primeiro passo, o cálculo PSNR é realizado para criar rapidamente uma lista de fatias com a melhor correspondência com a imagem de referência. No segundo passo, é calculado SSIM apenas para as melhores fatias classificadas que resultaram na reordenação do ranking de acordo seus resultados de similaridades mais altos.

A.4.3 Avaliação do algoritmo de dois passos para correlação de imagens

Montamos um protocolo de teste para avaliar e comparar o desempenho e eficácia entre a nossa abordagem (2-STEP), PSNR e SSIM para a detecção de similaridade. Nossa preocupação é que os algoritmos sejam rápidos e obtenham os melhores resultados possíveis na detecção de semelhança entre imagens. Todos os dados relevantes são registrados em tempo de execução e salvos automaticamente para análise. O hardware utilizado durante

os testes foi um computador com um processador Core I7-2600 rodando a 3.4GHz e uma GeForce 560Ti com 384 núcleos CUDA rodando a 880MHz.

A.4.3.1 Protocolo experimental

Para fins de teste, executamos o teste muitas vezes, cada um com uma imagem referência diferente. Elas são obtidas a partir de um dataset referência. Assim, a comparação é realizada entre todas as fatias de um conjunto de dados com o outro. As pilhas de imagens são previamente alinhadas de tal maneira que, por exemplo, a correspondência correta para a fatia 42 no primeiro conjunto de dados é a fatia 42 no segundo. Em cada conjunto de dados comparado, nós armazenamos a fatia mais semelhante e seus n vizinhos classificados em ordem decrescente. Esta informação permite calcular a distância (delta) para a correspondência exata. Além disso, ele permite saber se os vizinhos da correspondência correta estão dentro do ranking gerado. O número de posições armazenadas no ranking influencia diretamente o segundo passo do algoritmo.

A.4.3.2 Variáveis de teste

As variáveis independentes utilizados em nosso teste são: resoluções do conjunto de dados (512x512x140, 256x256x140, 128x128x140), planos de orientação ortogonais (Axial (XY), Coronal (XZ) e sagital (YZ)), algoritmos (2-STEP, PSNR, SSIM), fases de contraste radiológico (arterial, venosa e tardia).

As variáveis dependentes coletadas são: tempo de execução (em segundos), distância para a correlação correta (em fatias) e ranking (quantidade de fatias na distância máxima de dez fatias).

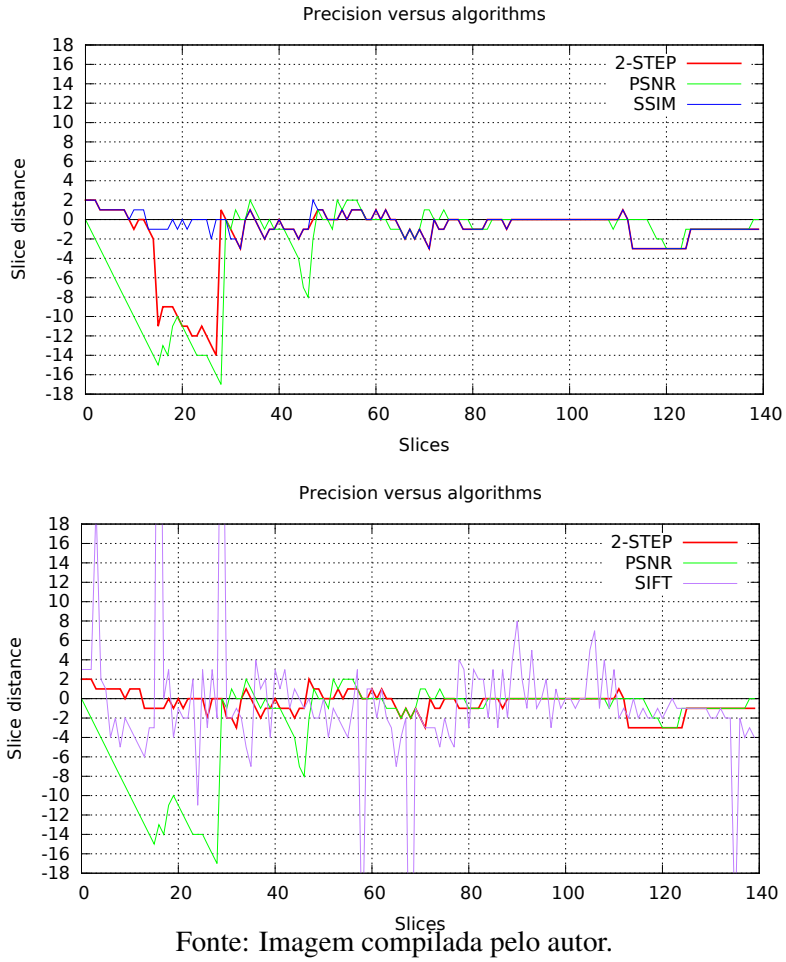
A.4.3.3 Avaliação de precisão

Para avaliar a precisão, nos escolhemos arbitrariamente a comparação do dataset de fase venosa contra o de fase tardia, ambos com resolução de 256x256x140 e alinhados ao plano axial. No gráfico da Figura A.2, o PSNR tem uma elevada instabilidade no início da comparação (fatias 1-30). Neste caso, o conjunto de fatias melhor classificados não contém a correspondência correta. Por os resultados do segundo passo do algoritmo dependerem do primeiro passo, os resultados de similaridade ficam limitados as fatias que estão contidas no ranking e, portanto nunca representaram o casamento correto, embora os resultados sejam ligeiramente melhores na zona de instabilidade. A solução para esse problema de instabilidade foi aumentar o número de entradas no ranking para aumentar as chances de um casamento correto estar contido. Essa modificação aumentou o tempo de computação do algoritmo, uma vez que tem de realizar mais comparações. O ranking foi aumentado de 10 para 20 comparações e o tempo de execução aumentou de 6,9 para 12,89 segundos (87%). A técnica SIFT é muito instável, gera diversos pontos que divergem do resultado ótimo e seu tempo de execução é mais alto (264.4s executando em CPU). O tempo de execução total para este teste é mostrado no gráfico da Figura A.3. A conclusão interessante aqui é que o 2-Step executa muito mais rápido do que SSIM e mantém o mesmo nível da precisão.

A.4.3.4 Avaliação variando a resolução dos datasets

Outra hipótese é que haveria pouca diferença nos resultados de precisão utilizando datasets com resoluções menores. Por outro lado, o tempo de execução cairia de forma quadrática como as resoluções. Este teste compara o tempo de execução e precisão na

Figure A.2 - Avaliação de precição. Comparação entre os datasets de fase venosa com o de fase tardia orientados no plano axial e com resoluções 256x256x140. no gráfico acima, a instabilidade gerada pelo cálculo utilizado o PSNR afeta a precisão do algoritmo 2-STEP, pois as fatias corretas não estão no ranking gerado pelo PSNR no primeiro passo. Após o aumento das posições no ranking para 20 a instabilidade é resolvida (gráfico abaixo). Os resultados para a técnica SSIM não são mostrados pois o gráfico é exatamente igual ao 2-STEP. A técnica SIFT apresenta os piores resultados.



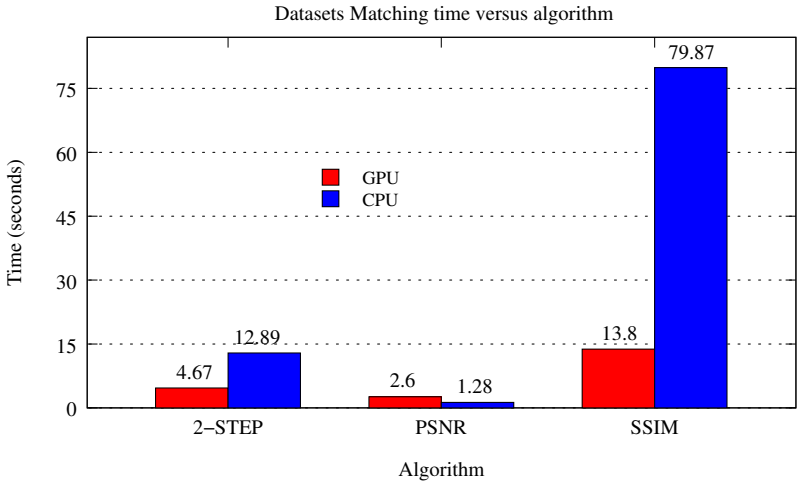
Fonte: Imagem compilada pelo autor.

similaridade obtida com o algoritmo 2-STEP para todas as combinações de datasets para cada resolução. Neste teste o plano de orientação foi fixado na perspectiva axial. O ranking utilizado para o 2-STEP foi o maior (20 comparações no segundo passo). Os resultados estão expostos no gráfico da Figura A.4. De 512x512 para 256x256 o tempo de execução diminuir em 45% e para 128x128, 76%. Precisão e tempo resultados correspondem a compilação dos resultados de todas as combinações de conjunto de dados. Note que, quando o tempo de execução cai bastante, a perda de precisão é pequena. A resolução 256x256 obteve o melhor equilíbrio entre desempenho e precisão. Na resolução mais baixa, a implementação em GPU não tem tantas vantagens quando comparada com a implementação em CPU. Apenas com resoluções maiores a paralelização faz diferença.

A.4.3.5 Avaliação variando a orientação dos planos

Tínhamos como hipótese que o 2-STEP iria executar com precisão semelhante nos três planos de corte ortogonais avaliados. As variáveis independentes para esse teste

Figure A.3 - Tempo de execução dos algoritmos para o testes de precisão (Fig. A.2). O 2-STEP foi avaliado com ranking de 20 posições. Na barra GPU (em vermelho) todos os algoritmos foram executados apenas na GPU.



Fonte: Imagem compilada pelo autor.

foram: soma das três combinações de comparação dos datasets, resolução 256x256x140 executando do algoritmo 2-STEP. Diferentemente da nossa hipótese, os conjuntos de dados de orientação axial apresentou um taxa de precisão inferior, mais de 40% quando comparado com o mais preciso, sagital. Uma das causas para esse comportamento são as grandes áreas da imagem que não contêm informações (partes pretas vistas na Fig. A.1).

A.5 Algoritmo de busca multidimensional de planos para correlação 2D-3D

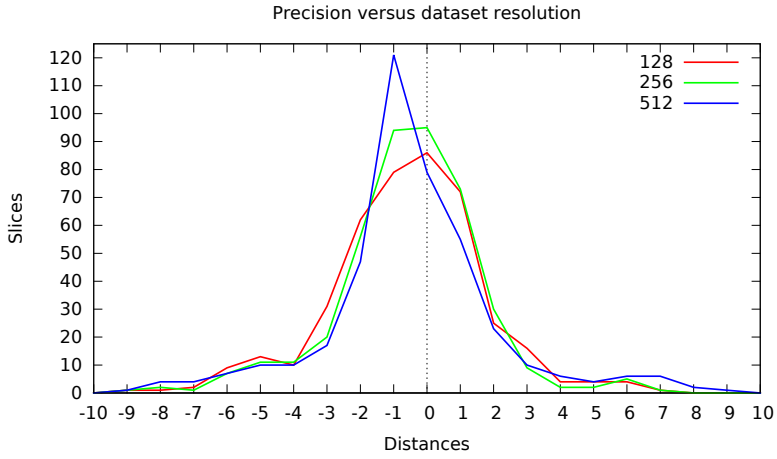
A abordagem em duas fases (seção A.4) resolve o problema de detecção de similaridade entre imagens alinhadas em um dos três planos ortogonais: axial, coronal e sagital. No entanto, na prática diária, a aquisição tende a estar fora dos planos principais. Isto ocorre especialmente quando a captura de imagem é dependente do usuário, por exemplo, durante o exame de ultrassom.

Nesta seção, apresentamos todas as etapas e métodos de nossa abordagem para encontrar a localização de uma imagem planar em posição e orientação arbitrárias dentro de um volume de dados. Pode-se imaginar uma solução por força bruta que compara a imagem com todas as secções de um volume usando algum critério similaridade. A seção mais semelhante corresponderá a localização de imagem planar no interior do volume. Como o custo computacional de tal busca exaustiva é proibitivo, propomos aqui é uma abordagem mais inteligente que minimiza o número virtualmente infinito de possíveis orientações da seção transversal.

A.5.1 Entradas do Algoritmo

Dois elementos compõem os dados de entrada para o nosso algoritmo: Um dataset, conjunto de dados 3D; uma imagem plana de qualquer tamanho gerado como um subconjunto do volume anteriormente mencionado É gerado um corte aleatório de um dos datasets. Em primeiro lugar os voxels do volume são mapeados no intervalo de -1 a 1 em x, y e z. Em seguida, gera-se um plano de imagem arbitrário selecionando três pontos não

Figure A.4 - Precisão atingida com três resoluções distintas. Os resultados foram combinados para todas as combinações de datasets para cada resolução. O plano de orientação avaliado foi o axial.



Fonte: Imagem compilada pelo autor.

colineares, p_1, p_2 e p_3. O vetor normal n é calculado como $\vec{n} = \hat{v}_1(p_2, p_1) \times \hat{v}_2(p_3, p_1)$. Para completar a equação do plano, a distância do plano para a origem do sistema de coordenadas, o coeficiente d é calculado como: $d = -(\vec{n} \cdot p_1)$. Um sampling 2D utilizando nearst-neighbor é aplicada para gerar a imagem final.

A.5.2 Busca de imagens em volumes

O algoritmo (Fig. A.5) trabalha localmente avaliando similaridade entre pequenas sub-imagens e sub-volumes, provenientes dos dados de entrada originais. Um tamanho s determina a resolução da sub-imagem ($SI_{s\times s}$) e sub-volume ($SV_{s\times s\times s}$). A similaridade de cada sub-imagem amostrada é avaliado contra nove planos orientados do sub-volume (três planos cardinais mais seis planos intermediários) usando MSE. Os nove planos são classificados por similaridade e, se o plano mais semelhante é abaixo de um determinado limiar, a posição do voxel é armazenado. Além da posição do voxel encontrado, valor de semelhança e a orientação do plano mais semelhante também são registrados para uso posterior.

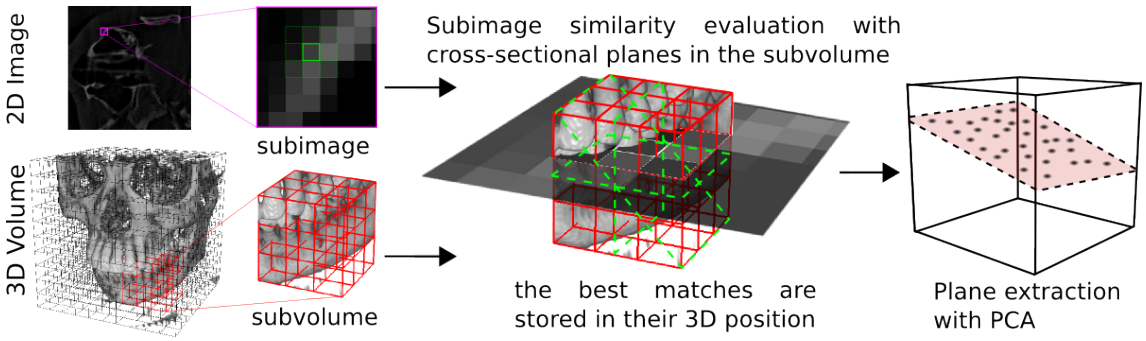
A.5.3 Plano Adequado

Após os cálculos de similaridade precisamos extrair o melhor plano que engloba todos os voxels armazenados na lista das maiores similaridades L^k. O método baseia-se no plano adequado 3D de primeira ordem proposto por BERKMANN; CAELLI (1994). O problema de determinar a equação de plano de L^k é uma estimativa por mínimos quadrados para o plano adequado. O plano é representado como um ponto x e um vector normal, $\vec{n}$. O centróid de L^k é definido como:$\bar{p} = \frac{1}{k}.\sum_{i=1}^{k} p_i$.

A solução para $\vec{n}$ é determinado por meio da análise de componentes principais de L^k. Os auto valores e auto vetores da matriz de covariância C, expressa em eq. A.4, são então calculados.

$$C = \frac{1}{k}\sum_{i=1}^{k}(p_i - \bar{p}).(p_i - \bar{p})^T, C.\vec{v_j} = \lambda_j.\vec{v_j}, j \in \{0, 1, 2\} \tag{A.4}$$

Figure A.5 - Para cada pixel e voxel uma sub-imagem e um sub-voluem quadrados são extraídos. Nove planos de imagem com diferentes ângulos são gerados apartir do sub-volume e sua similaridade é comparada com a sub-imagem. Para cada voxel, o sub-plano mais similar entre as nove orientações é armazenado. Após todas as comparações, PCA é computado para extrair o plano adequado a partir da lista de voxels L^k.



Fonte: Imagem compilada pelo autor.

A matriz C é simétrica e positiva semi-definida e seus autovalores (λ_j) são números reais. O auto vetor $\vec{v_j}$ com o menor autovalor é, por conseguinte, a aproximação da normal $\vec{n} = (n_x, n_y, n_z)$ ou $-\vec{n}$. O componente d da equação do plano é encontrado calculando o produto escalar de $\vec{n}$ e o $\bar{p}$ of L^k.

A.5.4 Otimizações do algoritmo

Duas estratégias de otimização foram aplicadas no algoritmo para evitar cálculos desnecessários. A primeira estratégia baseia-se no conhecimento sobre a aquisição de imagens médicas. As áreas circundantes do corpo são representadas como preto nos pixels na imagem. Essas porções não contêm informações relevantes e são excluídos dos cálculos. A segunda estratégia é redução da resolução dos dados volumétricos. A resolução original do volume é carregada na memória, um fator de redução da resolução é estipulado e o algoritmo executa os cálculos apenas nos voxels amostrados. Nosso algoritmo de busca de imagens em volumes ainda usa voxels vizinhos presentes na resolução original para construir sub-volumes.

A.5.5 Visualização da correlação entre imagem e volume

A técnica desenvolvida trabalha em conjunto com nosso sistema de visualização volumétrica de dados médicos. O objetivo é fatiar os dados volumétricos em uma secção transversal no lugar exato onde o plano arbitrário foi localizado.

Neste momento temos todos os elementos necessários para visualizar o plano da imagem resultante, que são: a imagem arbitrária de entrada, o volume V 3D e a equação do plano p que representa a posição e orientação da imagem arbitrária dentro do volume. A visualização de volume em tempo real é realizada pelo rendering direto de texturas desenvolvido por ENGEL et al. (2004). Mapeamento 3D de textura permite usar interpolação trilinear apoiado pelo hardware gráfico e fornece uma taxa de amostragem consistente. O volume é segmentado exatamente onde o plano calculado atravessa. Podemos descartar os voxels superiores ao plano ou inferiores com $V_k \cdot pe > 0$ e $V_k \cdot pe < 0$, respectivamente. Em seguida, a imagem original de entrada pode ser exibida naquele local, fundindo o 3D com os dados 2D.

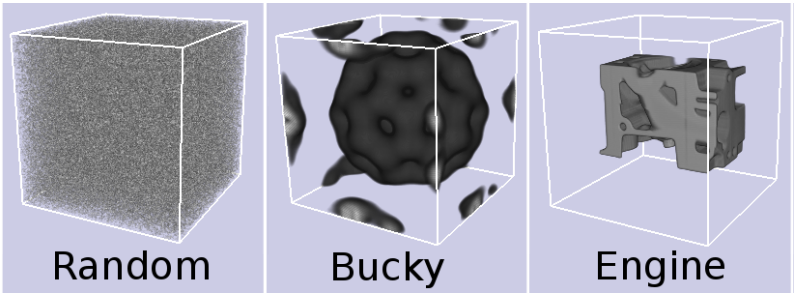
A.6 Experimentos e resultados

Montamos um protocolo de teste para avaliar e comparar o desempenho e eficácia de nossa técnica entre os diferentes conjuntos de dados, resoluções e fatores de amostragem. Nesta seção, nós detalhamos as métricas utilizadas. Nossa preocupação é que o algoritmo execute o mais rápido e entregue os melhores resultados possíveis em termos de encontrar o melhor plano que representa a imagem arbitrária.

A.6.1 Configuração do experimento

Os conjuntos de dados avaliados são apresentados na Fig. A.6. Bucky, Engine e Random são três datasets de uso geral. Random é composto com valores de intensidade randômicos e é dificilmente optimizada devido à ausência de regiões com voxels com valor de intensidade quase nulo. Thorax e Slull são datasets de TC de um ser humano do sexo masculino. Thorax foi convertido a partir do protocolo DICOM para dados brutos, excluindo todos os metadados existentes. Bucky, Engine e Skull podem ser obtidas on-line (FAU-3D, 2014). Todos os conjuntos de dados são armazenados em um canal de cor de 8 bits que representa valores de luminância. As resoluções do conjunto de dados e fatores de exemplos usados em todos os testes são listadas abaixo:

Figure A.6 - Representação volumétrica dos dados utilizados no experimento.



Fonte: Imagem compilada pelo autor.

- **Resoluções**: 128x128x128 (original), 64x64x64 and 32x32x32;
- **Factores de escala**: 1x (original), 2x and 4x para a imagem XY, para volume XY e volume Z.

Definimos uma notação para uma consulta rápida: quando dizemos Bucky 128_142, isso significa que a imagem de entrada tem 128x128 pixels e o volume é reduzido de 128x128x128 para 32x32x64, ou seja, cada uma das 64 fatias tem resolução de 32x32 pixels. Essa redução é realizada em tempo de execução. Note que todos os datasets originais tem resolução de 128x128x128 exceto Bucky que tem 32x32x32 e foi super amostrado para equalizar as comparações. O objetivo do nosso algoritmo é encontrar o melhor plano em um volume que represente da melhor forma uma fatia 2D dada como entrada. No entanto, para fins de teste, nós executamos os testes com um conjunto de 32 imagens de entrada construídas a partir de um plano axial e 31 planos arbitrários em ângulos diferentes. As imagens amostradas em planos arbitrários são os mesmos para todos os testes para permitir a comparação entre os diferentes datasets.

Nós armazenamos as equações do plano de cada plano arbitrário utilizado como entrada nos testes e os usamos como referência para avaliação de precisão. O erro entre a entrada p

e o plano q encontrado é medido com duas variáveis: a média da distância euclidiana entre os vértices correspondentes em cada plano, Eq. A.5 e o ângulo entre as normais de p e q, Eq. A.6. Quanto mais os resultados convergem para zero, mais correta é a correlação entre imagem e volume. O tempo de execução do algoritmo nos testes é medido em segundos e o resultado é a média de execução para a correlação de 32 imagens arbitrárias.

$$d(p,q) = \frac{1}{N}\sum_{i=1}^{N}(p_i - q_i) \tag{A.5}$$

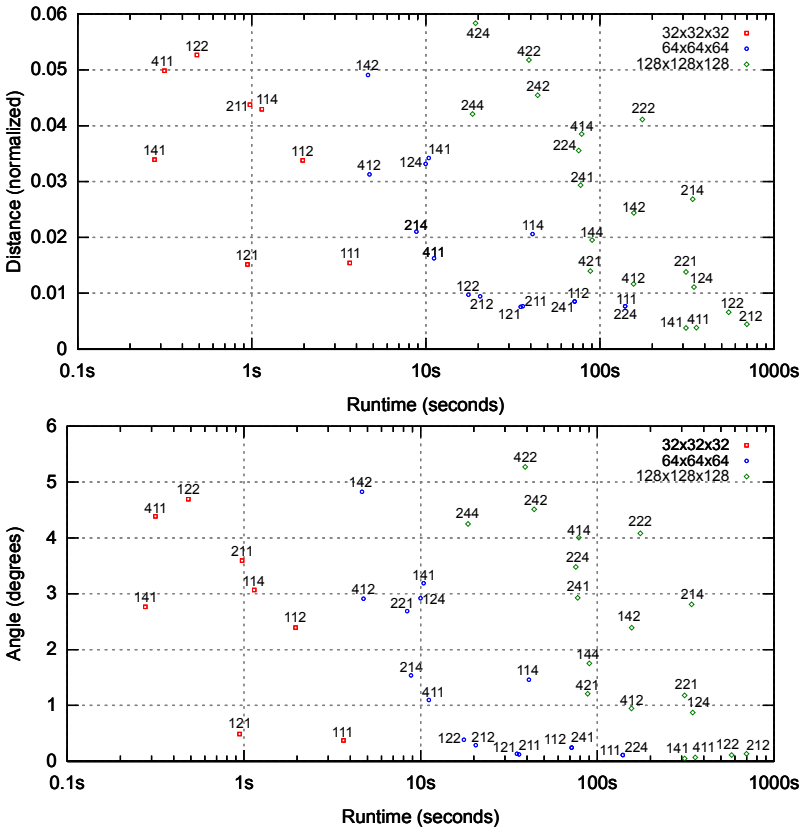
$$cos\alpha = \frac{\hat{n_p} \cdot \hat{n_q}}{|\hat{n_p}|\,|\hat{n_q}|} \tag{A.6}$$

Todos os testes foram realizados utilizando um computador com a seguinte configuração: processador Intel Core i7-2600@3.4GHz com 8GB de RAM. O algoritmo foi desenvolvido em C++, compilado com o GCC 4.8.2. O algoritmo executa em paralelo utilizando todos os núcleos do processador.

A.6.2 Avaliação no equilíbrio entre precisão e desempenho

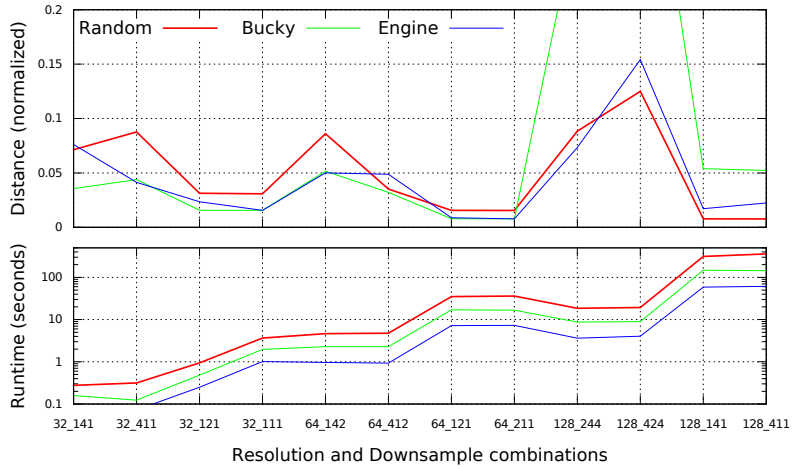
Devido ao elevado número de dimensões de dados que são avaliadas, nosso algoritmo tem uma complexidade computacional $O(N^7)$. No entanto, nossa hipótese de que é possível obter resultados consistentes trabalhando com um conjunto muito reduzido de amostras a partir dos dados originais de entrada. Além disso, acreditamos que há uma combinação ideal de resolução e fator de amostragem que é suficientemente preciso e rápido. Como alguns dados são mais homogêneos do que outros, e precisão de execução, também pode ser dependente da natureza do conjunto de dados. O experimento descrito a seguir testa algumas dessas hipóteses. Em primeiro lugar, para descobrir a melhor combinação amostragem, nós realizamos um teste com o conjunto de dados Random. Os gráficos da Figura A.7 mostram todas as combinações de resolução e amostragem que se encaixam no intervalo de tempo. Percebemos que os erros de distância e ângulo estão correlacionados, permitindo-nos concentrar em distância e sugerindo uma possível unificação métrica a ser investigada em trabalhos futuros. A resolução 128, sem amostragem é muito precisa, mas executa em tempo inviável (1200s, fora do alcance do gráfico). Resoluções menores e amostragens mais grosseiras, no entanto, ainda proporcionam uma elevada precisão. Como exemplos, localizar pontos de 32_121, 64_122, 128_141 no gráfico. Para um conjunto de dados 200x200x200mm, o erro normalizado 0,016 obtido para 32_121 é equivalente a 3,2 milímetros, e levou apenas um segundo para ser encontrado. Em seguida, para descobrir como as diferenças de conjuntos de dados interferem na precisão, foram selecionados as duas combinações de resolução e amostragem mais precisas e as duas mais rápidas para análise. Estas combinações foram avaliadas para todos os datasets e os resultados são mostrados na Fig. A.8 As combinações são classificadas a partir do menor para o maior tempo de execução em cada resolução. Como esperado, o conjunto de dados Random é mais lento. Distâncias entre planos (erros) são consistentes entre os conjuntos de dados, a não ser, em parte, para Bucky que apresentou um erro maior em 128_424. Nós acreditamos que isso ocorreu devido a falta de informação, pois Bucky foi super amostrado.

Figure A.7 - Acima: Erro entre distância e tempo de execução em escala logaritmica. Abaixo: Erro entre ângulo e tempo de execução em escala logaritmica.



Fonte: Imagem compilada pelo autor.

Figure A.8 - Tempo de execução para as 12 combinações de resolução e fator de escala. O erro e o tempo de execução são consistentes para todos os datasets, com o Skull sendo o mais rápido. Note a ótima relação tempo de execução e precisão atingida pela combinação 32_121.



Fonte: Imagem compilada pelo autor.

A.7 Conclusões

Nesta dissertação investigamos meios para estabelecer correlação e fusão de imagens 2D e volumes 3D de dados médicos de mesma modalidade de aquisição baseados em apenas intensidades de pixel/voxel. O trabalho foi dividido em dois ciclos de desenvolvimento. Em primeiro lugar, abordamos o problema da correspondência 2D e 3D de dados alinhados com os planos anatômicos. Nós aperfeiçoamos a busca de correlação integrando duas medidas de similaridade em uma técnica. Esta abordagem de dois passos provou ter um bom equilíbrio entre o desempenho e eficácia para correlacionar dados médicos de mesma modalidade, mesmo com protocolos de aquisição distintos, especialmente quando paralelizado em GPU. No segundo ciclo de desenvolvimento, utilizamos os conhecimentos adquiridos na investigação anterior e expandimos a pesquisa de semelhança para compreender planos de imagem arbitrariamente alinhados. Propomos uma abordagem para reduzir o conjunto das comparações e manter precisão na correlação entre imagem e volume. A técnica funciona localmente avaliando semelhança em sub-imagens e sub-volumes subamostrados. Uma nuvem de pontos tridimensional é gerada com apenas os mais altas semelhanças encontradas. Ao realizar a análise de componentes principais, podemos extrair o melhor plano que compreende todos os pontos da nuvem. Nós minimizado o custo computacional esperado trabalhando em um conjunto muito reduzido de amostras, com uma combinação de imagem e volume subescalados e subamostrados. Outra contribuição deste trabalho é o conjunto de experimentos realizados com os dados experimentais e médicos. Eles revelaram que há uma combinação ótima de redução de escala e redução da resolução que podem ser aplicadas para alcançar um equilíbrio entre precisão e desempenho. O ajuste dos planos encontrados é bom o suficiente para corresponder corretamente dezenas de planos orientados arbitrariamente com precisão maior do que os encontrados na literatura. O trabalho desenvolvido tem aplicações potenciais na área médica. Acreditamos que, no futuro, a nossa técnica poderia ser integrada no contexto médico e contribuir na análise e procedimentos operatórios com uma visualização multimodal. No entanto, devido à sua formulação geral, a técnica pode ser estendida e modificada para beneficiar outros campos e outras estruturas de dados.

A.7.1 Trabalhos futuros

Generalizar os algoritmos para lidar com dados médicos de diferentes fontes de aquisição. Com base nos trabalhos relacionados, consideramos a mudança dos critérios de similaridade para uma abordagem da teoria da informação, como a informação mútua e simulação de uma modalidade para outra, realizada por [1] o que nos permite continuar a utilizar abordagens de intensidade pixel/voxel. Para atingir um melhor desempenho, otimizações mais sofisticadas têm de ser abordadas no futuro. Algumas estratégias são: Multi-resolução adaptativa, como método de Powell, gradiente descendente ou quasi-Newton e espaço de subdivisão por KD-tree. Além disso, a técnica pode ser acelerada com paralelização em GPU. Pretendemos explorar em profundidade uma possível métrica de erro para avaliar similaridade entre planos orientados arbitrariamente unificando posição e orientação em uma única equação.