HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation
Department of Information and Computer Science

Tuomas Mäkelä

DENTAL X-RAY IMAGE STITCHING ALGORITHM

Master's thesis for the degree Master of Science (Technology)
Submitted for inspection in Espoo 2.10.2009

Thesis supervisor:	Prof. Erkki Oja
Thesis instructor:	M.Sc.(Tech.) Martti Kalke

HELSINKI UNIVERSITY OF TECHNOLOGY

Author: Tuomas Mäkelä

Title: Dental x-ray image stitching algorithm

Date: 2.10.2009 Language: English Number of pages: 7+64

Faculty: Faculty of Electronics, Communications and Automation

Professorship: Computer and Information Science Code: T-61

Supervisor: Prof. Erkki Oja

Instructor: M.Sc.(Tech.) Martti Kalke

Size restriction of the reading device of the phosphor imaging plate used in intraoral radiography prevented occlusion imaging. The solution was to use two overlapping plates to gain partially same imaging into both images. Images could be stitched into one, larger image by software.

The solution for the stitching algorithm has been presented in this thesis. It is based on the mutual information method and the adjustment of the images acquired by the system for suitable form prior to the stitching.

Functionality of the software was tested by a set of image pairs. Due to the overlapping phosphor plates and the properties of x-radiation, one of the images acquired has lesser contrast and weaker signal-to-noise ratio. Around the teeth the image registration was successful. Information on the palate area is less distinguishable and the registration was less accurate, but nonetheless, decent for the application. In the beginning of the thesis, there is a short review on x-radiography and image registration.

Keywords: Image registration, stitching, mutual information, dental, x-ray imaging

TEKNILLINEN KORKEAKOULU

Tekijä: Tuomas Mäkelä

Työn nimi: Limittäiset hammasröntgenkuvat yhdistävä algoritmi

Päivämäärä: 2.10.2009 Kieli: Englanti Sivumäärä: 7+64

Tiedekunta: Elektroniikan, tietoliikenteen ja automaation tiedekunta

Professuuri: Informaatiotekniikka Koodi: T-61

Valvoja: Prof. Erkki Oja

Ohjaaja: DI Martti Kalke

Suunsisäisessä röntgenkuvauksessa käytettävän fosforikuvalevyn lukulaitteen kokorajoitus esti hampaiden okluusiokuvauksen. Ratkaisu oli käyttää kahta kuvalevyä limittäin, jolloin niille tallentui osittain sama näkymä hampaista. Kuvat voitaisiin ohjelmallisesti yhdistää limittäisten osien avulla yhdeksi, suuremmaksi kuvaksi.

Tässä diplomityössä on esitetty kuvat yhdistävän algoritmin ratkaisu. Se perustuu keskinäisinformaatioon sekä laitteiston tuottamien kuvien muokkaamiseen tarvittavaan muotoon ennen niiden yhdistämistä.

Ohjelmisto testattiin testikuvapareilla. Johtuen kuvalevyjen päällekkäisyydestä ja röntgensäteilyn ominaisuuksista, toisella kuvista on heikompi kontrasti ja signaali-kohina-suhde. Hampaiden kohdalla kuvien kohdistus onnistui hyvin. Kitalaen alueella selkeästi erottuvaa informaatiota on vähemmän ja kohdistus oli hieman epätarkempi, joskin riittävä kyseiseen sovellukseen. Työssä on myös lyhyt katsaus röntgenkuvaukseen ja kuvien kohdistamiseen.

Avainsanat: Kuvien kohdistus, kuvien yhdistäminen, keskinäisinformaatio, hammaslääketiede, röntgenkuvaus

Preface

This thesis was carried out at PaloDEx Group Oy, an internationally operating designer and manufacturer of dental imaging equipment.

I would like to thank the instructor of the thesis, M.Sc.(tech.) Martti Kalke from PaloDEx, for helping me out with this very exciting work. Martti could always be trusted to have a suggestion or two when ever asked. Also, greetings for the rest of the PaloDEx crew, for your interest and help on my work.

Finally, I would very much like to thank my parents, Arto and Leena Mäkelä for their unselfish support and everlasting trust on me. Never did they question my choices, and my only wish is they can be proud of me.

Otaniemi, 2.10.2009

Tuomas T.J. Mäkelä

Contents

Symbols and Abbreviations

Symbols

A and B	Image matrices A and B
H	(Shannon) entropy
i, j	Pixel coordinates, row i, column j
n	Intensity value of a pixel
N	Number of possible intensity values

Abbreviations

CT	Computed Tomography
IP	Imaging Plate
MRI	Magnetic Resonance Imaging
PSP	Photo-Stimulated-Phosphor
ROI	Region of Interest

1 Introduction

X-ray imaging is a widely used form of medical imaging. It is based on x-rays (or Röntgen rays), which are a form of electromagnetic radiation. In the spectrum of electromagnetic radiation, they have higher frequency than ultraviolet rays but (usually) lower than gamma rays. X-rays are capable of passing through human tissue more or less unaltered. Some tissue—notably bones—absorb more radiation than other—say, softer tissue—while air in the mouth or in bodily cavities has practically no effect at all on the radiation. Therefore, a film placed on the other side of the patient in view of the x-ray source will record an image of the bones and other tissue according to the transmission coefficients of various tissues.

Nowadays the film is usually replaced by a digital sensor or a phosphor imaging plate, which both have the functionality of the film, but the data on them can be erased. Therefore they can be used repeatedly unlike the disposable film. Other than that, they are ready in digital form. A digital image has many advantages over traditional film. It is effortless to distribute and image manipulation processes (gray value correction, sharpening etc.) are easier to implement. A film, an imaging plate and a digital sensor are all fundamentally based on the same phenomenon and are therefore interchangeable among each other.

In intra-oral imaging—i.e. imaging within the mouth—film, imaging plate or digital sensor is placed inside the patient's mouth. X-rays are irradiated outside the mouth and the "shadows" of the teeth and other tissues are cast on the recording medium, which stores the dental x-ray image.

Occlusion imaging is used when one wants to take a picture of either all—or significant part of—upper or lower teeth together. A recording medium is placed horizontally between the patient's upper and lower teeth, and x-rays are beamed either from above the nose or below the jaw, for upper or lower teeth, respectively.

Occlusion imaging obviously requires rather large recording medium compared to the other intraoral imaging views in order to be able to cover the whole area of the teeth.

PaloDEx Group Oy is an internationally operating designer and manufacturer of dental imaging equipment. They use—among other dental imaging methods—several technologies in intraoral imaging. In one of them, the photo-stumulated-phosphor (PSP) imaging plate (IP) is used to record the x-ray image. The size of the imaging plate is restricted by the device used for reading the data, and is not large enough to be used in occlusion imaging. To bypass this, two images are taken and stitched together in order to gain a larger image.

A certain container is used to house two imaging plates and the occlusion image can be taken with only one exposure. The plates are slightly overlapping, so one part of the image is recorded into both plates.

The topic of this thesis is to design and implement an algorithm to integrate those two images. The algorithm stitches the images on account of the information of the overlapping area of the plates, for that area has the imaging of the same area of the mouth.

Imaging plate absorbs some of the x-ray radiation when the image is formed while the rest of the radiation passes trough. Since the two imaging plates are overlapping, the plate behind the other receives less radiation in the overlapping area during the exposure. For this reason, the intensity levels in that area of the image are most likely different from those of the imaging plate which was on top. Nonetheless, the overlapping area of both plates presents the same area of the mouth. Stitching of the images is based on the information of those areas of the images, so the chosen method must not rely on absolute intensity values.

The container for the imaging plates is designed to hold the plates always in the same direction. However, it is allowed for both plates to move on an

area about 1 mm larger in height and width than the size of the IP. Thus, if we consider the location of one of the plates being fixed, the other one may move ±1 mm from the given point in both vertical and horizontal direction. This may also cause small rotation between the plates if different ends of the plates move to different directions. The algorithm must manage such movements and rotations.

The algorithm is designed and tested with Matlab and is later ported into C++ language to be integrated with the rest of the device driver software.

The 2 Chapter 2 consists of two sections. In section 2.1 is presented a short review on the history of the x-ray imaging, followed by a small review of modern dental imaging. Section 2.2 is a literary research of the field of image registration and image stitching.

Chapter 3, 3, reveals testing arrangements. Chapter 4 is about the algorithm itself, and due to it's size it has been divided into several smaller sections each focusing on a specific step in the algorithm process.

Results of various tests are presented in Chapter 5, followed by discussion and conclusions in Chapter 6.

2 Background

2.1 X-ray Imaging

2.1.1 History and Technology of X-ray Imaging

X-rays (or roentgen rays) were found by a German physics professor, Wilhelm Conrad Röntgen, on November 8, 1895. Röntgen, Director of the Physical Institute of the University of Würzburg, was interested in work of Hertz and Lenard and many others on electrical discharges in vacuum tubes.[4] He set up his own apparatus and followed and repeated the work of predecessors, namely the work done by Hertz and Lenard.

They had been carrying out experiments with Hittorf-Crookes tube, one kind of vacuum tube. The Hittorf-Crookes tube is a partially evacuated glass envelope with two electrodes separated by a distance of a few centimeters. When a potential difference of few thousand volts is connected between the electrodes, the partially ionized, rarefied gas in the tube is accelerated by the electric field. Due to the high voltage, the ions accelerate and hit the cathode (negative electrode) with such energy, that they manage to release electrons from the surface of the cathode.

As electrically charged particles, the electrons are accelerated in the electric field away from the cathode and towards the anode (positive electrode). Should the voltage between the electrodes be huge enough, some of the accelerated electrons might overshoot, or go through the anode and strike the glass wall of the tube, emitting x-rays, though this wasn't known at the time.

X-rays are part of the same electromagnetic radiation as visible light and radio waves, ranging from frequencies of 30×10^{15} Hz to 30×10^{18} Hz. In the spectrum of the electromagnetic radiation they are between lower frequency ultraviolet and higher frequency gamma-rays, although sometimes the frequencies of x-rays and gamma-rays overlap and the only difference between the two is the method the rays were generated. Gamma-rays are formed by transi-

tions within atomic nuclei or matter-antimatter annihilation, while x-rays are generated when high-speed electrons are decelerated in matter.

Electrons decelerating in matter was what happened in the Röntgen's tube when the overshot electrons hit the glass, and when x-rays were emitted. [16] While carrying out his experiments with cathode rays, Röntgen made a discovery of fluorescence of a paper screen covered with barium platinocyanide crystals. The paper screens were used to detect whether there were cathode rays present or not. To use these papers, a special kind of tube with aluminum window was needed to pass the cathode rays outside the tube. This time, however, there was fluorescence even when working with a glass tube which shouldn't pass cathode rays. Röntgen realized he had found a new kind of radiation, and, unaware of the true nature of the radiation, called it the "x-ray".

Röntgen quickly experienced more with the rays and made a proceeding on them. The medical potential was understood soon and the first skeletal radiographs of a living hand were taken less than two months after the discovery of the radiation.

A modern dental x-ray tube is ultimately similar to the tube Röntgen used on his experiments. Figure 1 presents the tube. Electrons are emitted from filament that is heated by electric current. Voltage difference between a cathode and an anode forces the electrons to travel to the anode, where a tungsten target is located. X-rays are emitted when electrons decelerate in the target.

X-ray images were first recorded by a film. One of the properties of the film is that, the more radiation there is, the darker the image becomes. Therefore softer tissue in x-ray images show darker than bones, as more radiation passes through it. X-ray images are still shown in the same manner (as negative images), even if recorded by some other recording medium.

X-radiation is ionizing radiation, which means it has energy so high it

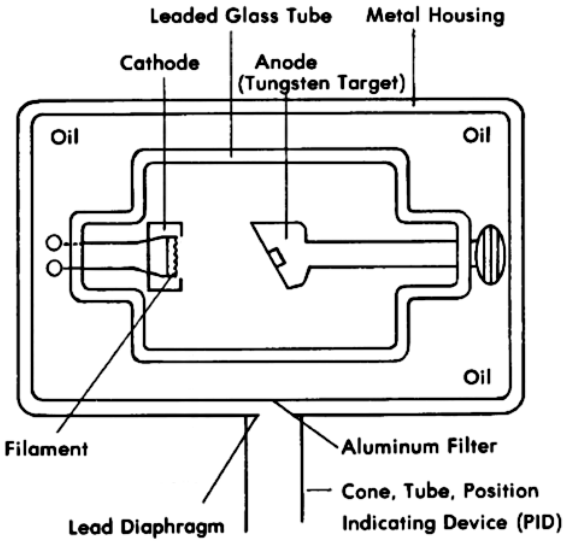


Figure 1: Modern deltal x-ray tube. Potential difference forces the electrons emitted by *filament* to travel from the *cathode* to the *anode*. Electrons decelerate in the *tungsten target*, which emits x-rays. The angle of target guides the radiation downward. *Aluminum filter* removes low-energy beam which is unwanted, whereas the *lead diaphragm* allows the radiation exit only to the desired direction, out of the *tube*. Figure from [6].

is capable of detaching an electron from the electron shell of the atom. If the quantity of the radiation is great enough, it can has undesired effect on chemistry of the cell. Higher amounts of radiation will lead to death of the cell.

Ionization of the DNA might lead to mutation. This kind of damage is cumulative and therefore people whose work involve x-rays are monitored for dosage they get from their work. Radiation can be reduced by using lead walls to block x-ray from escaping to unwanted areas as lead is known to efficiently attenuate radiation. Also, maintaining a distance from x-ray source helps, as on the spherical surface the radiation decreases to one fourth when ever the

distance doubles.

2.1.2 Modern Dental Imaging

Modern dental radiography is divided into three fields. *Intraoral radiography* was the first dental imaging method. In intraoral radiography, a certain recording medium is placed inside the patient's mouth (hence the name *intraoral*). X-ray tube—the source of x-rays—is located outside the head so that the radiation passes through the object and hits the recording medium. Recording medium might be either (now almost obsolete) film or in more modern devices either a reusable phosphor imaging plate or an image sensor. Recording medium of various sizes are used. The size of the medium is a trade-off between the area of the imaging, and the comfort of the patient due to the limited space in the mouth.

There are several *views* used in intraoral radiography. They are used for different needs but are fundamentally similar. *Periapical view* means the recording medium is located in the mouth so that it records an image of whole tooth including the crown and root. This view might be used to determine the need for endodontic therapy, or to look for aching tooth.

In *bitewing view* the recording medium is placed so that it records the image of the crowns of the teeth, which are usually the region of interest. One exposure records evenly the crowns of both maxillary (upper) and mandibular (bottom) teeth.

Lastly, *occlusal view* is used to get an image either from all maxillary or all mandibular teeth. The recording medium is placed between patient's upper and lower teeth. For upper teeth the x-ray tube is located above the nose, and for bottom teeth it is located below the jaw. The recording medium for occlusal view is larger that the one used for periapical or bitewing views.

One of the first mentions of tomography imaging is in the patent from the year 1922, owned by M. Bocagen.[11] In tomography imaging the recording

medium is located outside the patients mouth and is hence in the group of *extraoral imaging* methods in dental imaging. The x-ray tube and recording medium are at the opposite sides of the object to be scanned. The tube and film both rotate—or move otherwise, e.g. on linear or spiral path—in opposite directions around a fixed point, which determines the location of the imaging layer. (Figure 2.) Imaging layer is a predetermined plane which gets recorded sharply in the tomography. In intraoral imaging, all the matter—not only the desired one—between the tube and film ends up to the image. For tomography imaging the same holds, however, because of the movement of both x-ray tube and recording medium, only one layer is scanned sharply. Tissue far from this desired layer will leave a widely spread, faint haze to the image, which will be seen as noise in the final image.

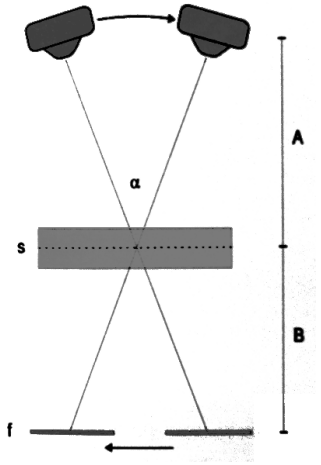


Figure 2: Sketch of the movement the x-ray tube and recording medium participate in tomography imaging. Tube and medium (denoted by f) move to opposite directions around a center point which will determine the location of the imaging layer (s), the plane of tissue to be shown sharp in final image. Image from [11].

During the years 1954–1960, Y. V. Paatero evolved the idea and finally,

after few stages of development, introduced an *orthopantomography* where the focus point follows the teeth during the scan with rather narrow beam. (Figure 3.) Narrowed beam means only a small vertical slice of the film will be exposed at the time, thus recording an image of the teeth at the current focus point only. As the focus point slowly moves, the film slides in the sledge and the imaging of the new focus point gets recorded to the newly revealed part of the film.

After the whole round, a panoramic image of the teeth is recorded. Other that teeth, both chin and sinus are also visible in the image. (See image 4.)

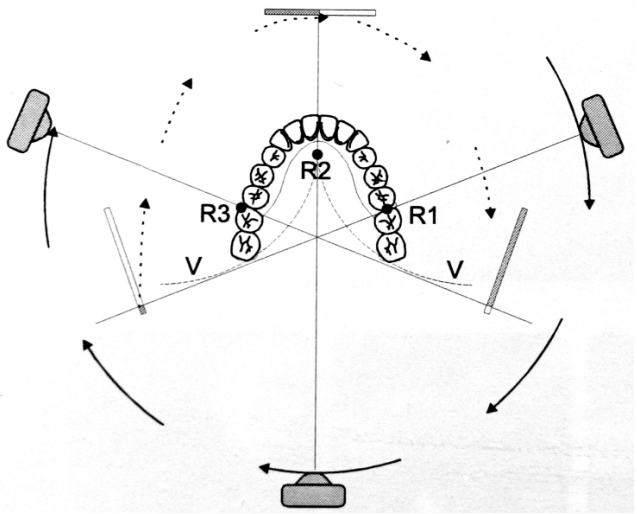


Figure 3: A sketch of the rotation of the x-ray tube and the film casing in the orthopantomography device. The focus point follows the presumed curve of human teeth thus recording a sharp image of all teeth. Image courtesy of [11].

On more modern, digital orthopantomography device, the film is replaced by a digital sensor. The film size on non-digital devices varies between different apparatus but might be e.g 15×30 cm[3]. Digital sensor, however, might be significantly smaller. Like in the case of film device, the beam is narrow and only a small vertical slice of teeth and skull is exposured at the time. Recorded

data is read from the sensor at certain intervals and the sensor is reset to zero. This is equivalent to sliding a film in a sledge so that the beam next hits an unexposured part of the recording medium, and there is no need to slide the sensor as is in the case of a film device. Panoramic image of teeth and skull is composed of these narrow slices.

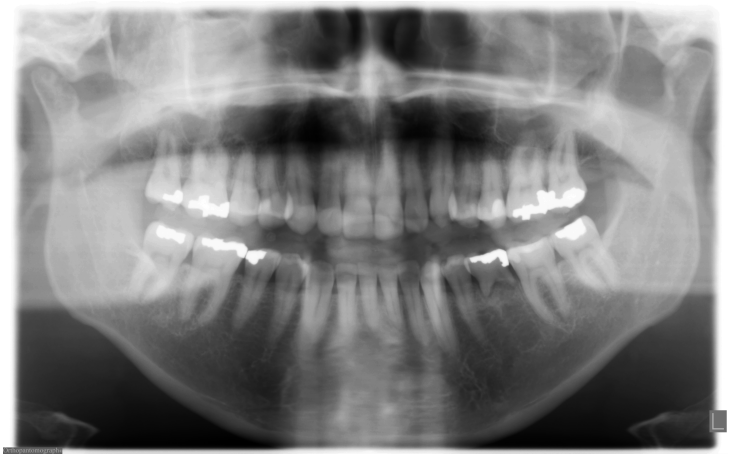


Figure 4: An example of panoramic image taken with digital orthopantomography device. Besides teeth, also jawbone and sinus are visible. Image courtesy of PaloDEx Group Oy.

The third and most modern way to utilize x-rays in dental imaging is *computed tomography* or CT. For this application, frequent exposures are made from different directions of one imaging layer at the time. The data is run through a *back projection* process with computer which tries to conclude what sort of tissue there is and how it is distributed along the object.

The result is either a series of slices orthogonal to one axis or true three-dimensional image of the scanned volume.

From the 3D-data collected by CT device, it is also possible to generate images similar to orthopantomography and intraoral imaging.

2.2 Image Registration

Image registration is a process where two or more images are transformed in some geometrical manner so that the coordinates of the images become parallel and the images can be matched.

In the following two sections, a short description is given in both general image registration and medical image registration.

2.2.1 In General

Image registration is needed when two (or more) images are to be merged into one. In the process the source images are transformed so that their coordinates match with each other. Transformations may be as simple as shifting, rotating or scaling, or they may be more complex—e.g. perspective, lens distortion or other kind of corrections. Images are of the same scene, but conditions of exposure may vary.

B. Zitová et al.[19] divided image registration applications into four groups according to the image acquisition manner. The following division gives some examples of why image registration is used.

Different viewpoints The first group consists of applications which acquire their source images from different viewpoints, i.e. the sensors are at different locations looking at the same thing. These applications usually try to *stitch* partially overlapping images in order to gain a wider view. The algorithm and application of this thesis also falls into this class as the focus is to stitch two adjacent x-ray images to gain a larger one.

In this group fall also applications which try to make a 3D model out of several 2D images. Humans (and presumably other animals with forward directed eyes) form 3D models of the objects they see with their two 2D eyes.

Different times For the second group of applications, images have been taken from the same view point but at different times. Aim of these applications is to find and evaluate changes in the scene over the time or between different conditions. In medical imaging some applications might e.g monitor healing or tumor evolution.

Different sensors Third group obtains images with different kind of sensors. In literature this sort of image registration is also referred as *multimodal image registration*. Multimodal image registration is used in medical imaging to get combinations of images from different kind of sensors, like magnetic resonance imaging (MRI), ultrasound or x-ray, for instance.

Although the images obtained for the application of this thesis are not from different kind of sensors, overlapping imaging plates have an influence on each other some what similar.

Scene to model registration The final group of applications does scene to model registration. They try to register recorded scene with preformulated model. The aim is to localize the model in acquired image or compare them, e.g. to see if some anatomical structure of a patient differ from normal, or to register satellite images with maps.

Regardless of the division of the applications, Barbara Zitová et al. found in their survey that majority of the methods consists of four steps: First, in *feature detection*, distinctive structures are detected from the images. After that, in *feature matching* step the found features of two images are matched before mapping functions try to do *transform model estimation*. When suitable estimation has been done, images are *resampled and transformed* to match each other. Steps are walked through in the following.

Feature detection Methods in the first step of the registration can be divided into two main groups. In *feature-based methods* significant struc-

tures are searched and detected. Different features are searched for different kind of images and applications. *Region features* are larger areas of some constant variable—intensity, color, etc.—in the image which are separated from each other with high contrast boundaries.

Line feature can be any sort of line, or segment of one. They are searched by different kind of edge detection methods.

Points might be also considered as detectable features. They might be intersections of lines, corners or they might be searched with some derivate based method.

These methods are usable—and often recommended—if there are enough distinctive and detectable objects in the images, which is usually the case in general photography. However, there are images from certain fields which generally lack such details and *area-based methods* are used instead. In area-based methods no features are searched, but rather, every single pixel in the image is considered to be a *feature*, and the whole image is sent to the next step of the algorithm.

Feature matching Once the definite features from the source images has been mapped into feature space, they are turned over and over in order to find a match. The division between feature-based and area-based methods holds in the *feature matching* step also.

For matching extracted features, there are number of methods to choose from. Methods might be based on e.g. spacial relations, where the information about the distances between found features are exploited, or some invariant descriptors of the features themselves are compared and the best matches are taken to present the same object in the scene.

Area-based methods are used if no features were extracted. These include e.g. correlation based methods where the intensity levels of the images are compared. If intensity values of the pixels are not expected to be the

same, the mutual information based methods might be used. Mutual information is a measure of statistical dependency between the images.

Area-based methods are usually time consuming compared to the feature-based method. Also, portions of images containing smooth area without any prominent details are at high risk to be misregistered with some other smooth areas.

Transform model estimation After the features are matched, a mapping function to transform the images is constructed. Mapping functions are based on the assumed geometric deformation of sensed images. Functions can be divided into two categories according to the amount of feature points they use. Global models use all available features and usually preserve the angles and curvatures of the images.

Local mapping models, however, divide the image into smaller pieces and form functions for those smaller parts independently thus allowing more complex transforms.

Image resampling and transform After the transform functions have been formed, the image or both are transformed and thus the images are registered. There are also numerous methods for resampling the data in order to maintain it's visual quality, however, the bilinear interpolation is most commonly used as it offers probably the best trade-off between accuracy and computational load.

2.2.2 In Medical Imaging

Registration of medical images usually involve multiple modalities, i.e. images acquired with different kind of sensors which are sensitive to different kind of tissue. Typical image sources are x-ray, magnetic resonance imaging (MRI), ultrasound (US) or some nuclear medicine methods (SPECT, PET). Even the scope of the dimensionality of the registration is wide covering all possibilities

of 2D/2D, 2D/3D and 3D/3D registration, with or without time as an extra dimension.[8]

Many different methods for registering medical images have been introduced, even some that are not based on images themselves but rather on calibrating coordinate systems by some other means. Some methods relay on searching for mounted markers which will guide the registration.

When it comes to registering images on account of the content of the image itself, one must remember that medical images, as a rule, lack salient objects, which would be searched by numerous feature extraction methods.[19] Therefore, area-based methods are used instead.

B. Zitová *et al.* introduced three area based methods in their survey. *Correlation-like methods* are based on cross correlation and its modifications. Cross correlation (1) is a measure of similarity where *sliding dot product* of two images is calculated.

In *digital image* each pixel is associated with one or more values which defines the color and intensity of the pixel. For the sake of clarity, and the fact that x-ray images are grayscale, the images in the following are taken to be grayscale, i.e. there is only one value, intensity, associated for each pixel.

Let's take an image matrix A, where each pixel can be represented by $A_{i,j}$, where i and j are the coordinates, i being the row and j the column of the pixel. The cross correlation between image matrices A and B can be calculated by the following equation.

$$CC = \frac{\sum_i \sum_j (A_{i,j} - mean(A))(B_{i,j} - mean(B))}{\sqrt{\sum_i \sum_j (A_{i,j} - mean(A))^2 \sum_i \sum_j (B_{i,j} - mean(B))^2}} \tag{1}$$

Correlation based methods have been around for a long time and they are well known. One advantage is easy hardware implementation, which makes them useful for real-time applications. On the other hand, they may give poor results on noisy images and usually require intensity levels to be similar by some linear transformation, i.e. multimodal images may give poor results.

In correlation-like methods. like in all area based methods, source images are transformed before similarity is calculated. The value is then compared to values of earlier transformations, and the best transformation is chosen.

Fourier methods work in the frequency domain. The Fourier transformations are compared and best match selected. These methods are used if images were acquired under varying conditions or there is frequency dependent noise. These methods are faster than correlation based methods, especially with larger images.

Mutual information methods are considered to be the leading technique in multimodal registration. They are able to register multimodal images because they don't compare directly intensity values as the correlation-like methods do, but instead measure the statistical dependency of the values between two images. The drawback of these methods is the same as for the correlation methods, they both require a lot of computation.

In the case of this thesis work, the images to be registered differ a lot by the intensity values of the pixels. Therefore correlation-like methods are not chosen. Since the calculation cycles can be reduced by taking into account the specification of the problem, there is no frequency dependent noise and the images are not outstandingly large, Fourier methods bring no extra value. Instead, mutual information based method is used. Theory is reviewed in the following sections.

2.2.3 Entropy

Mutual information method is an area based image registration method. It has its roots on information theory.

The measure of information—*entropy* or *information entropy*—was introduced by Hartley[5] in 1928, and advanced later by Shannon[13]. It is a measure of *information* or *uncertainty* in the signal. Usually, and in this thesis, by information entropy is meant the *Shannon entropy*.

The general definition for entropy H of some data sequence is

$$H = -\sum_{n=1}^{N} p(n) \log_b p(n) \tag{2}$$

where N is the number of different possible values or symbols the signal might have, sometimes referred as the length of alphabet. b is the base of the logarithm and $p(n)$ is the probability for a certain signal value n.

Entropy describes the statistical properties of the signal. Let's assume we've got a data sequence D, where every distinct sample $D(t)$ has the same, fixed value v, so that

$$D(t) = v \qquad \forall t \tag{3}$$

This means the probability for that very value is $p(v) = 1$ and

$$p(n) = 0 \qquad \forall n \neq v \tag{4}$$

We can divide the sum in equation 2 into two:

$$\begin{aligned} H &= -\sum_{n=v} p(n) \log_b p(n) - \sum_{n \neq v} p(n) \log_b p(n) \\ H &= -p(v) \log_b p(v) - \sum_{n \neq v} p(n) \log_b p(n) \end{aligned} \tag{5}$$

Now, from

$$\lim_{x \to 0} x \log x = 0 \qquad ,$$

and (4), we can write

$$\sum_{n \neq v} p(n) \log_b p(n) = 0$$

and the entropy for such data becomes

$$H = -p(v) \log_b p(v) = -\log_b 1 = 0 \tag{6}$$

This means there is no entropy or no uncertainty in data at all and we can always be sure the next value in the sequence is v. This also means there

is really no information in the data either since receiving constantly one and single value v is not of use at all.

Another example might be a sequence of white noise. In white noise the values of the signal are uniformly distributed. If there are N possible values for the next sample in data D, the probability for the value to be a certain value n is

$$p(n) = \frac{1}{N} \qquad \forall n$$

This time the entropy (equation 2) may be written as

$$H = -N\frac{1}{N}\log_b \frac{1}{N} = \log_b N \tag{7}$$

Choosing base b for the logarithm is only a matter of scaling. This time there is uncertainty in the data sequence (assuming $N > 1$). This means we can't be sure what the next value is, which means the next value in the signal contains some *information*.

Data sequences with different probability distributions yield various values for entropy.

A histogram of the image is a presentation of the gray value distribution of the image. Let's take an image matrix A, where each pixel can be represented by $A_{i,j}$, where i and j are the coordinates, i being the row and j the column of the pixel. Let's also say the length of the alphabet for signal is N, or there are N different values the intensity of the pixel can be presented.

If we count the number of pixels associated with certain intensity value n in the image matrix, and divide it by the number of all pixels in the matrix, we get the proportion of pixels with value n. We might even say that this equals to the probability for a pixel to contain value n should we pick a random pixel from the image matrix. This probability may be written as

$$p(n) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\delta(A_{i,j} - n) \qquad , \tag{8}$$

where I and J defines the size of the image matrix A, and n denotes the gray value. The Dirac's Delta function $\delta(x)$ is defined (for digital signals) as

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

which states in equation 8 that when the pixel has value $A_{i,j} = n$, the delta function triggers becoming equal to 1, which adds up in the sum.

If we count the probability for each and every value n, and place the results in an array in the order of value n, we get the probability distribution of gray values of the image, which is also known as histogram of the image. An example of a histogram of some grayscale photograph is presented in figure 5.

It is possible to calculate entropy of the image using the probabilities from the values of histogram as entries to the equation 2. The mutual information method utilizes entropy.

2.2.4 Mutual Information

In the mutual information method for image registration, two images are moved, rotated, scaled or transformed in even more sophisticated ways in view of each other in order to find the best match between images.

Viola and Wells[17] were among the first to use mutual information in its current form. There was, naturally, previous work, of which one was carried by Collignon *et al.*[2], who worked with joint entropy.

Joint entropy is one part of mutual information method. Whereas an entropy of an image is calculated from a histogram, a joint entropy is calculated from a joint histogram.

Joint histogram binds the two images together. It is a 2-dimensional presentation of distribution of gray value *pairs*. In the joint histogram matrix, the coordinates refer to the gray values, one coordinate indicating the intensity in the first image and the other coordinate in the second one.

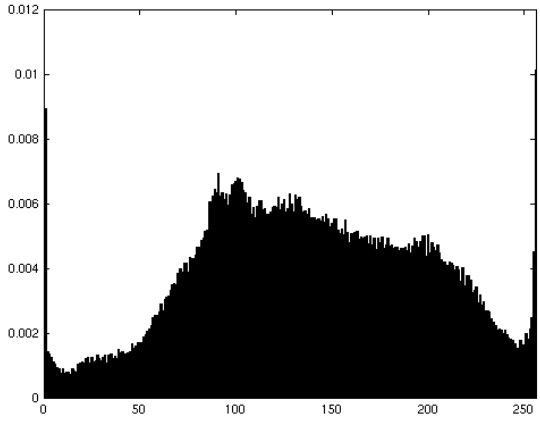


Figure 5: An example of a histogram of a image. On the far left ($n = 0$) is the amount of black pixels while in the far right ($n = 255$, 8-bit presentation, $N = 2^8 = 256$) can be seen the amount of white pixels. In between are the proportions of other pixel intensities. From this histogram can be read e.g. that the proportion of pixels with value 100 is 0.0068, which is also the probability to get gray value 100 when choosing a random pixel.

If n_A is some gray value in the image matrix A, and n_B is in the B, the figure at point (n_A, n_B) in the joint histogram tells the proportion of pixels that in the matrix A has the value n_A *and* in matrix B has the value n_B when the coordinates of those pixels are taken to be the same across the image matrices.

To calculate the joint histogram, the two images need to be of the same size, i.e. there must be the same number of rows and columns in both images. The intensity resolution however may differ. If the number of different values the pixels in the first image may have is N_A, and the number for the second image is N_B, the size of the joint histogram matrix is $N_A \times N_B$.

Let's say we've got two same sized image matrices, A and B. Since the matrices are of the same size, we can form pixel intensity pairs $A_{i,j}, B_{i,j}$, where

the coordinates i, j are the same between the images. Joint histogram is a distribution of the probabilities of these intensity pairs. A single entry for the probability distribution can be calculated by equation 10.

$$p(n_A, n_B) = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \delta(A_{i,j} - n_A)\delta(B_{i,j} - n_B) \tag{10}$$

where n_A is the gray value for the first image and n_B is for the second. After computing all of the $p(n_A, n_B)$ probabilities, the joint histogram is formed.

The joint histogram gives important clues whether the two images match or not. Let's take a case where a joint histogram is calculated from two unrelated images. If we searched for all coordinates (i, j) in matrix A, where $A_{i,j}$ equals to certain value n_A, and took a look for pixel values in matrix B in the very same coordinates, we would most likely find them to be random. This would mean the row n_A in the joint histogram matrix would be noisy. If we them went through all pixel values n in matrix A and by the same time formed the joint histogram of the two, unrelated images, we would find the whole joint histogram matrix to be noisy. (See figure 6 (a).)

In the case of identical image matrices, where $A = B$, things are different. Let's inspect first the joint histogram when $n_A \neq n_B$. This means we are to calculate the probabilities of pixel pairs $A_{i,j}, B_{i,j}$ in coordinates (i, j) where matrix A has value n_A and matrix B has $n_B \neq n_A$. Now that matrices are identical, i.e. $A = B$, Delta functions become

$$\delta(A_{i,j} - n_A)\delta(A_{i,j} - n_B) \tag{11}$$

and in the case of $n_A \neq n_B$ they can be written as

$$\delta(A_{i,j} - n_A)\delta(A_{i,j} - n_B) = 0 \qquad \forall n_A \neq n_B \tag{12}$$

and joint histogram (equation 10) is

$$p(n_A, n_B) = 0 \qquad \forall n_A \neq n_B \tag{13}$$

This means majority of the joint histogram (every n_A, n_B pair outside the diagonal) is zero-valued because choosing two times the same pixel from the same matrix ($A = B$) can't give different result ($n_A \neq n_B$ can't be valid).

Let's then take the case where $n_A = n_B$. This means we are calculating probability of those pixels (i, j) in matrix A that has value n_A whereas the same pixels in matrix $B = A$ has value $n_B = n_A$—notice that this is always true. Now the Dirac's Delta functions become simply

$$\delta(A_{i,j} - n_A)$$

and equation 10 for joint histogram can be written as

$$p(n_A, n_B) = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \delta(A_{i,j} - n_A) \qquad \forall n_A = n_B \tag{14}$$

which would show as a straight line in the joint histogram all the way from pairs $0, 0; 1, 1; 2, 2$ to $N - 1, N - 1$. (See figure 6 (c)).

Between two totally random images and two identical images, there are numerous cases where the two images are almost the same. They are of the same scene but shifted by some pixels or rotated by some small angle in view of each other. Such cases show in the joint entropy as dense clusters. In the figure 6 (b) is a demonstration of joint histogram of two identical images except for the shift of one pixel in horizontal direction.

Like in the case of perfect match, a crisp line or curve can be seen also when the images have been modified by some gray level correction. Figure 6 (d) is a joint histogram between an original and a gamma corrected version of an image.

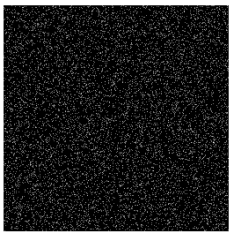
Like the entropy of the image can be computed from it's histogram, a joint entropy $H(A, B)$ of two images can be computed from the joint histogram of the two (15).

$$H(A, B) = -\sum_{n_A=0}^{N_A-1} \sum_{n_B=0}^{N_B-1} p(n_A, n_B) \log_b p(n_A, n_B) \tag{15}$$

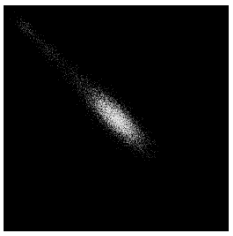
where $p(n_A, n_B)$ is the probability from the joint histogram.

Entropy is the measure of *uncertainty* in the signal. If we want the two image matrices to match, we want to minimize the *uncertainty* between them—to minimize the *entropy*. For the measurement we could use the additive inverse $-H(A, B)$ and try to maximize it.[2]

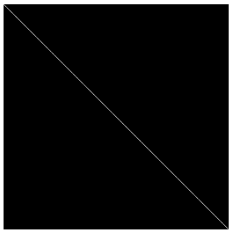
However, choosing to maximize the additive inverse of joint entropy, $-H(A, B)$,



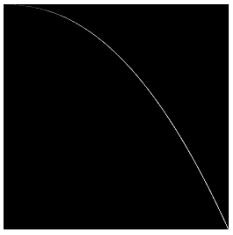
(a) Two random noise images.



(b) Identical images shifted by 1 pixel in horizontal direction.



(c) Two identical images.



(d) The other image modified by gamma correction.

Figure 6: Joint histograms. (a) A joint histogram of two random noise images. As the images are not same in any way, the histogram resembles noise. (b) Two images are the same, but the other one is shifted by 1 pixel in horizontal direction. A structure begins to form. (c) Perfect match. A joint histogram of two identical images shows a crisp line. (d) Otherwise the same as (c), but the second image has gone through gamma correction.

as the measurement isn't always quite enough. Joint entropy might give false minimum e.g. if backgrounds or other uniform areas are the only things overlaid. E.g. in the case of overlaying backgrounds there would be only one sharp point in the joint entropy matrix at the point n_A, n_B, where n_A and n_B are the background gray levels of matrices A and B, respectively. One sharp peak in joint entropy gives low value for joint entropy, which would lead to false maximum of the measurement.

There is, however, a measure to discriminate such areas of the overlaying images in favor of areas where the "proper" data is, and that measure is the entropy of the single image. Entropy for an image matrix consisting of only one pixel value is zero (as found in the example at page 17), whereas entropy for varying data is higher. Therefore, adding the entropies of the separated images to the equation should help us avoid false maximum. Now it can be written

$$I(A, B) = H(A) + H(B) - H(A, B) \tag{16}$$

where $H(A)$ and $H(B)$ are entropies of separated images, $H(A, B)$ is the joint entropy of the two, and $I(A, B)$ is the *mutual information* measure which we want to maximize.[9]

Mutual information has been studied a lot, and a normalized version, called NMI (17), has been proposed[15] and it is used in this thesis.

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)} \tag{17}$$

Although mutual information is mostly used to register multimodal images from different kind of sensors, there is also previous work involving alignment of two x-ray images. Sanjay-Gopal *et al.* exploited mutual information method to find lesion from the mammograms.[12] They used MI first for registering the temporal mammograms from the same patient. After the registration they used mutual information once again to find the lesion structure of the latest image from the previous images to obtain a final estimation of its location.

2.2.5 Image Rotation

As the phosphor imaging plates are practically at the same distance from the teeth and x-ray tube, and they are in the same plane in view of each other, the stitching problem of this thesis can be considered as rigid registration problem.

Rigid registration means that the images are only rotated and shifted spatially without any more complex transformations. Subpixel accuracy is not sought, so shifting the image matrices is an easy task. Rotating digital images, however, is trickier.

Image matrix is rotated using a rotation matrix. Two dimensional clockwise rotation matrix is

$$R(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{18}$$

where θ denotes the angle.

Image matrix is rotated by mapping pixels (i, j) into new pixel coordinates $(\hat{i}, \hat{j})$, which are calculated with rotation matrix (equation 19).

$$\begin{bmatrix} \hat{i} \\ \hat{j} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} \tag{19}$$

In this "forward method", the existing pixels (i, j) are mapped into a new pixels $(\hat{i}, \hat{j})$. Because the pixel coordinates must be integer, this approach can produce holes and overlaps in the new image due to the rounding errors. Therefore, the "backward method"—where the pixel coordinates of the *new* image are taken to be integer and coordinates of the image-to-be-rotated are calculated from them—is used instead (equation 20).

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \hat{i} \\ \hat{j} \end{bmatrix} \tag{20}$$

Newly calculated coordinates (i, j) are not neccessarily integers, although in the image matrix they are. In the *nearest neihgbour method* the coordinates

(i, j) in equation 20 are simply rounded and that pixel is mapped into the new image.

More sophisticated alternative is the *bilinear interpolation method*. Let's say we've computed the coordinate (i, j) with equation 20, and we then identify the four nearest integral coordinates as $Q_{11} = (i_1, j_1)$, $Q_{21} = (i_2, j_1)$, $Q_{12} = (i_1, j_2)$ and $Q_{22} = (i_2, j_2)$.

The value v for the pixel $(\hat{i}, \hat{j})$ in the rotated image matrix can be calculated by linear interpolation in both i and j directions (equation 21).

$$\begin{aligned} v(\hat{i}, \hat{j}) \quad &= \tfrac{v(Q_{11})}{(i_2-i_1)(j_2-j_1)}(i_2 - i)(j_2 - j) \\ &+ \tfrac{v(Q_{21})}{(i_2-i_1)(j_2-j_1)}(i - i_1)(j_2 - j) \\ &+ \tfrac{v(Q_{12})}{(i_2-i_1)(j_2-j_1)}(i_2 - i)(j - j_1) \\ &+ \tfrac{v(Q_{22})}{(i_2-i_1)(j_2-j_1)}(i - i_1)(j - j_1) \end{aligned} \tag{21}$$

To rotate the image around it's center point, the coordinates must be declared so that the center point is at $(0, 0)$. Since in the matrices the coordinate are defined by the top-left corner, the equation 20 must be rewritten as

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \hat{i} - i_c \\ \hat{j} - j_c \end{bmatrix} + \begin{bmatrix} i_c \\ j_c \end{bmatrix} \tag{22}$$

where (i_c, j_c) is the center point of the image matrix.

3 Materials

3.1 Imaging Plate

3.1.1 Phosphor Imaging Plate

The topic of this thesis, stitching two x-ray images, doesn't depend on the means the images are acquired. However, because the algorithm is going to to be used with a certain device which uses phosphor imaging plates only, these plates are used as solely means to obtain test images.

Imaging plate is a flexible plate with a photostimulable phosphor compound coating which is capable of storing the energy of x-rays among other radiations.[14] The most important feature of the phosphor compound coating is that, when stimulated by visible or infrared light, it emits light corresponding to earlier absorbed energy—which, in other words, means the phosphor imaging plate records x-rays.

The advantage of imaging plate over conventional film is its reusability. After the image has been read from the plate, the committed energy can be released fully by exposing the plate to bright light. This resets the plate and it is ready to be exposured by another x-ray dosage.

Besides film and imaging plate, one might take x-ray images with a digital sensor. With digital sensor the operator doesn't need to use external device to read the data out of the imaging plate. However, the sensor is usually thicker and not flexible at all, making the plate easier to place into a patient's mouth.

Imaging plates are cut from larger pieces, and the edges of the plates might not be exactly straight but might curl in the pixel scale. This must be taken into account in the program.

The size of the IP used in this thesis is 7.0×3.1 cm. The corners of the plates are rounded.

3.1.2 Imaging Plate Reading Device

After the imaging plate has recorded the x-ray radiation, it is inserted into the plate reading device. The plate is attached to a sledge, which moves the plate in one direction. While the sledge and the plate moves, a laser beam sweeps from side to side. The movement of laser beam is transverse to the movement of the plate.

When the laser hits the phosphor compound coating, the coating emits a visible light with energy corresponding to the amount of x-ray radiation the very spot on the plate was exposed to. Emitted light goes through light amplifier tube and is then recorded by light sensor.

Recorded data is digitized by 14 bit A/D converter. A computer then recombines the image from the sequence read from the imaging plate. During the tests, the resolution setting is 25 pixels per millimeter yielding image size of 1750×775 pixels.

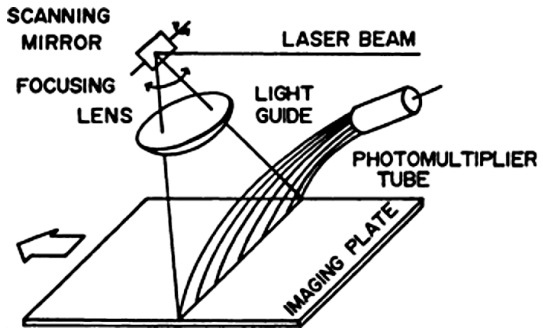


Figure 7: Construction of the image reader. The imaging plate moves at constant speed to one direction. A laser beam sweeps the surface of the plate from side to side in transverse direction. A light emitted from the phosphor compound coating is amplified and recorded by photomultiplier tube. Image courtesy of [14].

The designer of the plate reading device must decide what is the maximum size of the imaging plate, as it cannot be larger than the area the laser beam

is able to sweep. This limitation is the reason for this work. There is a maximum width for the imaging plate, and it is not large enough to cover a significant part of the teeth for occlusion imaging. This is overcome by using two overlapping imaging plates, hence creating the need for this algorithm.

3.1.3 Imaging Plate Container

Due to the size limitation of the phosphor imaging plate size (see section 3.1.2), two plates are needed to cover the area necessary for the occlusion imaging.

Earlier attempts to gain larger imaging area in clinics included putting two imaging plates side by side. While this truly doubles the area obtained with a single exposure, the two images where still considered to be separated by the viewing software on the computer. Positioning them would be based only on the educated guess of the doctor.

To automatically merge the images into one, single image, the plates are placed partly one upon the other. This will give the algorithm two images where a section of the images is similar.

To secure that the plates are overlapping sufficiently for successful stitching while not overlapping too much to needlessly reduce the combined surface area, a certain container is used to hold the plates in proper position.

The container is made of carbon fiber. Two imaging plates can be slided in from the other end of the container (figure 8). Before imaging plates are slided in, they are covered with thin, plate sized carton, which both protects the delicate coat of phosphor on the plate, and holds the plate inside the case.

The size of the imaging plate used in testing is 7.0×3.1 cm. In the container the longer sides of the imaging plates overlap for about 0.5 cm, thus making the total surface area of the overlapping plates to be around 7.0×5.7 cm.

The container will be covered by a removable protection bag during the scan to ensure sterility between patients.

For testing purposes, one container is equipped with four metal balls, which

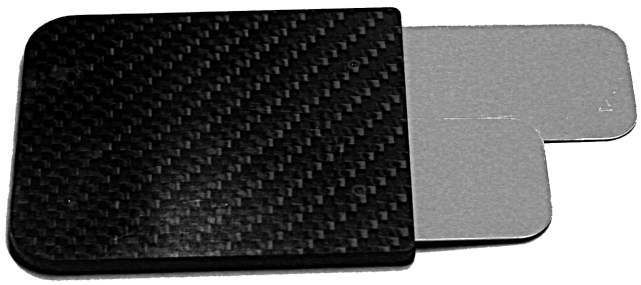


Figure 8: A prototype of the imaging plate container with two imaging plates partially slided into the container. In the clinic, imaging plates are first covered with thin carton to protect the phosphor surface, and the whole container is put in a protection bag to ensure sterility between patients.

will show as bright spots in the image. The distances of the spots can be used as a numerical measurement of the success of the image stitching, as the distances of the spots should last between different images. Balls are placed so that their images won't participate in image stitching process.

3.2 Dental X-ray Images for Testing

A set of intraoral image pairs was taken for testing purposes. Imaging plates were placed into the plate container which would hold them in right position in view of each other. The phosphor plates were a standard *Digital Imaging Plates* from *CAWO Photochemisches Werk GmbH* except from the special size of 31×70 mm.

To minimize the radiation dose for living tissue, no humans were exposed. Instead, images were taken from a human skull. The skull was equipped with a handle to change the position of jawbone, thus allowing the teeth to bite and hold the plate container between teeth.

Images were taken from both upper and lower teeth. The case was also rotated in random angles in horizontal plane to make the testing data more uncorrelated.

Plates were exposured using the *Instrumentarium Focus*^{TM} intraoral x-ray tube from *PaloDEx Group Oy*.

After exposure, the recorded images from the plates were read by *Soredex Digora®Optime* reading device, also from *PaloDEx Group Oy*.

4 Design of the Algorithm

This section covers the design of the algorithm. The flow diagram of this algorithm is presented in figure A1 at appendix A. Each of the blocks in the diagram are described in the following subsections.

4.1 Preprocessing Images

Preprocessing block of the algorithm, as the name indicates, processes the images before they are sent to the actual image fusion algorithm.

When the image is read, the plate reading device scans a certain area. To make sure all data in the plate is read, this area is larger than the surface area of the plate. This scanned area defines the size of the *image canvas*. Image canvas can be considered as an area where the actual image is laid within. The size of the canvas may be larger or equal to the size of the image in it.

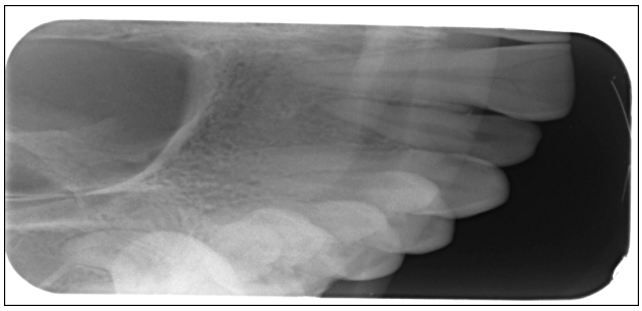


Figure 9: An example of the actual image within the image canvas. The image data received from the imaging plate reading device (the image and surrounding whites inside the border line) is larger than the actual image (the gray imaging of the mouth).

In our case the canvas of the image file is larger than the actual imaging of the teeth recorded by the imaging plate (see figure 9). As the data passes through the software, the surrounding white area is removed for two reasons.

First, it carries no information at all, and second, it would degrade the quality of the stitching process as surrounding whites would be mixed with teeth and other lighter colored areas in the actual image. The easiest way to remove surrounding whites is simply to crop off any pixel line containing more than preset amount of white.

Depending on how the operator of the device manages to place the plate into the track of the reading device, the actual image may be more or less rotated. This may cause cropping some of the actual image data if the angle between the plate and the track is significant enough. Figure 10 shows a sketch of such situation.

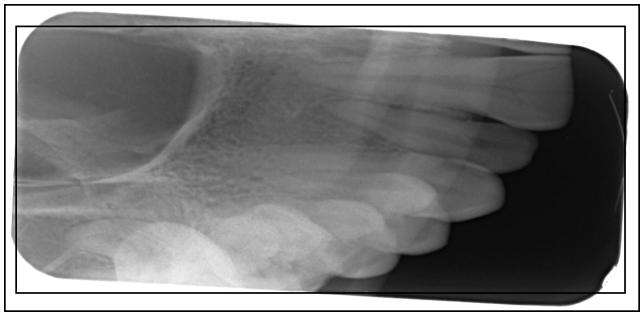


Figure 10: A sketch of cropping the image due to the significant angle between the imaging plate and the reading device. The outline represents the canvas of the image data where the actual imaging of the teeth is within. The smaller rectangle represents the border of the resulting image after rather simple removal of the surrounding whites. Other than whites, it also crops off pixels with proper information.

Normally in intraoral imaging, a small rotation—or a small cutting of the edges of the image caused by rotation of the image—wouldn't matter as usually the most interesting objects are set to the middle of the imaging plate. In image stitching process however, the edges are needed, for now the most important information—the information needed to combine the two images—is at the

edge areas of the images. Nonetheless, the whites must be removed or otherwise marked before image data can be sent to next step in the algorithm. Should this not be done, the surrounding whites would have negative impact on fusion process as structures (surrounding whites) found in the first image would be in totally different area in the second image.

Since the problems of cropping whites are caused when the actual image in the image canvas is more or less rotated, a self-evident correction is to counter-rotate the image so that the pixel lines of the edges of the actual image align with the pixel grid.

Before the angle to counter-rotate the image can be calculated, the edge of the image is searched. The surrounding area of the image canvas is known to be bright white, whereas the imaging of the teeth—unless underexposed—is darker. This can be seen in the histogram of the image (figure 11). Histogram presents the gray level distribution of the image. Each bar in the histogram represents the amount of pixels with corresponding gray value.

The threshold to decide whether some pixel belongs to the surrounding white or to the proper image is searched from the histogram. Pixels with gray level higher than threshold are considered to be part of the surround.

Some sample images were examined in image editing software to find a clue of suitable threshold value where surrounding whites turn into actual image. It turned out that the threshold is to be found from the band of the histogram between the great mass of the actual image and far white end of surroundings. For instance in image 11 the actual image has values $2300 \ldots 14500$ and the far white end of the surroundings has $15500 \ldots 16383$, where gray values may have $2^{14} = 16384$ different values in 14-bit gray images.

However, numbers stated here cannot be trusted to be the same from image to image since they, especially the values of the actual imaging, depend on varying conditions such as used voltage and current and the distance between the x-ray tube and imaging plate.

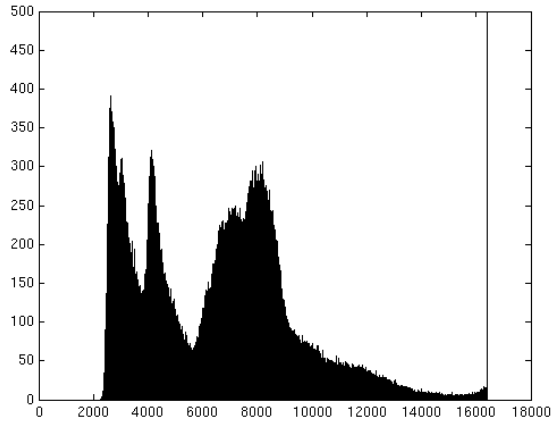


Figure 11: A sample histogram of the x-ray image. The high bar on the right (the bright end) tells us there are significant portion of white in the image. We already know that it is (mostly) the surrounding white area of the image canvas around the actual x-ray image. The threshold gray level for removing the whites is searched from the histogram, and it is the local minimum just before the figure rockets up in the white end.

Therefore a suitable threshold is searched from the histogram. Since the histogram is separated in two sections, a fit value for threshold is assumed to be a local minimum between the sections.

If for some reason the bit accuracy of the image has been decreased and is lesser than accuracy used to determine the histogram, or for some other reason the histogram appears to be comb-shaped, the histogram is "low-pass filtered". In mathematical point of view there really is no such operator for histograms. What is done is a low-pass filtering of the series containing the gray levels of the histogram. This will remove the comb-shape from the histogram, a necessary action if one wishes to avoid false minima in the series. Conveniently this will also reduce noise and small variations in the series, which is nice since we hope to find local minimum of the tendency only.

A simple way to apply low-pass filter is to count average value of the nearby values as presented in equation 23, where $h(l)$ is the portion of pixels with gray value l from the actual histogram, and $h_L(l)$ is low-pass filtered histogram. The constant C in the equation determines the number of histogram entries $(2C+1)$ used for calculating the average.

$$h_L(l) = \sum_{i=l-C}^{l+C} h(i), \qquad l = 0 \ldots L-1 \tag{23}$$

For the parts, where $i = l - C < 0$ or $i = l + C > L - 1$ and the histogram is undefined, the last know value, $h(0)$ or $h(L-1)$, is used, respectively.

After the histogram has been smoothened, the algorithm searches for the last local minimum before the white end. This is assumed to be a gray value which discriminates surrounding whites (values above) and actual image (values below). It is possible there are pixels with gray values above the threshold within the actual image, specially around the teeth, but the minimum is considered to be decent approximation nonetheless.

Next, each pixel row is scanned and the coordinate where the pixel value drops from high (surrounding white) to below the threshold (actual image) is stored. The edge of the image is searched by fitting a line on account of these coordinates (x_i, y_i) using least squares fitting.

For a linear fit we've got a function of form

$$y(a, b) = a + bx$$

whose parameters a and b can be calculated from the data points x_i, y_i. Using least square fitting we get

$$b = \frac{(\sum_{i=1}^{n} x_i y_i) - n\overline{xy}}{(\sum_{i=1}^{n} x_i^2) - n\overline{x}^2} \tag{24}$$

and

$$a = \overline{y} - b\overline{x} \tag{25}$$

where $\overline{x}$ and $\overline{y}$ are averages of values x_i and y_i, respectively.[18]

The edge of the image is now determined by values a and b. There is no use for the figure a, but from b we can calculate the angle of the estimated edge of the proper image. Since this figure is the amount of units the slope of the rotated edge grows against the movement of one unit in desired direction of the edge, the angle θ of the image is simply

$$\theta = \arctan b \tag{26}$$

Now the image can be rotated by angle $-\theta$ using the backward method and bilinear interpolation as described in section 2.2.5. The border of the actual imaging should now align with the border of the image matrix, or the pixel grid.

After the image has been straightened, the surrounding whites can be cut off by removing any line with white pixels exceeding the preset amount. Because of the straightened image matrix, there is no fear of losing any vital information of the images.

4.2 Identifying "Upper" and "Lower" Image

Since the imaging plates in the container are only slightly overlapping each other, there is really no need to try to register the images by searching the whole area of the pictures. Instead, to reduce calculation time it is wise to search only from the overlapping parts of the images. To predefine whether certain parts of the images—say certain sides of the images—are overlapping or not, we need information about the container and the placement of the imaging plates. Information for estimating a suitable size for the overlapping area of the image could be used too, but in this particular application it is not needed as the necessary information is actually stored in the images.

Let's say $a_{i,j}$ corresponds to the amount of radiation that is about to hit corresponding pixel i, j of the imaging plate. (Strictly speaking there are no

pixels in phosphor imaging plate, but we can assume that the energy of certain surface area of the plate is in the process converted to equivalent energy of corresponding pixel.) If there is no imaging plate above, this pixel on the plate gains the whole radiation $a_{i,j}$. However, should there be another imaging plate blocking the rays, our pixel will only gain the radiation of $Ta_{i,j}$, where T is a transmission coefficient indicating the proportion of radiation passing through the first plate ($0 < T < 1$).

Since an x-ray image is a negative image of recorded data, the value $v_{i,j}$ of each pixel in the actual image will be

$$v_{i,j} = \begin{cases} 1 - a_{i,j} & \text{if no plate is above} \\ 1 - Ta_{i,j} & \text{if another plate is above} \end{cases} \tag{27}$$

where the values of v and a are both between $0 \ldots 1$.

We can also state that expected value of $a_{i,j}$ is average of all $a_{i,j}$ (equation 28). From this and equation 27 we get expected values (equations 29 and 30) for pixel values v in two different cases.

$$E(a_{i,j}) = \frac{1}{IJ} \sum_{i,j} a_{i,j} = \overline{a} \tag{28}$$

$$E(v_{i,j}, \text{ no plate above}) = 1 - \overline{a} \tag{29}$$

$$E(v_{i,j}, \text{ another plate above}) = 1 - T\overline{a} \tag{30}$$

Since $0 < T < 1$, we can say that

$$E(1 - T\overline{a}) > E(1 - \overline{a}) \qquad \text{, or} \tag{31}$$

$$E(v_{i,j}, \text{ another plate above}) > E(v_{i,j}, \text{ no plate above})$$

or, in other words, gray values of pixels shaded by another imaging plate during the exposure are higher than those not shaded.

The change in the overall gray levels between fully exposured (not shaded) and partly exposured (shaded) areas in the image of the lower plate can be seen

clearly. (See the right side on figure 12). Because of the more or less rectangle shape of the plate with other side longer than the other, the imaging plate can be inserted in only two ways into the reading device. As a result, there are only two sides in the image where the border between different exposure amounts may lie, and due to the design of the container, that border is on either of the long sides of the image. Also, the border is more or less aligned with the side of the image because of the design of the container doesn't allow there to be significant angle between the plates.

Figure 12: Sample image of the imaging plate showing a distinct vertical border—or change in the overall gray levels—on the right side of the image. It has been caused by another imaging plate on top this one during the exposure.

To make decision on which one of the images was on top and which one was underneath the other one at the time of the exposure, the algorithm tries to find if there is an area of higher gray values in the image as expected by equation 31. The area of lesser exposure should be on the long side of the image, which means, moving along the i-coordinate should not have an influence whether we are on the fully exposured area or not, assuming the image (matrix) is in upright position and i is taken to be the row coordinate of the matrix.

The values of each row are added up (32). The series from (32) are differentiated and since we are interested in the absolute change only rather than direction of the change, the absolute value of the series is taken (33).

$$s(j) = \sum_{i=1}^{I} v_{i,j} \tag{32}$$

$$\Delta s(j) = \|s(j) - s(j-1)\| \tag{33}$$

Differentiated series Δs shows a distinct peak where the overall gray values of the image changed in the "lower" image. In the "upper" image however, there is no such peak in the differentiated series since the plate has not been shadowed by any other imaging plate. During this thesis, by "lower" image is meant the image acquired from the imaging plate that was underneath during the exposure. By the "upper" image, naturally, is meant the image of the plate that was above the "lower".

The images are identified as "upper" and "lower" image according the absence or the presence of this peak respectively. (See figure 13)

In practice one must notice that there might be false peaks near the ends of the series. These are formed because the edge of the physical plate is not always sharp, and there might be white pixels as the remainings of the canvas. The Δs series shows a peak where the empty space changes to actual image since there is a also a change in the overall gray levels.

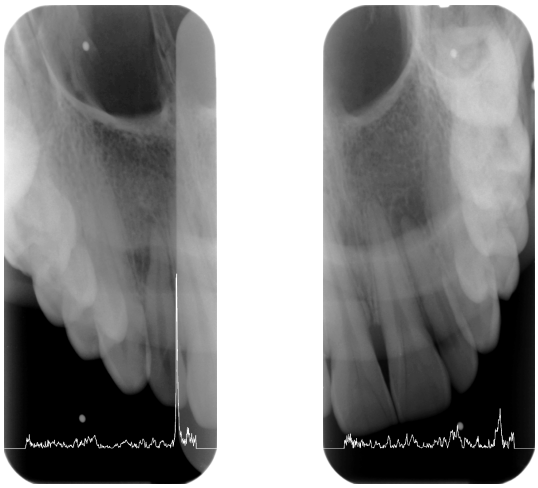
When searching the peak from the series, the application must omit those false spikes. In this application it is done by excluding circa 10% of both ends of the differentiated $\Delta s(j)$ series since the overall change in gray levels should not be there according to the plate container design. Ends of the series has been set to zero in figure 13.

The series $\Delta s(j)$ of both lower and upper image are compared and identified so that

$$\max \Delta s_L > \max \Delta s_U$$

is satisfied, where Δs_L presents the series of the lower image and Δs_U of the upper image.

As seen in figure 13 (b), even when teeth (in this case molar teeth)—which show in the image as lighter areas—are situated in a line at the side of the image, the diagram shows no prominent peak which would be mixed with the peak of overall change in gray values in the other diagram.



(a) Image of the lower imaging plate overlaid with diagram.

(b) Image of the upper imaging plate overlaid with diagram.

Figure 13: Images of the lower (a) and upper (b) imaging plate along with overlaid diagrams. The diagrams represent the values of series $\Delta s(j)$ (equation 33) at each column j of the image matrix. A distinct peak can be seen in diagram of the lower image at the very spot where there is a border between fully and partially exposured areas of the image. Diagrams are in the same scale. There is no peak at the diagram of the upper image. The ends of the diagrams are set to zero to avoid possible false peaks near the edges of the images.

4.3 Tentative Aligning of Images

The plate container allows there to be only small rotations between the plates, and to reduce calculation cycles only small angles are to be gone through in the iteration process later in this algorithm. Therefore the images are first aligned tentatively so that they are not upside down or mirrored in view of each other. The pool of image processing techniques included in this part of the process are only rotation of 180° and vertical and horizontal mirroring of the image.

The preprocessing block (section 4.1) has already turned the image to upright position if the image was initially wider than higher, thus ensuring that the edge of the "lower" image with lesser exposure is either on the left or the right side of the image.

The "lower" image is aligned (i.e. rotated 180° degrees if necessary) so that the edge with lesser exposure is located to the right. Once again the change in the overall gray levels (equation 33) is used here, now as an indicator of alignment of the image. Nothing is done if the maximum value of series Δs is already found on the right side of the image, i.e.

$$\underset{j}{\operatorname{argmax}}\, \Delta s(j) > \frac{J}{2}, \tag{34}$$

where J is the width of the image matrix, and thus half of it denoting the middle of the image in view of coordinate j. Otherwise, if

$$\underset{j}{\operatorname{argmax}}\, \Delta s(j) < \frac{J}{2} \tag{35}$$

is true, the "lower" image is rotated 180° to get the less exposured area to the right side.

Choosing to rotate the image this way is totally arbitrary and one might choose the opposite as well with no effect on the functionality of the algorithm at all, if the necessary changes are applied to the rest of the algorithm. This is done to simplify the later process of the algorithm. The arbitrary choice made

here only means that the image with lesser exposure—i.e. the lower image—is going to be located on the left in the resulting image.

Theoretically there might be also situations where

$$\underset{j}{\operatorname{argmax}}\, \Delta s(j) = \frac{N_j}{2}. \tag{36}$$

This case is omitted in this application as it is assumed that the change in gray levels caused by another imaging plate is closer to the edge of the image than the middle of the image. This assumption is based on the foreknown shape of the container for imaging plates.

Next, a section of the image is taken from the lesser exposure part of the image. The size of this *window* is determined by the earlier found border between fully and partially exposured areas of the "lower" image. All columns between coordinates $j = \operatorname{argmax}_j \Delta s(j)$ and $j = J$ are read to the window.

Equally sized sections are taken from both sides of the "upper" image. (See figure 14.)

Imaging plates have round corners which means the images have them too. On the overlaying parts of the images the corners however are horizontally mirrored in view of each other. The area outside round corners is more or less bright white (white surrounding) and would have a negative effect on the image registration as different parts of the imagings of the same view would be rendered in white. Therefore the undefinable data near the corners is excluded.

The two sections from the "upper" image are rotated and mirrored according to different alignments the plates can have, and then compared to the window from the "lower" image by calculating the mutual information between them.

The highest mutual information figure denotes the most likely match of two windows. Earlier the "lower" image has been rotated so that the under exposured part faces to the right. The "upper" image is rotated or mirrored so that, when placed on right in view of the "lower" image, the matching side

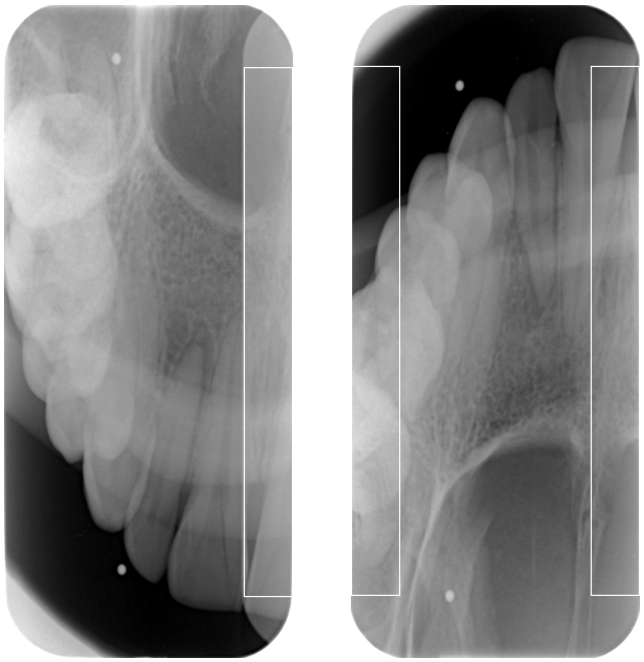


Figure 14: Windows from the images are taken in order to tentatively align the images. Window is taken from the lesser exposure part of the image. (White rectangle in the left side image.) From the other image, sections are taken from both sides. (Two white rectangles in the right side image.)

faces left. Now the overlapping parts of the images are next to each other like the example in figure 15.

4.4 Defining Region of Interest

The algorithm now knows which parts of the images were overlapping. When it comes to registering the images, those are the only interesting parts of the image. These regions of interest (ROI) of the images are cut off to separated image matrices for closer inspection.

ROI might be selected by the content of the image, but in this application the region of lesser exposure and corresponding region from the “upper” image

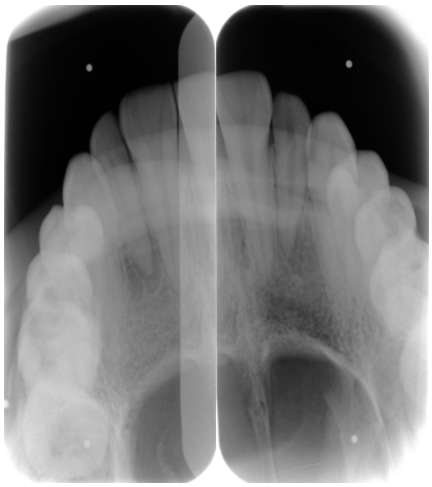


Figure 15: An example of tentatively aligned images placed right next to each other. The overlapping areas are right next to each other in the middle of the image.

are chosen.

The round corners are excluded from the region of interest for the reasons mentioned in the earlier section. Because of the little inexact edge of the phosphor coating in the imaging plate (see section 3.1), some margin is to be cut off from the long sides also to make sure the undefinable data is excluded.

Sizes of margins are determined by studying the images in some image viewing software. By removing about 10% of the width of the underexposured part of the image from both left and right sides of the image will remove the fuzzy edges caused by the inexact border of the phosphor coating. (Fig. 16) Round corners were removed already in the previous step of the algorithm.

A ROI from the "upper" image is also defined, and its size is determined by the ROI previously selected from the "lower" image. The size is almost the same, except now small margins are added so that this region becomes larger than the other thus allowing the smaller region to move and rotate inside the bigger one.

Figure 16: A top of the underexposured part of the image from which the margins are removed to define region of interest. White rectangle illustrates the border of ROI. Circa 10% of original height of the image is removed from both top and bottom of the image. (Bottom not seen here.) Circa 10% of the width of the underexposured image is also removed from left and right sides of the image.

Size of this bigger ROI might either be fixed and only slightly larger—which is suitable if angle and shift are searched by going through all possible combinations—or more loosely determined, even the size of the original image, if some optimization routine is used to find the angle and shift.

As the prime aim of this work was to find robust means to register two images, no optimization is involved in the early phase of the final product. Not using optimization in rigid medical image registration is not rare choice as searches are rather fast.[19] Instead all possible combinations of angles and coordinate shifts are going to be searched for, and the "upper" region of interest is chosen to be only little larger than the region of the "lower" image.

Based on the plate container design, movement of one IP is ± 1 mm in view of the other. This predefined information is used when calculating the size for ROI of "upper" image. If we know the resolution R of the sensor system (pixels per millimeter), we can compute that distance $s_{\text{mm}} = 1$ mm is equivalent to

$$s_{\text{pixels}} = R s_{\text{mm}}$$

where s_{pixels} is the distance of 1 mm in pixels.

The window taken from the "upper" image will be $2 \times s_{\text{pixels}}$ pixels larger than the "lower" window in both width and height.

4.5 Preprocessing Gray Levels of the ROIs

The response of the imaging plate reading device is not exactly smooth along the image. There is a tendency for the pixel values to become lighter when approaching the edge of the longer sides of the image. In the selected regions of interest the phenomenon is almost linear, however, changing in the opposite directions.

The mutual information method doesn't handle non-statistical changes well and the gradients must be eliminated. Few images were examined with Matlab and a suitable linear function was formed to remove the gradients. The function is of form

$$v_n(j) = \left(a + b\frac{j}{J}\right) v(j)$$

where a and b are coefficients, v is the original value of the pixel and v_n is the new. The column coordinate j of the pixel is used to further determine the coefficient when the size of the selected ROI is J.

To reduce the calculation cycles, the bit accuracy of the images is decreased. In image manipulation this is referred to color quantization of the image windows. The color quantization means the number of different possible gray levels denoting the amount of x-rays is reduced, or, the accuracy to determinate the amount of x-rays is decreased. Images acquired with Soredex OpTime plate reading device are stored with bit depth of 14 bits, yielding 2^{14}, or $16,384$ different gray level values.

A certain, arbitrary figure for bin count could be chosen, but Amankwah Anthony and Otmar Lofffeld gained good results with their tests on bin count chosen by Freedman and Diaconi's rule.[7]

The Freedman and Diaconi's rule is

$$W = 2(IQ)N^{-\frac{1}{3}} \quad , \tag{37}$$

where W is the number of bins remaining after quantization. Quartiles are the three values which divide the histogram into four equal parts, so that each part represents one fourth of the pixels or samples. *Interquartile range* IQ is the range between the third and the first quartile, or in other words, the gray value range where 50% of the midmost pixels lay. N in equation 37 is the number of pixels in the image, or rather, in the region of interest.

For example, one underexposured part of the image (ROI of the "lower" image) has the range $IQ = 898$. The number of pixels is $N = 87,264$ and so, from (37) we get $W \simeq 40$ bins. To make it easier for computer calculation, we then choose the nearest exponent of 2.

$$2^b = W = 40$$
$$b = \frac{\ln 40}{\ln 2} \simeq 5$$

So, the bit depth of the ROI will be reduced and there are going to be only 2^5, or 32 different gray values. This decreases cycles needed to compute mutual information without decreasing the quality of registration. Values are calculated separately for "upper" and "lower" ROI.

The region for the "lower" image is chosen from the area of the imaging plate which was underneath the other and therefore has lesser contrast with rather narrow histogram. If the reduction of bins were done directly to the "lower" ROI, majority of the pixels would fall into just few bins. As this is not wanted, the gray values of the region are adjusted. The histogram is stretched along all possible values, so that the lowest pixel values become 0 and the highest $2^{14} - 1 = 16,383$ just before the quantization.

Although in the case of the "upper" ROI, the color quantization wouldn't be so violent, its values are adjusted also.

In the field of digital image manipulation, there exist many methods for quantization of colors while keeping the quality of the image as good as possible for human eye. These methods include, among others, use of color palettes or dithering. In this application, however, we are not interested in the subjective quality of the windows of the images and we don't use such methods. Instead, a plain nearest color algorithm is used for quantization.

If the number of bit accuracy to be is b, and the current value of the pixel is v, the new value v_n can be calculated by the following equation

$$v_n = round\left(v\frac{2^b}{2^{14}}\right) \tag{38}$$

where $round$ rounds the parameter into nearest integer.

The size of the joint histogram of the mutual information method is defined by the values W_{lower} and W_{upper}, which are calculated by equation 37 for lower and upper image window.

4.6 Finding Spatial Location and Angle

Now that original source images has been preprocessed, the algorithm can proceed to the core section. This part of the program finds the angle θ between the images and the spatial shift (i.e. the i, j coordinates) needed to register the images and fuse them together to form a single image.

To reduce the computation load, the match between images is not searched through whole images, but rather from previously chosen regions of interest (see section 4.4) which are known to overlay each other. These overlaying parts of the images are then rotated and shifted in view of each other.

The imaging plate defines the maximum angle there can exists between the images. The images themselves have been straightened up earlier in section 4.1. This, however, only nullifies the angle between single imaging plate and sledge of the plate reading device at the time of image reading, not the angle between the plates when they were exposured. If location of the other plate is

considered to be fixed, the other is allowed to move ±1 mm in any direction. Different ends can move to different directions and thus contribute an angle between the plates.

To calculate the largest possible angle, we need to know the height h of the imaging plate. In case of the largest angle, both ends on the non-fixed plate move 1 mm to different direction, which means, that we've got a triangle where hypotenuse is $h/2$ and opposite side is 1 mm. Since sine function is

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

we can write

$$\theta = \arcsin\frac{\text{opposite}}{\text{hypotenuse}}$$

and, assuming the length h of the plate is given in millimeters, we get an equation 39 for the the maximum angle $\theta_{\max}$.

$$\theta_{\max} = \arcsin\frac{1}{h/2} = \arcsin\frac{2}{h} \tag{39}$$

For a 70 mm length imaging plate equation 39 yields an angle of 1.64°.

The ROI of the "lower" image is rotated with various angles between $-\theta_{\max}$ and $\theta_{\max}$. The number of different angles to choose is a tradeout between computational speed and accuracy of the stitching.

Whenever the ROI from the "lower" image is rotated, some undefined data from outside the border is bound to appear. Depending on the implementation of the rotate algorithm, appearing data might be pixels of e.g. white or black. In either way this data is undesired and must be removed. (Almost same thing emerged earlier in the process when surrounding whites were removed from the initial images in section 4.1.)

Since the maximum angle $\theta_{\max}$ corresponds to 1 mm tilt of both plate ends, the maximum row or column count needed to remove is the number of pixels on the distance of 1 millimeter. This way, even when rotated by $\theta_{\max}$, all undesired pixels will get cut out. The amount of pixel rows and columns been

equivalent to 1 mm length on the plate is removed from all four sides of the image matrix every time, regardless of the actual angle the image has been rotated. This ensures the pixels count in the rotated and cropped image is always the same and doesn't effect the entropy calculations.

The search for right angle and spatial shift is done by the following loop.

Rotation The "lower" region of interest is rotated by an angle θ. Undefined data outside the original ROI will appear, but it is removed by cropping the rotated window. Cropped ROI of the "lower" image becomes the "lower" window.

Spatial transition A same sized window from the "upper" ROI is selected. The spatial coordinates of this window go through all possible shifts. This way the windows are moved in vertical and horizontal ways in view of each other.

Calculating mutual information Mutual information is calculated between the two windows. First the joint entropy of the two images is formed, from which the mutual information is evaluated as descripted in section 2.2.4.

Comparison of the calculated figure Newly calculated figure is compared to the previous found maximum value of the mutual information. If the new value is higher than the previous maximum, it is stored in the memory along with the values of angle and spatial shift.

After the comparison of the MI, the windows are shifted to the next spatial coordinates and a new mutual information figure is counted. When all possible shifts have been gone through, the original "lower" ROI is rotated by the next angle and all allowed spatial shifts are to be examined. Finally, after all angles has been tested, the proper angle and shift coordinates should be known.

No optimization is used when searching right angle and spatial shift, as both are restricted by rather small limits and there are no image transformations involved. Computation time stays reasonable even when searching through all combinations.

4.7 Rotate "Lower" Image

The offset of the two imaging plates at the time of exposure is now known. In the previous step of the algorithm the offset was found by rotating and moving the window taken from the less-exposured area of the "lower" image.

To counter the rotation, the "lower" image is rotated by opposed angle. By the "lower" image here, it is meant the image after preprocessing step (see section 4.1) where the surrounding whites are removed but where no color quantization is done.

The size of the image matrix is allowed to grow as needed so that no pixel from the image will be cut off.

4.8 Relocate "Upper" Image

The spatial shift—coordinates i, j—between the two images were found in section 4.6. The image rotating method used here rotates the image matrix around it's center point. This means the "lower" region of interest used in section 4.6, and the entire "lower" image will be rotated around different center points. This must be taken into account before the images are stitched.

Let's say the center point of the entire "lower" image is i_c, j_c. Coordinates i, j denote the point where the top-left corner of the "upper" image should be in view of the "lower" image, which was found in section 4.6. Now, when "lower" image is rotated, the point i, j point to another, wrong pixel. This can be bypassed by calculating a new point $\hat{i}, \hat{j}$ for the top-left corner with equation 22.

4.9 Combining Images

At this point of the algorithm, images are ready to be finally merged. The "lower" image has been counter rotated by an angle that the imaging plates were in at the moment of exposure. The right coordinates for the "upper" have been also found.

Corners of the plates are round, and show up in the images as white background. If these white background areas of the "upper" image were to be placed over the "lower" image, they would needlessly cover the image behind, as they contain no information at all. Therefore, in this step of the process, corner areas of the images are considered to be transparent.

Other that that, the "upper" image is simply laid over the "lower" one. For the overlapping area, where there is information from both images, the one from the "upper" image is used. The "upper" is preferred because of the better image quality over the under-exposured part of the "lower" image.

5 Results

The quality of the algorithm was tested by a set of intraoral x-ray images ($n = 33$). The x-ray images were obtained by inserting two phosphor imaging plates in a container which would hold the plates in right position, making the other plate partially cover the other.

Usually the result of the image registration is evaluated by experts.[1] There are different measures for the quality of the registration (e.g. speed or memory usage), but here only the precision of the registration is used. In our case the experts to evaluate the process would have been dentists, but we weren't able to use any. Therefore the evaluation of the registration was made entirely relying on the markers on the images, which is one way to estimate the quality of the algorithm, as surveyed by Ximiao Cau *et al.*[1].

The container for imaging plates was equipped with four small metal balls. The balls were attached to surface of the casing so that they could not move, hence the distance of the balls would remain fixed between different exposures.

Metal absorbs x-rays, so those four metal balls cast a visible, round shadows into the image (see figure 17). Measuring the distances reveals whether the right position has been found or not.

Before the distances can be measured, the dots must be extracted from the rest of the image. Raimondo *et al.*[10] used a *Top-hat filtering* to detect carcinoma nuclei from a image. Top-hat filter tries to remove all but spots of desired size from the image. From the filtering result Raimondo *et al.* generated a binary mask using a suitable threshold from which they could count the number of dots (nuclei).

Applying such dot extraction to the result images of this algorithm generated binary masks where the "shadows" of the metal balls were found, although the filtering didn't always remove all other structures. Therefore a predefined masking was also applied to ensure only proper data was collected by removing structures from the parts where the metal balls could not be.

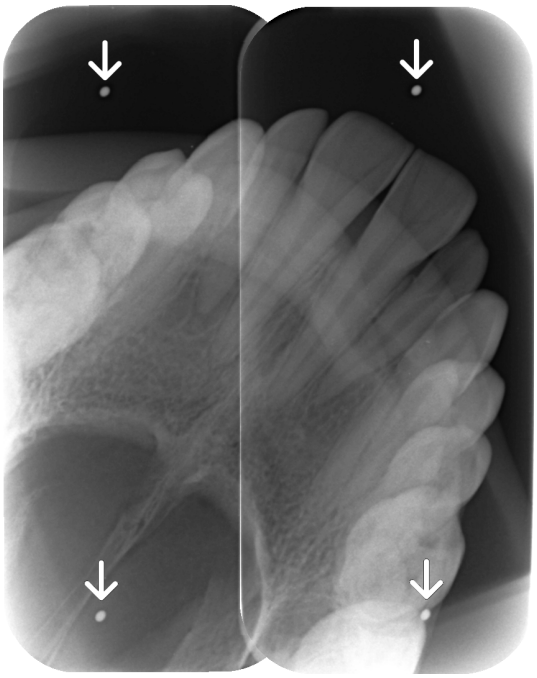


Figure 17: A sample result image of stitching process. Four metal balls cast round shadows into the image (See added arrows). These figures are searched and the distances are calculated. Since metal balls are at fixed distances, the distances of the figures in the image should always be the same. Measuring the distances will reveal the quality of the algorithm.

There were now only four circular regions in the binary image. The center point (center of gravity) of each region was calculated. From the center points, two distances were sought. (These measurements can be seen in image 17, being the euclidean distances between two upper balls and between two lower balls.) The two measurements are not comparable with each other as the metal balls were manually mounted to the the surface of the container and hence were at different distance from each other.

The exact pixelwise distances of the metal balls were unknown due to the manual mounting and the high spatial resolution of 25 pixels per mm. This,

Table 1: Results for distance d_1. Distances between the upper metal balls in the set of images.

distance (pixels)	≤ 499	500	501	502	503	504	≥ 505
count	0	8	4	16	4	1	0

Table 2: Results for distance d_2. Distances between the lower metal balls in the set of images.

distance (pixels)	≤ 521	522	523	524	525	526	527	528	≥ 529
count	0	2	2	11	3	7	4	4	0

however, was considered not to be a problem. If the measured distances would be the same or close to each other in every image, the algorithm would register source images properly.

Let d_1 and d_2 be the measured euclidean distances of the metal balls. Tables 1 and 2 reveal the distribution of the space (in pixels) between the shadows of the metal spheres. The same results are also visible in figures 18 and 19.

Results with registration error of one pixel can still be considered good.[1] Measure d_1 is pretty close to that with the standard deviation of $\sigma_1 = 1.09$. Distance d_2, however, whose standard deviation is $\sigma_2 = 1.99$, doesn't fit in this criteria, and must be considered as a less accurate result.

After examining the result images, it was found that between the dots, from which the distance d_1 was measured, there was practically always teeth and other higher contrast structures in the image. The space between the other dots was occupied by roof of the mouth, which usually lack distinct structures. This might explain the difference in accuracy between the measurements.

The worst result for registration was three pixels away from the average. Considering the resolution of 25 pixels per mm, this would mean a displacement of 0.12 mm in the mouth, which might be tolerable. Otherwise the mutual information method seems to find the right registration decently enough.

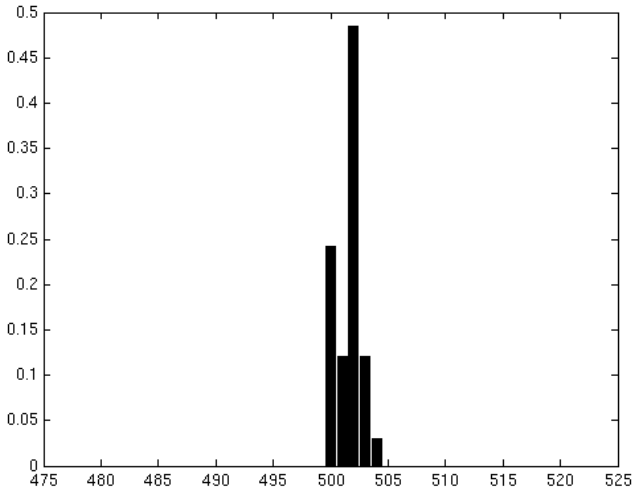


Figure 18: Results for the distance d_1. There is a slight variation in the measure. Standard deviation is $\sigma_1 = 1.09$.

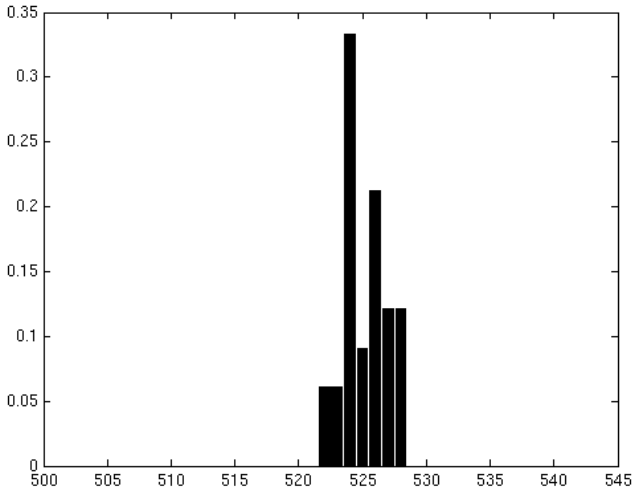
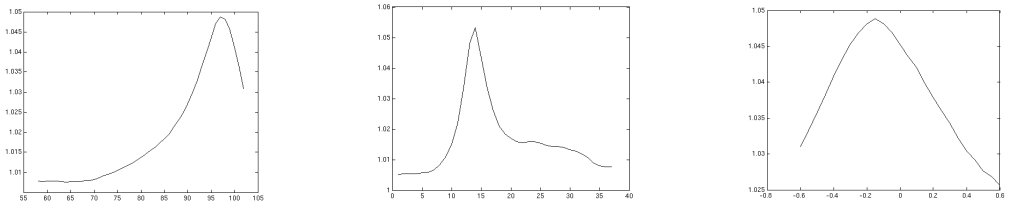


Figure 19: Results for the distance d_2. There is a bit greater variation with this measure than with the distance d_1. The standard deviation is $\sigma_2 = 1.99$.



(a) MI in function of coordinate i

(b) MI in function of coordinate j

(c) MI in function of angle θ

Figure 20: Mutual information in function of different variables. The proper shift and angle between the images was calculated beforehand. Then the mutual information was calculated again when two of the three variables (i, j, θ) were kept constant while the remaining coordinate swept through all allowed values.

6 Discussion and Conclusions

A special-purpose algorithm for registering and stitching two overlapping dental x-ray images, taken by a certain device, was designed and implemented. The amount of overlapping area was fixed.

Image pairs were acquired at the same time, with two phosphor coated imaging plates. Plates were inserted into special container, which would hold them in position during the exposure. Due to the attenuated x-radiation behind the first imaging plate, the intensity values on the second image were notably different from the first one. This is why the mutual information method—originally introduced for multimodal image registration—was chosen. Mutual information method is an area-based, statistical measure of the similarity of the images.

Testing images were obtained by the same device that the software was going to be used with. Total of 33 x-ray image pairs were taken for the testing of the algorithm. Images were taken with various voltage and current settings for the x-ray tube. Images were taken from both upper and lower teeth of one human skull with different positioning for each time to minimize the correlation between the image pairs.

Although the mutual information method itself was well known and easy to implement, the preprocessing of the data caused some extensive work. There were e.g. a small gradients at the edges of the images, which happened to be opposite directed and must be eliminated to void misregistration.

One of the requirements for the algorithm was full automation, so the first step of the algorithm was to identify and tentatively align the images. Also, some image manipulation processes had to be implemented to bypass features of the existing device as they would have reduced the quality of the registration.

Algorithm testing was arranged by inserting four metal balls on the surface of the container which would hold the imaging plates. Metal balls would cast a shadow into the image. The balls were inserted in such locations they would

not participate in the stitching process. Since they were mounted permanently to the surface of the container, they could not move and the shadows in the images should always be at fixed distance from each other on properly stitched result image.

Distances of the shadows were searched from every image. They were compared with each other to gain measurement of the registration of the images. It was found out that on the area, where there are more salient structures in the image, the algorithm was able to register the images properly with the standard deviation of 1.09. However, on the area with less structures, the algorithm performed a bit less accurately. The standard deviation was 1.99. The worst registration error was three pixels. This might be too much on the areas of the teeth, but since these errors emerged only in the areas of the palate, the results could be considered sufficient enough.

The amount of overlapping area of the imaging plates is predetermined to be approx. 0.5 cm. Generally mutual information method gives more accurate results when calculated from larger area as smaller areas are easier to got misregistered. If more accuracy is wanted for the stitching software, one approach might be increasing the overlapping area of the plates.

The gradients at the edges of the imaging plates might also be studied with greater detail, although good results were gained even with the correction by linear approximation.

Details in the under-exposured area of the image were very faint. Even when the mutual information should be quite insensitive to poor signal-to-noise ratio, choosing the parameters—especially the bin count for color quantization—was very crucial.

However, after the right parameters chosen, the performance of the algorithm is satisfying.

References

[1] Ximiao Cao and Quiqi Ruan. A survey on evaluation methods for medical image registration. In *IEEE/ICME International Conference on Complex Medical Engineering*, pages 718–721, 2007.

[2] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3d multi-modality medical image registration using feature space clustering. In *CVRMed '95: Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 195–204, London, UK, 1995. Springer-Verlag.

[3] Eastman Kodak Company, http://www.kodakdental.com/film-and-anesthetics/extraoral-film//media/F5EBCCD33F2F4B4892C5D1119286A98B.ashx. *Kodak Dental Extraoral Film Screen Combinations*. Visited 31st March 2009.

[4] Otto Glasser. *Wilhelm Conrad Röntgen and the Early History of the Roentgen Rays*. Norman Publishing, San Francisco, 1993.

[5] R. V. L. Hartley. Transmission of information. *Bell System Technical Journal*, 7:535–563, 1928.

[6] David L. Heiserman. *Fundamentals of Dental Radiology*. http://www.waybuilder.net/sweethaven/MedTech/Dental/DentalRad/. Visited 19th August 2009.

[7] J. Lin, Z Gao, B. Xu, Y. Cao, and Yingjian Z. The affection of grey levels on mutual information based medical image registration. In *26th Annual International Conference of the IEEE EMBS*, pages 1747–1750, 2004.

[8] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[9] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual information based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[10] F. Raimondo, M. Gavrielides, G. Karayannopoulou, K. Lyroudia, I. Pitas, and Kostopoulos I. Automated evaluation of her-2/neu status in breast tissue from fluorescent in situ hybridization images. *IEEE Transactions on Image Processing*, 14(9):1288–1299, 2005.

[11] Jukka Rosberg. *Hammaslääketieteellinen radiologia—Tekniikka ja diagnostiikka*. Oulun yliopistopaino, Oulu, 2000.

[12] S. Sanjay-Gopal, H.P. Chan, T. Wilson, M. Helvie, N. Petrick, and B. Sahier. A regional registration technique for automated interval change analysis of breast lesion on mammograms. *Medical Physics*, 26(12):2669–2679, 1999.

[13] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[14] M. Sonoda, M. Takano, J. Miyahara, and H. Kato. Computed radiography utilizing scanning laser stimulated luminescence. *Radiology*, 148(3):833–838, 1983.

[15] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[16] Michel M. Ter-Pogossian. *The Physical Aspects of Diagnostic Radiology*. Harper & Row, New York, Evanston, and London, 1967.

[17] P. Viola and W.M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[18] Weisstein Eric W. *Least Squares Fitting.* MathWorld—A Wolfram Web Resource, http://mathworld.wolfram.com/LeastSquaresFitting.html. Visited 19$^{\text{th}}$ August 2009.

[19] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

Appendix A: The Flow Diagram of the Algorithm

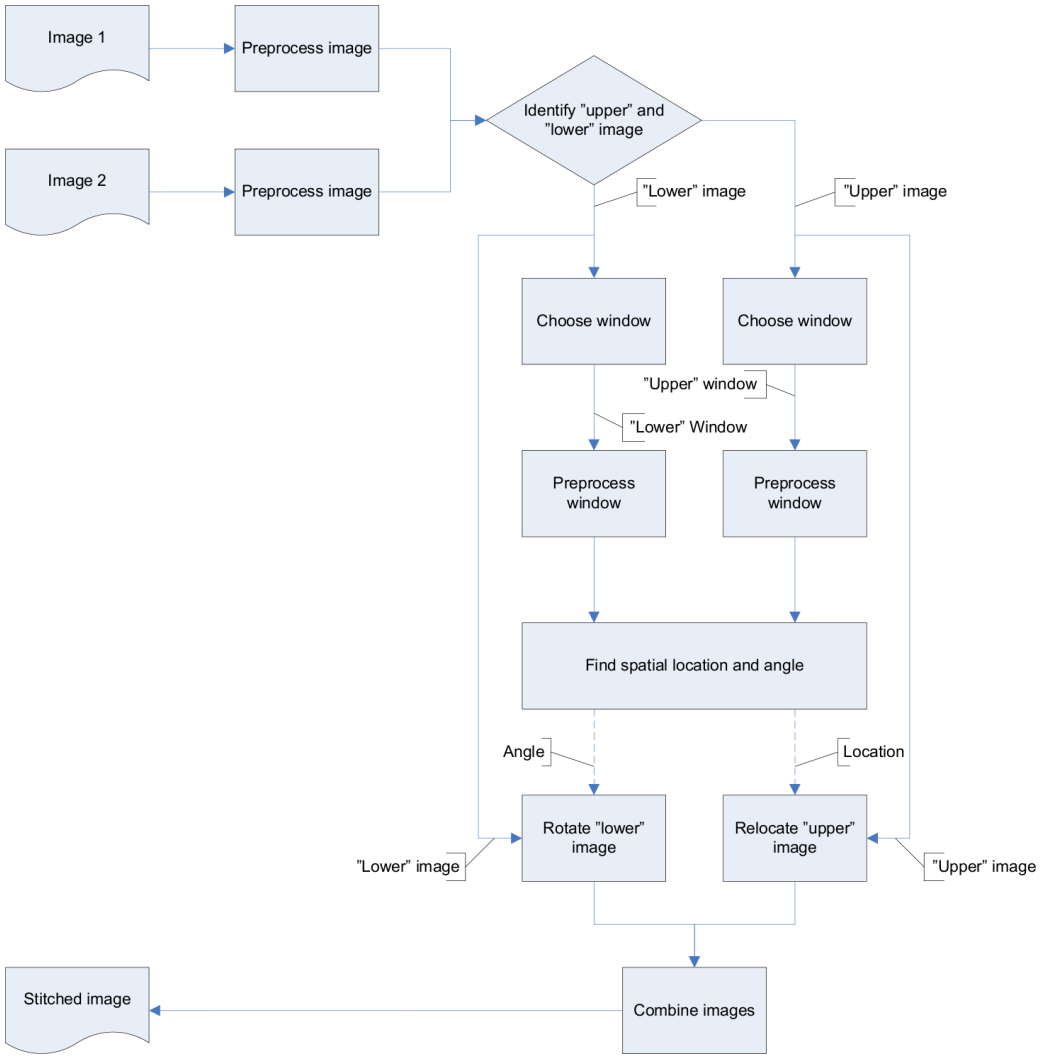


Figure A1: The Flow Diagram of the Algorithm.