Respiratory Motion Correction in PET/CT with Gated Data

J.Eugenio Iglesias

December 9, 2005

Abstract

The introduction of the combined PET/CT has been one of the major achievements in the last decade in cancer diagnose. The hybrid PET/CT, being two separate detectors connected in conjunction and sharing the same bed, can hence acquire CT images pre- or post acquisition of the PET data, but suffers from the fact that the acquisitions are not done fully simultaneously. Furthermore, while the acquisition of CT is done in a fraction of second, the PET acquisition is done over a period of approximately 20 minutes, during which the patient breathes, resulting in a certain diaphragmatic motion and hence image blurring. If the obtained images are then fused, they are not fully co-registered.

The purpose of this master's thesis is to evaluate a novel method for correcting of this unwanted blurring by means of a software correction. CT images are acquired for several respiratory phases and co-registered using the deformable "demons" algorithm. The PET data is also gated, and the contribution from every phase to the image is modified with transforms calculated from the CT images so that they are all referred to the same phase (an "optimal" one, in which with most injury to the tumor and least to the healthy tissue might be achieved when applying radiotherapy).

A superposition of gated PET images transformed to a reference phase according to parameters extracted from CT data, and an uncorrected version, estimated as the average of the gated images, were compared. Not that many sample images were available, though, and they needed to be aligned by software (as the CT and PET were acquired by different machines).

The results show that the method partially corrects the tumor motion. The lesion's center of gravity position error is reduced between a 60% and a 70% and the lesion's volume increment in a 33%, approximately. The maximum voxel value (proportional to the standard uptake value, SUV) is also partially corrected, as the a tumor's maximum comes back to around 4300 after having fallen from 4500 in the gated, sharp images to around 4000 in the blurry ones. The results are expected to improve when the images are acquired by a combined PET/CT scanner and thus, "hardware" aligned.

Acknowledgements

I would like to thank my parents in first place for their support during my studies, as well as the rest of my family. I would also like to thank Sharok Kimiaei for giving me the opportunity of writing this master's thesis, and also Anders Brahme for accepting me in the Radiation Physics department at the Karolinska Hospital.

I cannot skip all those people who selflessly helped me with my job, among them E.Damen and J.Wolthaus for providing me with their sample images and W.Segars for his phantom.

And last, and not least at all, I would like to say thank you so, so much, to all the nice people that have made me feel like at home during this two years in Sweden. I could write a two kilometers long list, but you know perfectly who you are. Tack very mucho!

Contents

Chapter 1

Introduction

This chapter relates this master's thesis's topic to the current medical imaging background. The basics of the image types used in this work are introduced and discussed. The concept of "image resgistration", which is one of the most important tools used in this work, will be introduced as well. Special care will be taken when using medicine concepts, as this thesis is also meant to be read by non-medicine-specialized engineers.

1.1 Background

Cancer is one of the main death causes in industrialized countries. According to the US's Centers for Disease Control and Prevention [1], cancer ranks second only behind heart diseases in death causes in that country. These two death causes (cancer and heart diseases), with 685000 and 555000 deaths, lead with a big difference against those that come next: cerebrovascular and respiratory diseases, with 158000 and 126000 deaths respectively.

Just within cancer, this year's American Cancer Society's statistics [2] reveal that lung and bronchus cancers represent 31% of the cases among men in the US, and 27% among women, being by far the most usual kind of cancer case, well ahead of breast cancer in women (15%).

It is thus a matter of capital importance to improve all the ways of fighting against this disease. Medical images of the thorax let physicians diagnose this type of cancer and are also very useful in dose planning in case of external radiation therapy being applied as a treatment. Radiotherapy is the typical option when surgery is not feasible in order to remove a tumor, usually because of the proximity to other vital-importance organs [3].

Conformal radiotherapy is a complex process that begins with the creation of an individualized, digital data set of a patient's tumors and normal

adjacent anatomy. This data set is then used to generate computer images and to develop complex plans to deliver highly conformed (focused) radiation to the tumors while sparing normal adjacent tissue. Because higher doses of radiation can be delivered to cancer cells while significantly reducing the amount of radiation received by surrounding healthy tissues, the technique should increase the rate of tumor control while decreasing side effects.

Sharp, high resolution images help the physicians delineating the tumor more precisely and hence taking the best decision for the patient in terms of which treatment should be applied. They also let the specialists design a very efficient conformal radiotherapy plan.

1.2 Medical Images

Medical imaging modalities can be divided into two subgroups: anatomical and functional. The goal of both of them is to provide physicians with an image of the internal organs in a non-invasive fashion. While the first one (for instance, Computerized Tomography, CT, or magnetic resonance imaging, MRI) is a representation of the internal anatomy of the body, the latter one (like positron emission tomography, PET) marks the places in the patient's organism with a higher biologic activity. In this thesis, CT and PET images will be utilized. These two modalities are now further explained.

1.2.1 CT

CT (computerized tomography) imaging is the evolution of X-rays [4] [5]. Classic X-ray imaging is a two-dimensional representation of a three-dimensional body part. When an x-ray beam comes into contact with a body part, it can travel through it or be absorbed by it. The beams that exit the patient's body will interact with a film and generate an image on it. The image intensity on each point depends on the properties of the media (all the different layers) the beam travels through and interacts with until reaching the film (apart from the film's detecting properties, of course). A typical x-ray image is depicted in figure 1.1).

Hence, the final image is a summation of anatomic shadows, depending on thickness, form, and atomic number of the different tissue layers the x-ray beam has travelled through. The main disadvantage is the loss of depth and contrast, and that opaque structures can hide the tissue behind them. That is the reason why different images from different perspectives are usually acquired in order to provide the physician with enough information to make a correct diagnose.

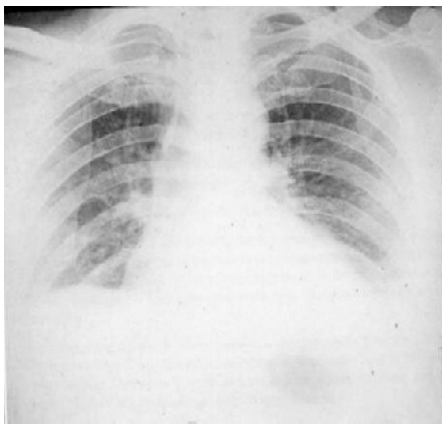

Figure 1.1: Typical x-ray image from the thorax.

In CT, in order to study a thin slice (a volume can then be constructed by stacking slices acquired at different planes), a collimated beam from the radiation cone from the X-ray tube is used. Through this tomographic technique, it is possible to project just a thin slice of the body part to study. The detected one-dimensional signal is saved. Both the x-ray tube and the detector rotate then around the patient (as shown in figure 1.2), acquiring multiple linear projections. These projections allow a computer to reconstruct the planar two-dimensional image in the studied plane, as it is explained in section 1.2.3.

1.2.2 PET

PET (positron emission tomography) imaging [4] [5], is a nuclear medicine technique in which a radiopharmaceutical product is injected into the patient. The most spread one is 18-fluorodeoxyglucose (18-FDG), which is a tracer composed of radioactive fluorine (18-F, with a half-life of 110 minutes) and a sugar (deoxyglucose).

18-FDG has two important properties. One is that it is a radioactive product that emits positrons that annihilate with electrons in the tissue in a relative short range (1-3 mm; this is a limiting factor for the resolution). This process results in two gamma photons being emitted in opposite directions with an energy of 511 keV each. It is perfectly possible to detect this gamma photons. The second property is that 18-FDG is taken up and retained by tissues with high metabolic activity, such as the brain, the liver, and most types of malignant tumors.

The two described properties are the base of the PET imaging technique (with 18-FDG). After waiting for approximately 45 minutes, so that 18-FDG

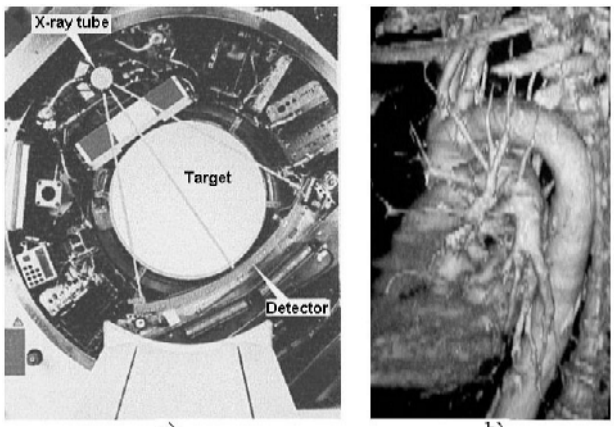

Figure 1.2: a) CT detector and b) CT image from the aorta artery: a 3-D volume is built by stacking several transverse slices and a vertical slice is then extracted.

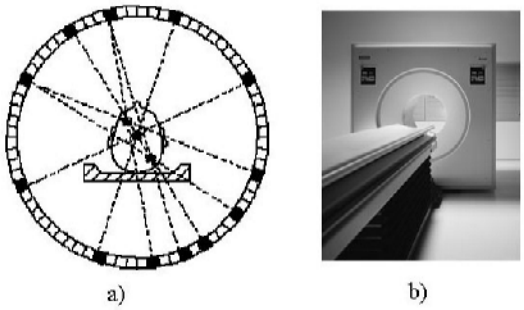

Figure 1.3: a) PET detector principle and b) commercial PET detector from Siemens

gets concentrated around tissues with high metabolic activity, the patient is placed into the PET scanner, which consists of one or more rings of detectors. Every time that a pair of gamma photons are detected in coincidence in the rings, a positron must have annihilated in the straight line that joins the two detection points. That means that there has been an annihilation in that line (figure 1.3). Different reconstruction techniques take us back from coincidence pairs to a three-dimensional image representing the metabolic activity in a volume in the patient's body. Figure 1.4 shows an example of a PET image.

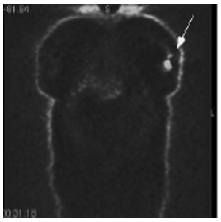

Figure 1.4: PET image from a breast cancer

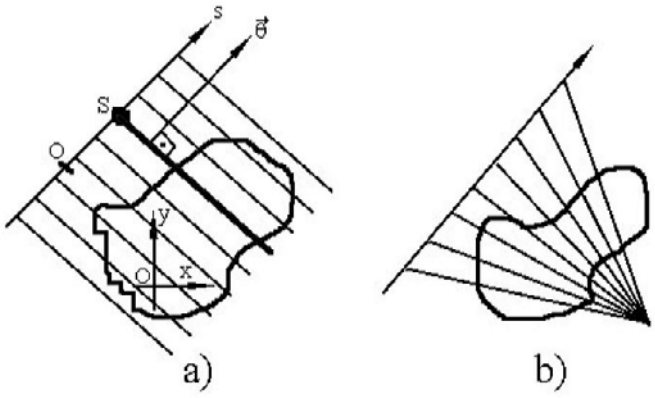

Figure 1.5: a) Illustration of the Radon transform and b) Fan-beam geometry

1.2.3 Reconstruction techniques

The mathematics of the 2D reconstructions from a set of projections are based on the Radon transform [6]. This transform maps a two-dimensional spatial function to its projections along lines in all the possible angles (figure 1.5):

$$p(\vec{\theta}, s) = \int_{\mathbb{R}^2} d\vec{x}\delta(\vec{x} \cdot \vec{\theta} - s) f(\vec{x})$$

The reconstruction problem consists of finding back $f(\vec{x})$ from $p(\vec{\theta}, s)$. Some methods can be found in [4]. They can be classified into direct Fourier methods, signal space convolution and frequency space filtering, iterative methods and series methods. FBP (filtered back-projection) can be cited as one of the most popular ones. The lack of the complete set of line integrals (due to discrete nature of the machine that acquires the projections) leads to inaccuracies in the reconstructed image, also due to non-linearities and insufficient data [7].

There are some other more modern techniques for reconstructing the image. The expectation maximization (EM) algorithm for PET images [8] is

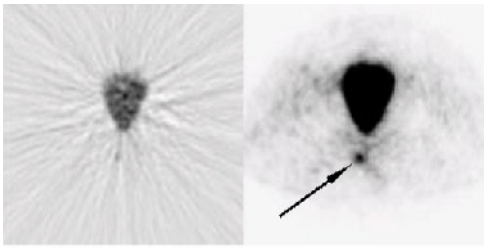

Figure 1.6: a) FBP reconstruction of an image showing a carcinoma and b) OSEM reconstruction of the same data. The artifacts in the FBP one makes it difficult to notice the lesion

an iterative procedure that takes into account the stochastic nature of the positron emission process. It has some interesting properties:

1. There are modified versions for accelerated convergence (like ordered subsets, described below).
2. The images produced by the algorithm are non-negative (negative emission densities are meaningless).
3. The algorithm is self-normalizing in the sense of that the sum of the events in the image is equal to the sum of the counts in the data.
4. It allows the incorporation of many physical factors, such as:
   (a) Attenuation correction information.
   (b) Accidental coincidence corrections.
   (c) Time-of-flight information, that is, the difference in time of arrival of the gamma particles for a better estimation.
   (d) Variation in spatial resolution.

An improved version of EM is the OSEM (Ordered Subsets EM) algorithm [9]. EM has the problem of being computationally expensive. OSEM, on the other hand, divides the data within each iteration in blocks, which accelerates the convergence by a factor proportional to the number of subsets. A comparison between a classic FBP PET reconstruction and a OSEM one is shown if figure 1.6.

In the case of CT images, there are also improvements to the FBP algorithm. They are not based on the emission's statistics (as there are not emissions any longer), but on trying to reduce motion artifacts [10].

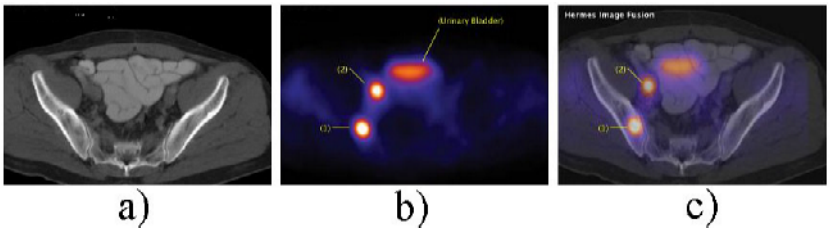

Figure 1.7: Prostate cancer images a) CT image: there is no track of cancer b) PET image: there are signs of tumor c) PET-CT image: the situation of the lesion is now clear.

The complete discussion of reconstruction techniques is outside the scope of this thesis. However, it is important to remark that it is also possible to reconstruct 3-D volumes both in PET and CT imaging. The third dimension can be obtained by stacking adjacent transverse sections [11]. There are also algorithms that consider information from other slices [12].

1.2.4 Image fusion

The two imaging techniques, functional and anatomical, are a complement to each other: while a structural image represents the proton density, hence imaging the anatomy,a functional one can provide information about physiology that the first one lacks.

This is the reason for hybrid modalities like PET/CT or PET/MRI increasing popularity: a combined image image (which is created by superimposing two images) provides the physicians with both anatomical and physiological data, improving the detecting ability of malignant lesions and staging treatment of cancer. In these techniques, two different images are combined either by software or hardware. An example of a hybrid image is shown in figure 1.7.

The hybrid PET/CT, being two separate detectors connected in conjunction and sharing the same bed [13], can hence acquire CT images pre- or post acquisition of the PET data. Even though the accuracy of the acquired images is substantially superior to those of software co-registered fully separated modalities, they suffer from the fact that the acquisition is not done fully simultaneously. Furthermore, while the acquisition of CT, specially with the last generation multi-slice CT:s with up to 64 slices, is done in a fraction of second (for example in breath hold position, where the separation between internal organs is in average larger, the image usually clearer, and a more efficient radiotherapy dose plan can in general be designed), the PET acquisition on the other hand, due to the detection geometry and the inherent

acquisition nature of PET imaging, can not be done instantly.

Once the medical imaging modalities involved in this thesis have been presented, the most used image processing technique in this project, image registration, is introduced.

1.3 Image registration

The image registration concept will now be introduced and the different involved steps reviewed. Some methods and the possibilities of applying each of them to PET and CT images will be studied.

1.3.1 Introduction

Image registration is the process of relating two or more images of the same scene taken in different conditions: different viewpoints, different sensors, different times, or perhaps a combination of more than one of them. Another discipline consists of registering an image to a template, which is known as "template matching". The idea is to find the mathematical transformations that convert one image ("sensed" or "moving" image) into the other ("reference" image).

One of the possible formal definitions of image registration is "a mapping between two images both spatially and in pixel/voxel intensity values". If the images are assumed to be three-dimensional, this can be expressed:

$$I_2(x, y, z) = g(I_1(f(x, y, z)))$$

The function f represents a transformation for the volume elements' (voxels) positions and g is an one-dimensional intensity mapping. This intensity transformation is not necessary in many cases, but is still of big importance in the case of different sensors, as the typical voxel values in each image will in general be very different. It is for example very important in the case of registering a PET and a CT image.

An example of a transformation function could be, in 2-D:

$$I_2(x, y) = 2 \cdot (I_1(x - 20, y))$$

where the image has been shifted 20 pixels in the x-axis direction and made 2 times brighter, as shown in figure 1.8-a. In 1.8-b, a CT-PET example is shown. The spatial deformation f cannot be expressed as a simple mathematical function, and the intensity mapping function g is not trivial either.

Registering a pair of images is a way of comparing them. The need for this comparison has arisen in may different problems in different fields. One

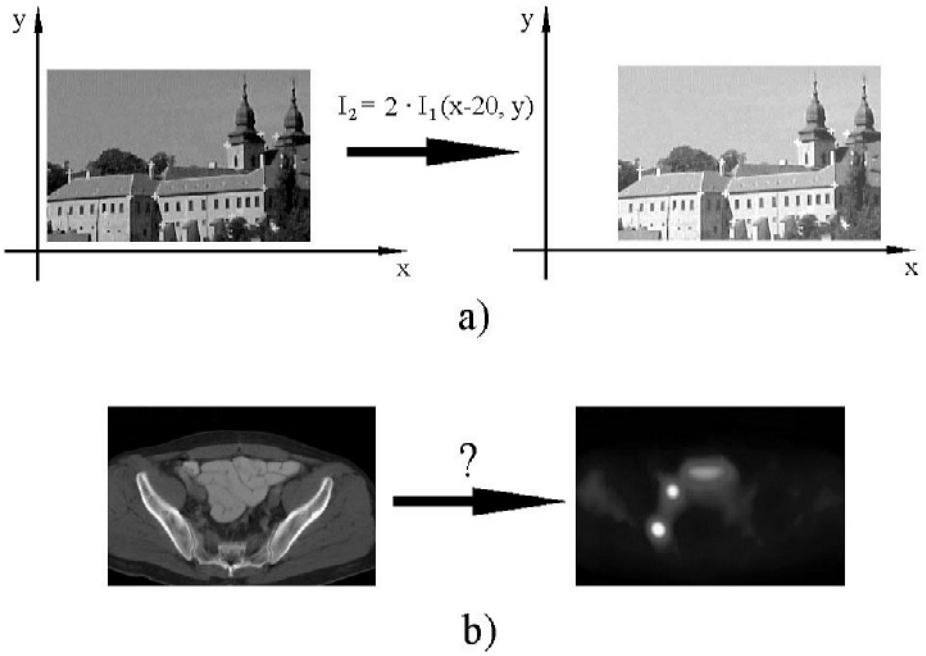

Figure 1.8: Mapping function in two cases: a) simple translation and intensity amplification and b) complex CT-PET registration problem

may for example want to integrate the information acquired by two different sensors (for example, a CT and a PET image, as already discussed). It may also be desirable to find the spatial changes between two images taken at different times to characterize the movement (something also done in this thesis). Interested readers may further read [14] and [15].

Independently of the application, the majority of the registration methods consist of the following four steps [16], which are illustrated in figure 1.9 (from the same source):

1. Feature detection: distinctive objects, such as edges, contours or corners, are manually or automatically detected. For further processing, these objects are represented by their point representatives (endings of lines, centers of gravity), which are called control points (CPs) in the literature. In an extreme case, every pixel in the images can be a CP.

2. Feature matching: the correspondence between the features detected in the two images is determined.

3. Transform model estimation: the type and parameters of the function that "translates" the sensed image's features into the reference image's

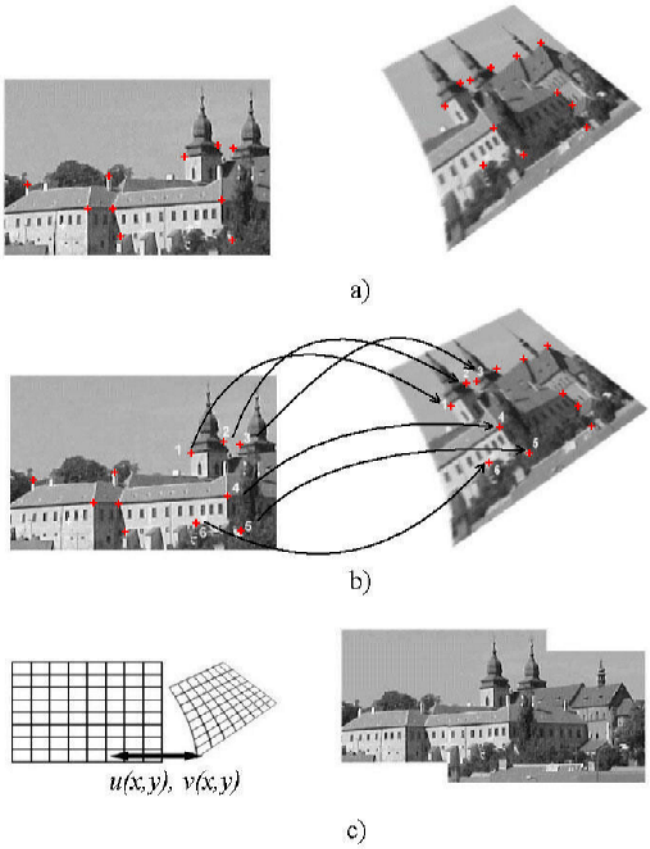

Figure 1.9: Registration steps: a) feature detection b) feature mapping c) transform model estimation and image transformation

ones are estimated. Different deformation functions and similarity measures to maximize can be used to achieve it. The idea is to try to find a function that depends on a certain number of parameters and that can potentially represent the transformation between the two images. The values for these parameters that maximize a similarity function (based for example in the sum of the squared voxel differences, or in the crossed correlation) are then sought.

4. Image resampling and transformation: the sensed image is finally transformed by the mapping function. Image values in non-integer coordinates must be interpolated.

Apart from the four steps above, each registration method can be seen as a combination of choices for the following components [17]:

1. A feature space: e.g. voxel values, edges...

2. A search space: the function that maps one image to the other.

3. A search strategy: the different ways of trying to find out the best parameters for the transformation function.

4. A similarity metric: a quantitative measure of how similar two images are. The search strategy's goal is to try to maximize this value for the fixed image - registered image pair.

The following section describes these different components in more detail.

1.3.2 Registration components

Feature space

The first step when registering images is to decide the feature space that will be used. As previously mentioned, edges, contours, surfaces, corners, points... may be used. The decision of the feature space can affect the registration's performance dramatically. For example, if the case of having two images that are exactly equal but with scaled voxel intensity levels, a raw pixel comparison would result in a huge difference, whereas if the images' edges were detected in a preprocessing step, the results would be almost equal. That is why one must consider which features remain in common in the two images. The decision can also be influenced by possible limitations in the computational cost: need to preprocess or not, number of detected features to compare...

Features can be extracted both manually or automatically. The latter is of course preferred, but is not always possible or reliable. In the manual or semi-manual case, an expert user must choose equivalent structures or points in both images manually. One of the problems with manual extraction is that it is irreproducible. Slightly different selections may lead, especially if the algorithm is not robust, to completely different results.

In automatic detection, in order to help the computer to make its work as good as possible, image enhancement techniques can be used prior to the the process's start. For example, image smoothing can be used to remove high-frequency noise, or image sharpening can be used to highlight edges if they are supposed to be detected afterwards. However, it is difficult to know the filters' parameters (for example, cut-off frequencies) a priori, automatizing the enhancement process.

The simplest feature detection technique consists of directly working with the **raw intensity data**. This is the feature space that contains most information. It saves computations in the preprocessing step, but usually leads to a more costly matching process, especially if all the voxels are considered in the calculations. Pyramidal processing, which consist of working with downsampled versions of the image to estimate solutions at higher resolutions and thus save iterations, can help to speed the process up. This will be further explained and discussed later.

Another possibility for detecting features is to extract **region features**. They are usually projections of high contrast closed-boundary regions that can be detected through segmentation. The regions may be represented by their centers of gravity, which have the good property of being invariant to rotation and scaling. This representation is also very robust against noise and intensity variation. This technique will only lead to good results with images from the same modality, though.

Another type of feature space is the one of **line features**. Object contours, elongated anatomic structures, roads, etc are well represented by lines. Edge detection algorithms are used in order to detect them, and they represent much of the intrinsic structures of an image. Another advantage of this scheme is that they are quite robust against several variations (rotation, translation, scaling...). Once the edges are detected, it is possible to choose prominent features that are easily distinguished, such as corners, line intersections or points of maximum curvature.

There are yet another kind of features: **point features**. They usually provide the user with accurate positioning. One must be careful with the number of detected points though, because the computational costs increase along with them. There are several proposed methods for an efficient selection of a subset of points [18] [19] [20].

It is also possible to use **higher level features**, that are good for inexact matching (for example, when different types of images are registered). Examples of these are structural features, graphs of pattern configurations [21], or syntactic features, like grammars composed from patterns [22].

Finally, it is worth mentioning that frequency domain features are not considered in this master's thesis. The reason is that they do not offer any kind of spatial local possibilities.

Search space

The spatial transformation from the sensed image to the reference one is assumed to be performed according to a model, which depends on a set of parameters. The search space is the class of transformation parameters from

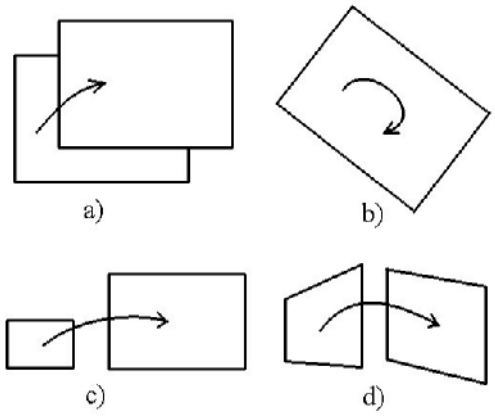

Figure 1.10: Typical distortions: a) translation b) rotation c) scaling d) change of viewpoint

which the optimal values to align the images are searched.

The main goal when choosing the search space is to select the type of transformation that can compensate the spatial distortion between the sensed and the reference image. A priori knowledge of the kind of phenomena that has originated the distortion (a rotation, a translation, a change of viewpoint..., view figure 1.10) can result in a reduction of complexity of the search space. If one knows for example that an image is a translated version of the other one, a simple three-parameter translation transform would be suitable and give satisfactory results. Otherwise methods that do not make any assumption about the kind of deformation are required. These methods will require in general a much larger computation time.

A first general classification distinguishes between global and local transformations. In the first case, a single equation maps the entire image. In the second one, the mapping function depends on the spatial location. This can handle locally deformed images. If a global method is applied on one of this images, it will just average the local geometric distortion over the whole image. Local methods can achieve much more powerful transformations. In some cases, the best approach is to find a global transformation and then refine it with local transformations in the places where the result is not sufficiently good.

One of the simplest global models for two-dimensional images is the **similarity transform**:

$$u = s(x \cdot cos\theta - y \cdot sin\theta) + t_x$$

$$v = s(x \cdot sin\theta - y \cdot cos\theta) + t_y$$

where (u, v) are the transformed coordinates from (x, y). This transformation is capable of handling rotation (of an angle θ), translation (t_x in the horizontal direction and t_y in the vertical one) and scaling (with a factor of s). Three control points (six equations) are required to determine the parameters. More points are usually taken into a least square fit in order to make the algorithm less sensitive to errors in the control point positions.

A more complex transform would be the **affine transform**, which is capable of mapping a parallelogram onto a square:

$$u = a_0 + a_1 x + a_2 y$$

$$v = b_0 + b_1 x + b_2 y$$

All these methods can be generalized to higher dimensions. An even more complex method consists of using **higher order polynomials**, which can handle more complicated deformations. Given a set of control points, it is possible to adapt the coefficients of the polynomials to minimize the error, usually measured in a least square fashion. As explained above, it is always preferable to have as much data as possible in order to have more reliable statistics. It would be possible to have just twice as coefficients as control points (each point leads to two equations, still assuming 2D), but this would lead to very high order polynomials with several undulations that would create artifacts in the image. These transformations can in general be expressed as:

$$u = \sum_{i=0}^{m} \sum_{j=0}^{i} a_{ij} x^i y^{j-i}$$

$$v = \sum_{i=0}^{m} \sum_{j=0}^{i} b_{ij} x^i y^{j-i}$$

When two images are misaligned by an unknown transformation, (which is the case in this thesis), one is forced to use yet another kind of mappings. This happens, for example, when the transformation includes complex deformations (that would require a too high order if one tries to use polynomials). If the distortions become local, it is difficult to perform a satisfactory registration with a global mapping function. All of the above described global methods can however be applied piecewise in order to create local transformations.

It is usually desirable to constrain the global transformation to be smooth. This can be done through introducing a penalty term in the cost function

that determines the quality of the transformation. This penalty term will be higher when the transformations parameters change rapidly. An illustrative example for spline interpolation can be found in [23], and it looks like:

$$v = \frac{1}{V}\int_0^X \int_0^Y \int_0^Z [(\frac{\delta^2 \vec{T}(x,y,z)}{\delta x^2})^2 + (\frac{\delta^2 \vec{T}(x,y,z)}{\delta y^2})^2 + (\frac{\delta^2 \vec{T}(x,y,z)}{\delta z^2})^2 + \dots$$

$$\dots + 2(\frac{\delta^2 \vec{T}(x,y,z)}{\delta x \delta y})^2 + 2(\frac{\delta^2 \vec{T}(x,y,z)}{\delta y \delta z})^2 + 2(\frac{\delta^2 \vec{T}(x,y,z)}{\delta x \delta z})^2]dxdydz$$

The main point of this penalty term is that it gets larger as the absolute values of the derivatives increase, that is, as the parameters tend to change quickly.

The last approach that will be discussed is the most general one: not using any parametric mapping functions at all. This is often called **elastic registration**. The images are regarded as pieces of rubber in which external forces stretch the image trying to bring it into the reference one with the minimum possible amount of deformation. The result will be a different displacement vector for every voxel in the image, without fitting any preestablished model. The values of these displacement vectors are usually calculated according to laws from physics. There are for example methods based on diffusion processes, optical flows... The voxel values are usually interpreted as potential functions.

A very powerful registration method is the "demons" algorithm, by Thirion [24]. In this model, the object boundaries in the reference image are considered semipermeable membranes and the sensed image is considered deformable. The voxels in the latter diffuse through these interfaces by the actions of effectors (Maxwell called them demons to explain concepts about thermodynamics) situated within the membranes. The optical flow equation is used to estimate the forces in this points. For a point P, if s is the intensity in the static image S and m the intensity in the moving image M, the displacement $\vec{u}$ required for P to match the corresponding point in M is:

$$\vec{u} = \frac{(m-s)\vec{\nabla} s}{|\vec{\nabla} s|^2 + (m-s)^2}$$

Thirion proposed in his work [24] three different algorithms. In the first one, every voxel is a demon. In the second one, demons are only present in contours. The third one works with already segmented images. As expected, the first one provides the best results, as the displacement force is calculated for every single image element, but at the cost of computational

power. An improvement to this algorithm is introduced in [25], where an active force based on the moving image's gradient information and a multiresolution approach are introduced, making the convergence much faster and saving computation time.

$$\vec{u} = \frac{(m-s)\vec{\nabla}s}{|\vec{\nabla}s|^2 + (m-s)^2} - \frac{(s-m)\vec{\nabla}m}{|\vec{\nabla}m|^2 + (s-m)^2}$$

Search strategy

Due to the large computational costs associated with most of the registration techniques, it is often unfeasible to go along all of the possible combinations in the feature space searching the best one. This process would always lead to the optimum solution (lowest cost or maximum similarity according to a established criterion), but this is only affordable in very simple cases. For example, if one intends to just adjust a translation, all displacements over a reasonable range of distances can be tested. The bigger the feature space (more control points and more model parameters), the more complex the search becomes.

There are many different search strategies, each one with its advantages and disadvantages. Some of them can even be used in conjunction with another one. The choice of the search strategy is based on the deformation model, as well as it can be influenced by computational and storage limitations. Some of the most popular techniques (mostly based on gradient descent algorithms) are briefly described in the ITK toolkit documentation [26], and are implemented in this package. Some of them will be tested in this master's thesis work.

Similarity metric

The choice of the similarity measure is closely related to the selection of the matching features, as their alignment is the registration process's goal. Some measures focus on spatial differences between the locations of the features in each image, while others try to match high-level structures, for example graphs.

The simplest voxel intensity based cost function is the **squared sum of intensity differences (SSD)**:

$$SSD = \frac{1}{n}\sum (I(t_0) - \mathbf{T}(I(t)))^2$$

where $I(t_0)$ denotes the reference image and $I(t)$ the sensed one, while n is the total number of voxels. This function has the advantage of being

computed very quickly, but it is only valid for images from the same modality (sensor) with properly normalized intensities. It is appropriate for finding matches with no local distortions. The **normalized correlation coefficient (NCC)**, slightly more computationally expensive, is not tolerant to local distortions either. Its peak is somehow harder to find than the SSD's, but it has the advantage of allowing a linear relationship between the images' intensities:

$$NCC = \frac{\sum(I(t_0) - \overline{I}(t_0))(\mathbf{T}(I(t)) - \mathbf{T}(\overline{I}(t)))}{\sqrt{\sum(I(t_0) - \overline{I}(t_0))^2 \sum(\mathbf{T}(I(t)) - \mathbf{T}(\overline{I}(t)))^2}}$$

where $\overline{I}(t_0)$ and $\overline{I}(t)$ refer to the average values, and $T()$ represents the transform. Another possibility is to measure the **entropy of the difference image**, that has the advantage of each pixel having the same weight independently of its intensity. This represents and advantage over the previous two methods, in which a single pixel with a large difference in intensity could contribute too much to the cost function. On the other hand, it is computationally more complex.The cost function to be minimized is:

$$H(s) = -\sum_x p(x) log[p(x)]$$

where x represents the different intensity values in the difference image, and $p(x)$ is the discrete probability distribution function, that is, the number of voxels of value x divided by the total number of voxels. If the image becomes zero or constant (the ideal case, perfect registration achieved), its entropy would be zero.

Another method based on entropies is the **mutual information** one. This method is appropriate in inter-modality registration, as a linear relationship between the pixel values is not assumed any longer:

$$H(s) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

where x refers to the intensity values in the reference image and y to those in the sensed image. In this case, $p(x,y)$ represents the joint discrete probability distribution function, that is, the number of voxels that have a value of x in the reference image and y in the sensed one, divided by the total amount of voxels.

There are also more complicated cost functions. Some of them include penalty terms for transformation parameters discontinuities, some of them

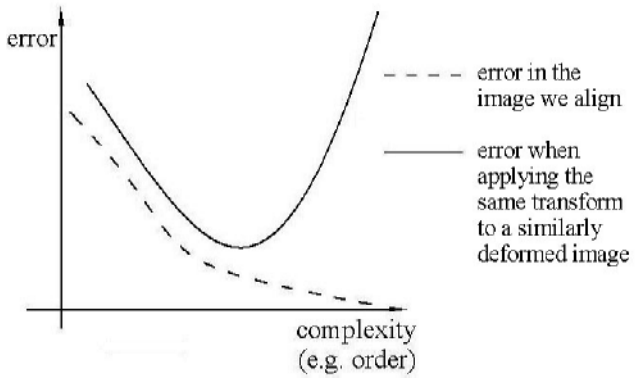

Figure 1.11: Typical error-model complexity curve shape

use the images' gradients, other functions consider the number of sign changes... Some of the most popular for medical image registration are reviewed in [27].

1.3.3 Practical considerations

There are some practical issues that deserve being remarked. The first one is that one must be careful with the transform's flexibility. A too complicated transform can lead to undesired results, as it may be capable of creating false structures and other artifacts.

If the same very complex transformation was applied to another image physically deformed in the same way (somehow equivalent to the original moving image), the results may not be what one could expect (that is, something equivalent to the fixed image). As that is exactly, as it will be explained afterwards, the goal of this thesis (applying the transformations estimated from a CT to PET images), one must be careful with this phenomena.

The typical error behavior with increasing complexity of a model is shown in figure 1.11. The alignment in the data always improve with the complexity, assuming a perfect optimization process. If polynomials were used, for example, the results would always improve if the order is increased (assuming a perfect coefficient optimization, as already mentioned). However, there is a point beyond which the new data's fitting starts to impair, due to "overfitting" to the original data. Care of trying not to go beyond this point will be taken in this thesis.

The following example clarifies this effect. Assuming that two images of the same pair of lungs from the same patient are available, a transformation that maps every point from the left lung to its symmetrically equivalent point in the right one (assuming that the lungs are symmetrical, which they are

not) could be found. The result would give a high similarity between the fixed and registered images, but would be far from representing the reality: that there is no movement between them. If this transform was applied to a corresponding PET image showing a tumor in the right lung, the tumor would actually be moved to the left lung.

Another practical consideration is that it is a rule in image registration to start with a simple model and increase the degree of complexity as required. Is is better to increase the complexity level stepwise rather than applying a complex transform initially. Sometimes it also helps to perform an image enhancement preprocessing prior to the registration algorithm. A gaussian low-pass filter may for example be useful to smooth the error, reducing the noise and thus and improve the final result. However as mentioned previously the filter parameters may be difficult to set automatically.

It is also important to mention that the similarity measure used in the registration process may not to be the one used in the evaluation of the transformation. It is possible to use a more complex but also more accurate one that one cannot afford calculating for every optimization iteration.

And last, but not least, the importance of pyramidal processing must be mentioned. The complexity of matching two images grows with an order of $O(n^3)$ or even $O(n^4)$ in some cases, depending on the used algorithm. Downsampling the image to quickly calculate the transformation parameters and use this solution as an initial value at the next resolution would improve the algorithm's efficiency. As 512x512 slices will be used, it would for example be possible to begin with 64x64 and going on with 128x128 and 256x256 before finishing with the original 512x512 resolution.

Pyramidal processing has also another big advantage: preliminary results are quickly available. That lets the doctor perform an initial evaluation of the results while the calculations in the higher resolutions levels are still running.

A last important advantage of pyramidal processing is that it is much more difficult to fall in a local minimum of the cost function. This function is somehow smoothed when the resolution is made lower, so the local minima tend to disappear in the lower pyramidal levels. This is illustrated in figure 1.11. If registering in these first levels leads to a point close to the global minimum, using this point as starting point in the higher levels will not lead to a local minimum, but to the global one.

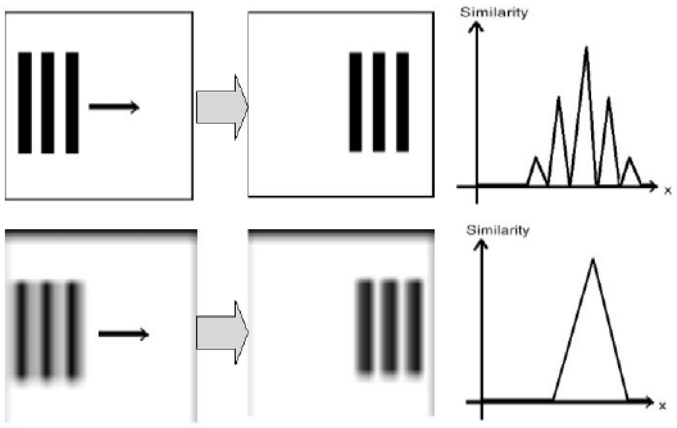

Figure 1.12: Example of how a lower resolution image can lead to a better parameter choice: in the low resolution version, there are not local minima in which the optimization algorithm may fall. The graphs show the similarity metric evolution when the image is shifted to the right.