

https://dspace.org/

Automatic Minimisation of Patient Setup Errors in Proton Beam Therapy

Ransome, Trevor Malcolm

2006-11-14T09:19:26Z
http://hdl.handle.net/10539/1667

Downloaded from a DSpace repository. https://dspace.org/

Automatic Minimisation of Patient Setup Errors in Proton Beam Therapy

Trevor Malcolm Ransome

A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering.

Johannesburg, March 2006

Declaration

I declare that this dissertation is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Master of Science in Engineering in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

Signed this __ day of _____ 20__

_____

Trevor Malcolm Ransome.

Abstract

Successful radiotherapy treatments with high-energy proton beams require the accurate positioning of patients. This paper investigates computational methods for achieving accurate treatment setups in proton therapy based on the geometrical differences between a double exposed portal radiograph (PR) and a reference image obtained from the treatment planning process. The first step in these methods involves aligning the boundary of the radiation field in the PR with a reference boundary defined by the treatment plan. We propose using the generalised Hough transform (GHT), followed by an optimisation routine to align the field boundaries. It is found that this method worked successfully on ten tested examples, and aligns up to 82% of reference boundary points onto the field boundary. The next step requires quantising the patients anatomical shifts relative to the field boundary. Using simulated images, a number of intensity-based similarity measures and optimisation routines are tested on a 3D/2D registration. It is found that the simulated annealing algorithm minimising the correlation coefficient provided the most accurate solution in the least number of function evaluations.

Acknowledgements

I would like to thank my university supervisors, Dr D. Rubin and Prof T. Marwala, and Evan de Kock who served as my *de facto* external supervisor. Through my two years of masters study, they have provided me with ideas, guidance, support, feedback, constructive criticism and lots of encouragement. Without each of them, my masters would not have been possible.

I would also like to thank iThemba LABS, for making this project available, and for their financial support.

Contents

Declaration i

Abstract ii

Acknowledgements iii

Contents iv

List of Figures viii

List of Tables x

Paper 1

1 Introduction 2

2 Materials 4

2.1 Image Acquisition 4

3 Methods 6

3.1 Field Boundary Alignment 6

3.2 Anatomical Registration 9

3.3 Optimisation Routines 12

4 Results 13

4.1 Field Boundary Alignment 13

4.2 Intensity-based Registration 15

5 Future Work 19

6 Discussion and Summary 19

7 Conclusion 20

References **23**

Appendix **28**

A Notation **28**

B DRR Calculation Algorithm **30**

1 Introduction 30

2 Initialisation 31

3 CT Data Calibration Curve 33

4 Ray and Plane Notation 33

5 Incremental Method 34

6 Results 36

7 Conclusion 36

References **38**

C Existing Field Boundary Alignment Algorithms **39**

1 Introduction 39

2 Field Edge Detection and Alignment 40

2.1 Thresholding Techniques 40

2.2 Radon Transform 41

3 Conclusion 42

References **43**

D Existing Anatomical-Body Alignment Algorithms **44**

1 Introduction 44

2 Feature-based Registration 44

3 Intensity-based Registration 45

4 Conclusion 45

References **47**

E Classical Optimisation Methods **49**

1 Nelder-Mead Simplex 49

2 Quasi-Newton Algorithm 50

3 Powell-Brent Procedure 51

References **52**

F Population-Based Optimisation Methods **53**

1 Genetic Algorithm (GA) 53

2 Particle Swarm Optimisation (PSO) 56

3 Simulated Annealing Process (SA) 57

References **61**

List of Figures

1 An example of portal (*left*) and DRR (*right*) images. 5

2 Histograms of a portal image obtained from previously documented solutions (*left*) [4] and iThemba LABS (*right*). 6

3 An illustration for the layout of orientation and offset parameters for an arbitrary shaped boundary. Adapted from [17]. . . . 8

4 The alignment of field boundaries for the GHT algorithm 14

5 Similarity measures (vertical graph-axis) versus x translation parameter (horizontal graph-axis). Each row of graphs represents the same similarity measure, with the first, second, third and fourth representing the ID, CC, χ^2 and MI measures respectively. Each column plots the similarity measure versus variations in the x translation parameter. The first column varies only the x translation parameter. The second column, alters the y and z translation parameters and plots the similarity measures versus a range of x translation parameters. The third column alters all DRR parameters and plots the similarity measures versus a range of x translation parameters. Furthermore, the co-ordinates of the global minima/maxima are marked on the illustrations. 22

B.1 Configuration of a portal x-ray imaging device 32

B.2 Intersections of a ray with 2D planes in the x and y directions . 34

B.3 Comparison of computational times between Incremental method (solid) and Siddon's method (dashed) for various DRR sizes . . 36

E.1 A simple reflection of a two-dimensional Simplex 50

F.1 Genetic Algorithm Flow Chart 58

F.2 Particle Swarm Algorithm Flow Chart 59

F.3 Simulated Annealing Flow Chart 60

List of Tables

1 An example of a R-Table 8

2 Average optimisation results for 10 radiation field boundary alignments 15

3 Average results for 10 registrations having maximum translations and rotations equal to $\pm 4mm$ and $\pm 4^{\circ}$ respectively. . . . 17

4 Average results for 10 registrations having maximum translations and rotations equal to $\pm 0.5mm$ and $\pm 0.5^{\circ}$ respectively. 18

Automatic Minimisation of Patient Setup Errors in Proton Beam Therapy

Trevor Malcolm Ransome

Abstract- Successful radiotherapy treatments with high-energy proton beams require the accurate positioning of patients. This paper investigates computational methods for achieving accurate treatment setups in proton therapy based on the geometrical differences between a double exposed portal radiograph (PR) and a reference image obtained from the treatment planning process. The first step in these methods involves aligning the boundary of the radiation field in the PR with a reference boundary defined by the treatment plan. We propose using the generalised Hough transform (GHT), followed by an optimisation routine to align the field boundaries. It is found that this method worked successfully on ten tested examples, and aligns up to 82% of reference boundary points onto the field boundary. The next step requires quantising the patients anatomical shifts relative to the field boundary. Using simulated images, a number of intensity-based similarity measures and optimisation routines are tested on a 3D/2D registration. It is found that the simulated annealing algorithm minimising the correlation coefficient provided the most accurate solution in the least number of function evaluations.

1 Introduction

Conformal, high energy proton beams are often used in radiotherapy for treating lesions that are near to critical or radiosensitive organs [1]. The accuracy of the patient's position relative to the proton beam is critical in these treatments. The accurate positioning of the patient requires reliable methods to first measure the patient's setup errors and then apply corrections to minimise the errors.

The process of minimising the patient's setup errors can be defined as minimising the differences between the patient's actual position and the required treatment position, the latter being specified during the treatment planning process [2]. Information about the patient's actual position can be retrieved from a portal radiograph (PR), while the required treatment position is characterised by a digitally reconstructed radiograph (DRR). The DRRs are generated during the planning process by forward projecting discrete rays through the Computed Tomography (CT) data set for the patient and computing the line integrals of the linear x-ray attenuation coefficients along each ray [3].

Calculating the geometrical difference between images (PR and DRRs) is commonly known as image registration. The registration requires the completion of two concurrent steps. The first step identifies and aligns the radiation field boundary with a reference field boundary, where the reference boundary is given by the treatment plan [4]. The alignment of the boundaries requires the quantification of four unknowns, namely, three translations and one in-plane rotation. The second step requires the alignment of the patient's anatomy, relative to the treatment field, in the PR and DRR images [4]. This step requires the determination of seven unknowns to completely quantify the patient's anatomical misalignments. The unknowns are: the rotation of the field edge around the beam axis and the 3D translations and rotations of the patient with respect to the desired treatment position.

At iThemba LABS [5], this process is done manually using measuring instruments to minimise the distances between anatomic or artificial landmarks. These manual techniques calculate the in-plane geometrical differences between the PR and a single DRR and fails to make use of the entire 3D CT data set. As a result, the registration is limited to only calculating in-plane

translations and rotations up to an accuracy of 1.8mm [6]. To improve this accuracy, it is necessary to incorporate the full CT data set into the registration and thus allow for the calculation of out-of-plane rotations. Manually identifying and quantifying out-of-plane rotations surpasses the human recognition capability and thus a semi/fully automatic registration method is sought.

The aim of this paper is to present and test the success of automated methods to minimise the errors in the patient's treatment setup. The proposed field boundary alignment algorithm, for double exposed portal radiographs, is to use the generalised Hough transform (GHT) algorithm, followed by an optimisation routine. For anatomical alignment, a number of gradient and population-based optimisation methods are tested on various intensity-based similarity measures for a 3D registration. This study investigates if intensity-based registration is a viable approach for PR and DRR registration. Furthermore, it shows what optimisation or sequence of optimisation routines are the best for 3D image registration. In previously proposed solutions, the testing of optimisation routines is omitted from the study [4, 7]. However, it is felt that the optimisation routine is the main contributor to the speed and accuracy of the final solution.

This study extends the methods presented in [8] for field boundary alignment and anatomical registration. In [8], it is proposed to follow the GHT algorithm by an optimisation routine that maximises the number of reference points lying on the field boundary. This paper tests the proposal using a number of optimisation routines. Furthermore, for anatomical registration the testing of the intensity-based registration is extended to out-of-plane rotations, and the optimisation routines are extended from a 2D image registration to a full 3D/2D registration of DRR images.

This paper is organised as follows: Section 2 provides information pertaining to the images and image formats used in the registration. Section 3 presents the methods for field boundary alignment and anatomical registration. Test results for the methods are presented in Section 4 and discussed in Section 6.

2 Materials

Currently at iThemba LABS [5], work is in progress to develop a new robot-based patient positioning system for proton radiotherapy [9]. This system will provide precise verification of the patient's orientation with respect to the treatment beam by using similar methods proposed in this study to minimise patient setup errors. All patient image data was rendered completely anonymous and was acquired for the purpose of manual position verification during routine clinical treatments.

2.1 Image Acquisition

Reference Field Boundary

The reference field boundary is determined from a treatment planning session. During this session, a radio-oncologist examines the tumour and determines a field boundary that encapsulates it. The resulting field boundary is defined in the treatment plan and is given as a set of vertical and horizontal co-ordinates at a certain depth down the beam's-eye view.

Portal Radiograph (PR)

The PR is routinely acquired by taking a double exposure portal radiograph in-line with the proton beam. The first exposure is taken with the collimator out of the beam path, followed by a second exposure with the collimator in the beam path. The resulting image has the boundary of the collimator aperture projected on the patient's anatomy. An example of a PR of a patient's head is illustrated in the left image of Figure 1. In the illustration the top rectangle is merely inserted to block-out patient information that was written on the x-ray.



Figure 1: An example of portal (*left*) and DRR (*right*) images.

Digitally Reconstructed Radiograph (DRR)

A DRR is computed by forward projecting discrete rays through a CT data set and integrating the linear x-ray attenuation coefficients along each ray [3]. In radiotherapy, the DRR images generated during the treatment planning session are used to verify the treatment setup and possibly to correct or minimise the errors in the patient's treatment setup position. One of the outputs from the treatment planning process is a reference DRR that illustrates the desired treatment position of the patient as seen from the beam's-eye-view or the treatment field. Patient setup errors are detected by calculating the anatomical misalignment between the reference DRR and PR. This involves iteratively minimising the difference between the PR and reference DRR, whose view is altered with each iteration. To achieve the highest accuracy in the registration and minimise memory usage, it is necessary to compute DRR images on-line in real time. Hence, an incremental implementation of Siddon's method, see Appendix B, is applied for fast DRR calculation. The implementation extends Jacobs *et al's* 2D incremental method in [10] to 3D for DRR calculation. It was found that the method was able to calculate a 256×256 DRR in 9 seconds for a $256 \times 256 \times 58$ CT data set. The algorithm is implemented using MATLAB© [11] and is not optimised for computational efficiency. Converting the implementation from MATLAB© to a C implementation reduces the DRR calculation time from 9 seconds to 1.7 seconds [12].

3 Methods

3.1 Field Boundary Alignment

The first step in the treatment verification procedure is the alignment of the radiation and reference field boundaries. This involves identifying the radiation field boundary in the PR and aligning it with the reference field boundary, given by the treatment plan [4]. The automatic alignment of field boundaries requires the determination of four unknowns: scale, two translations and one rotation. Previous methods, see Appendix C, use thresholding techniques to identify the field edges and centre-of-mass calculations to align the edges [4, 13, 14]. These methods work successfully for single exposure x-ray images that have large discrepancies between the intensities in the patient's anatomical data and surrounding field boundary. However, for double exposure portal radiographs the contrast between the patient's anatomy inside and outside the field boundary is relatively low. The reason being that the exposure with the collimator inserted into the beam-path is low in order to keep the x-ray dose received by the patient below reasonable limits. To avoid increasing this exposure, a more computationally expensive algorithm, compared to thresholding techniques, is presented. This algorithm is known as the generalised Hough transform (GHT) [15]. The image histograms for single (left illustration) and double (right illustration) exposure PIs are illustrated in Figure 2.

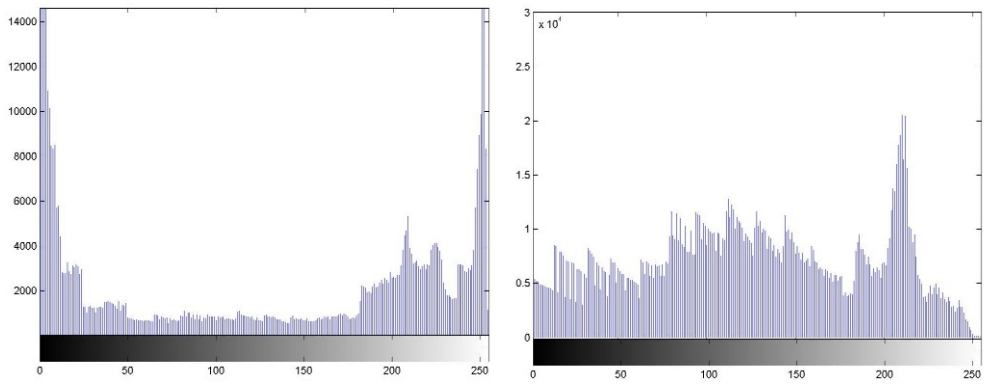


Figure 2: Histograms of a portal image obtained from previously documented solutions (*left*) [4] and iThemba LABS (*right*).

Generalised Hough Transform (GHT)

The GHT is a powerful object detection and alignment algorithm that is able to align arbitrary-shaped semi-occluded edges [15]. The algorithm sequentially or randomly steps through the orientation (slope) of all edges in the target image (PR) at a specific rotation and scale [15]. If the orientation of the edge in the target image is the same as an edge in the reference image, an offset co-ordinate in the Hough transform is accumulated. The resulting Hough transform has localised high intensity regions, where the region with the highest intensity corresponds to the location of the target point.

The first step in the GHT algorithm is to create a look-up table, which is commonly known as the R-Table [16]. Table 1 describes each edge on the reference boundary with a specific orientation (ϕ) and offset (Δ) [17]. The orientation of each edge ($e(x, y)$) on the boundary is calculated using:

$$\phi = \arctan\left(\frac{h_r(i,j)}{v_r(i,j)}\right) \tag{1}$$

where v_r and h_r are the vertical and horizontal gradients for the reference boundary respectively. To calculate v_r and h_r, it is necessary to first convert the reference boundary co-ordinates to a binary image (bi_r) using a region-of-interest filling algorithm. The parameters: v_r and h_r are calculated by filtering bi_r in the vertical and horizontal directions using a gradient operator (e.g. Sobel, Prewitt, Roberts or Laplacian of Gaussian). The orientation of the reference boundary can also be determined by calculating the slope between two adjacent co-ordinates on the boundary ($r_{x,y}$) using:

$$\phi = \arctan\left(\frac{\Delta r_x}{\Delta r_y}\right) \tag{2}$$

However, the success of the GHT algorithm is dependent on finding the same orientations in the target image. Hence, it is preferable to determine the orientation using (1) to ensure that the calculation between the boundaries is consistent. The offset is defined as the difference between a boundary edge point and target point ($t(x, y)$) co-ordinates, which is denoted as follows:

$$\Delta_n(x, y) = t(x, y) - e(x, y) \tag{3}$$

For an arbitrary-shaped boundary, the layout of ϕ and Δ with respect to a point on the boundary ($p(x, y)$), is illustrated in Figure 3. Once the R-Table

Table 1: An example of a R-Table

Slope (ϕ)	List of Offset Pairs
-90	$\Delta_1(x_1, y_1)$, $\Delta_5(x_5, y_5)$
-89	Nil
$\vdots$	$\vdots$
90	$\Delta_m(x_m, y_m), \ldots, \Delta_n(x_n, y_n)$

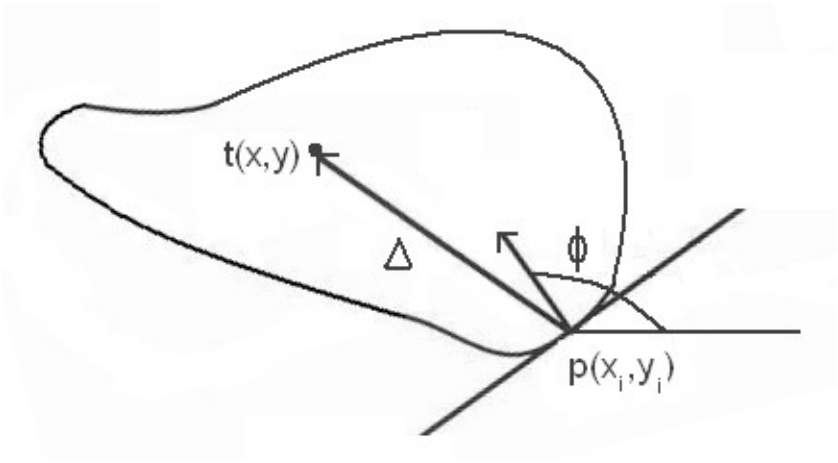


Figure 3: An illustration for the layout of orientation and offset parameters for an arbitrary shaped boundary. Adapted from [17].

has been created, the next step in the algorithm is to step through each edge in the target image and accumulate the Hough transform for a specific scale (s_m) and rotation (R). The edges in the target image are calculated by thresholding (ζ), the squared sum of the target image's vertical (v_t) and horizontal (h_t) gradients using:

$$E(i,j) = \zeta \left(h_t(i,j)^2 + v_t(i,j)^2 \right) \tag{4}$$

For each edge in the target image, the radiation edge's orientation is determined by adding R to the edge's orientation given by 1. Using this value as the look-up index into the R-Table, the Hough transform is accumulated at the following co-ordinates:

$$(x_{in}, y_{in}) = \lfloor s_m(\Delta_k(x,y) \times R_{2D}(r_n)) + p(x_i, y_i) \rfloor \tag{5}$$

where R_{2D} is a two-dimensional rotation matrix and $\lfloor . \rfloor$ is a round down operator. The resulting Hough Transform for a specific R and s_m has localised

non-zero intensity regions. The highest intensity region in the Hough transform indicates the location of the target point in the portal radiograph and reference boundary transformations (R and s_m). The centre of the highest intensity region is calculated by convolving the Hough transform with a square matrix of ones. The resulting matrix has its highest intensity index value representing the centre of the highest intensity region and the indices of this point represents the co-ordinates of the target point.

Cost Function

To improve the accuracy in the alignment, the GHT is followed by an optimisation routine. The optimisation routine maximises the number of reference boundary points lying on the radiation field edge given by equation (4). The cost function (ζ) denotes the percentage of points lying on the boundary and is defined as follows:

$$\zeta(t_x, t_y, s, \theta) = \sum_x \sum_y edge[s(r_{x,y} \times R_{2D}(\theta)) + [t_x, t_y]] \tag{6}$$

3.2 Anatomical Registration

Anatomical alignment quantifies the errors in the placement of the patient's anatomy relative to the treatment beam. This requires calculating the 3D geometrical misalignments between a PR and a reference DRR. This involves calculating seven unknowns, namely, rotation of field edge around the beam axis, three translation values (x, y and z), one in-plane rotation value (θ) and two out-of-plane rotation values (ψ and φ). Calculating these parameters requires the registration of the PR to either a series of pre-calculated DRRs, or to DRRs that are generated on-line.

Numerous methods are available for measuring the misalignment between a DRR and PR (see Appendix D). The bulk of the algorithms can be classified as being landmark, [2, 18–25], feature-based, [26–30] or intensity-based methods [4, 7]. Landmark based registration requires the user to manually select corresponding structural points between the two images. This method has been shown to be successful for the registration of two 2D images [31].

However, for the registration of multiple images obtained from a 3D data set, this method becomes labour intensive. Feature-based methods are defined as the alignment of line segments, curves or patterns that are common to both images [4]. These patterns are automatically extracted by the evaluation of the change in the image's pixel intensity surface. Intensity-based registration, also known as voxel-based registration, is based on the similarity between corresponding pixel intensities. This, unlike feature-based registration, avoids the extraction of corresponding features between the images and thus its accuracy is not limited by segregation errors [32]. Intensity-based registration methods can be split into two groups: *viz.* intra-modality and inter-modality similarity measures. Intra-modality measures are most successful for images obtained from the same modality, and the measure relies on the assumption of linear correspondence between corresponding pixel intensities [32]. Popular intra-modality measures include intensity difference and correlation coefficient measures. The assumption of linear correspondence between the images often fails for multi-modality images. This is overcome using inter-modality measures that are based on the images marginal and joint histograms [32, 33]. Popular inter-modality measures include Mutual information and Chi-squared measures.

Intensity Difference (ID)

ID is the sum of the squared intensity differences between the target image (T) and the scene image (S) and is defined as [33]:

$$ID = \frac{1}{N} \sum_{x} \sum_{y} (T(x, y) - S(x, y))^2 \tag{7}$$

where x and y are the indices of the horizontal and vertical pixels respectively.

It is necessary to divide the squared difference by the number of overlapping pixel intensities (N), as N might vary with each parameter estimation. It has been shown by Viola in [34], that this measure is optimum if the images only differ by Gaussian noise [33]. In practical applications, both intra-modality and inter-modality, this assumption is seldom true. Furthermore, the measure is sensitive to large intensity differences [33]. This property can be reduced by replacing the squared difference by an absolute difference calculation [33].

Correlation Coefficient (CC)

CC avoids the assumptions made by (7), but is replaced by a less strict assumption that the image's intensities are linearly correlated [33]. The calculation includes the image means in order to prevent high intensity differences dominating the registration. The correlation coefficient is defined as follows [33]:

$$CC = \frac{\sum_x \sum_y \left(T(x,y) - \mu_T\right)\left(S(x,y) - \mu_S\right)}{\left[\sum_x \sum_y \left(T(x,y) - \mu_T\right)^2 \sum_x \sum_y \left(S(x,y) - \mu_S\right)^2\right]^{\frac{1}{2}}} \tag{8}$$

where μ_T and μ_S are the means for the target and scene images respectively.

Mutual Information

The assumptions made by the ID and CC measures most often fail for images obtained from different modalities. These assumptions are overcome using inter-modality measures that are based on the images' marginal and joint histograms [32]. The joint histogram ($p_{T,S}$) is a square matrix with *bin* number of elements in each row and column. The matrix $p_{T,S}$ is initially a zero matrix and is calculated by iteratively stepping through each gray scale element in T and S (normalised to a maximum pixel intensity of 255). With each iteration, the matrix values of $p_{T,S}$ are incremented at horizontal (a) and vertical (b) indices given by:

$$a = bin \times \frac{T(x,y)}{255} \tag{9}$$

$$b = bin \times \frac{S(x,y)}{255} \tag{10}$$

The marginal histograms (p_T and p_S) are calculated by summing $p_{T,S}$ in its respective directions and are defined as:

$$p_T\left(i\right) = \sum_j p_{T,S}(j,i) \tag{11}$$

$$p_S\left(i\right) = \sum_j p_{T,S}(i,j) \tag{12}$$

where i ranges from 0 to *bin* in increments of 1.

An illustration of joint histograms for registered and unregistered images is presented in [32]. Comparing the histograms in the illustration, it can be seen

that the registered joint histogram has localised regions with higher intensities compared to the unregistered histogram. From this observation, it is deduced that the aim of inter-modality registration, using intensity-based methods, is to maximise the intensity of localised regions in the joint histograms.

Using the joint and marginal probabilities, the Mutual Information measure, also commonly known as relative entropy, is calculated using [7, 32]:

$$MI = \sum_{x}\sum_{y} p_{T,S}(x,y)\, log_2 \frac{p_{T,S}(x,y)}{p_T(x)p_S(y)} \tag{13}$$

Chi-square (χ^2)

Chi-square similarity measure is also calculated using the joint and marginal probabilities of the images defined in equations 9-12. The measure is given by [7]:

$$\chi^2 = \sum_{x}\sum_{y} \frac{\left(p_{T,S}(x,y) - p_T(x)p_S(y)\right)^2}{p_T(x)p_S(y)} \tag{14}$$

3.3 Optimisation Routines

Assuming that the correct registration method is chosen for the particular application, the main consideration in the final registration's accuracy and computational speed is the optimisation routine. In general, optimisation is the task of determining a set of parameters in order to minimise or maximise a function, subject to certain constraints [35].

Optimisation methods can broadly be broken into two groups, *viz.* classical and population-based methods. Classical optimisation methods, also known as local optimisation algorithms, probe the search space in one dimension using gradient or function evaluation information. These methods start at an initial guess ($\widetilde{xo}$) in the parameter space (also called the search space) and with each iteration $\widetilde{xo}$ is updated in the direction of the local minimum. This process is continued until the stopping criteria are reached. Population-based methods are different to classical methods in that the search space is probed in multi-dimensions. This enables the algorithm to determine a global minimum solution when the function is prone to have local minima.

4 Results

4.1 Field Boundary Alignment

It is required to first identify the edge of the radiation field in the PR and then to align this boundary with the reference boundary given by the treatment plan.

For double exposure portal radiographs, it is proposed to use an edge detection algorithm to identify the field edges. The edge detection algorithm filters the PR in the vertical and horizontal directions using a gradient operator at a predefined threshold value. By evaluating the edges of all the gradient operators applied to a PR, it was found that the Sobel operator with horizontal mask defined in equation (15) gave the best results. The vertical mask is simply the transpose of the horizontal mask:

$$Hor = \begin{bmatrix} -1 & -2 & -3 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix} \tag{15}$$

Testing the GHT algorithm on various examples, it was found that the success of the algorithm is very sensitive to the orientation calculations. Hence, if the target image has gone through a slight affine transformation, the algorithm often fails. To improve the robustness of the algorithm error bars are added to the orientation calculations. Error bars can be included by increasing the step size of the orientations in the R-Table.

The PR shown in Figure 4 provides a representative example of a field boundary to illustrate the typical performance of the GHT algorithm. The size of the PR was decreased, in order to improve the GHT algorithm's overall computational speed and success rate. The rotation parameter was set to vary between $-18°$ and $18°$ in steps of $2°$ and the scale parameter varying between 0.7 and 1.3 in steps of 0.1. All tested alignments provided an accurate solution and took on average 60 seconds to compute. The algorithm was tested using Matlab© on an Intel 4, 2.8GHz processor. The results for a single example is illustrated in Figure 4.

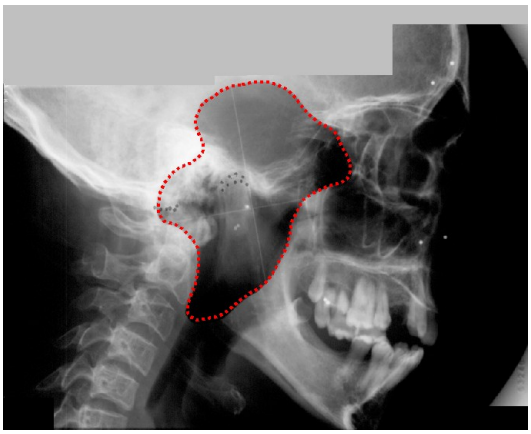


Figure 4: The alignment of field boundaries for the GHT algorithm

To improve the boundary alignment illustrated in Figure 4, the GHT algorithm is followed by an optimisation routine that maximises the cost function defined in equation (6). The cost function calculates the percentage of reference points lying on binary edges in the PR, which is given by binary masking equation (4). In the initial GHT alignment, illustrated in Figure 4, the cost function evaluates to 43% of reference points lying on the radiation field edge.

A number of optimisation algorithms were tested on the cost function, namely, the Nelder-Mead simplex, genetic algorithm, particle swarm, and simulated annealing algorithms. The Nelder-Mead simplex implementation is a standard function provided by MATLAB© [11]. The genetic algorithm [36], offers a number of parent selection, cross-over and mutation functions. All variations of the functions were tested and it was found that the normalised geometric selection function, arithmetic cross-over and non-uniform mutation with values of 0.6, 10 and 20 respectively and a population size of 600, provided the best results for the alignment. The PSO algorithm [37] is implemented with both the "cognitive" and "social scaling" parameters set equal to 2 and with the "weight inertia" diminishing, with each iteration, from 0.95 down to 0.4. A swarm size of 30 was used for the testing. The SA algorithm is based on an implementation presented in [35]. The implementation is altered to allow the algorithm to restart after a solution has been found. Restarting the algorithm with the previous best point being a vertex, minimises the chance of the algorithm getting stuck at local minima. This strategy is recommended in [35].

The results for the optimisation algorithms are presented in Table 2.

Table 2: Average optimisation results for 10 radiation field boundary alignments

Method	$\zeta(\%)$	n	Time (s)
Simplex	82.4	179	0.155
GA	79.2	729	0.781
PSO	78,9	1001	0.922
SA	78,3	976	1.032

ζ Evaluation of the cost function defined in (6).

n Number of function evaluations.

4.2 Intensity-based Registration

Once the field boundaries have been aligned, the next step in determining the setup errors is to align the patient's anatomy in the PR and DRR images.

Similarity Measures

This section tests a number of intensity-based similarity measures on DRRs with both in-plane and out-of-plane transformations. The similarity measures tested are: the intensity difference (ID), correlation coefficient (CC), mutual information (MI), and chi-squared (χ^2) measures.

Ideally the testing of the similarity measures should be done by registering a PR with DRRs. However, the transformation parameters necessary to align the PR and DRRs is unknown and thus an accurate evaluation of the similarity measures cannot be made. To solve this problem, it is chosen to approximate the PR by a reference DRR and thus allow for the transformation parameters between the reference and transformed DRRs to be known.

A DRR is calculated using the following transformation parameters: [x, y, z, θ, ψ, φ], where x, y and z are the translation parameters, θ is the in-plane rotation and ψ and φ are the out-of-plane rotation parameters. For this example, the

reference DRR is calculated with transformation parameters equal to [0, 0, 0, 5, -1, 180].

Firstly the similarity measures are tested by varying each transformation parameter independently of the other parameters. For the x-translation parameter, the transformed DRR is calculated using parameters equal to [k, 0, 0, 5, -1, 180], where k is chosen to vary between -3mm and 3mm. The results for the variation of the x-translation parameters, independent of the other parameters, is illustrated in the first column of Figure 5. Similar results were obtained for the variation in the other parameters.

Secondly the similarity measures are tested with a change in all the translation parameters (x, y and z). To maintain a 2D plot, the x-translation parameters are varied between -3mm and 3mm with a single change in the other parameters. The transformed DRR is calculated using [k, 2 -2, 5, -1, 180]. The results for this test are shown in the second column of Figure 5.

Finally the similarity measures are tested with a change in all the parameters. The transformed DRR is calculated using [k, 2, -2, 7, -0.5, 177] and these results are shown in the third column of Figure 5.

The illustrations in columns two and three show a few random variations in the other parameters, besides the x-translation parameter. It was felt that a single variation in each was sufficient to illustrate the presence of local minima. The local minima are further illustrated by giving the co-ordinates of the determined global minima. It should be clear that the true global minima should occur when the x-translation parameter is equal to zero and the similarity measure evaluates to the minimum/maximum value, which is indicated in the first column of Figure 5. However, in columns two and three the global minima are shifted away from the translation position equal to zero, and thus indicates the presence of local minima in all similarity measures.

Optimisation Routines

A number of classical and population-based optimisation methods are implemented on a 3D/2D DRR registration. The registration requires the geometrical alignment of a 2D reference DRR with a transformed DRR, where six

degrees of freedom are involved in the calculation of the transformed DRR.

In Section 4.2 it was found that all similarity measures were prone to local minima. Hence for large transformations with translations and rotations greater than $0.5mm$ and $0.5°$), a number of population-based methods are tested on DRR images having maximum translation and rotations between $[-4mm, 4mm]$ and $[-4°, 4°]$ respectively. The population-based methods are tested on the correlation coefficient similarity measure, which provides a smooth response compared to mutual information and chi-squared measures (see Figure 5). The population-based methods are the genetic algorithm, particle swarm and simulated annealing algorithms. Refer to Appendix F for a detailed description of each algorithm. For the GA, all parent selection, cross-over and mutation functions were tested. It was found that the normalised geometric selection function, simple cross-over and non-uniform mutation with values of 0.6, 10 and 20 respectively and a population size of 80 with 20 generations, provided the best results for the registration. For the PSO algorithm, the population size is set to 25 with a maximum number of iterations equal to 20. The scaling parameters and weight inertia are set to the same values listed in Section 4.1. The SA algorithm's cooling schedule decreases from 80 down to 0 in steps of 10, with the following number of samples per temperature: [60,40,25,20,15,20,25,40,60]. The results for the registrations are presented in Table 3.

Table 3: Average results for 10 registrations having maximum translations and rotations equal to $\pm 4mm$ and $\pm 4°$ respectively.

Method	ID	n	$\iota_t(mm)$	$\iota_r(°)$
GA	230.19	515	0.06	0.0496
PSO	2379	500	0.1529	0.3175
SA	17.12	482	0.0072	0.0102

ID Intensity difference similarity measure denoted in (7).
n Number of function evaluations.
ι_t average translation distance to global min (mm)
ι_r average rotation distance to global min (°)

For smaller translations ($< 0.5mm$) and rotations ($< 0.5°$), both classical and

population-based methods are tested. The classical optimisation methods are the Nelder-Mead Simplex, Quasi-Newton algorithm and Powell-Brent procedure. These algorithms are presented in Appendix E. The Quasi-Newton algorithm is provided by the NETLAB toolbox [38]. The Powell-Brent procedure is obtainable from [35]. The method requires a set of direction vectors to be initially defined and are updated each iteration, through the routine. The direction vectors are defined by running an initial image registration with an arbitrary direction vector. These results were then used for the other registrations. Limits were placed on the number of iterations in both Powell's and Brent's methods. For the GA, the population size is decreased down to 35, with 6 generations. For the PSO algorithm, the swarm size is decreased down to 15, with a maximum number of iterations equal to 10. The cooling schedule for the SA algorithm is altered to decrease from 60 down to 0 in steps of 10, with the following number of samples per temperature: [20,18,15,10,15,18 20]. The results for the registrations are presented in Table 4. Various alterations could be implemented on the optimisation algorithms to improve their efficiency and thus alter the results in Tables 3 and 4. However, these amendments surpass the scope of the project. The time for each optimisation routine is omitted from the results, as the time is proportional to the number of function evaluations, and is dependent on the time required to calculate a DRR.

Table 4: Average results for 10 registrations having maximum translations and rotations equal to $\pm 0.5mm$ and $\pm 0.5^\circ$ respectively.

Method	ID	n	$\iota_t(mm)$	$\iota_r(^\circ)$
Simplex	2856	150	0.2393	0.2393
Q-N	1635	172	0.1334	0.1306
P-B	5173	146	0.2586	0.2598
GA	173	161	0.0331	0.048
PSO	985	150	0.1084	0.1233
SA	99	180	0.0195	0.0355

5 Future Work

Before the anatomical-body registration algorithms can be implemented in practical applications it is necessary to validate the consistency of the intensity-based measures with the registration of DRRs and the PR. This will involve altering the DRR calculation parameters in order for it to match the PR.

6 Discussion and Summary

The success of proton radiotherapy depends on the accurate positioning of the patient relative to the treatment beam. Estimating the errors in the treatment setup involves calculating the 3D transformation parameters between a 2D portal radiograph and DRRs that are calculated from a 3D CT data set. The first step in the process requires the alignment of radiation and reference field boundaries. This creates a common frame of reference between the PR and CT data. Previously proposed methods are based on single exposure portal radiographs [4, 13, 14]. However, double exposed radiographs do occur in practise, e.g. at iThemba LABS. The first exposure has the collimator out of the beam path, followed by a second exposure with the collimator in the beam path. The resulting image has the field boundary projected onto the patient's anatomy.

This paper proposes a unique solution for field boundary that is based on the field edges in the PR. The field edges are first aligned with the reference boundary using the *generalised Hough transform* (GHT) followed by the Nelder-Mead Simplex method maximising the number of reference points lying on the field edge. It was found that the method successfully aligned the tested field boundaries, provided the search space (PR) was decreased. Given the configuration of the PR, the location of the field boundary should be approximately known thus decreasing the search space, without obscuring the field boundary, should be easy. Decreasing the PR, results in fewer false edges being tested, and thus improves the algorithm's computational time. To further improve the computational time, it is recommended to use a parameter reduction implementation of the GHT algorithm [39, 40]. However, if sufficient accuracy is applied in the positioning of the field boundary, the GHT algorithm can be omitted form

the process and a population-based algorithm maximising the cost function, denoted in (6), can be used.

The next step in the process requires the determination of the errors in the patient's position relative to the treatment beam by registering the patient's anatomy in a number of transformed DRRs to the anatomy portrayed in the PR. It is proposed to register the PR and DRR images using intensity-based registration. A number of intensity-based similarity measures were tested on DRR images having different transformations. It was found that the intensity difference and correlation coefficient provided a consistent measure for in-plane and out-of-plane transformations. However, all the measures (intra and inter-modality) were prone to local minima. It is concluded that intensity-based registration is a viable approach for PR and DRR registration, provided that the DRR approximates the PR correctly.

To determine the transformation parameters of the DRR that best matches the PR, a number of classical and population-based optimisation methods were tested on a 3D/2D DRR registration. For small and large transformations, it was found that the simulated annealing algorithm provided the most accurate solution. For translations and rotations less than $4mm$ and 4° respectively, the SA algorithm obtained the most accurate solution in the least number of function evaluations.

7 Conclusion

This paper presents and tests algorithms that can be used to estimate the errors in the patient's treatment setup for proton radiotherapy. For field-boundary alignment, the first step in the procedure, it is proposed to use the generalised Hough transform (GHT) followed by an optimisation routine to align the reference and radiation field boundaries. It was found that the GHT algorithm successfully aligns reference and radiation field boundaries of a double exposure portal radiograph. However, if sufficient accuracy is applied in the positioning of the field boundary (collimator), the GHT algorithm can be avoided, and the boundaries can be perfectly aligned using the optimisation routine.

For anatomical registrations, a number of intensity-based similarity measures and optimisation routines are tested on a 3D/2D registration of DRR images. The intensity-based similarity measures are: intensity difference, correlation coefficient, mutual information and chi-squared measures. It was found that the correlation coefficient provided the most consistent measure for both small and large transformations. This paper includes the testing of classical (Nelder-Mead Simples, Powell Brent procedure and Quasi-Newton algorithm) and population-based (genetic algorithm, particle swarm optimisation and simulated annealing) optimisation routines on the anatomical registration (omitted in previously proposed solutions). It was found for both small and large transformations that the simulated annealing provided the most accurate solution. For translations and rotations, less than $4mm$ and 4° respectively, the SA algorithm obtained the most accurate solution in the least number of function evaluations.

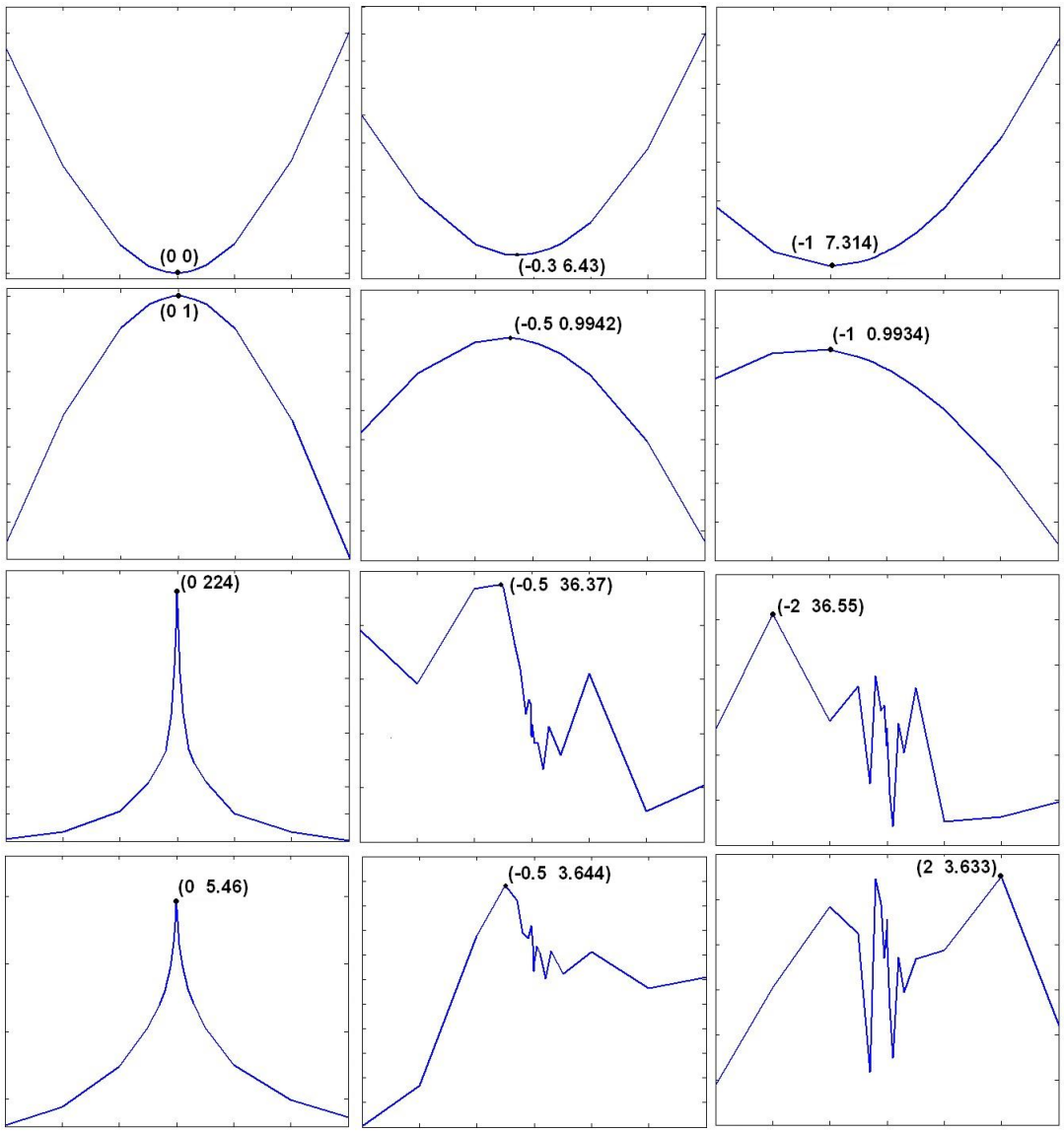


Figure 5: Similarity measures (vertical graph-axis) versus x translation parameter (horizontal graph-axis). Each row of graphs represents the same similarity measure, with the first, second, third and fourth representing the ID, CC, χ^2 and MI measures respectively. Each column plots the similarity measure versus variations in the x translation parameter. The first column varies only the x translation parameter. The second column, alters the y and z translation parameters and plots the similarity measures versus a range of x translation parameters. The third column alters all DRR parameters and plots the similarity measures versus a range of x translation parameters. Furthermore, the co-ordinates of the global minima/maxima are marked on the illustrations.

References

[1] P. Yushkevich, D. Fritsch, S. Pizer, and E. Chaney, "Towards automatic, model-driven determination of 3d patient setup errors in conformal radiotherapy," University of North Carolina at Chapel Hill, Tech. Rep., 1998.

[2] J. Bijhold, "Three-dimensional verification of patient placement during radiotherapy using portal images," *Med. Phys.*, vol. 20, pp. 347–356, 1993.

[3] R. Siddon, "Fast calculation of the exact radiological path for a three-dimensional ct array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, March/April 1985.

[4] L. Dong and L. Boyer, "A portal image alignment and patient setup verification procedure using moments and correlation techniques," *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.

[5] D. Jones, "Nac - the only proton therapy facility inthe southern hemisphere," in *LINZ U (ed) In Ion Beams in Tumor Therapy.* Chapman and Hall, 1995, pp. 350–359.

[6] K. Gilhuijs and M. van Herk, "Automatic on-line inspection of patient setup in radiation therapy using digital portal images," *Med. Phys.*, vol. 20, pp. 667–677, 1993.

[7] S. Clippe, D. Sarrut, C.Malet, S. Miguet, C. Ginestet, and C. Carrie, "Patient setup error measurement using 3d intendsity-based image registration techniques," *Int. J. Radiation Oncol. Biol. Phys.*, vol. 56, no. 1, pp. 259–265, 2003.

[8] T. Ransome, D. Rubin, T. Marwala, and E. de Kock, "Optimising the verification of patient positioning in proton-beam therapy," *Proceedings*

of IEEE 3rd Int. Conf. on Computational Cybernetics, ISBN 963 7154 388., April 2005.

[9] E. de Kock *et al*, "Integrating an industrial robot and multi-camera computer vision systems into a patient positioning system for high-precision radiotherapy," *International Symposium on Robotics*, March 2004.

[10] F. Jacobs, E. Sundermann, B. D. Sutter, M. Christiaens, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *Journal of Computing and Information Technology*, vol. CIT 6, no. 1, pp. 89–94, 1998.

[11] *Matlab, The language of technical computing*, July 2002, 6.5 Release 13.

[12] L. v. d. Bijl, E. A. d. Kock, and N. Muller, "Influence of DRRs on accuracy and efficiency of automatic patient positioning during conformal radiotherapy with protons," *Proc. 16th PRASA Conf.*, November 2005.

[13] H. Shu, Y. Ge, L. Luo, J. Wang, W. Yu, and X. Bao, "An orthogonal moment-based method for automatic verification of radiation field shape," *Phys. Med. Biol.*, vol. 45, pp. 2897–2911, 2000.

[14] J. Bijhold, K. Gilhuijs, M. van Herk, and H. Meertens, "Radiation field edge detection in portal images," *Phys. Med. Biol.*, vol. 38, pp. 1493–1502, 1993.

[15] D. Ballard, "Generalized hough transform to detect arbitrary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, pp. 111–122, 1981.

[16] P. Fung, W. Lee, and I. King, "Randomized generalized hough transform for 2-d grayscale object detection," *Proc. 13th Int. Conf. Pattern Rec.*, vol. 2, pp. 511–515, 1996.

[17] M. Ulrich, C. Steger, and A. Baumgartner, "Real-time object recognition using a modified generalized hough transform," *Pattern Recognition*, vol. 36, pp. 2557–2570, 2003.

[18] H. Meertens, J. Bijhold, and J. Strackee, "A method for the measurement of field placement errors in digital portal images," *Phys. Med. Biol.*, vol. 35, pp. 299–323, 1990.

[19] W. Lam, M. Herman, K. Lam, and D. Lee, "On-line portal imaging: computer-assisted error measurement," *Radiology*, vol. 179, pp. 871–873, 1991.

[20] K. Lam, R. T. Haken, D. McShan, and A. Thornton, "Automated determination of patient setup errors in radiation therapy using spherical radio-opaque markers," *Med. Phys.*, vol. 20, pp. 1145–1152, 1993.

[21] G. Ende, H. Truer, and R. Boesecke, "Optimisation and evaluation of landmark-based image correlation," *Phys. Med. Biol.*, vol. 37, pp. 261–271, 1992.

[22] J. Michalski, J. Wong, W. Bosch, D. Yan, A. Cheng, R. Gerver, M. Graham, D. Low, R. Valicenti, and J. Piephoff, "An evaluation of two methods of anatomical alignment of radiotherapy portal images," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 27, pp. 1199–1206, 1993.

[23] B. McParland, "Uncertainty analysis of field placement error measurements using digital portal and simulation image correlations," *Med. Phys.*, vol. 20, pp. 679–685, 1993.

[24] K. Gall, L. Verhey, and M. Wagner, "Computer-assisted positioning of radiotherapy patients using implanted radiopaque fiducials," *Med. Phys.*, vol. 20, pp. 1153–1159, 1993.

[25] S. Li, C. Pelizzari, and G. Chen, "Unfolding patient motion with biplane radiographs," *Med. Phys.*, vol. 21, pp. 1427–1433, 1994.

[26] K. Eilertsen, A. Skretting, and T. Tennvassas, "Methods for fully automated verification of patient setup in external beam radiotherapy with polygon shaped fields," *Phys. Med. Biol.*, vol. 39, pp. 993–1012, 1994.

[27] K. Gilhuijs, P. van de Ven, and M. van Herk, "Automatic three-dimensional inspection of patient setup in radiation therapy using portal images, simulator images and computed tomography data," *Med. Phys.*, vol. 23, no. 3, pp. 389–399, 1996.

[28] J. Stoeckel, F. Vos, P. Vos, and A. Vossepoel, "An evaluation of ridge extraction methods for portal imaging," in *International Conference on Pattern Recognition*, vol. 3. IEEE, 2000, p. 3433.

[29] S. Pizer, D. Fritsch, V. Johnson, and E. Chaney, "Segmentation, registration and measurement of shape variation via image object shape," University of North Carolina, Tech. Rep., 1996.

[30] S. Pizer, D. Fritsch, K. Low, and J. Furst, "2d and 3d figural models of anatomic objects from medical images," *Kluwer Computational Imaging and Vision Series*, pp. 139–150, 1998.

[31] T. Ransome, "Point-based methods for rigid and non-rigid body registration for mri/ct and pet images," University of Witwatersrand, Johannesburg, Tech. Rep., 2003.

[32] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[33] D. Hill, P. Batchelor, M. Holden, and D. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, pp. R1–R45, 2001.

[34] P. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, Massachusetts Institute of Technology, 1995.

[35] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C++, Second Edition.* Cambridge University Press, 1998.

[36] C. Houch, J. Joines, and M. Kay, "A genetic algorithm for function optimization: a matlab implementation," North Carolina State University, Tech. Rep., 1996.

[37] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE International Conference on Neural Networks Proceedings*, pp. 1942–1948, 1995.

[38] I. Nabney, *Netlab: algorithms for pattern recognition.* Springer, 2000.

[39] D. Ma and X. Chen, "Hough transform using slopes and curvature as local properties to detect arbitrary 2d shapes," *Proceedings of the 9th International Conference on Pattern Recognition*, vol. 88, pp. 511–513, 1988.

[40] A. Kassim, T. Tan, and K. Tan, "A comparitive study of efficient generalised hough transform techniques," *Image Vision Comput.*, vol. 13, no. 2, pp. 737–748, 1999.

Appendix A

Notation

The following notation is used to describe the co-ordinates of a point in 2D and 3D space:

p_{x_i} = the x co-ordinate for the point (p) in two or three-dimensions.

$p_{x,y}$ = a point in two dimensions.

$p_{x,y,z}$ = a point in three dimensions

In two dimensions, the point ($p_{x,y}$) is rigidly transformed by a translation ($q_{x,y}$) and rotation (θ) about the axes using:

$$\Im(q_{x,y}, \theta) = q_{x,y} + R_{2D}(\theta)\ p_{x,y} \tag{A.1}$$

where $R_{2D}(\theta)$ is the rotation matrix defined as follows:

$$R_{2D} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{A.2}$$

For a point in three dimensions ($p_{x,y,z}$), it is rigidly transformed by three dimensional translation ($q_{x,y,z}$) and rotation ($\psi_{x,y,z}$) vectors using:

$$\Im(q_{x,y,z}, \psi_{x,y,z}) = q_{x,y,z} + R_{3D}(\psi_{x,y,z})\ p_{x,y,z} \tag{A.3}$$

where $R_{2D}(\psi_{x,y,z})$ is the net product of the three elementary rotation matrices defined as follows:

$$
\begin{aligned}
R_x(\psi_x) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\psi_x & -sin\psi_x \\ 0 & sin\psi_x & cos\psi_x \end{pmatrix} \\
R_y(\psi_y) &= \begin{pmatrix} cos\psi_y & 0 & sin\psi_y \\ 0 & 1 & 0 \\ -sin\psi_y & 0 & cos\psi_y \end{pmatrix} \\
R_z(\psi_z) &= \begin{pmatrix} cos\psi_z & -sin\psi_z & 0 \\ sin\psi_z & cos\psi_z & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}
\tag{A.4}
$$

Appendix B

DRR Calculation Algorithm

1 Introduction

In various medical imaging applications, it is required to calculate digitally reconstructed radio-graphs (DRR) in real time. One example, is to use DRR images to quantify patient setup errors in radiotherapy. This is done by registering the DRR images with a digital portal x-ray image. The registration is done through an optimisation routine, which requires the determination of seven unknowns. Two of the unknowns require DRR images to be calculated at different out-of-plane rotation angles. This entails many DRR images to be calculated and can be a time consuming process if the DRR calculation method is slow. Time is an important factor in this application, as the patient would be fixed to their position through the registration. One solution to speed up the registration is to calculate the DRR images beforehand (off-line). This method limits the accuracy in the final registration and requires a large amount of memory to save the DRR images for each patient. A better method is to implement a fast DRR algorithm to enable the DRR images to be calculated on-line.

A DRR image is computed by forward projecting discrete rays through a CT-data set. Each CT value ($CT(i,j,k)$) represents a voxel space and is assigned a Hounsfield number given by:

$$H = 1000\left(\frac{\bar{\mu}_m - \bar{\mu}_w}{\bar{\mu}_w}\right) \tag{B.1}$$

where $\bar{\mu}_m$ and $\bar{\mu}_w$ are the attenuation coefficients for an x-ray passing through the material and water respectively. As the ray propagates through the CT data, its intensity increases through a line integral calculation given by:

$$\zeta_{gh} = \sum_{i,j,k} l(i,j,k) CC(CT(i,j,k)) \tag{B.2}$$

where $l(i,j,k)$ is the length of the ray passing through a voxel and CC is a calibration curve used to map the Hounsfield number to its respective attenuation coefficient. Solutions for the line-integral calculation are based on a method presented by Robert L. Siddon in [1]. Siddon's method calculates the radiological path lengths ($l(x,y,z)$) by determining the intercepts of the ray with the CT voxel planes.

Modifications have been made to Siddon's method to improve the overall calculation time for a DRR image. This is done by replacing computationally expensive calculations with incremental steps. The incremental modifications were originally proposed by Jacobs *et al*, [2], for 2D data sets. This document extends the equations to 3D, thus allowing for DRR calculation. Furthermore, the computational results are compared to a standard Siddon method implementation.

2 Initialisation

A DRR algorithm models a portal x-ray as a series of discrete rays that forward propagate from the x-ray's start point, through the CT data, to a pixel on the DRR image. To fully describe the portal x-ray configuration, seen in Figure B.1, the DRR algorithm requires the x-ray's start point, orientation and distance to the central point on the DRR image to be known.

The treatment plan, provided by iThemba LABS, describes the orientation of the treatment beam (equivalent to the central x-ray beam) by defining the co-ordinates of the beam's entry ($e_{x,y,z}$), target ($t_{x,y,z}$) and exit ($e'_{x,y,z}$) points. Using these co-ordinates and defining the lengths of the x-ray's start and end points from $t_{x,y,z}$, the ray's start point ($b_{x,y,z}$) is calculated using:

$$b_x = t_x + \frac{l_{bt}}{l_{et}} (e_x - t_x) \tag{B.3}$$

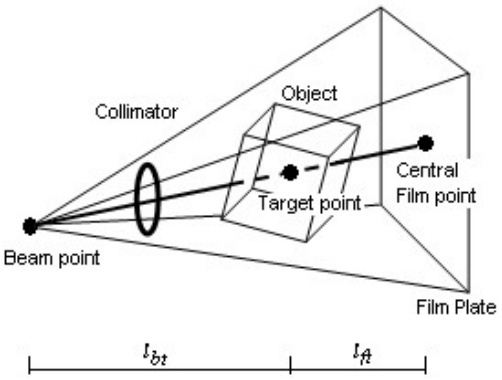


Figure B.1: Configuration of a portal x-ray imaging device

where l_{bt} is the distance from the x-rays start point to the target point and l_{et} is the distance from $e_{x,y,z}$ to $t_{x,y,z}$ denoted as:

$$l_{et} = \|t_{x,y,z} - e_{x,y,z}\| \tag{B.4}$$

Furthermore, the orientation $(\psi_{x,y,z})$ of the central x-ray beam is calculated using:

$$\begin{aligned} \psi_x &= \arctan\left(\triangle bt_y / \triangle bt_z\right) \\ \psi_y &= \arctan\left(\triangle bt_x / \triangle bt_z\right) \\ \psi_z &= 0 \end{aligned} \tag{B.5}$$

where:

$$\triangle bt_x = t_x - b_x \tag{B.6}$$

The distance from the x-ray's start point to the central point on the DRR (l_{bf}) is calculated using:

$$l_{bf} = l_{bt} + l_{ft} \tag{B.7}$$

Using the above equations, the co-ordinates of each pixel on the DRR image$(f_{x,y,z})$ can be calculated. This is done by transforming a 2D plane, situated at a distance l_{ft} along the z-axis, by $\Im(t_{x,y,z}, \psi_{x,y,z})$, where $\psi_{x,y,z}$ is given by B.5 and $t_{x,y,z}$ is the target point.

The x-plane co-ordinates are given by:

$$\begin{aligned} f'_x(i) &= \left[-\frac{pix_x}{2} + 1,\ f'_x(i-1) + \frac{1}{res}\right. \\ &\quad \left., \ldots, \frac{pix_x}{2}\right] \end{aligned} \tag{B.8}$$

with a similar expression for the y-plane co-ordinates, where pix_x is the number of horizontal DRR pixels.

3 CT Data Calibration Curve

The CT Data values, known as CT-numbers, are given in Hounsfield units. To calculate a DRR, it is necessary to convert each CT-number (CT(i,j,k)) to its respective attenuation coefficient ($\mu(E)$) using a calibration curve. A simple calibration curve is defined as follows:

$$\mu(E) = k\left(1 + \frac{CT(i,j,k)}{1000}\right) \tag{B.9}$$

where k equals 0.1805 or 2, depending on whether CT(i,j,k) is negative or positive respectively.

4 Ray and Plane Notation

The ray starting at $b_{x,y,z}$ and progressing through the CT data to $f_{x,y,z}$ is defined by the following parametric equations:

$$\begin{aligned} x(\alpha) &= b_x + \alpha(f_x - b_x) \\ y(\alpha) &= b_y + \alpha(f_y - b_y) \\ z(\alpha) &= b_z + \alpha(f_z - b_z) \end{aligned} \tag{B.10}$$

The CT data is arranged as a series of voxel planes in the x, y and z directions. These planes are denoted:

$$\begin{aligned} X_{plane}(i) &= X_{plane}(1) + (i-1)d_x \;\; (i = 1, ..., N_x) \\ Y_{plane}(j) &= Y_{plane}(1) + (j-1)d_y \;\; (j = 1, ..., N_y) \\ Z_{plane}(k) &= Z_{plane}(1) + (k-1)d_z \;\; (k = 1, ..., N_z) \end{aligned} \tag{B.11}$$

where d_x, d_y and d_z are the distances between the x, y and z voxels and N_x, N_y and N_z are integer values, one greater than the number of voxels in the x, y and z directions respectively.

5 Incremental Method

Incremental methods replace computationally expensive calculations in Siddon's method by incremental steps. The amendments are made in the calculation of the parametric values(α) and the voxel indices(i, j, k). The parametric values correspond to the intersections of the ray with the CT voxels and i, j and k give the corresponding voxel intensity. These parameters can be incrementally determined as the intersections of the ray with the planes occurring at set increments. This is illustrated in Figure B.2, for a 2D example.

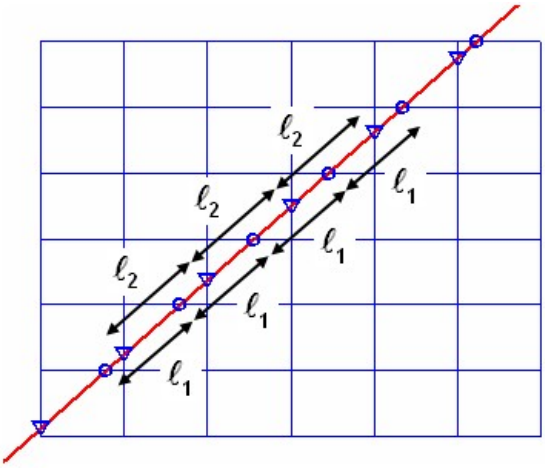


Figure B.2: Intersections of a ray with 2D planes in the x and y directions

The first step in the incremental method is to calculate the minimum and maximum parametric values (α_{min}, α_{max}) for the ray. These parametric values are determined by intersecting the ray with the sides of the CT array as follows: If $b_x \neq f_x$,

$$\begin{aligned} \alpha_x(1) &= (X_{plane}(1) - b_x)/(f_x - b_x) \\ \alpha_x(N_x) &= (X_{plane}(N_x) - b_x)/(f_x - b_x) \end{aligned} \tag{B.12}$$

with similar expressions for $\alpha_y(1)$, $\alpha_y(N_y)$, $\alpha_z(1)$ and $\alpha_z(N_z)$. However, if $b_x = f_x$, the ray is perpendicular to the x-axis and $\alpha_x(1)$ must be set to a number below 0 and $\alpha_x(N_x)$ to a number above 1. Using the parametric values determined above, the quantities α_{min} and α_{max} are given by:

$$\begin{aligned} \alpha_{min} \;=\; & max\,\{0, min[\alpha_x(1), \alpha_x(N_x)], min \\ & [\alpha_y(1), \alpha_y(N_y)], min[\alpha_z(1), \alpha_z(N_z)]\} \end{aligned} \tag{B.13}$$

$$\begin{aligned}\alpha_{max} &= min\left\{1, max[\alpha_x(1), \alpha_x(N_x)], max\right. \\ &\quad \left.[\alpha_y(1), \alpha_y(N_y)],\ max[\alpha_z(1), \alpha_z(N_z)]\right\}\end{aligned}$$

Using α_{min}, the minimum parametric values in the x, y and z directions ($\alpha_{min}^{x,y,z}$) and indices ($i_{min},\ j_{min},\ k_{min}$) can be calculated using:
If $b_x < f_x$:

$$\alpha_{min}^{x} = \left(d_x \left\lceil x(\alpha_{min})/d_x \right\rceil - b_x\right) / \left(f_x - b_x\right) \tag{B.14}$$

$$i_{min} = \left\lfloor N_x - \left(X_{plane}(N_x) - x(\alpha_{min})\right)/d_x \right\rfloor \tag{B.15}$$

else:

$$\alpha_{min}^{x} = \left(d_x \left\lfloor x(\alpha_{min})/d_x \right\rfloor - b_x\right) / \left(f_x - b_x\right) \tag{B.16}$$

$$i_{min} = \left\lceil N_x - 1 - \left(X_{plane}(N_x) - x(\alpha_{min})\right)/d_x \right\rceil \tag{B.17}$$

with similar expressions for the parametric and index values in the y and z directions. The expressions: $\lfloor . \rfloor$ and $\lceil . \rceil$ denotes a round down and up operations respectively.

The final step in the method is to incrementally determine the parametric values(α) corresponding to the plane intersections and voxel indices($i,\ j,\ k$). This involves iteratively updating the following parameters until $\alpha_c > \alpha_{max}$:

If $\alpha_x < \alpha_y < \alpha_z$

$$\begin{aligned} l &= (\alpha_x - \alpha_c) d_{bf} \\ \zeta_{gh} &= \zeta_{gh} + l \times CC(CT(i_n, j_n, k_n)) \\ i_n &= i_n + i_u \text{ if } l > 0 \\ \alpha_c &= \alpha_x \\ \alpha_x &= \alpha_x + \alpha_{xu} \end{aligned} \tag{B.18}$$

where the following parameters are initialised:

$$\begin{aligned} \alpha_x &= \alpha_{min}^{x} \\ d_{bf} &= \|f_{xyz} - b_{xyz}\| \\ \zeta_{gh} &= 0 \\ \alpha_{xu} &= \frac{d_x}{|f_x - b_x|} \end{aligned} \tag{B.19}$$

$$
\begin{aligned}
i_n &= i_{min} \\
i_u &= \begin{cases} 1 & if\ f_x > b_x \\ -1 & else \end{cases}
\end{aligned}
$$

with similar expressions for the parameters in the y and z directions.

6 Results

The incremental method is compared to Siddon's method, with respect to the DRR calculation times. The tests are implemented on a $256 \times 256 \times 58$ CT data set and DRR images having varying sizes. The results are illustrated in Figure B.3.

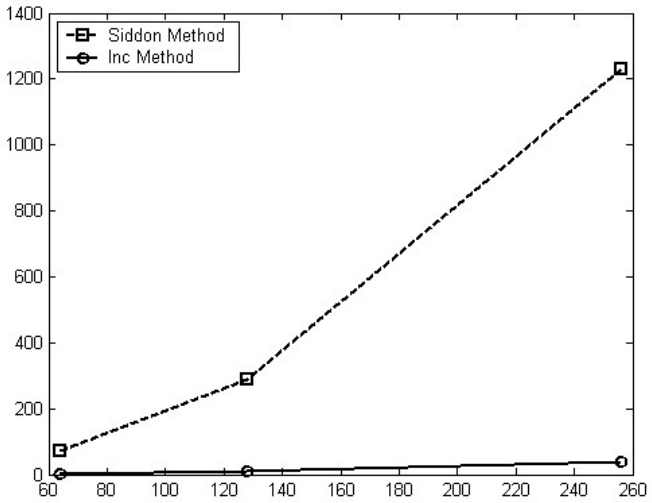


Figure B.3: Comparison of computational times between Incremental method (solid) and Siddon's method (dashed) for various DRR sizes

The DRR calculation methods were implemented using MATLAB© on an Intel Pentium 4 processor operating at 2.8GHz.

7 Conclusion

Many applications require DRR images to be calculated in real time. This paper presents an incremental implementation of Siddon's method for fast

DRR calculation. The incremental method was provided by F. Jacobs *et al* [2]. However certain alterations were made to the equations to extend it from a 2D to a 3D data set implementation and thus allow for DRR calculation. It was found that the incremental method was able to calculate a 256×256 DRR in 9 seconds for a $256 \times 256 \times 58$ CT data set.

References

[1] R. Siddon, "Fast calculation of the exact radiological path for a three-dimensional ct array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, March/April 1985.

[2] F. Jacobs, E. Sundermann, B. D. Sutter, M. Christiaens, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *Journal of Computing and Information Technology*, vol. CIT 6, no. 1, pp. 89–94, 1998.

Appendix C

Existing Field Boundary Alignment Algorithms

1 Introduction

In radiotherapy, the patient's treatment position is verified with respect to a pre-defined position. The patient's treatment position, with respect to the treatment beam is given by a portal x-ray (PR), which is acquired by placing a digital x-ray imaging device in-line with the treatment beam path. The resultant image has the patient's anatomy surrounded by the field boundary. However, with the use of smaller tumour-conforming fields in radiotherapy, it has become necessary to depict the patient's anatomical data past the field boundary. This is done by taking a double exposure x-ray, with the first exposure having the collimator out of the beam path, followed by a second with the collimator in the beam path. The resulting image has the collimator boundary projected onto the patient's anatomy.

Both single and double exposure x-rays require the alignment of the radiation field boundary with a reference field boundary given by the treatment plan. This requires the completion of two concurrent steps: the first step identifies the radiation field edges in the PR and the second step aligns these edges with the reference field edges. In medical literature, many algorithms have been proposed for the completion of both steps. The solutions are based on single

exposure PRs and certain adjustments must be made to extend the solutions to double exposure PRs.

2 Field Edge Detection and Alignment

2.1 Thresholding Techniques

Detection

Thresholding Techniques for field edge detection was originally used by Bijhold *et al* in [1] and has been extended by Dong and Boyer in [2]. Dong and Boyer's method uses a binary mask to create an image with the highest intensity values corresponding to the pixels within the field boundary and lowest to surrounding pixels. The mask is denoted:

$$bi_{pi}(i,j) \quad = \quad \begin{cases} 1 & if\, pi(i,j) \leq I_{threshold} \\ 0 & else \end{cases} \tag{C.1}$$

where $(pi(i,j))$ is the portal radiograph's pixel intensity at position (i,j) and $I_{threshold}$ is the threshold value.

This technique works well for a PRs having high differences in pixels intensities between the field boundary and anatomical data. For double exposure portal x-rays, increasing the intensity differences can be done by increasing the exposure of the second x-ray. However, this results in a higher radiation dose received by the patient.

Alignment

The binary image given by C.1, is aligned with a binary image of the reference boundary (bi_{ref}), which is created using a region-of-interest filling algorithm. Aligning the images requires aligning the centroids of the images that are calculated using geometric or orthogonal moments. A two-dimensional geometric moment for an image (Im) is calculated as follows:

$$m_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} i^p j^q Im(i,j) \tag{C.2}$$

where M and N are the number of pixels in the horizontal and vertical directions respectively and p and q are the orders of the moments.

Using C.2, the centroids of the image in the x and y directions can be calculated by:

$$\bar{x} = m_{10}/m_{00} \tag{C.3}$$

$$\bar{y} = m_{01}/m_{00} \tag{C.4}$$

Aligning the centroids of reference and radiation field boundaries is done by:

$$\Delta x = \bar{x}_{pi} - \bar{x}_{ref} \tag{C.5}$$

$$\Delta y = \bar{y}_{pi} - \bar{y}_{ref} \tag{C.6}$$

The scaling factor (sf) between the radiation and reference field boundaries is calculated using:

$$sf = \sqrt{(m_{00})_{pi}/(m_{00})_{ref}} \tag{C.7}$$

Once the centroids have been aligned and sf applied to the reference boundary, the edges can be correctly aligned using an optimisation procedure to minimise:

$$SM = \sum_x \sum_y (bi_{pi} - bi_{ref}(t_x, t_y, r, s))^2 \tag{C.8}$$

where t_x, t_y, r and s are the x, y translation, rotation and scale values respectively.

2.2 Radon Transform

Detection

The Radon Transform's use in field boundary identification was proposed by Eilertsen *et al* in [3]. The method firstly requires the edges in the PR to be identified using an edge detection algorithm (Canny, Sobel, Roberts, Prewitt or Laplacian of Gaussian). This is followed by identifying the radiation field edges using a Radon transform, where the co-ordinates of the highest intensities in the transform give the field edges.

In this method, the Radon transform is used as a line detection algorithm and thus it only works well if the edges of the radiation field are large.

Alignment

Given the equations of the lines representing the edges of the field boundary, the co-ordinates of the corner edges ($a(x, y)$) are calculated by determining the intersections of the neighbouring lines. These co-ordinates can be aligned with the co-ordinates of the reference field boundary ($b(x, y)$) by minimising the following expression:

$$\begin{aligned} E(x_k, y_k, s, \theta) &= \|a(x, y) - b((|b| \cos(\alpha + \angle b)s + x_k), \\ &\quad (|b| \cos(\alpha + \angle b)s + y_k))\| \end{aligned} \tag{C.9}$$

where x_k, y_k, s and θ are the x, y translation, scale and rotation parameters respectively.

3 Conclusion

Two previously proposed methods were presented for field boundary alignment, namely, thresholding techniques and the radon transform. Thresholding techniques use thresholding methods to identify the radiation field boundaries and geometrical or orthogonal moments to align the boundaries. The radon method identifies the edges using the Radon transform and aligns the co-ordinates of the corner edges using an optimisation routine.

It was found, for double exposure portal x-rays, that the thresholding technique will work successfully if there is a large intensity difference between the pixels enclosed by the field boundary and the patient's anatomical data. This will require increasing the exposure of the second x-ray. Alternately, the radon transform will work successfully if the radiation field edges can be determined using an edge detection algorithm and the length of the edges is large.

References

[1] J. Bijhold, K. Gilhuijs, M. van Herk, and H. Meertens, "Radiation field edge detection in portal images," *Phys. Med. Biol.*, vol. 38, pp. 1493–1502, 1993.

[2] L. Dong and L. Boyer, "A portal image alignment and patient setup verification procedure using moments and correlation techniques," *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.

[3] K. Eilertsen, A. Skretting, and T. Tennvassas, "Methods for fully automated verification of patient setup in external beam radiotherapy with polygon shaped fields," *Phys. Med. Biol.*, vol. 39, pp. 993–1012, 1994.

Appendix D

Existing Anatomical-Body Alignment Algorithms

1 Introduction

In radiotherapy, the patient's position with respect to the treatment beam is verified by geometrically aligning or registering a digital portal x-ray image (PR) with a prescribed digitally reconstructed radiograph (DRR). For the registration, numerous methods are available in medical literature. The bulk of the algorithms can be classified as being either feature based or intensity-based methods. Feature-based methods measure the similarity between previously extracted features or shapes from the two images, while intensity-based methods measure the similarity between geometrically corresponding pixel intensity values.

2 Feature-based Registration

For the verification of the patient's position in radiotherapy, various solutions use feature-based registration methods to align the patient's anatomy.

Feature-based registration requires the completion of two steps. The first step extracts corresponding features from the PR and DRR images. This is followed

by aligning the features using a least squares regression analysis [1]. Gilhuijs *et al* in [2], successfully extract features from the PR image using a top-hat transformation [3] or a multi-scale medial axis enhancement [4], while the features in the DRR are extracting using a high pass filter to mask the image. The medial axis is later extended by Yushkevich *et al* in [5] using methods presented in [6]. Eilertsen *et al* and Leszczynski *et al*, use a high pass filter or Difference of Gaussians (DoG) to mask the high intensity bone structures in the images [7].

Once the features are extracted they are aligned using either chamfer matching [8–10], a correlation measure [11] or a 2D ray-tracing measure [2].

3 Intensity-based Registration

Intensity-based registration, also known as voxel-based registration, measures the correlation between two images using a similarity measure. The method has various advantages over feature-based methods. Furthermore, it avoids the extraction of corresponding features and the similarity measure is straightforward, consistent and easy to interpret.

Dong *et al* [11], propose to use the cross correlation measure to calculate the misalignment between the PR and DRR images. This is extended by Clippe *et al* [12], to test inter-modality measures for the registration. Clippe *et al's* study showed that the cross correlation measure provided the most consistent output compared to the chi-squared, mutual information and correlation ratio measures.

4 Conclusion

Previously proposed feature-based and intensity-based measures are presented for patient position verification in conformal radiotherapy. It was found that intensity-based methods have various advantages over feature-based methods. Firstly, it avoids the extraction of corresponding features between the DRR

and PRs. This results in a fully automatic method, where the medical personnel avoid having to verify and edit the extracted features. Furthermore, the registration's accuracy is not limited by segmentation errors.

References

[1] L. Brown, "A survey of image registration techniques," Columbia University, New York, Tech. Rep., 1992.

[2] K. Gilhuijs, P. van de Ven, and M. van Herk, "Automatic three-dimensional inspection of patient setup in radiation therapy using portal images, simulator images and computed tomography data," *Med. Phys.*, vol. 23, no. 3, pp. 389–399, 1996.

[3] J. Serra, *Image Analysis and Mathematical Morphology, 2nd ed.*, 1988, vol. 1.

[4] D. Fritsch, S. Pizer, B. Morse, D. Eberly, and A. Liu, "The multiscale medial axis and its apllications in image registration," *Pattern Recognition Letters*, vol. 15, pp. 445–452, 1994.

[5] P. Yushkevich, D. Fritsch, S. Pizer, and E. Chaney, "Towards automatic, model-driven determination of 3d patient setup errors in conformal radiotherapy," University of North Carolina at Chapel Hill, Tech. Rep., 1998.

[6] S. Pizer, D. Fritsch, K. Low, and J. Furst, "2d and 3d figural models of anatomic objects from medical images," *Kluwer Computational Imaging and Vision Series*, pp. 139–150, 1998.

[7] K. Eilertsen, A. Skretting, and T. Tennvassas, "Methods for fully automated verification of patient setup in external beam radiotherapy with polygon shaped fields," *Phys. Med. Biol.*, vol. 39, pp. 993–1012, 1994.

[8] K. Leszczynski, S. Loose, and S. Boyko, "An image registration scheme applied to verification of radiation therapy," *Brit. J. of Rad.*, pp. 413–426, April 1998.

[9] K. Gilhuijs and M. van Herk, "Automatic on-line inspection of patient setup in radiation therapy using digital portal images," *Med. Phys.*, vol. 20, pp. 667–677, 1993.

[10] D. Frisch, E. Chaney, A. Boxwala, M. McAuliffe, S. Raghavan, A. Thall, and J. Earnhart, "Core-based portal image registration for automatic radiotherapy treatment verification," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 33, pp. 1287–1300, 1995.

[11] L. Dong and L. Boyer, "A portal image alignment and patient setup verification procedure using moments and correlation techniques," *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.

[12] S. Clippe, D. Sarrut, C.Malet, S. Miguet, C. Ginestet, and C. Carrie, "Patient setup error measurement using 3d intendsity-based image registration techniques," *Int. J. Radiation Oncol. Biol. Phys.*, vol. 56, no. 1, pp. 259–265, 2003.

Appendix E

Classical Optimisation Methods

1 Nelder-Mead Simplex

The Nelder-Mead simplex method is documented as the most popular direct search method for unconstrained optimisation problems. Initially, the algorithm creates a simplex having N+1 vertices's (x_i) in a N dimensional space. A simplex of size a is initialised around x_o using the following rule [1]:

$$x_i = x_o + pe_i + \sum_{k=1 \neq i}^{n} qe_k,\ i = 1, n \tag{E.1}$$

where e_i are the unit base vectors and p and q are defined as follows:

$$\begin{array}{rcl} p & = & \frac{a}{n\sqrt{2}}\left(\sqrt{n+1} + n - 1\right) \\ q & = & \frac{a}{n\sqrt{2}}\left(\sqrt{n+1} - 1\right) \end{array} \tag{E.2}$$

The simplex vertices co-ordinates are altered using *reflection*, *expansion* and *contraction* operators. The process is described as follows: for each iteration in the optimisation routine the vertex with the worst fitness measure is replaced by a new vertex. The co-ordinates of the new vertex is determined by reflecting the old vertex's position about the remaining vertices. A simple reflection of a two dimensional simplex is illustrated in Fig. E.1. If the fitness measure of the new vertices is lower than the previous removed vertex's fitness, the dimensions of the simplex is minimised. If not, it is enlarged. This process is continued until the vertices function evaluation values become similar, which

is measured using the following inequality:

$$\sqrt{\sum_{i=1}^{n+1} \frac{(f_i - \overline{f})^2}{n}} < \varepsilon \tag{E.3}$$

where ε is a small positive scalar and $\overline{f}$ is calculated using:

$$\overline{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f_i \tag{E.4}$$

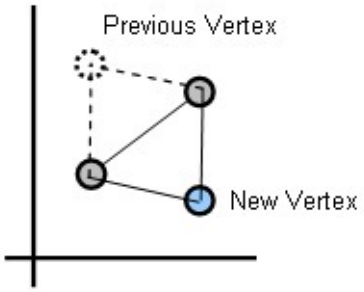


Figure E.1: A simple reflection of a two-dimensional Simplex

2 Quasi-Newton Algorithm

The Quasi-Newton optimisation algorithm is an effective, robust and quadratically convergent gradient-based optimisation method. The algorithm is a derivation of the Newton-Raphson algorithm, with the difference being that the inverse of the second derivative is updated using a one-dimensional or multi-dimensional Hessian approximation method. The Newton-Raphson algorithm is denoted as follows:

$$\tilde{x}_{n+1} = \tilde{x}_n - \eta \frac{\dot{f}(\tilde{x}_n)}{\ddot{f}(\tilde{x}_n)} \tag{E.5}$$

where $\dot{f}(\tilde{x}_n)$ is the Jacobian (first-derivative) of $f(x)$ and $\ddot{f}(\tilde{x}_n)$ is the Hessian (second-derivative).

For one-dimensional functions, $\ddot{f}(x_n)$ is updated using the following equation:

$$\ddot{f}(\tilde{x}_{n+1}) = s_n / y_n \tag{E.6}$$

where:

$$s_n = \tilde{x}_{n+1} - \tilde{x}_n \tag{E.7}$$

$$q_n = \dot{f}(\tilde{x}_{n+1}) - \dot{f}(\tilde{x}_n) \tag{E.8}$$

For multi-dimensional problems, the most popular method for approximating the Hessian is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The BFGS approximation of the inverse Hessian H_{n+1} is given by:

$$H_{n+1} = H_n + \frac{q_n q_n^T}{q_n^T s_n} - \frac{H_n^T s_n^T s_n H_n}{s_n^n H_n s_n} \tag{E.9}$$

3 Powell-Brent Procedure

The Powell-Brent Procedure is a robust, quadratically convergent direction-set optimisation method that successfully minimises non-differential functions [2]. The algorithm attempts to determine a function's minimum by altering a set of direction vectors. Initially the direction vectors are pre-defined and are updated each iteration to be mutually conjugate (non-interfering). The algorithm avoids the direction vectors becoming linearly dependent by discarding the direction with the largest decrease.

References

[1] M. Luersen and R. L. Riche, "Globalized neldermead method for engineering optimisation," *Computers and Structures*, 2004.

[2] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C++, Second Edition.* Cambridge University Press, 1998.

Appendix F

Population-Based Optimisation Methods

1 Genetic Algorithm (GA)

The genetic algorithm is based on the Darwinian principle of reproduction, survival of the fittest and naturally occurring genetic operations such as crossover and mutation [1]. In order to explain the algorithm, it is necessary to define the following terms:

- *Population* is an array of possible function solutions that are initially randomly selected. Using the GA algorithm, the array is evolved to finally represent the strongest solution.
- *Individuals/Chromosomes* is a value in the population that represents a candidate solution to the problem.
- *Fitness Function* is commonly known as the error function and is a measure of the discrepancy between the desired and current output responses. The GA finds the strongest solution (maximum). Therefore for minimisation, the function must be inverted.
- *Parent* is a chromosome used in the reproductive process to create new chromosomes in the population.

- *Child* is a new chromosome that is produced from the reproductive process.
- *Crossover* is part of the reproductive process, where the bits of two parents are combined to produce a child.
- *Mutation* is used with crossover, where a random bit in the chromosome is altered.

The aim of the algorithm is to find the strongest solution for a problem by genetically breeding a population of individuals over a series of generations [1]. Initially a random population is created, where each individual/chromosome in the population represents a candidate solution to the problem. In each iteration, sets of individuals (parents) are selected using *normal geometric, roulette wheel* or *tournament* selection functions. Using two of the selected parents a new chromosome (child) is produced using crossover and mutation functions. Popular crossover and mutation functions are *simple, uniform, arithmetic* and *heuristic* crossover and *boundary, uniform, non-uniform* and *Gaussian* mutation functions respectively. The strength of each chromosome (existing and new) is quantified using the fitness function. The weakest chromosomes are removed and a new population is established. The process of removing the weakest chromosomes, so that a stronger population is established is known as evolution. This process is further illustrated in Figure F.1

The parent selection functions are defined as follows:

- *Normal geometric* uses a normalised geometric distribution to select a subset of parents from the population.
- *Roulette wheel* selects parents from the population with probabilities proportional to the chromosomes fitness.
- *Tournament* uses N successive roulette wheel selections to produce a subset of chromosomes from the population. This applies additional selective pressure over an individual roulette wheel selection.

The mutation and crossover functions used to evolve the population are explained as follows:

- *Simple crossover* swaps one or two bits of two selected parents to create two new offspring.

- *Uniform crossover* uses a uniform distributed value to determine the number of bits to swap between the two parents.

- *Arithmetic crossover* linearly combines two parents (p_1 and p_2) to create new offspring (o_1 and o_2) using the following equations:

$$\begin{aligned} o_1 &= a \times p_1 + (1 - a) * p_2 \\ o_2 &= p_1 \times (1 - a) + a \times p_2 \end{aligned} \tag{F.1}$$

  where a is a random weighting factor.

- *Heuristic crossover* uses the parents with the highest (p_{high}) and lowest (p_{low}) fitness values to create new offspring. The offspring are calculated using the following equations:

$$\begin{aligned} o_1 &= p_{high} + r(p_{high} - p_{low}) \\ o_2 &= p_{high} \end{aligned} \tag{F.2}$$

  where r is a random number between 0 and 1.

- *Boundary mutation* replaces a randomly selected chromosome with either the upper or lower bound.

- *Uniform mutation* replaces a selected chromosome with a randomly uniform distributed value that falls within the user denoted bounds.

- *Non-Uniform mutation* decreases the probability of a mutation as the generation number (iterations) increases. This allows the algorithm to initially randomly search across the search space and as the generations increase, the solution is fine tuned.

- *Gaussian mutation* adds a unit Gaussian distributed random value to the selected chromosome.

In optimisation, the fitness value is generally the error function, which is a measure of the discrepancy between the desired and current output responses. The GA finds the strongest solution (maximum). Therefore for minimisation, the function must be inverted

2 Particle Swarm Optimisation (PSO)

Particle swarm optimisation is a population-based stochastic optimisation technique and is based on the simulation of social behaviour [2]. It was initially developed by J. Kennedy and R. Eberhart in 1995 by trying to simulate the social behaviour of birds flocking and fish schooling. PSO is similar to GA, where a initial population is randomly generated and is iteratively updated to converge to the best solution (strongest fitness) [2]. However, unlike GA, PSO has no evolutionary functions such as crossover and mutation [2]. PSO methods rely on the exchange of information between individuals (particles) in the population (swarm). In effect, the particles adjust their trajectory (position and velocity) towards the previous best position obtained by any member in its surrounding neighbourhood [2]. Each particle's position (x^j) in the swarm is updated as follows:

$$x_{k+1}^j = x_k^j + v_{k+1}^j \tag{F.3}$$

With a velocity (v^j):

$$v_{k+1}^j = w_k v_k^j + c_1 r_1 \left(p_k^j - x_k^j\right) + c_2 r_2 \left(p_k^g - x_k^j\right) \tag{F.4}$$

Where:

j = each particle in swarm

k = increment

r_1 & r_2 = random generated number $\in$ [0 1].

c_1 & c_2 = cognitive and social scaling parameters.

p_k^j = best (lowest fitness) position recorded by j-th particle.

p_k^g = best (lowest fitness) global position from one of the particles in swarm.

w_k = weight inertia.

The algorithm is illustrated in Figure F.2 This communication between particles, allows for the global sharing of information [2]. Hence particles can benefit from the discoveries and previous experience of other particles.

3 Simulated Annealing Process (SA)

Simulated annealing is a powerful stochastic search method and was first introduced by Kirkpartick et al [3] in 1983. SA is based on the cooling process (annealing) of metal [4], where a metal is heated-up to a temperature near its melting point and then slowly cooled-down according to a predefined schedule. This results in the metal having a better chance of forming a perfect crystal, which is the global minimum configuration of the system.

The SA algorithm starts with a Metropolis Monte Carlo simulation at a predefined temperature. After a sufficient number of Monte Carlo iterations, the temperature is decreased according to a defined schedule. At the new temperature, the Metropolis Monte Carlo simulation is continued. This process is continued until the final temperature is reached. The Metropolis Monte Carlo algorithm randomly searches the landscape for downhill and uphill movements. If the step in the search is downhill (current step's fitness is less than previously determined lowest fitness) the step is accepted. However if the step is uphill, it is accepted if its transition probability is less than a random number [4]. The transition probability is defined as follows:

$$p(T) = e^{-\frac{\triangle F}{T}} \tag{F.5}$$

Where:

$p(T)$ = transition probability

$\triangle F$ = reduction in fitness between current step and previously determined lowest fitness

T = current temperature

The acceptance of uphill movements prevents the SA algorithm from being trapped in a local optimum [5]. Hence it makes it superior to previously implemented search methods. The process is illustrated in Figure F.3.

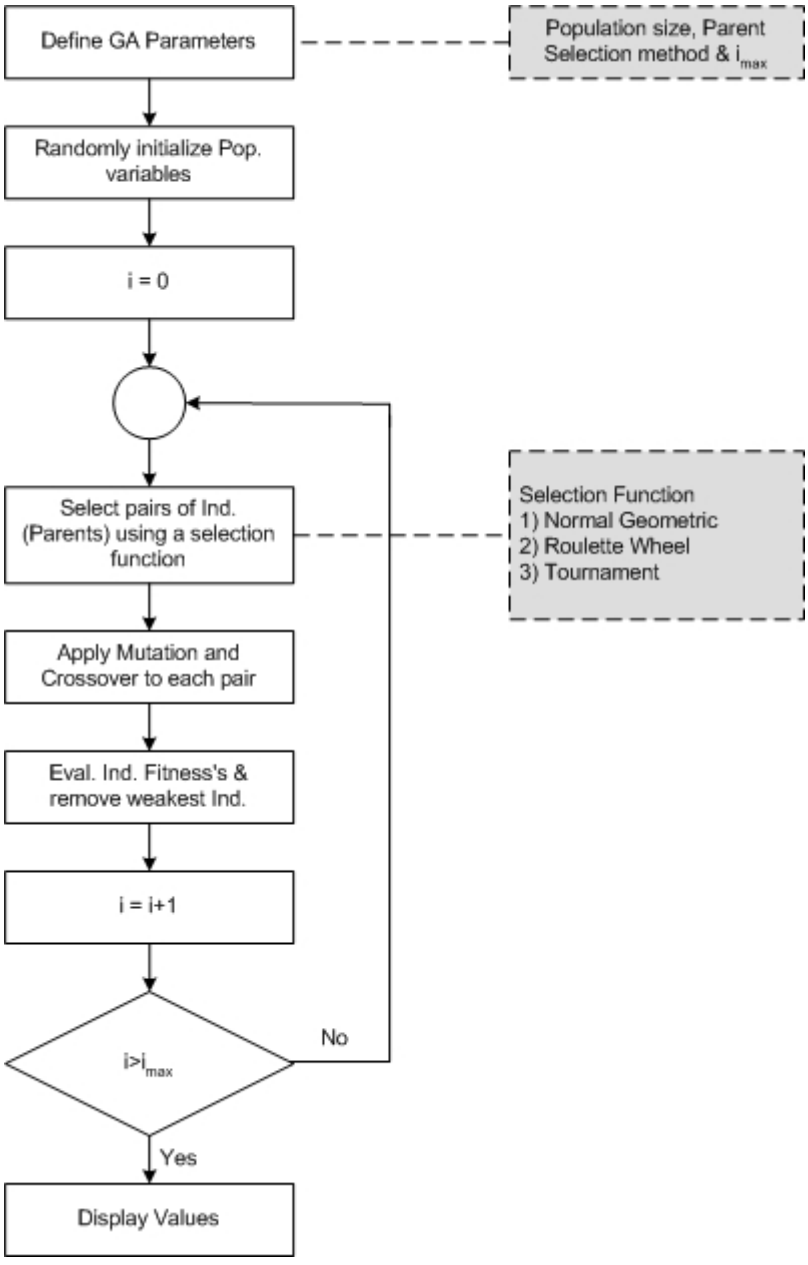


Figure F.1: Genetic Algorithm Flow Chart

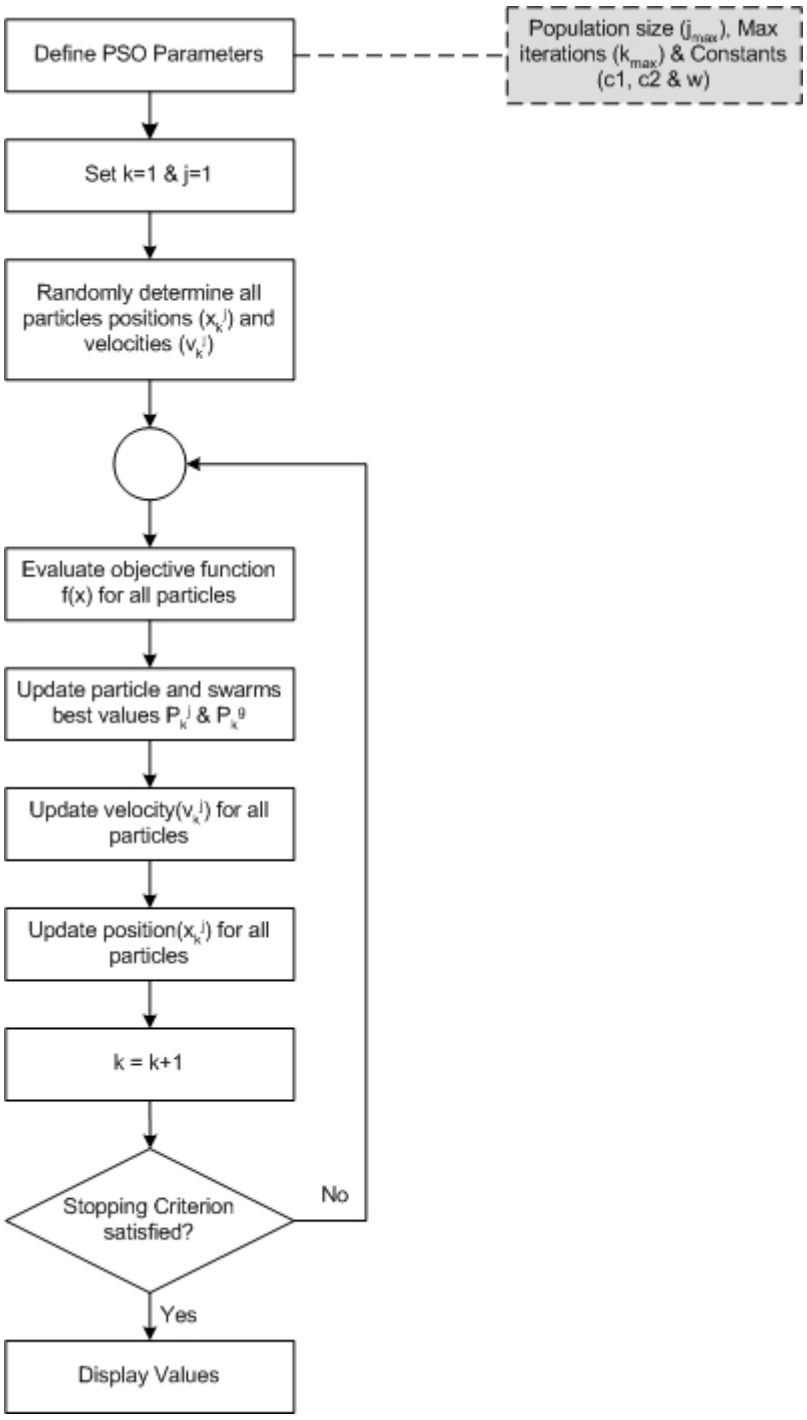


Figure F.2: Particle Swarm Algorithm Flow Chart

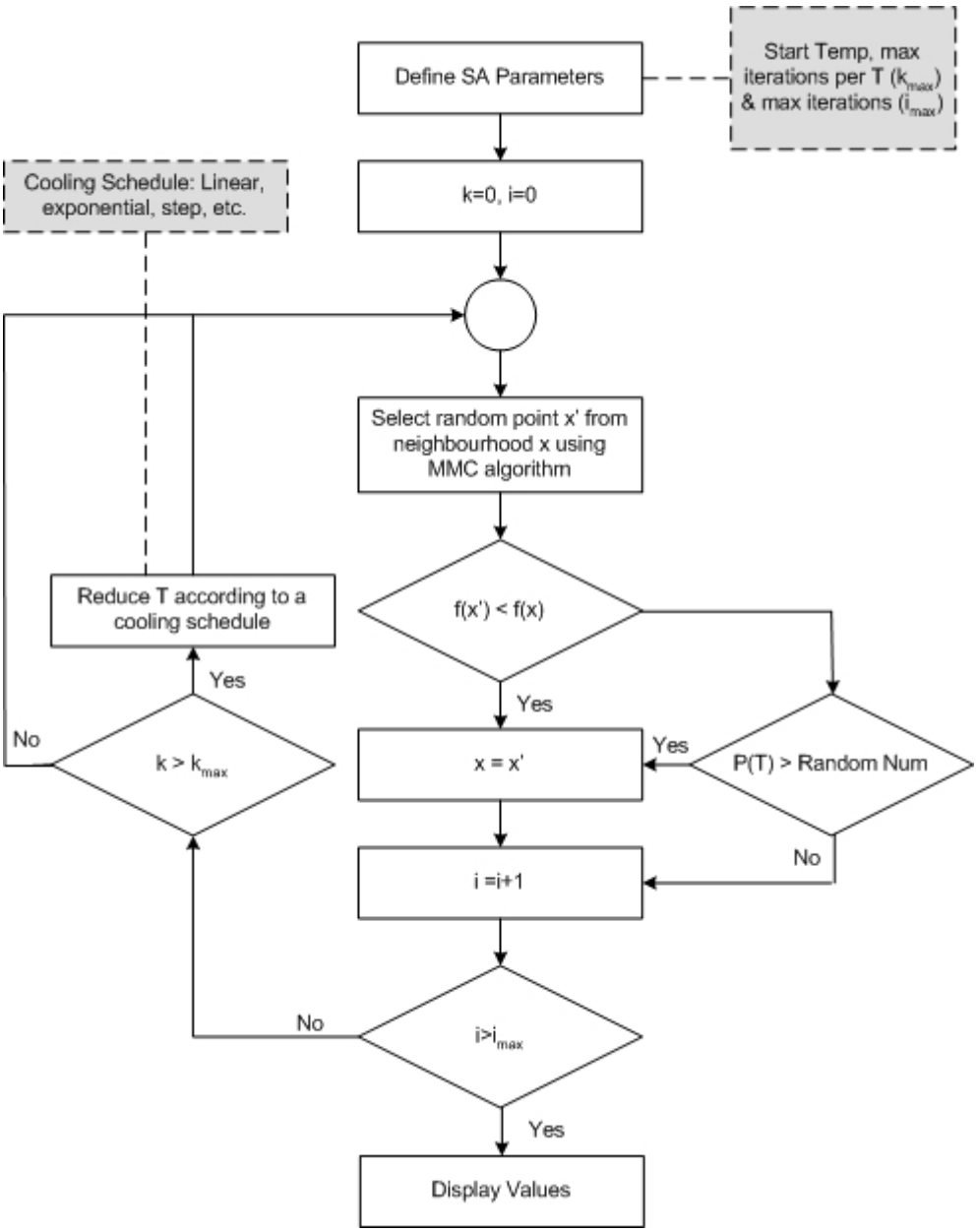


Figure F.3: Simulated Annealing Flow Chart

References

[1] J. Koza, "Genetic programming," Stanford University, Tech. Rep., 1997.

[2] K. Parsopoulos and M. Vrahatis, "Particle swarm optimization method for constrained optimization problems," University of Patras, Greece, Tech. Rep.

[3] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimisation by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–681, 1983.

[4] L. Nolle, D. Armstrong, A. Hopgood, and J. Ware, "Simulated annealing and genetic algorithms applied to finishing mill optimisation for hot rolling wide steel strip," *Int. J. of Knowledge-Based Int. Eng. Sys.*, vol. 6, no. 2, 2002.

[5] H. Yildiz, "Simulated annealing and applications to scheduling problems," Bilkent University, Turkey, Tech. Rep., 2000.