Deformable Area-based Template Matching with Application to Low Contrast Imagery

Martin Berger
Image Science Group
Swiss Federal Inst. of Technology
8092 Zürich, Switzerland
berger@vision.ee.ethz.ch

Guido Gerig
Dept. of Computer Sciences
University of North Carolina
Chapel Hill, NC 27599-3175
gerig@cs.unc.edu

Abstract

This paper reviews the mathematical formalism of the least squares template matching (LSM) and presents a framework for automatic quality control of the resulting match. LSM is an iterative and area-based fitting method which replaces the conventional multi-step procedure of feature extraction followed by discrete parameter search. The technique is especially suitable for attaining very high precision or for processing low-contrast, noisy and blurred imagery as it takes into account the full information of the image signal. The automatic quality control—a component often missing in commonly used image matching methods—is achieved by self-diagnostic measures supervising the iterative procedure.

The development is driven by applying image analysis in order to control patient position in radiotherapy treatment. The exact positioning of patients during radiotherapy is essential for high precision treatment. Accurate information about patient position is gained by the automated matching of electronic portal images acquired during treatment sessions. However, the particular problem of such megavoltage X-ray imagery is its extremely low contrast, rendering reliable feature extraction a difficult task and thus favoring the LSM approach. The presented system includes both field edge alignment and 2D anatomy displacement measurement based on sets of template regions of stable anatomic features selected by physicians. A very promising success rate of over 90% was achieved in clinical test data which consisted of roughly 70 image series with totally 500 portal images.

Using digitally reconstructed radiographs (DRRs) as simulated portal images with known ground truth, the measurement errors were analyzed in more detail. In particular, the systematic error caused by measuring only the projection of the patient position was estimated, resulting in error bounds for the additional influence of small out-of-plane patient motion. Furthermore, promising results for the multi-modal match between a DRR reference image computed from the patient’s CT and therapeutic portal images are presented. This demonstrates the feasibility of a direct link between CT volume data with its associated treatment plan and electronic portal images, which can significantly increase the accuracy of a treatment.

The implemented measuring software including a graphical user interface is running on a Workstation in the University Hospital of Zürich. This enables clinicians to carry out medical studies, which should eventually lead to an improved quality assurance in radiotherapy.

Keywords—portal imaging, image matching, deformable templates, self-diagnosis
1 Introduction

Accurate motion measurements in images are essential to solve numerous problems in computer vision. For medical imagery in particular, precise position measurements and registration of image series represent two important applications. Various methods have been applied depending on the specific problems to be solved. The most important difference between these methods lies in the type of image features used, leading to a classification into two groups: area-based and feature-based methods. Area-based methods take into account the full information of the image signal since they directly compare the greyvalues, whereas feature-based methods rely on previously extracted features.

However, most algorithms of both groups assume either a fair image quality or a rather restricted parameter search space in order to function reliably. Thus, low-contrast, noisy, and distorted imagery poses problems for most algorithms. Feature-based methods especially often fail since the step of feature extraction will most probably be unreliable. The limiting factor in area-based methods is usually the restricted degree of freedom of the geometric transformation.

Furthermore, the strong requirements for quality assurance in quantitative medical applications demand a rigorous error propagation analysis and self-diagnostic capability. But the complexity of many algorithms render the computation of such statistical measures difficult.

The specific goal in this work is the exact positioning of patients during radiotherapy, which is essential for high-precision treatment. This involves automatically measuring patient setup deviation between or even during treatment sessions. The most promising sensor is an electronic portal imaging device (EPID), which delivers images of the exit dose distribution during treatment. Unfortunately, the contrast of these megavoltage X-ray images is very low due to the high energy beam, and, since we are dealing with projected images, parts of a rigid 3D motion must be estimated by evaluating projected 2D images.

In [Berger and Danuser 1997] we propose the area-based LSM algorithm to find displacements between two portal images, yielding a fast and accurate image matching procedure. Extending this approach, we further exploit the self-diagnosis capabilities. The variation of image quality of portal images inhibits the estimation of a 2D affine transformation for all cases. An adaptive scheme based on self-diagnostic measures allows for an automatic reduction of the parameter set where the full parameter set is not determinable. Furthermore, we include the multi-modal matching of portal images against digitally reconstructed radiographs computed from the CT volume data.

1.1 Previous work on portal images

So far, several methods for portal image registration have been implemented, all of which lack either the robustness or the accuracy necessary to be used reliably in daily hospital routine. Furthermore, most state of the art techniques require too much user interaction. Algorithms like point-to-point (or landmark) matching greatly depend on the exact localization of the landmarks by the physician. This is not only time-consuming, but also varies for different operators. Less user interaction is required by the chamfer matching algorithm where significant ridges are manually outlined in a reference image and matched onto the detected features of the treatment image [Gilhuijs and van Herk 1993]. A similar approach in the sense that it also uses binary features, namely cores, is described in [Fritsch et al. 1995].

Since portal images are inherently noisy and low in contrast, it is difficult to robustly detect features like edges, ridges or cores. Therefore, an area-based match promises to be superior to a feature-based algorithm. Greyvalue correlation techniques are described in [Dong and Boyer 1996, Moseley and Munro 1994, Radcliffe et al. 1994]. Their limitations lie in the restriction to a translation or in a coarse search grid for computational reasons.

None of the papers mentioned show an error propagation analysis for the measured displacement. This would be crucial for any kind of automated quality control. In current clinical practice, the result must be checked visually by a physician.
1.2 Previous work on least squares template matching

The method of least squares template matching (LSM) with deformable templates meets the requirements of being an area-based approach with rigorous error propagation, self-diagnosis and thus, minimized user interaction. Early work in this field was presented by Lucas and Kanade [1981], who published an iterative image registration scheme base on LSM. Among the first papers that discussed the concept of exploiting the full information of the statistical models for robust template matching are Grün [1985] and Förstner [1987]. Bergen et al. [1992] describe basically the same algorithm for motion estimation. It was further developed by Lindeberg [1995] using a multi-scale approach.

Following and extending the work of Grün, Danuser and Mazza [1996] achieved highly accurate results at the resolution limit of a light microscope. The high accuracy of this technique even in the case of low-contrast imagery is extensively exploited in Danuser [1996]. The paper reports on high accuracy positional measurements of a calibration grid used to calibrate a stereo light microscope. Compared to this application, additional problems arise in portal images from the higher complexity and variability of the image scene, and from the effects of out-of-plane rotations on projected images. On the other hand, the requirements on accuracy are not as high in portal imaging.

A similar technique for the registration of medical image series is reported by Unser et al. [1995], where each image is matched to the reference image based on a global greyvalue difference measure. In contrast to their work, our framework does not rely on one global template, but on several small templates each containing a significant image structure. Thus, the inclusion of distinct but insignificant image features which vary between the data of one sequence is avoided and the impact of global greyvalue errors such as intensity inhomogeneity is reduced.

2 Least squares template matching

LSM is an area-based matching algorithm. It replaces the conventional multi-stage approach where feature extraction is followed by thresholding, binarization and a discrete search. Thus, LSM does not depend on the extraction of binary (also called non-iconic) image features. This is an important advantage in low-contrast and blurred imagery, where feature extraction is mostly unreliable. Furthermore, unlike in most correlation methods, the optimum transformation is not searched on a discrete grid, but approached using an optimization scheme. Assuming that a fair initial guess can be supplied, this is not only faster but also more accurate.

The image signal of a template is fitted into the search image, minimizing the least squares error between the greyvalues of the two regions. The fitting procedure has to account for both radiometric and geometric transformations. In order to avoid ambiguities, these two transformations have to be estimated separately. This is achieved by estimating the parameters for the radiometric transformation based on a global measure within the template region before each iteration. Thus, they are not directly included in the actual least squares optimization.

In the first three sections, the generic framework of solving least squares problems is reviewed including error propagation and self-diagnosis. The less general case of template matching is presented in detail starting with section 2.4.

2.1 General least squares framework

Least squares estimation is widely dealt with in standard literature. In this section, a short summary is given about the parts relevant to LSM. For more in-depth information the reader may for instance refer to Höpcke [1980], where the emphasis is on adjustment theory, or Koch [1988] for parameter estimation.

Least squares estimation problems are commonly described with a set of relations between unperturbed measurements, denoted by \( \ell \), and the unknown model parameters \( \zeta \). The most general model...
consists of a set of functions \( F_k \), which we collect in a vector valued function

\[
F(\zeta, \ell) = 0 .
\]  

(1)

The unperturbed measurements \( \ell \) are unknown before the optimization. The actual measurements \( l \), also called observations, are always subject to errors and only their expected value is equivalent to \( \ell \).

\[
\ell = E[l]
\]

In order to account for these errors, the expected value \( \ell \) is replaced by \( \ell = l + e_l \). The vector \( e_l \) is called residual vector. Estimates of the measurements and parameters must not be confounded with the actual unknown value and are denoted by \( \hat{\ell}, \hat{\zeta} \), respectively.

The measurements are often at least partly separable from the parameters. Denoting the separable observations with \( l_2 \) and the others with \( l_1 \), we write

\[
F_1(\zeta, l_1 + e_{l1}) + F_2(l_2 + e_{l2}) = 0 .
\]  

(2)

The residuals can also be introduced in the space of the function vector \( F \) instead of the observation space of \( l \). However, in some circumstances there are substantial disadvantages and a few thoughts are necessary before such simplification. Solving (2) in the least squares sense results in a best linear unbiased estimator\(^1\) for \( \zeta \). If we further assume normally distributed observations, it turns into a maximum likelihood estimator (MLE). These important properties are lost when changing the residual model to functional residuals. Nevertheless, under certain assumptions about the functions \( F_i \), the additional bias can be neglected. One important prerequisite is that \( F_i \) must be linear in \( l_i \), which we will always assume within this work. For a more detailed discussion about differences between these two residual models we refer to [Danuser and Stricker 1998]. Exploiting these assumptions, equation (2) simplifies to

\[
F_1(\zeta, l_1) + e'_{l1} + F_2(l_2) + e'_{l2} = 0
\]

\[
F_1(\zeta, l_1) + F_2(l_2) + e_l = 0
\]

(3)

where we combine the residuals \( e_l := e'_{l1} + e'_{l2} \).

At this point we also introduce the notion of the weight matrix \( P_{\ell\ell} \) and the cofactor matrix \( Q_{\ell\ell} \). Denoting the covariance matrix of the measurements \( \ell \) with \( \Sigma_{\ell\ell} \), their weight matrix is defined as multiple of the inverse of \( \Sigma_{\ell\ell} \).

\[
P_{\ell\ell} = \sigma_0^2 \Sigma_{\ell\ell}^{-1}
\]

(4)

The factor \( \sigma_0^2 \) is called the variance of unit weight, since with \( P_{\ell\ell} := I \) it follows \( \Sigma_{\ell\ell} = \sigma_0^2 I \). The cofactor matrix \( Q_{\ell\ell} \) is defined as the inverse of \( P_{\ell\ell} \)

\[
Q_{\ell\ell} = P_{\ell\ell}^{-1} = \frac{1}{\sigma_0^2} \Sigma_{\ell\ell} .
\]

(5)

Defining the covariance matrix of the measurements is an important modeling step. Usually, it is not possible to fully define \( \Sigma_{\ell\ell} \), as the factor \( \sigma_0^2 \) is often unknown a priori. Nevertheless, the weight matrix \( P_{\ell\ell} \) is specified after an a priori analysis of the estimation problem. Therefore, the weights of the optimization are based on physical properties instead of heuristics, which is an important advantage over other optimization methods. In particular, the knowledge of the relative uncertainty of the input data renders data fusion and error propagation possible.

The least squares objective function aims at minimizing the squared sum of the residuals considering the different weights of the observations. Thus, we can formulate the goal function as \( \Theta = e_l^T P_{\ell\ell} e_l \).

\(^1\)Strictly, the estimator is only unbiased for linear problems. Applied to nonlinear models, the random errors cause an additional statistical bias [Box 1971], which can be neglected in most cases.
Notice that for uncorrelated observations, \( P_\ell \ell \) is a diagonal matrix. Minimizing this goal function leads together with equation (3) to an unconstrained nonlinear least squares (NLS) problem

\[
\begin{aligned}
\text{minimize } & \quad e^T \ell P_\ell \ell e_\ell \\
\text{subject to } & \quad F_1(\zeta, l_1) + F_2(l_2) + e_\ell = 0.
\end{aligned}
\]

This optimization problem can either be solved using the Levenberg-Marquardt algorithm (cf. for instance [Press et al. 1994, pp. 683]) or simply by applying a Newton-Raphson scheme. Both methods iteratively solve this equation using a linear approximation for equation (3).\(^2\) This linearization is further discussed in section 2.2.

Constraints on the model parameters \( \zeta \) are formulated analogous to equation (1). Again, the constraint functions \( G_k \) are collected in a vector valued function \( G \):

\[
G(\zeta) = 0.
\]

The constraint NLS problem defined by equations (1) and (6) is either solved using a Lagrange formalism or by introducing the constraints as so-called zero or pseudo observations. In this work, we will concentrate on the latter interpreting the constraints as zero observations with very small variances.

\[
G(\zeta) + e_0 = 0,
\]

denoting the residuals by \( e_0 \) analogous to \( e_\ell \). Such constraints are called soft or spring constraints, in contrast to hard constraints when applying the Lagrange formalism.

In particular, constraints can be used to suppress the estimation of a certain parameter by fixing it to an a priori value. In this case, a constraint function \( G_k(\zeta) = \zeta_i - \bar{\zeta}_i \), where \( \bar{\zeta}_i \) is the a priori value for parameter \( \zeta_i \). Inserting into equation (7) it follows

\[
\zeta_i - \bar{\zeta}_i + e_{0_k} = 0.
\]

Another important case is tying two parameters together. It is straightforward to derive the corresponding constraint equation, which will force parameter \( \zeta_i \) and \( \zeta_j \) to be equal after the optimization

\[
\zeta_i - \zeta_j + e_{0_k} = 0.
\]

Using such constraints, implementation problems due to the varying number of parameters are avoided and the computation time needed for reparametrization is saved. This of course at the cost of a slightly less stable equation system and larger matrices.

All constraints \( G_k \) are added to the original observations using large corresponding weights in the augmented weight matrix \( P \). These weights correspond to the relative accuracy compared to the variances of the estimated parameters as will be discussed in section 2.2. Each constraint \( G_k(\zeta) = 0 \) will then hold through the optimization with this specified accuracy. The augmented function vectors and residual vector are denoted by

\[
\begin{aligned}
F^*_1 := & \begin{bmatrix} F_1 \\ G \end{bmatrix}, & F^*_2 := & \begin{bmatrix} F_2 \\ 0 \end{bmatrix}, & e := & \begin{bmatrix} e_\ell \\ e_0 \end{bmatrix},
\end{aligned}
\]

leading to a formally unconstrained NLS problem

\[
\begin{aligned}
\text{minimize } & \quad e^T P e \\
\text{subject to } & \quad F^*_1(\zeta, l_1) + F^*_2(l_2) + e = 0.
\end{aligned}
\]

Matrices \( P \) and \( Q \) represent the augmented weight matrix and cofactor matrix, respectively:

\[
\begin{aligned}
P := & \begin{bmatrix} P_\ell \ell & 0 \\ 0 & P_0 \end{bmatrix}, & Q := & \begin{bmatrix} Q_\ell \ell & 0 \\ 0 & Q_0 \end{bmatrix},
\end{aligned}
\]

with \( Q_0 = P_0^{-1} \) being the cofactor matrix of the constraints.

\(^2\)The Levenberg-Marquardt algorithm is actually based on the gradient and Hessian matrix of the least squares goal function. However, the second derivatives are usually neglected, thus it is equivalent to a combination of Newton-Raphson and steepest descent.
2.2 Linearization and Error Propagation

One major advantage of the least squares framework is the possibility of rigorous error propagation. Equation (10) defines a NLS estimation problem which can be linearized around the current estimate \( \hat{\zeta}^0 \). Note that we already assumed \( F^*_i \) to be linear in \( l_i \) when the residual \( e \) was introduced. Thus, the derivatives \( \nabla_l F^*_i \) reduce to constants and are included in \( e \).

\[
F^*_1(\hat{\zeta}^0, l_1) + A \Delta \zeta + F^*_2(l_2) + e = 0
\]

(11)

Matrix \( A \) represents the Jacobian \( \nabla_\zeta F^*_1 \) and \( \Delta \zeta = \hat{\zeta} - \hat{\zeta}^0 \). Denoting the functional residuals by \( w = -F^*_1(\hat{\zeta}^0, l_1) - F^*_2(l_2) \) and solving for the residuals \( e \), the following expression for the goal function is found

\[
e^T P e = \Delta \zeta^T A^T P A \Delta \zeta + 2 \Delta \zeta^T A^T P w + w^T P w,
\]

(12)

which is solved by setting the first derivative with respect to the variable \( \Delta \zeta \) to zero

\[
A^T P A \Delta \zeta + A^T P w = 0.
\]

(13)

Thus, the estimates for the parameter change and the residuals are given by

\[
\hat{\Delta} \zeta = -(A^T P A)^{-1} A^T P w
\]

\[
\hat{e} = -A \hat{\Delta} \zeta - w
\]

The statistics of the parameter estimate \( \hat{\zeta} := \hat{\zeta}^0 + \hat{\Delta} \zeta \) in the adjusted system is the same as the statistics of the parameter update \( \hat{\Delta} \zeta \). Therefore, we can directly compute the cofactor matrix \( Q_{\zeta \zeta} \). To propagate the covariances from the observations to the estimated parameters, the following error propagation theorem is used:

**Theorem (error propagation).** The covariance matrix \( \Sigma_{yy} \) of a linearly transformed random vector \( y = H x + c \) is given by \( \Sigma_{yy} = H \Sigma_{xx} H^T \) (cf. for instance [Koch 1988, pp.115ff]).

Thus, we derive the cofactor matrix \( Q_{\zeta \zeta} \) using the properties \( Q_{ww} = Q = P^{-1} \) and \( Q^T = Q \)

\[
Q_{\zeta \zeta} = (A^T P A)^{-1} A^T P \left( (A^T P A)^{-1} A^T P \right)^T
\]

\[
Q_{\zeta \zeta} = (A^T P A)^{-1}
\]

(14)

and a similar expression is found for the cofactor matrix of the residuals \( Q_{\hat{e} \hat{e}} \):

\[
Q_{\hat{e} \hat{e}} = Q - A (A^T P A)^{-1} A^T.
\]

2.3 Diagnostic Measures

The strict least squares framework allows for a number of diagnostic measures, for instance to judge the goodness of fit or the parameter determinability. Hence, various tests are applied to supervise the iteration progress and the final result, yielding self-diagnosis of the framework. The following section outlines the employed measures. Details and proofs about statistics and hypothesis testing in general are found in many textbooks, for example [Papoulis 1991] or [Koch 1988] for hypothesis testing in linear models.

**Global model test**

After the completed adjustment, the a posteriori noise estimate \( \hat{\sigma}_0 \) is computed using

\[
\hat{\sigma}_0 = \frac{\hat{e}_l^T P_{\hat{e}l} \hat{e}_l}{n - r},
\]

(15)
where $n$ is the number of observations and $r$ is the number of parameters or degrees of freedom. It can be shown that $\hat{\sigma}_0$ is an unbiased estimator of $\sigma_0$. Since $\hat{\sigma}_0$ also includes errors caused by an inaccurate model, we wish to test the hypothesis $\hat{\sigma}_0 = \sigma_0$ using the random variable $q$ as test statistic:

$$H_0: \hat{\sigma}_0 = \sigma_0 \text{ against } H_1: \hat{\sigma}_0 \neq \sigma_0 \text{ based on } q = \frac{(n - r) \hat{\sigma}_0^2}{\sigma_0^2}.$$  

Under the assumption that the observations are normally distributed, the test statistic $q$ is chi-square distributed with $n-r$ degrees of freedom: $q \sim \chi^2(n-r)$. The hypothesis $H_0$ is accepted with a significance level of $\alpha$ if and only if

$$q > \chi^2_{\alpha}(n-r).$$

**Estimation of parameter accuracy**

Parameter estimation in linear least squares problems are extensively discussed in standard literature on parameter estimation theory (e.g. [Koch 1988]). The iterative solution of equation (3) is an unbiased estimate for the unknowns with a stochastic variance expressed by the diagonal elements of the covariance matrix

$$\Sigma_{\xi\xi} = \hat{\sigma}_0 \cdot Q_{\xi\xi}. \quad (16)$$

The value $\hat{\sigma}_0$ denotes the a posteriori noise estimate computed using equation (15) and $Q_{\xi\xi}$ is the cofactor matrix as defined in the equation (14).

**Determinability analysis**

The determinability of a parameter $i$ is tested using its relative contribution $\delta_i$ to the trace of the cofactor matrix $Q_{\xi\xi}$

$$\delta_i = \frac{\text{tr}[Q_{\xi\xi}] - \text{tr}[Q_{i\xi\xi}]}{\text{tr}[Q_{\xi\xi}]}, \quad (17)$$

where $Q_{i\xi\xi}$ is the cofactor matrix with parameter $i$ excluded. $\delta_i$ describes the influence of each parameter upon the others. Weakly determinable parameters cause quasi-singular normal equations, thus a large change in the trace of the cofactor matrix.

An efficient implementation of (17) is achieved by using the Kalman-Bucy filter technique (cf. [Koch 1988]), computing the partial cofactor matrix $Q_{i\xi\xi}$ directly from $Q_{\xi\xi}$. Applying this framework to (17) yields an expression for the contribution

$$\delta_i = \frac{\sum_j q_{ij}^2}{q_{ii} \sum_j q_{jj}^2}, \quad (18)$$

where $q_{ij}$ denotes the elements of the full cofactor matrix $Q_{\xi\xi}$.

If a contribution $\delta_i$ of the shape parameter $i$ is high, this parameter strongly correlates to one or more parameters. One should either exclude parameter $i$ or combine it with the correlating parameters by applying the formalism for parameter constraints described in section 2.1.

Strictly, the contribution value is only meaningful in the adjusted state, that is after optimization. Still it is useful before or during the iteration procedure employing a multistage approach. The application of such a procedure to LSM is explained in section 2.9.

**Cross correlation value**

In the adjusted state, i.e. after completing the optimization, the normalized cross-correlation between the model and the observed data represents a criterion for the goodness of fit. At the correct estimate, this value is very close to 1.0 and therefore is suitable to discriminate wrong results from correct estimates.
2 Least squares template matching

2.4 Unconstrained LSM

The LSM includes two observations, the template image \( f[\cdot] \) and the search image \( g[\cdot] \), called patch. While the template is independent of the model parameters \( \zeta \), the patch \( g[\cdot] \) cannot be separated. Applying strictly the least squares framework of equation (2), this leads to the equation

\[
F(\zeta, g + e_g) + f + e_f = 0,
\]

where \( f \) correspond to \( l_2 \) and \( g \) to \( l_1 \). In the case of LSM, where the function \( F \) consists of a geometric and a linear radiometric transformation of \( g \), the abovementioned simplification of applying the residuals in the transformed observation space does not considerably influence the result. Hence, equation (3) is used instead:

\[
F(\zeta, g) + f + e = 0. \tag{19}
\]

The main modeling step lies in the definition of the function \( F \). The simultaneous estimation of both types of transformations would lead to an overdetermined system, since it is not possible to distinguish them locally (cf. figure 1). In order to overcome this problem, the parameters of the radiometric transformation are estimated based on a global measure within the template region apart from the actual least squares optimization. The resulting radiometrically adjusted patch \( \bar{g}[\cdot] \) is then used for the next optimization step. In the case of a linear transformation this can be written as

\[
T_R : g \rightarrow \bar{g}, \quad \text{where} \quad \bar{g}[u] = \alpha + \beta g[u], \tag{20}
\]

where \( u \) stands for the discrete image coordinates. The radiometric parameters \( \alpha \) and \( \beta \) can then be treated as known values and hence equation (19) becomes

\[
\bar{F}(\xi, \bar{g}) + f + e = 0, \tag{21}
\]

writing \( \xi \) for the remaining parameters.

The geometric relation between the original template and the matched area is defined by an arbitrary transformation. Depending on the type of the chosen transformation, this allows for displacement, rotation and/or deformation of the template. In the general case, the image coordinates \( u \) are transformed using the parameter vector \( \xi \) to

\[
x = \psi(\xi, u). \tag{22}
\]

Thus, the function \( \bar{F} \) is defined as

\[
\bar{F}(\xi, \bar{g}) := -\bar{g}(\psi(\xi, u))
\]

which leads together with equation (21) to the final form of the LSM observation equations

\[
f[u] + e[u] = \bar{g}(x). \tag{23}
\]
Figure 2: Coordinate system transformation between template and search image.

Equation (23) represents a relation between each greyvalue within the template and its corresponding image intensity in the search image. Notice that square brackets denote functions defined on a discrete grid. The functions $g(.)$ and $\tilde{g}(.)$ simply represent the continuous versions of $g[.]$ and $\tilde{g}[.]$, respectively. Hence, the template greyvalues $f[u]$ are defined on the grid of the reference image while $\tilde{g}(x) = \tilde{g}(\psi(\xi, u))$ fall between the grid of the search image. Interpolating the greyvalues $\tilde{g}(\psi(\xi, u))$ for a given $\xi$ we substitute $\tilde{g}_\xi[u] := \tilde{g}(\psi(\xi, u))$.

Based on a coordinate list $u[k]$, equation (23) is reordered into a vector notation

$$f + e = \tilde{g},$$

building a series of $n$ equations, where $n$ is the number of pixels included in the template and $k = 1 \ldots n$. Together with the least squares objective function $e^T Pe$ this defines an unconstrained NLS problem. Following the framework described in section 2.2, equations (11) to (14), this nonlinear problem is iteratively solved using a Newton-Raphson scheme. A new estimate $\hat{\xi} = \xi^o + \Delta \xi$ is computed linearizing the observation equations around the current estimate $\xi^o$ or $x^o = \psi(\xi^o, u)$, respectively:

$$f[u] + e[u] = \tilde{g}(x^o) + \nabla_\xi \tilde{g}(x^o) \Delta \xi$$

$$f + e = \tilde{g}^o + A \cdot \Delta \xi.$$  

Matrix $A$ is the $n \times r$ Jacobian matrix with respect to the parameter vector $\xi$ and $r$ denotes the number of parameters.

If we neglect the stochastic nature of the Jacobian matrix, this linear problem corresponds to a Gauss-Markov model with full rank. Note that only under this assumption, the observations are separated from the parameters. The linear problem (26) is solved analytically setting the first derivative of the least squares goal function $e^T Pe$ to zero, which yields the normal equation system

$$A^T P A \cdot \Delta \xi = -A^T P (\tilde{g}^o - f)$$

$$N \cdot \Delta \xi = -A^T P w.$$  

Notice that if $P$ is a diagonal or a band-diagonal matrix, the matrix $A$ does not have to be computed and stored as a whole. The weight matrix $P$ is diagonal if and only if the observations are independent of each other. In the case of LSM, this assumption usually holds\(^3\) and it is sufficient to compute $A$ row by row.

\(^3\)Strictly, the assumption is fulfilled only for the original observations $f$ and $g$, and not for the transformed and interpolated patch $\tilde{g}$. But the influence of neglecting this becomes only significant for very high precision measurements.
row in order to build the normal matrix $N$. This is an important property, since the size of $A$ increases with the number of pixels included in the templates.

As long as $A$ has no row deficiency, the $r \times r$ normal matrix $A^T P A$ is always positive definite and symmetric and hence the Cholesky decomposition can be applied to solve equation (27). After each iteration step, matrix $A$ must be recomputed using the updated set of parameters $\xi^{t+1} = \xi^t + \Delta \xi$. When the parameter change $\Delta \xi$ falls below a specified numerical resolution the iteration process is stopped.

### 2.5 Multi template extension

There are several ways to extend the standard LSM to multiple templates. For all of them, the equation (23) has to be adapted to include multiple templates and their corresponding patches. The most straightforward extension is to keep one single transformation for all patches such that $\psi^K \equiv \psi$ with the same parameter set shown in equation (30), leading to

$$f[u^K] + e[u^K] = \tilde{g}^K(x^K),$$

(29)

where $\tilde{g}[:] = \alpha^K + \beta^K g[:]$ and $x^K = \psi(\xi, u^K)$, analogous to equations (20) and (22).

Formally, this procedure is equal to defining one large template with several scattered regions of interest. However, since the radiometric parameters $\alpha^K$ and $\beta^K$ may vary between the templates, it is possible to compensate for global greyvalue differences like bias fields.

Using multiple template instead of one large template allows significant and stable regions to be selected without including regions unsuitable for matching. This is an important aspect for the application to megavoltage X-ray imagery and will be discussed in section 3.

### 2.6 Affine transformation as geometric transformation

So far no assumptions have been made on the dimensionality of the problem and on what type of transformation is used. In the following, the case of a two dimensional affine transformation is presented. The corresponding parameter vector consists of six variables $\xi = [t_1, t_2, m_1, s_1, s_2, m_2]^T$ and the coordinate transformation is written as

$$x = \begin{bmatrix} t_1 \\ t_2 \\ m_1 \\ s_1 \\ s_2 \\ m_2 \end{bmatrix} u.$$  

(30)

The derivative $\nabla_\xi \tilde{g}(x)$ is then calculated explicitly using the chain rule:

$$\nabla_\xi \tilde{g}(x) = \left( \frac{\partial}{\partial t_1}, \frac{\partial}{\partial t_2}, \frac{\partial}{\partial m_1}, \frac{\partial}{\partial s_1}, \frac{\partial}{\partial m_2}, \frac{\partial}{\partial s_2} \right) \tilde{g}(x)$$

$$= \left[ \frac{\partial \tilde{g}}{\partial t_1} \frac{\partial \tilde{g}}{\partial t_2} \frac{\partial \tilde{g}}{\partial m_1} \frac{\partial \tilde{g}}{\partial s_1} \frac{\partial \tilde{g}}{\partial m_2} \frac{\partial \tilde{g}}{\partial s_2} \right].$$

In vector notation, this leads to the $n \times 6$ Jacobian matrix $A$ (cf. equation (26)), each line $A_k$ representing the derivatives at $x_k = \psi(\xi, u[k])$. Denoting the derivatives by $\tilde{g}_{x1}(\cdot)$ and $\tilde{g}_{x2}(\cdot)$, we write

$$A_k = [\tilde{g}_{x1}, \tilde{g}_{x2}, \tilde{g}_{x1}u_1, \tilde{g}_{x1}u_2, \tilde{g}_{x2}u_1, \tilde{g}_{x2}u_2],$$

leaving out the parameter $x[k]$ of the functions $\tilde{g}_{x}(\cdot)$ for better readability. Using the resampled patch image $\tilde{g}[:]$ instead of $\tilde{g}(\cdot)$, the final form for the implementation is achieved

$$A_k = [\tilde{g}_{u1}, \tilde{g}_{u2}, \tilde{g}_{u1}u_1, \tilde{g}_{u1}u_2, \tilde{g}_{u2}u_1, \tilde{g}_{u2}u_2],$$

again omitting the parameter $u[k]$ of the functions $\tilde{g}_{u}[:]$.

As mentioned in the previous section, the normal matrix $N = A^T P A$ can be built computing $A$ row by row, as long as the weight matrix $P$ is diagonal. If the greyvalues of each pixel are considered independent this is fulfilled and $N$ is computed without the need to multiply large matrices.
2.7 Linear constraints

As we have shown in section 2.1, the least squares formalism allows one to introduce additional constraints in a simple and intuitive way. In addition to the observation equations, zero observations are included in the framework, which results in soft or spring constraints. Equation (8) and (9) give two examples of such constraint equations.

In this section, we will apply this technique to LSM. As an example serves the reduction of an affine to a similarity transformation. Instead of reparametrization we still employ equation (30) as transformation equation and add the following constraints to the parameter vector $\xi$:

$$
\begin{align*}
    m_1 - m_2 + e_m &= 0 \\
    s_1 + s_2 + e_s &= 0
\end{align*}
$$

The inclusion of the constraints in the normal equation system (28) strictly follows the framework shown in section 2.1. Analogous to the observation equations (23), the constraints are linearized around the current estimates $m_i^o$ and $s_i^o$. Thus, the matrix $A$ is augmented by the constraint vectors

$$
A_m = \begin{bmatrix} 0, 0, 1, 0, 0, -1 \end{bmatrix} \\
A_s = \begin{bmatrix} 0, 0, 0, 1, 1, 0 \end{bmatrix}
$$

and the corresponding residuals are given by $w_m = -m_1^o + m_2^o$ and $w_s = -s_1^o - s_2^o$.

2.8 Nonlinear constraints

Nonlinear constraints must be linearized in order to include them in the least squares framework. This is not trivial in the general case and often increases the risk of oscillation in parameter space. Still, this problem can be solved for special cases, in particular for the constraint

$$
m_1^2 + s_1^2 - 1 + e_c = 0 ,
$$

which is used together with constraints (31) for estimating a congruent transformation. The straightforward linearization around the current estimates $m_1^o$ and $s_1^o$ yields the linear constraint equation

$$
2 \Delta m_1 m_1^o + 2 \Delta s_1 s_1^o - (1 - m_1^o^2 - s_1^o^2) + e_c = 0 ,
$$

leading to the constraint vector $A_c = [0, 0, m_1^o, s_1^o, 0, 0]$ with a residual of $w_c = \frac{1}{2}(1 - m_1^o^2 - s_1^o^2)$.

The thin line in figure 3 illustrates the constraint line defined by this equation. Clearly, the simple linearized constraint is biased and thus often causes oscillation during optimization. An improved version is achieved by adding a correction term to the residual, computed by using simple geometric relations (figure 3, thick line). This adjustment results in the improved residual

$$
w_c = \sqrt{m_1^o^2 + s_1^o^2} - (m_1^o^2 + s_1^o^2) .
$$

2.9 Self-diagnosis within LSM

From the three diagnostic measures presented in section 2.3, we first focus on the determinability analysis. Since the statistics behind the self-diagnosis is valid only in the adjusted state, this measure can not be applied directly to the initial system. However, an upper bound for the determinability is computed matching the templates onto themselves before the actual optimization and analyzing the contribution values $\delta_i$ of this perfect match.

Based on this upper bound, a coarse result is computed using a restricted parameter set which still approximates the final parameter set sufficiently, for instance a congruent transformation. At this first
Figure 3: Graphical interpretation of the uncorrected linearized constraint (thin line) and corrected version (thick line). Oscillations are greatly reduced when using the corrected version.

\[ w_c = \sqrt{m_1^0 + s_1^0} - (m_1^0 + s_1^0) \]

\[ \frac{1}{2} (1 - m_1^0 - s_1^0) \]

\[ m_1^0 + s_1^0 = 1 \]

Figure 4: Flowchart of the diagnostic measures.
Figure 5: Overview of the different steps of radiotherapy treatment. The three steps CT+planning, simulator, and treatment yield three types of imagery: digitally reconstructed radiographs (a), diagnostic X-ray images (b), and portal images (c).

estimate, the full affine parameter set is tested for determinability. If none of the parameters show intolerably large contributions $\delta_i$, the optimization is continued with the full parameter set. A general flowchart is depicted in figure 4. More details on the contributions values are shown in the result section in figure 11 on page 18.

The global model test compares the a posteriori noise estimate $\hat{\sigma}_0$ with the a priori noise (or variance of unit weight) $\sigma_0$, which can be estimated based on the method presented by [Voorhess and Poggio 1987]. The model test is applied to each template and serves to locate possible errors or deficiencies in templates. However, it is not an efficient test to find misalignments, especially in low-contrast imagery.

More discriminative than the model test is the value of the cross-correlation in the adjusted state. The cross-correlation between the template $f[.]$ and the patch $\tilde{g}[.]$ — which is interpolated from the search image using the final parameter set $\xi$ — should be very close to 1.0. If multiple templates are used, the correlation value is computed for each template. Since a low correlation value indicates a mismatch for this region, excluding such a template and reoptimizing the geometric transformation usually results in a better parameter set.

3 Application to megavoltage X-ray imagery

One possible application of the LSM algorithm is the motion measurements in low-contrast imagery where feature-based methods often fail. In this case, the emphasis is less on the precision, but on the self-diagnostic capabilities of this method.

The driving project for this work is the accurate positioning of patients for radiotherapy treatment. In this section, we will first outline the current hospital practice and describe the imagery which we intend to use for checking the patient’s position. Second, various specific issues of this application are discussed.

3.1 Current procedure in high precision radiotherapy

In high precision conformal radiotherapy it is essential to accurately position the patient during the series of treatment sessions. Figure 5 gives an overview of the various steps of conformal radiotherapy. First,
3 Application to megavoltage X-ray imagery

A CT dataset is acquired. Based on this CT volume and the desired dose distribution, the irradiation directions and the field shapes — seen as white outlines in the megavoltage X-ray images — are defined in the planning step. At this stage, the theoretical dose distribution within the CT volume is computed. Furthermore, digitally reconstructed radiographs (DRR) can be computed from the CT data (figure 5a), which will be discussed in section 3.3.

In order to reach the theoretical dose distribution as exactly as possible, the patient must be positioned at each treatment session in accordance with the CT coordinate system. This is usually achieved using a therapy simulator and skin marks. During simulation, diagnostic X-rays are used instead of megavoltage X-rays, which allows standard X-ray images to be acquired (figure 5b). Based on these images, the patient is iteratively moved towards the optimum position. Skin marks are then used to be able to reproduce this position for each treatment session.

There are two main drawbacks of this conventional procedure. First, the patient positioning is only based on reference points outside the body. Second, these reference points are defined on the flexible skin, which introduces another source of error. For a more reliable alignment, features close to the tumor should be employed.

3.2 Electronic portal images

The most promising sensor for improving patient alignment is the electronic portal imaging device (EPID). This device delivers images of the exit dose distribution during treatment similar to a standard X-ray image. Figure 6 shows two typical portal images from the pelvis region.

Due to the high energy beam, the contrast of these megavoltage X-ray images is very low. Moreover, unstable features often appear as in figure 6b. The dark blur in the center of the image originates from air in the rectum and may not be used for matching. Also, the projective nature of the portal images should be taken into account when the 3D patient motion is to be estimated. This leads to the challenging task of estimating at least parts of the rigid 3D motion based on low-contrast 2D imagery.

The patient setup and four standard beam directions are depicted in figure 7. Since the position of the EPID is not completely fixed, the images usually differ slightly in position and scale. Thus, the field edge must be aligned in a first step in order to register a portal image series.

The portal images were acquired at the University Hospital of Zürich using a Varian accelerator and their electronic portal imaging device. This device delivers a distortion free image with a resolution of 256×256 pixel on an area of 32×32 cm² [van Herk and Meertens 1988].

3.3 Digitally reconstructed radiographs

As illustrated in figure 5a, artificial X-ray images or digitally reconstructed radiographs (DRR) can be computed from the CT volume [Sherouse et al. 1990, Chaney et al. 1995]. The therapy setup depicted in figure 7a is simulated by a ray-tracing based algorithm, correcting for the different absorption coefficients.
3.4 Choosing a reference image

In order to define an accurate reference image for checking patient position, there are mainly three possible choices.

1. The first or best portal image of a series, visually validated by a physician.
   Employing a portal image as reference image assures very similar greyvalue characteristics, hence yields the best match using an area-based matching algorithm. However, this portal image must be manually matched to the simulator image in order to establish a reference to the planning step. This visual validation is labor intensive and error-prone.

2. The diagnostic X-ray image from the simulator.
   When measuring patient motion visually using portal images, often the diagnostic X-ray image (see also figure 5b) is used as reference. Bony structures are easier to identify in this imagery, since the contrast is much higher. Nevertheless, it is not suitable for an area-based algorithm, since the different greyvalue characteristics strongly influence the quality of the measurements.

3. A digitally reconstructed radiograph.
   As described in section 3.3, the most accurate reference is a megavoltage DRR (see also figure 5a), which is directly computed from the planning CT. Thus, additional sources of positioning errors are avoided, for instance the step of therapy simulation. However, the remaining greyvalue differences pose additional problems to an area-based algorithm like LSM compared to simply match two portal images.

Figure 7: The therapy setup is illustrated in (a) including the patient coordinate system. The four main beam directions are shown in (b).
3.5 Field edge extraction and alignment

In general, the EPID is not in a fixed position and the image coordinate systems can differ between images. Thus, in a first step of the matching procedure, the field edge is aligned with LSM using small regions along the edge as templates. Since the field edge is a distinct image feature, it is feasible to extract the templates for the field edge match automatically based on the gradient magnitude. Figure 8 depicts examples of generated template configurations for various types of portal images. The absolute position of the field edge does not have to be known exactly, because the whole area flanking the edge is used for the alignment.

![Field edge templates](image)

Figure 8: Examples of the automatic definition of the field edge templates. The black lines outline the template regions, the dotted line represents the estimated field edge position.

In order to match the field edge, it is sufficient to estimate a similarity transformation, since there is no possibility of out-of-plane rotations. This restriction to four parameters and the fact that the field edge is an unambiguous feature makes the field edge match very robust. Hence, the results from the field edge match can be safely used as an initial guess for the anatomy match.

3.6 Selecting suitable anatomy templates

Suitable templates must consist of structures that are known to be stable over a series of images. However, due to artifacts and the presence of distinct but unstable features (for instance originating from air in the rectum), a fully automated template selection is beyond the possibilities of computer vision. Thus, the strategy is to implement an operator guided template selection. In the current implementation, the physician has to define regions containing significant structures that are known to be stable over a series of images. These regions are considered templates for the subsequent matching process. Figure 9 shows typical template configurations.

![Anatomy templates](image)

Figure 9: Images (a) and (b) show two typical template selections for AP pelvis fields. An example for a thorax image is given in (c).

In feature-based algorithms developed for this type of imagery, usually landmarks or contours must be manually defined. This is not only time-consuming, but also introduces a great variability between different operators. Since LSM is area-based, these problems do not arise and the physician only has to
3.7 Anatomy alignment

The resulting similarity transformation from the field edge match is used as an initial guess for the anatomy alignment. If determinable, an affine transformation is then estimated for the anatomy displacement. Figure 10 depicts an overview of the various algorithm parts and their results, which are marked with grey boxes.

The current implementation of the multi template LSM follows the strategy outlined in section 2.5, where one single affine transformation is introduced for all patches. The radiometric scaling parameter $\beta$ is kept the same for all patches. Only the constant parts $\alpha^K$ may differ between templates. Therefore, the total number of parameters using an affine transformation and $N$ templates amounts to $r = 6+N+1$. This strategy allows for inherent bias correction without adding too many degrees of freedom.

4 Results

The clinical data used for testing the algorithm consisted of roughly 70 portal image series with totally 500 images acquired during routine radiotherapy at the University Hospital of Zurich. In the first section, the performance of the statistical tests is analyzed. Then, results from the various parts of the matching procedure with and without known ground truth are presented.

The typical run time for the complete matching procedure including Gauss filtering of the search image, matching the field edge and matching the anatomy templates is about 4 seconds on a Sun Ultra 1 (167 MHz UltraSPARC CPU).
4.1 Statistical tests

The meaning of the contribution values $\delta_i$ in equation (18) is explained in figure 11. The $63 \times 63$ test image (figure 11a) contains a noisy corner with a signal to noise ratio of 10. It is Gaussian filtered with $\sigma = 1$. Such a feature allows for the estimation of the translation and the shear parameters but not the scaling parameters. Building the normal equation system by matching the template onto itself and using an affine transformation as defined by equation (30), the contribution values shown in figure 11b result. As expected, the contribution of the scale parameters $m_{1,2}$ is significantly higher than of the shear parameters $s_{1,2}$, which reveals the weak determinability of the two scale parameters.

$$
\begin{bmatrix}
\delta_{m1} & \delta_{s1} \\
\delta_{s2} & \delta_{m2}
\end{bmatrix} =
\begin{bmatrix}
0.4424 & 0.0922 \\
0.0958 & 0.3731
\end{bmatrix}
$$

Figure 11: Mesh view of the test image for the determinability analysis (a). Based on a match onto itself, the contribution values of all parameter are computed (b). As expected, the contribution of the scale parameters $\delta_{m_i}$ is significantly higher than of the shear parameters $\delta_{s_i}$.

The parameter precision is given by the diagonal elements of the covariance matrix $\Sigma_{\hat{\xi}\hat{\xi}}$ as defined in equation (16). In contrast to the overall accuracy, the parameter precision does not take into account external error sources. Thus, the precision represents an upper bound of the overall accuracy. Table 1 summarizes the range of typical parameter precision for the field edge and the anatomy match.

<table>
<thead>
<tr>
<th></th>
<th>translation [pixel]</th>
<th>rotation [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>field edge match</td>
<td>0.001 – 0.002</td>
<td>0.001 – 0.002</td>
</tr>
<tr>
<td>anatomy match</td>
<td>0.02 – 0.04</td>
<td>0.01 – 0.1</td>
</tr>
</tbody>
</table>

Table 1: Typical range of parameter precision for the field edge and the anatomy match. The parameter precision does not include external errors and thus represents an upper bound of the overall accuracy.

4.2 Field edge alignment

The displacement of the field edge is measured by estimating a similarity transformation. As expected, field edge alignment is very robust and LSM even finds larger displacements and rotations as occur in radiotherapy treatment. The number of necessary iterations typically lies between 5 and 15. Figure 12 shows the iterative computation of the optimum position and orientation starting from an initial guess, in this case an identity transformation. The search image was artificially rotated by -15° (clockwise). Still, a perfect match to the template was found after 15 iterations.

Furthermore, the diagnostic tools outlined in section 2.9—especially the cross correlation value—give a reliable feedback on the result. A few examples are given in figure 13, which show that a match of field edges with identical shape results in a cross correlation value well above 0.99, whereas small differences in the field edge shape already lower this value significantly.
### 4.2 Field edge alignment

Figure 12: Iteration series of an artificially rotated field edge of an AP pelvis image. The search image $(t = 0)$ is rotated by $-15^\circ$. The grey polygons outline the four automatically extracted templates and their placement on the search image. After 15 iterations a perfect match to the template is found.

![Field edge alignment iteration series](image)

Figure 13: Field edge alignment and its cross correlation values $\lambda$ for three typical cases. The top row shows image series with slight variation of the field edge shape. A match of field edges with identical shape results in a cross correlation value well above 0.99 (bottom row).

![Cross correlation values](image)
Figure 14: Setup of the self-consistency test. In (a), the original relation of two portal images (reference and search image) is shown. Diagram (b) depicts the situation after rotating the search image by -15°. The measurements on this rotated image are then compared with the expected values (see table below).

<table>
<thead>
<tr>
<th></th>
<th>(dx) [mm]</th>
<th>(dy) [mm]</th>
<th>(\text{rot}) [deg]</th>
<th>(\text{cross correlation})</th>
</tr>
</thead>
<tbody>
<tr>
<td>original (estimated)</td>
<td>-0.053 ± 0.0016</td>
<td>-6.888 ± 0.0013</td>
<td>0.047 ± 0.0013</td>
<td>0.9996</td>
</tr>
<tr>
<td>rotated (estimated)</td>
<td>-1.833 ± 0.0017</td>
<td>-6.637 ± 0.0014</td>
<td>-14.954 ± 0.0013</td>
<td>0.9996</td>
</tr>
<tr>
<td>rotated (ground truth)</td>
<td>-1.834</td>
<td>-6.640</td>
<td>-14.953</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Results of the self-consistency test, computed by estimating a similarity transformation. The estimated parameters (second row) almost perfectly agree with the expected values (third row).

In order to estimate the overall accuracy of the field edge match, a self-consistency test was carried out. Its setup is described in figure 14. After estimating the field edge displacement between two portal images (figure 14a), the search image was artificially rotated by -15° and the displacement was estimated again. Table 2 shows a detailed comparison of the results. The rotation measurements of 0.047° for the original images and -14.954° after the artificial rotation demonstrate perfect performance with respect to a ground truth.

### 4.3 2D patient displacement measurements

Due to much weaker contrast, the anatomy alignment is not as robust as the field edge match. Moreover, an affine transformation model is used for the anatomy match, which introduces two additional degrees of freedom. This also leads to a slightly higher number of iterations compared to the field edge match, usually lying between 10 and 20. The iterative optimization scheme is illustrated in figure 15 and 16 with an AP pelvis image. During the optimization, the resampled patch gradually rotates and translates to fit the greyscale values in the template.

In all tests, an affine transformation was used for matching and the initial position was set to the field edge position. Within the test data of 500 portal images, the success rate was over 90%. This means that in less then 10% of the displacement measurements, the result was either automatically or manually rejected by diagnostic measures or after visual validation, respectively.

The overall accuracy is difficult to estimate, since no ground truth is available for real portal images. Various external errors add to the parameter precision presented in table 1, in particular the projective nature of the imagery. However, visual validation and comparison with manual measurements indicate an accuracy of about 1 or 2 pixel for translation and below 1° for rotation measurements. In the following, two tests are presented in order to assess the reliability of the measurements.
Figure 15: Anatomy match example of an AP pelvis image. The search image is additionally rotated by 8° to test the iterative matching procedure (see the iteration series below). White polygons outline the template regions and the patches found, respectively. The black line represents the field edge of the reference image and is overlaid onto the search image to better visualize the patient displacement relative to the field edge.

Figure 16: Iteration series of the middle template in the above pelvis image. Notice that the contrast of these images is manually enhanced for printing purpose.
Sensitivity to template selection

In order to estimate the accuracy of the patient displacement measurements, the same images are matched using slightly varying set of templates. Figure 17 summarizes the results using 10 different sets of templates on a series of 6 images. The standard deviations of the x and y position are both below 0.3 pixel and for the orientation below 0.3°.

\[
\begin{align*}
\sigma_x & = 0.28 \text{ [pixel]} = 0.24 \text{ [mm]} \\
\sigma_y & = 0.22 \text{ [pixel]} = 0.19 \text{ [mm]} \\
\sigma_{\text{rot}} & = 0.292 \text{ [deg]}
\end{align*}
\]

Figure 17: The sensitivity to slight variation of the template regions is tested using 10 different template configurations. They are all overlaid onto the reference image in (a). The resulting standard deviations of the measurements are shown on the right (b) and represent an estimate of the overall accuracy.

Convergence radius

An important aspect in optimization schemes is the radius of convergence. The initial guess or starting point of the iterative optimization must be within this convergence area. The LSM algorithm starts to fail when the difference between the initial guess and the correct position amounts to more than half the template size. For a typical template configuration in a AP pelvis image, this corresponds to about 20 mm shift (22 pixel) or 10° rotation. However, these are at the same time reasonable upper limits for alignment errors in daily hospital routine.

4.4 Artificially generated data with known ground truth

In real datasets of portal images, the ground truth is always unknown. In order to test the algorithm on datasets with known ground truth, two artificial portal image series are generated by computing various megavoltage DRRs from the same CT volume.

In-plane translation and rotations

The following test series consisted of 35 simulated portal images with a maximum patient displacement of 20 mm in x and y direction and a maximum rotation of 10° (figure 18). An affine transformation model was used for the measurements. The results are summarized in table 3 and figure 20a. The standard deviations of the translation measurements are 0.25 pixel (0.23 mm) in x direction and 0.37 pixel (0.33 mm) in y direction. These systematic errors are caused by the unknown y position in the CT coordinate system of the template features (see figure 7 for an illustration of the CT coordinate system). Within the rotation measurement, these systematic error do not occur and the standard deviations are below 0.01°.

Including out-of-plane rotation

In order to test under more realistic conditions, a test series with small out-of-plane rotations was generated (figure 19). As with the in-plane dataset, we employed a planar affine transformation model. The
4.4 Artificially generated data with known ground truth

Figure 18: In-plane test series computed from CT volume. The CT is translated and rotated in the image plane yielding a test series of 24 translated and 10 rotated images.

Figure 19: Out-of-plane rotation test series computed from CT volume with totally 200 images rotated around the x and z axis, respectively.

<table>
<thead>
<tr>
<th></th>
<th>standard deviations</th>
<th>point error</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma_x$</td>
<td>$\sigma_y$</td>
<td>$\sigma_{rot}$</td>
</tr>
<tr>
<td>in-plane translation</td>
<td>0.23 mm</td>
<td>0.33 mm</td>
<td>0.19° 0.29 mm</td>
</tr>
<tr>
<td>in-plane rotation</td>
<td>0.05 mm</td>
<td>0.05 mm</td>
<td>0.01° 0.05 mm</td>
</tr>
<tr>
<td>out-of-plane rotation (2°)</td>
<td>1.4 mm</td>
<td>1.1 mm</td>
<td>0.28° 1.2 mm</td>
</tr>
<tr>
<td>out-of-plane rotation (5°)</td>
<td>3.3 mm</td>
<td>3.0 mm</td>
<td>0.59° 3.2 mm</td>
</tr>
</tbody>
</table>

Table 3: Standard deviations of the displacement measurements in the artificially generated test series.

Figure 20: Displacement measure errors in test series without (a) and including out-of-plane rotations of 2° (b). In (a) the error vectors are enlarged by a factor 10. The total point error amounts to 0.3 pixel (0.28 mm) without and 1.4 pixel (1.2 mm) with out-of-plane rotations.
area-based match still found the corresponding regions with a correlation well above 0.9. The systematic errors already encountered in the example above, which are an inherent problem of using projected images, are of course higher in this example. But the total point errors of 1.2 mm for 2° rotation and 3 mm for 5° still are promising results (see table 3 and figure 20b).

4.5 Multi-modal match with portal images

Matching portal images directly with a megavoltage DRR computed from the planning CT provides an accurate link between planning and treatment (cf. section 3.3 and figure 5a). Additional sources of positioning errors are avoided, thus it is a very interesting approach for increasing the accuracy of radiotherapy treatment.

We performed prototypical tests to check the feasibility of this multi-modal match using the LSM algorithm. Figure 21a shows a DRR and one set of templates including four validation lines, which are not used for matching. The DRR was computed from a CT volume with a voxel size of 2×2×3 mm³. The corresponding portal image series contained 22 images, two of which are depicted in 21b and c. All images were matched three times using different sets of templates. Of these 66 measurements, only one optimization failed, that is, did not find a minimum after 100 iterations. This was due to a rather large patient displacement (over 10 mm), for which one of the template sets was unsuitable. The remaining results were visually validated and all were accepted to be correct, which indicates a success rate of over 95%.

Figure 21: Multi-modal match between a DRR image computed from CT data (a) and a portal image series (b,c). The match is based on an affine transformation model. White polygons outline the template regions and patches respectively, black lines represent validation lines which are not used for matching.

5 Discussion and outlook

The LSM method with deformable templates is a versatile matching algorithm. Since it is an area-based method, the often unreliable step of feature extraction is circumvented. Especially in low-contrast imagery like megavoltage X-ray images (portal images), this is an important feature.

LSM has been successfully applied to matching megavoltage X-ray images of the same modality. The test data contained more than 500 portal images. Visual validation and comparison with manual measurements showed a success rate of over 90%, which is very promising compared to existing algorithms. The sensitivity to varying set of templates was examined and found to be below 0.3 pixel for translation and 0.3° for rotation. The estimated overall accuracy lies between 1 and 2 pixel for translation and below 1° for rotation measurements. As 2 pixel relate to less then 2 mm for a standard portal image, the achieved results lie within the required accuracy of 3 mm translation.

In two test series with generated data, the systematic error caused by estimating a 3D motion from projected 2D images was examined. This systematic error represents the main limitation of the overall accuracy for any 2D measuring method. The first test series contained only images with in-plane dis-
placements, leading to a total point error below 0.33 mm. When including minor out-of-plane rotations of 2°, the accuracy dropped to 1.2 mm, which still would be excellent for a clinical application.

The results presented in this paper also show the suitability for a multi-modal match between digitally reconstructed radiographs as reference images and the portal images acquired during radiotherapy. This is a highly interesting approach, since it provides a direct link between therapy planning and treatment. The LSM method proved robustness even when matching these slightly distorted patterns and succeeded in all but one of 66 measurements. This encouraging result will be validated more thoroughly in the near future.

The presented LSM algorithm is embedded in a graphical user interface and running on a Sun Workstation in the University Hospital of Zürich. Thus, it can be used by clinicians for medical studies and might be a step further in the direction of automatically controlling the patient position in radiotherapy. Applied in daily hospital routine, this should lead to an improved quality assurance in radiotherapy.

Future investigations will focus first on the problem of template selection as outlined in section 3.6 and second on the multi-template matching itself. The limitation of our current implementation is that only one global geometric transformation is estimated based on an ensemble of local template–patch relations. This approach relies on the assumption that the patient’s body undergoes a rigid in-plane motion and a minor out-of-plane rotation which can be compensated with one single affine transformation. Thus, perspective and higher order imaging distortions are neglected.

To account for these distortions, minor local transformations can be estimated for each template in addition to the global transformation. Another goal is to combine multiple 2D measurements from different directions to better estimate the 3D motion and to reduce the influence of the systematic error occurring in the 2D measurements.

Acknowledgments

We would like to thank Dr. U. R. Meier and Prof. Dr. U. M. Lütolf for providing the portal images and useful information concerning medical topics. This work was granted by the Swiss Cancer League and the Walter Honegger Foundation.

References


