Evaluation of optimization methods for intensity-based 2D-3D registration in x-ray guided interventions

I.M.J. van der Bom[a], S. Klein[b], M. Staring[c], R. Homan[d], L.W. Bartels[e], and J.P.W. Pluim[e]
[a] Departments of Radiology, University of Massachusetts Medical School, Worcester, MA, USA
[b] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[c] Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands
[d] Philips Healthcare, X-ray Predevelopment, Best, The Netherlands
[e] Image Sciences Institute, Departments of Radiology, University Medical Center Utrecht

ABSTRACT

The advantage of 2D-3D image registration methods versus direct image-to-patient registration, is that these methods generally do not require user interaction (such as manual annotations), additional machinery or additional acquisition of 3D data.
A variety of intensity-based similarity measures has been proposed and evaluated for different applications. These studies showed that the registration accuracy and capture range are influenced by the choice of similarity measure. However, the influence of the optimization method on intensity-based 2D-3D image registration has not been investigated. We have compared the registration performance of seven optimization methods in combination with three similarity measures: gradient difference, gradient correlation, and pattern intensity. Optimization methods included in this study were: regular step gradient descent, Nelder-Mead, Powell-Brent, Quasi-Newton, nonlinear conjugate gradient, simultaneous perturbation stochastic approximation, and evolution strategy. Registration experiments were performed on multiple patient data sets that were obtained during cerebral interventions. Various component combinations were evaluated on registration accuracy, capture range, and registration time. The results showed that for the same similarity measure, different registration accuracies and capture ranges were obtained when different optimization methods were used. For gradient difference, largest capture ranges were obtained with Powell-Brent and simultaneous perturbation stochastic approximation. Gradient correlation and pattern intensity had the largest capture ranges in combination with Powell-Brent, Nelder-Mead, nonlinear conjugate gradient, and Quasi-Newton. Average registration time, expressed in the number of DRRs required for convergence, was the lowest for Powell-Brent. Based on these results, we conclude that Powell-Brent is a reliable optimization method for intensity-based 2D-3D registration of x-ray images to CBCT, regardless of the similarity measure used.

Keywords: 2D-3D registration, x-ray guided procedures, optimization

1. INTRODUCTION

Intensity-based 2D-3D registration methods are by far the best known and most commonly used. These methods perform registration by seeking the digitally reconstructed radiograph (DRR) that corresponds to the 2D x-ray image. DRRs are generated from the 3D volume with different orientations and positions relative to the 2D image. The correspondence between x-ray image and DRR image is quantified using a similarity measure or cost function $\mathcal{C}$. Registration is obtained by optimization of the cost function. Ideally, the extremum that corresponds to successful registration is the global extremum of the cost function used and is enclosed by smooth surroundings. In reality however, cost functions may have multiple local extrema, which do not correspond to successful registrations. These local extrema may hamper the registration process since algorithms can converge towards erroneous results.
The most robust strategy to find the global optimum is to perform a full search. This is, however time consuming and computationally expensive especially when the cost function depends on multiple parameters.
Over the last decades, a variety of intensity-based similarity measures has been proposed and their performances

Further author information: Send correspondence to I.M.J. van der Bom (martijn.vanderbom@umassmed.edu)

in terms of accuracy and robustness for 2D-3D registration have been well evaluated in various comparison studies.[1,2] Penney *et al.*[3] have compared six similarity measures (cross correlation, entropy, mutual information, gradient correlation, pattern intensity, and gradient difference) on phantom data. Their conclusions were that entropy-type similarity measures are very sensitive to soft tissue deformations, and therefore worsen the registration accuracy. Pattern intensity, gradient difference, and gradient correlation showed the best performances. In a later study performed by Penney *et al.*[4] pattern intensity and gradient difference were evaluated on clinical data obtained during abdominal interventions. As in their previous study, both similarity measures tied in terms of accuracy and robustness.

Hipwell *et al.*[5] have compared the same six similarity measures as Penney *et al.* for the registration of (cerebral) x-ray digital subtraction angiograms (DSA) to magnetic resonance angiograms (MRA). Again, the best results in terms of accuracy and robustness were obtained with pattern intensity, gradient difference and gradient correlation.

Khamene *et al.* have evaluated eight similarity measures (mutual information, normalized correlation coefficient, correlation ratio, local normalized correlation, variance weighted correlation, gradient difference, gradient correlation, and pattern intensity) for patient positioning in radiation therapy.[6] Two portal images with 90° vergence were simultaneously registered to a CT volume. Results showed that the highest accuracy was obtained with local normalized correlation, gradient correlation, and pattern intensity, respectively. The accuracy of gradient difference was slightly lower than that of pattern intensity.

The success of registration is not only determined by the characteristics of the cost function, but also by the search strategy that is used. Since cost functions often show multiple local optima, optimization strategies should be able to ignore these, and converge towards the global optimum that corresponds to accurate registration.

For optimization of multidimensional functions, a diversity of search strategies has been developed and evaluated. Their competence with respect to rigid and non-rigid medical image registration has been explored by Maes *et al.* and Klein *et al.*, who evaluated several optimization methods for one similarity measure (mutual information) on accuracy, precision, and convergence.[7,8] Convergence rates and registration precision varied per search strategy. It is therefore crucial that an optimal combination of search strategy and similarity measure is used in order to obtain the best registration results.

In the comparison studies performed by Penney *et al.*, and Hipwell *et al.*,[4,5] a gradient ascent/descent strategy was used. Khamene *et al.* have tested three optimization routines: gradient ascent/descent, best neighbor search, and Powell-Brent, which in their experiments led to no significant difference.

To our knowledge, a study covering both the effects of search strategies and cost functions on the accuracy and robustness of 2D-3D image registration has not been reported.

In this research, we evaluate a variety of optimization routines in combination with three similarity measures for registration of 2D x-ray to 3D cone beam computed tomography (CBCT) in cerebral interventions. In previous studies, 2D-3D registration methods have often been evaluated by multi-view image data. In these studies, registrations were completed using two or more oblique x-ray images simultaneously, which improves registration accuracy. However, the availability of multiple x-ray images from different perspectives is not trivial during these procedures, especially when they are performed with a mono-plane c-arm system. We have therefore evaluated the various registration components using a single x-ray image. The performances were compared based on the evaluation method presented by Van de Kraats *et al.*[9] With this method, robustness and accuracy were determined by a high number of registration experiments using gold standard image data. In addition, total registration times of optimization methods were compared.

2. INTENSITY-BASED 2D-3D REGISTRATION

X-ray images acquired during an intervention and 3D volume data are spatially related by internal ($\boldsymbol{\nu}$) and external transform parameters ($\boldsymbol{\mu}$). The internal, or perspective, parameters consist of the source-to-image distance (SID), the pixel size (p_x, p_y), and the location of the 2D image relative to the 3D volume (c_x, c_y). The external transform parameters $\boldsymbol{\mu}$ describe the orientation and position of the 3D volume relative to the 2D image by three rotation (r_x, r_y, r_z) and three translation parameters (t_x, t_y, t_z) (see Figure 2.1). In 2D-3D image registration, we aim to find transform $\boldsymbol{T}$ with parameters $\boldsymbol{\mu}$ that defines the position and orientation of the 3D volume relative to the x-ray image, and assume the modality-dependent internal parameters to be known.

In order to quantify image correspondence during the registration process, DRRs are used.[10] DRRs are simulated

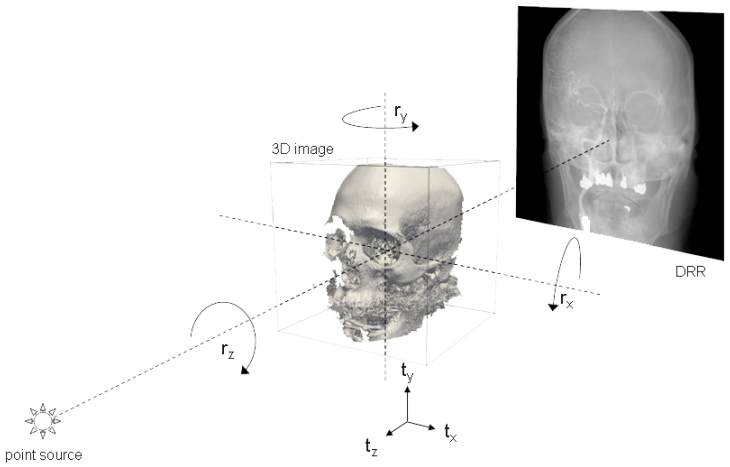


Figure 1: Geometric overview of 2D-3D registration. The external transform parameters $\boldsymbol{\mu}$ that describe the orientation (r_x, r_y, r_z) and position (t_x, t_y, t_z) of the 3D volume with respect to the 2D image are indicated in the figure.

projection images generated by casting rays through 3D image data. The intensities of the voxels that are crossed by a single ray are integrated, and mapped onto a 2D image in the location where the ray hits the 2D image.
If we consider the 2D image data as "fixed" and the 3D volume data as "moving" (since the latter is transformed to match the former), 2D-3D registration can be formulated as a minimization problem:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}(\boldsymbol{\mu}, I_F, I_M) \tag{1}$$

where the cost function $\mathcal{C}$ represents the negated similarity measure that is minimized. I_F and $I_M(\boldsymbol{\mu})$ are the fixed and moving images, respectively, $\boldsymbol{\mu}$ contains the external transform parameters that are applied to $I_M(\boldsymbol{\mu})$ before DRR, $I_{\mathrm{DRR}}(\boldsymbol{\mu})$, is generated, and $\hat{\boldsymbol{\mu}}$ represents the transform parameters that align the images. For clarity, henceforth the dependence of I_M and I_{DRR} on $\boldsymbol{\mu}$ are omitted in the equations.
To determine the relationship between a 2D x-ray and the 3D volume, the optimal set $\hat{\boldsymbol{\mu}}$ is generally obtained using an iterative strategy:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + a_k \boldsymbol{d}_k, \qquad k = 0, 1, 2, \ldots \tag{2}$$

Here, $\boldsymbol{d}_k$ represents the search direction in parameter space at iteration k and a_k is the gain factor which controls the step size during the iteration process. The value of the gain factor and the search direction are determined by the optimization method that is used.

3. SIMILARITY MEASURES

In this study, three similarity measures for 2D-3D registration that were successfully evaluated in previous studies were compared: gradient difference, gradient correlation, and pattern intensity.[3–6,11]

3.1 Gradient difference (GD)

With gradient difference, the similarity between an x-ray image and DRR image is calculated using the difference in intensity gradients. The intensity gradient images of x-ray and DRR images are generated using horizontal and vertical Sobel operators. Gradient difference GD is defined as:

$$GD = \sum_{i,j} \frac{\sigma_v}{\sigma_v + (D_v(i,j))^2} + \sum_{i,j} \frac{\sigma_h}{\sigma_h + (D_h(i,j))^2}, \tag{3}$$

with:

$$D_v(i,j) = \frac{\partial I_F(i,j)}{\partial i} - s_v \frac{\partial I_{\mathrm{DRR}}(i,j)}{\partial i}, \tag{4}$$

$$D_h(i,j) = \frac{\partial I_F(i,j)}{\partial j} - s_h \frac{\partial I_{\mathrm{DRR}}(i,j)}{\partial j}. \tag{5}$$

I_F and I_{DRR} represent the fixed and DRR image, respectively, with pixel indexing (i, j) and s_v and s_h denote scaling factors. σ_v and σ_h are the variances of the vertical, respectively, horizontal gradient images of the fixed image.

3.2 Gradient correlation (GC)

The term gradient correlation as used by Penney *et al.* refers to the normalized cross correlation between gradient images. Like with gradient difference, the gradient images are derived using Sobel operators. The resulting partial gradients are normalized by the mean gradient value. Subsequently, the cross correlations are calculated for the horizontal and vertical gradient images. Gradient correlation is defined as the sum of these two normalized cross correlation values. Let $\partial_i I = \frac{\partial I(i,j)}{\partial i}$ and $\partial_j I = \frac{\partial I(i,j)}{\partial j}$, then:

$$GC = \frac{\sum_{i,j}(\partial_i I_F - \overline{\partial_i I_F})(\partial_i I_{\mathrm{DRR}} - \overline{\partial_i I_{\mathrm{DRR}}})}{\sqrt{\sum_{i,j}(\partial_i I_F - \overline{\partial_i I_F})^2}\sqrt{\sum_{i,j}(\partial_i I_{\mathrm{DRR}} - \overline{\partial_i I_{\mathrm{DRR}}})^2}} + \frac{\sum_{i,j}(\partial_j I_F - \overline{\partial_j I_F})(\partial_j I_{\mathrm{DRR}} - \overline{\partial_j I_{\mathrm{DRR}}})}{\sqrt{\sum_{i,j}(\partial_j I_F - \overline{\partial_j I_F})^2}\sqrt{\sum_{i,j}(\partial_j I_{\mathrm{DRR}} - \overline{\partial_j I_{\mathrm{DRR}}})^2}}, \quad (6)$$

where $\overline{\partial_i I}$ represents the mean of $\frac{\partial I(i,j)}{\partial i}$. Using $\overline{I^i}_F = (\partial_i I_F - \overline{\partial_i I}_F)$ and $\overline{I^i}_{\mathrm{DRR}} = (\partial_i I_F - \overline{\partial_i I}_F)$, equation 6 has the form of the inner product between the normalized vectors:

$$GC = \sum_{i,j} \frac{\overline{\nabla I}_F \cdot \overline{\nabla I}_{\mathrm{DRR}}}{\| \overline{\nabla I}_F \| \| \overline{\nabla I}_{\mathrm{DRR}} \|}, \quad (7)$$

where ∇I represents the divergence of I and $\| \boldsymbol{a} \|$ is the norm of vector $\boldsymbol{a}$. Hence, registration is obtained by aligning the normalized gradient vectors.

3.3 Pattern intensity (PI)

Pattern intensity is a regional similarity measure that quantifies the amount of "structuredness" within a region with radius r of the difference image D between a fixed and a DRR image. The measure is weighted by σ, the standard deviation in I_F, which is related to the amount of noise in the image and should be defined by the user. Like with gradient difference, D is calculated using a scaling factor s. Pattern intensity PI is defined as:

$$PI = \sum_{i,j} \sum_{d^2 \leq r^2} \frac{\sigma^2}{\sigma^2 + (D(i,j) - D(v,w))^2} \quad (8)$$

with:

$$D(i,j) = I_F(i,j) - s \cdot I_{\mathrm{DRR}}(i,j) \quad (9)$$

$$d^2 = (i-v)^2 + (j-w)^2. \quad (10)$$

4. OPTIMIZATION METHODS

All three similarity measures described in the previous section were tested in combination with seven optimization methods: gradient descent, Nelder-Mead, Powell-Brent, quasi-Newton, nonlinear conjugated gradient, simultaneous perturbation stochastic approximation, and evolution strategy. In general, all optimization methods included in this study minimize equation $\mathcal{C}$ using the strategy in equation (2). The major difference between the methods is how parameters a_k and $\boldsymbol{d}_k$ are determined. Some of the optimization methods used in this study require the derivative $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}}$ to update the search direction $\boldsymbol{d}_k$. A finite difference approximation was used, because calculating the exact derivative of the cost function is a challenging task. In the following section, the various optimization methods are briefly discussed.

4.1 Regular step gradient descent (RSGD)

Gradient descent[12] takes downhill steps that are proportional to the local gradient of cost function $\mathcal{C}$:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \kappa a_k \frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k} \tag{11}$$

Various implementations of gradient descent have been reported that use different definitions of the gain factor. Here, we have chosen a variant where a_k is determined by the inner product of the derivatives at k and $k-1$ and is weighted by the relaxation factor κ (with $0 < \kappa < 1$).

4.2 Nelder-Mead (NM)

The Nelder-Mead method, also known as the downhill simplex method[13,14] minimizes an n-dimensional cost function by comparing the cost function values at $n+1$ vertices $(P_0, P_1, \ldots P_n)$ of a polytope, called a simplex. The simplex progresses towards a local minimum of $\mathcal{C}$ by replacing the vertex with the "worst" function value with a new point that can be obtained by reflection:

$$P^* = (1+\alpha)\bar{P} - \alpha P_h, \tag{12}$$

and subsequently by either expansion:

$$P^{**} = \gamma P^* + (1-\gamma)\bar{P}, \tag{13}$$

or contraction:

$$P^{**} = \beta P_h + (1-\beta)\bar{P}, \tag{14}$$

of the current vertex point. In these equations, P_h represents the vertex with the "worst" similarity value that is being replaced, $\bar{P}$ is the centroid of the simplex using all points except P_h. The positive constants, α, γ, and β are the reflection, expansion, and contraction coefficients, respectively, which can be considered as gain factors. Progression is limited to these three operations.

4.3 Powell-Brent (PB)

Powell's (conjugate gradient descent) method minimizes a cost function by performing line search optimization. Where most other methods progress a single step in direction $\boldsymbol{d}_k$ or perform a inexact line search each iteration, Powell-Brent performs an exact line search, i.e. exactly minimizes $\mathcal{C}$ along direction $\boldsymbol{d}_k$:

$$a_k = \arg\min_a \mathcal{C}(\boldsymbol{\mu}_k + a\boldsymbol{d}_k). \tag{15}$$

Each iteration, exact line minimization is commenced along $\boldsymbol{d}_k$, representing each of the transform parameters (conjugate) directions separately. Line minimization is achieved by bracketing the minimum, which is then found using Brent's method.[14,15]

4.4 Quasi-Newton (QN)

With Quasi-Newton methods,[12,14] the gain factor is weighted by L_k, an approximation to the inverse of the Hessian $[H(\boldsymbol{\mu}_k)]^{-1}$, evaluated at $\boldsymbol{\mu}_k$:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k L_k \frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k} \tag{16}$$

Various methods to update L_k have been proposed. In this study we have used limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS):[14,16]

$$L_{k+1} = (I - \frac{\boldsymbol{s}\boldsymbol{y}^T}{\boldsymbol{s}^T\boldsymbol{y}})L_k(I - \frac{\boldsymbol{y}\boldsymbol{s}^T}{\boldsymbol{s}^T\boldsymbol{y}}) + \frac{\boldsymbol{s}\boldsymbol{s}^T}{\boldsymbol{s}^T y} \tag{17}$$

where I represents the identity matrix, and:

$$\boldsymbol{s} = \boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k, \tag{18}$$

$$\boldsymbol{y} = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k, \tag{19}$$

where $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k}$ and $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_{k+1}}$ were replaced by $\boldsymbol{g}_k$ and $\boldsymbol{g}_{k+1}$, respectively. The limited memory BFGS is used in combination with the inexact line search routine proposed by Moré and Thuente,[17] which obtains a value for a_k by assuring that the "strong Wolfe conditions":

$$\mathcal{C}(\boldsymbol{\mu}_{k+1}) \leq \mathcal{C}(\boldsymbol{\mu}_k) + c_1 a_k (\boldsymbol{d}_k)^T \boldsymbol{g}_k \tag{20}$$

$$\left|(\boldsymbol{d}_k)^T \boldsymbol{g}_{k+1}\right| \leq c_2 \left|(\boldsymbol{d}_k)^T \boldsymbol{g}_k\right| \tag{21}$$

are satisfied. c_1 and c_2 are user-defined constants which must satisfy $0 < c_1 < c_2 < 1$. Initially, $a_k = 1$ is recommended[18] as long as the above conditions are met. If not, the gain factor is updated by the Moré-Thuente routine.

4.5 Nonlinear conjugate gradient (NCG)

For nonlinear conjugate gradient,[12,19–21] $\boldsymbol{d}_k$ in equation 2 is a combination of the local gradient $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k}$ and the search direction $\boldsymbol{d}_{k-1}$ of the previous iteration:

$$\boldsymbol{d}_k = -\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k} + \beta_k \boldsymbol{d}_{k-1} \tag{22}$$

where in our case $\beta_k = max(0, min(\beta_k^{DY}, \beta_k^{HS})$ is a combination of the Dai-Yuan and Hestenes-Stiefel expressions, respectively, as proposed in:[20]

$$\beta_k^{DY} = \frac{(\boldsymbol{g}_k)^T \boldsymbol{g}_k}{(\boldsymbol{d}_{k-1})^T (\boldsymbol{g}_k - \boldsymbol{g}_{k-1})} \tag{23}$$

$$\beta_k^{HS} = \frac{(\boldsymbol{g}_k)^T (\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{(\boldsymbol{d}_{k-1})^T (\boldsymbol{g}_k - \boldsymbol{g}_{k-1})} \tag{24}$$

where $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k}$ was replaced by $\boldsymbol{g}_k$. Similar to quasi-Newton, the gain factor is updated by the Moré-Thuente inexact line search routine.

4.6 Simultaneous perturbation stochastic approximation (SPSA)

Like gradient descent, simultaneous perturbation[22] progresses along the negative gradient of $\mathcal{C}$ as stated in equation 11. However, it uses approximations of the cost function ($\hat{\mathcal{C}}$) by applying a random perturbation, defined by vector $\boldsymbol{\Delta}_k$, of all parameters at once:

$$\hat{\mathcal{C}}_k^+ = \mathcal{C}(\boldsymbol{\mu}_k + c_k \boldsymbol{\Delta}_k) + \epsilon_k^+ \tag{25}$$

$$\hat{\mathcal{C}}_k^- = \mathcal{C}(\boldsymbol{\mu}_k - c_k \boldsymbol{\Delta}_k) + \epsilon_k^- . \tag{26}$$

Each iteration, the elements of $\boldsymbol{\Delta}_k$ are randomly assigned ± 1. The ϵ terms in equations 25 and 26 represent the approximations errors. The i-th element of the approximated derivative $\hat{\boldsymbol{g}}_k = \frac{\partial \hat{\mathcal{C}}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k}$ is determined with:

$$[\hat{\boldsymbol{g}}_k]_i = \frac{\hat{\mathcal{C}}_k^+ - \hat{\mathcal{C}}_k^-}{2 c_k [\boldsymbol{\Delta}_k]_i} \tag{27}$$

where $\boldsymbol{c}_k$ is a multiplication factor that defines the magnitude of the perturbation:

$$\mid \boldsymbol{c}_k \mid = \frac{c}{(k+1)^\gamma} \tag{28}$$

with γ a constant value. In this method, a_k is not a constant value as was the case with gradient descent, but rather a decaying function of user-defined variables a, A and α:

$$a_k = \frac{a}{(k+A)^\alpha} \tag{29}$$

4.7 Evolution strategy (ES)

As the name suggests, evolution strategies[23] are based on the principle of natural selection. A great variety of methods that use this principle have been proposed, of which covariant matrix adaptation (CMA) is considered the state-of-the-art implementation of this optimization technique.[24] The method comprises three evolutionary phases: offspring generation, selection, and recombination. During offspring generation, a set of λ trial search directions is obtained from a normal distribution $\mathcal{N}$

$$\boldsymbol{d}_k^\ell \sim \mathcal{N}(0, C_k), \qquad for\ \ell = 1, 2, \ldots, \lambda. \tag{30}$$

In this equation, λ defines the user-defined population size and C_k is the covariant matrix, which indulges search directions that were successful in previous iterations. For each trial search direction, the cost function is evaluated for:

$$\mathcal{C}(\mu_k + a_k \boldsymbol{d}_k^\ell). \tag{31}$$

During the selection phase, the $P \leq \lambda$ "best" trial directions with the lowest cost function values are selected. Subsequently, a weighted sum of the selected trial directions is calculated in the recombination phase:

$$\boldsymbol{d}_k = \sum_{p=1}^{P} w_p \boldsymbol{d}_k^{(p;\lambda)} \tag{32}$$

where $\boldsymbol{d}_k^{(p;\lambda)}$ and w_p represent the p^{th} selected trial direction and weight factor, respectively. Finally, the optimization is updated using equation 2.

5. EXPERIMENTS

To analyze the performance of the various registration components, 2D-3D registration experiments were performed on 3 flat-panel cone-beam computed tomography (CBCT) datasets and x-ray images obtained during cerebral interventions. Registrations were completed using a single x-ray image at a time. In the following sections, image data and evaluation methodology are described. Also, the parameter settings used during the registration experiments are briefly explained.

5.1 Image data

Three image datasets were used, which were acquired during cerebral interventions using a calibrated x-ray angiography c-arm system (Allura Xper FD20, Philips Healthcare, Best, The Netherlands) with an angiographic "head 3D" protocol. CBCT data (256^3 voxels with size 0.76^3 mm^3) were reconstructed from a set of 116 x-ray images (952^2 pixels with size 0.31×0.31 mm^2) obtained during a rotational motion of the x-ray source over approximately 200° around the patient. For each patient data set, an anterior-posterior (AP) and a lateral (LAT) x-ray image were selected from the x-ray datasets. Both images were registered to the corresponding CBCT data.

In order to minimize the influence of pixels outside the object on the registration, square regions of interest (ROIs) were defined in the x-ray images. All ROIs had a size of 180×180 pixels and were positioned in the image centers (see Figure 2).

5.2 Evaluation method

The results were evaluated by comparing the final registration errors and capture ranges. Registration errors were expressed in mean target registration error mTRE, defined as the mean distance between N predefined target points $\boldsymbol{p}_n$ in the 3D moving image, transformed with transformation $\boldsymbol{T}$, and with the gold standard transformation $\boldsymbol{T}_{\text{gold}}$:

$$\text{mTRE}(\boldsymbol{T}) = \frac{1}{N} \sum_{n=1}^{N} \left\| \boldsymbol{T}(\boldsymbol{p}_n) - \boldsymbol{T}_{\text{gold}}(\boldsymbol{p}_n) \right\|. \tag{33}$$

Gold standard transformation parameters between x-ray data and CBCT were available from calibration of the c-arm system. Target points $\boldsymbol{p}_n$ were located at the vertices of a cubic volume of interest (size $120 \times 120 \times 120$

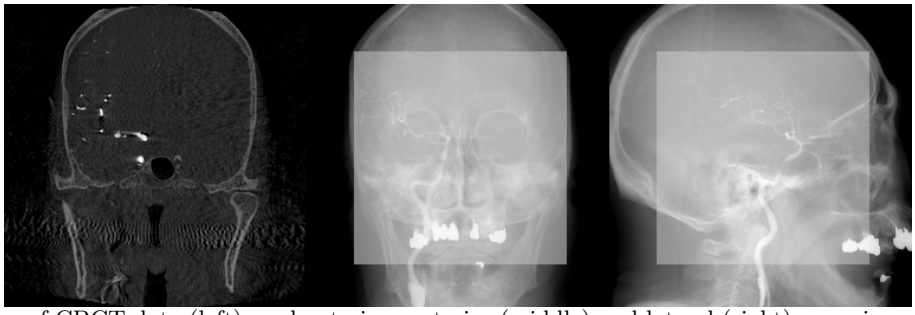


Figure 2: Examples of CBCT data (left), and anterior-posterior (middle) and lateral (right) x-ray images with ROIs that were used for the 2D-3D registration experiments.

mm^3) with its center aligned with the center of the 3D volume data. As a result, 1 mm mTRE corresponded to a rotation of 0.67° around one axis.

The capture range of a method is defined as the range of initial registration offsets for which the algorithm is capable of converging to correct alignment. In these experiments, both the capture range and the threshold for successful registrations are expressed in millimeter mTRE. We have used the same capture range criteria as used by Van de Kraats *et al.*,[9] where final mTRE ≥ 2 was considered misregistration and a total of 5% misregistrations within the capture range was allowed.

Registrations were performed using a single x-ray image and were initialized with randomly generated start transform parameters with initial mTRE varying from $0 - 10$ mm and with a frequency of 10 experiments per 1.0 mm mTRE, resulting in 100 experiments per x-ray image. Maximum allowed translational and rotational offsets were ±5.0 mm and ±5.0°, respectively.

In addition, for all optimizer-similarity measure combinations 10 registration experiments with start mTRE between 0.9-1.1 mm were performed on a single data set. Mean registration time and mean number of iterations were calculated for each combination. Since some methods require multiple cost function evaluations per iteration, the total number of DRRs required for registration were recorded.

5.3 Parameter settings

Parameter settings were based on recommendations given in the literature where available. Subsequently, settings were tested and optimized by trial-and-error experiments. For gradient difference, the scaling factors s_v and s_h in equations 4 and 5 were set to the ratios between the maximum vertical, respectively, horizontal gradient magnitude of the x-ray image and DRR. The noise constant in equation 8 was set to $\sigma = 100$, obtained by noise characterization within a homogenous region of interest in the x-ray data, r was set to 3 pixels and s was set to the ratio of the maximum intensity of the x-ray image and DRR.

To prevent a premature termination of the registration process by reaching the maximum number of allowed iterations, the maximum number of iterations for all optimization methods was set to 500 (exceptions given below). Value tolerance (tol) and minimum gradient magnitude (g_{min}) values define the minimum changes in similarity measure between two or more iterations and the minimum local gradient of the similarity measure, respectively, upon which convergence is declared. Optimization is continued until at least one of these values is reached. In order to obtain acceptable accuracy, we chose both values to be relatively small compared to the range of the similarity measures: $tol = 1.0 \times 10^{-6}$ and $g_{min} = 1.0 \times 10^{-9}$ for all methods. To assure that these parameters had similar influence on the similarity measures, gradient difference and pattern intensity values were scaled by a constant value, to bring their values in the same range with gradient correlation.

Regular step gradient descent specific parameters were: maximum step length $a_k^{max} = 0.1$, which defines a_k for $k = 0$, the minimum step length that is allowed before the registration is terminated $a_k^{min} = 0.001$, and the relaxation factor $\kappa = 0.5$. Experiments showed that larger relaxation factors led to insufficient decrease of the step size during convergence, which caused the optimization method to diverge from global minima.

Nelder-Mead requires a parameter that defines the initial size of the line segments of the simplex. This parameter was set to 2. The parameters $\alpha = 1$, $\beta = \frac{1}{2}$, and $\gamma = 2$ were defined as recommended in the literature.[13] Tuning experiments showed that Nelder-Mead generally required more than 500 iterations to convergence. Hence, the maximum number of iterations for this method was set to 700.

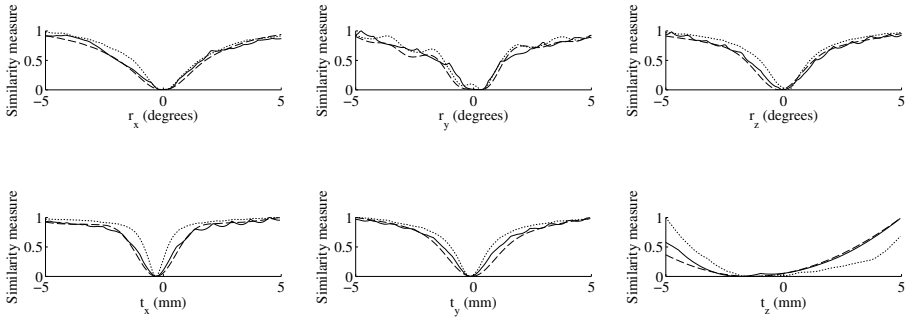


Figure 3: Gradient difference (solid line), gradient correlation (dashed line), and pattern intensity (dotted line) measures as a function of each of the six transform parameters, calculated using a single x-ray image.

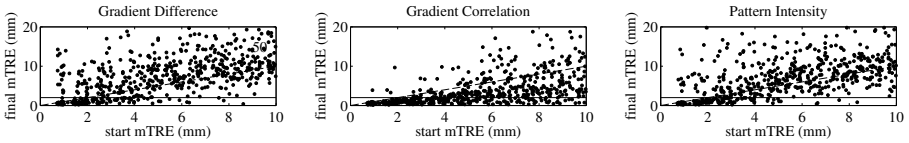
Besides *tol* and g_{min} only the minimum $a_k^{min} = 0.001$ and maximum step sizes $a_k^{max} = 2.0$ need to be defined for Powell-Brent. Since every iteration contains several line minimization iterations, typically the total number of iterations necessary for convergence is smaller than 20.
Parameters for Quasi-Newton optimization, $c_1 = 10^{-4}$, $c_2 = 0.9$, and $a_k = 1$ were all set as was recommended.[18] The limited memory implementation requires the user to define how many past iterations the algorithm uses to approximate the Hessian. This parameter was set to 5.
Because we chose the same inexact line search technique for nonlinear conjugate gradient as for Quasi-Newton, the exact same parameters were used for this optimization method.
Based on tuning experiments and recommendations, parameters for simultaneous perturbation stochastic approximation as defined in equations 28 and 29, were set to $c = 1$, $\gamma = 0.101$, $a = 25$, $A = 50.0$, and $\alpha = 0.602$. Experiments showed that this method requires a large number of iterations. Therefore, the maximum number of iterations was set to 1000.
Evolution strategy was used with $\lambda = 9$, $P = 4$, an initial step size of $a^0 = 1$, and a recommended definition for w_p.[24]
2D-3D registration software was developed in `elastix`.[25]

6. RESULTS

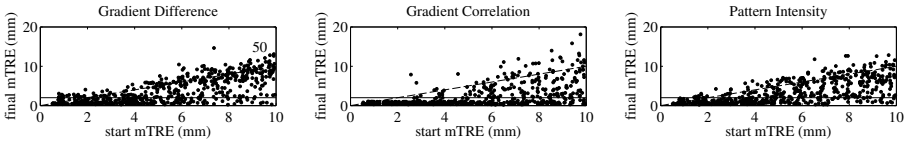
In Figure 3 gradient difference (solid line), gradient correlation (dashed line), and pattern intensity (dotted line) values are displayed as functions of the six transform parameters. The values were calculated using a single data set. For clarity, similarity values in Figure 3 were scaled between 0 - 1. All cost functions show a clear optimum in the proximity of the corresponding gold standard parameter value. Typically, translations along the axis of projection (t_z) result in minor changes in the cost function, which makes accurate registration in this direction difficult when only a single x-ray image is available, as can be seen in the graph.
The results of the 2D-3D registration experiments are summarized in Figures 4 and 5. Figure 4 shows the final mTRE (mTRE after registration) as a function of the start mTRE (initial random offset) for each optimization method (ordered by row) and each similarity measure (ordered by column). Each graph contains 3 (datasets) $\times$2 (x-ray images) $\times$100 (experiments per dataset) data points. The solid line indicates the threshold for success.

Table 1: Summary of the registration results using a single x-ray image. Mean registration errors ($\bar{\epsilon}$) for all successful registrations, and capture ranges (cr) are given in mm mTRE for all optimizer-similarity measure combinations.

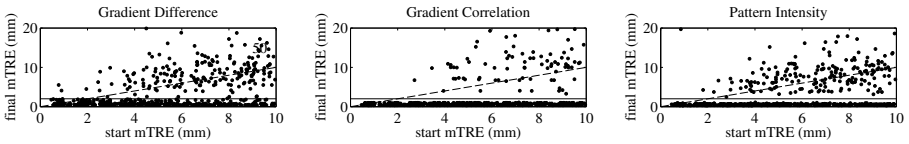
	GD		GC		PI	
	$\bar{\epsilon}(\sigma)$	$cr(\sigma)$	$\bar{\epsilon}(\sigma)$	$cr(\sigma)$	$\bar{\epsilon}(\sigma)$	$cr(\sigma)$
RSGD	0.90(0.08)	0.72(0.52)	0.93(0.01)	2.93(0.04)	0.87(0.07)	0.75(0.98)
NM	0.89(0.14)	0.77(1.87)	0.69(0.24)	5.05(0.78)	0.89(0.05)	1.70(0.91)
PB	0.79(0.41)	3.57(1.25)	0.61(0.26)	5.31(0.35)	0.54(0.10)	3.61(2.06)
QN	0.71(0.19)	1.37(1.45)	0.61(0.26)	3.81(1.22)	0.62(0.08)	1.93(0.64)
NCG	0.73(0.07)	2.05(0.42)	0.60(0.18)	4.56(2.27)	0.76(0.02)	1.89(0.55)
SPSA	0.91(0.08)	3.32(0.58)	0.95(0.00)	3.87(0.13)	0.97(0.04)	0.87(0.45)
ES	0.69(0.28)	0.49(0.44)	0.70(0.26)	4.59(3.11)	0.63(0.17)	0.55(3.56)



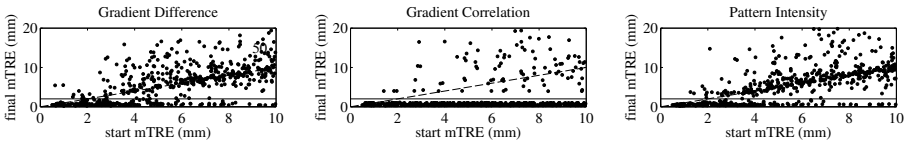
(a) Gradient descent ($o_{GD} = 18, o_{GC} = 3, o_{PI} = 15$)



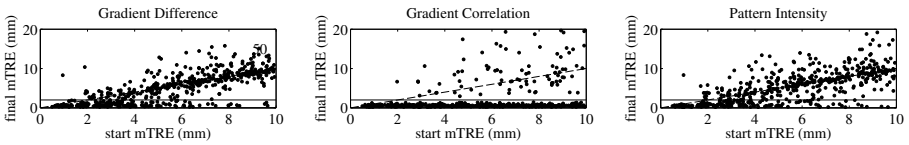
(b) Nelder-Mead ($o_{GD} = 0, o_{GC} = 0, o_{PI} = 0$)



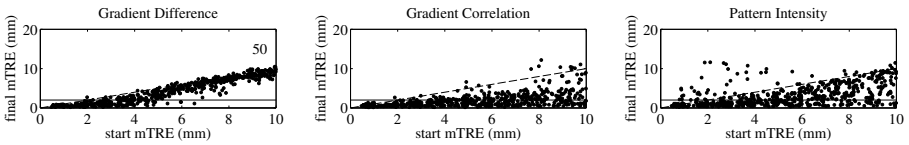
(c) Powell-Brent ($o_{GD} = 9, o_{GC} = 11, o_{PI} = 8$)



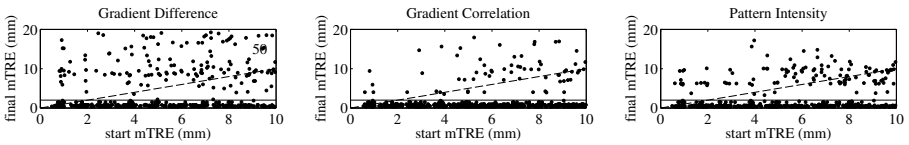
(d) Quasi-Newton ($o_{GD} = 9, o_{GC} = 18, o_{PI} = 4$)



(e) Nonlinear conjugate gradient ($o_{GD} = 11, o_{GC} = 31, o_{PI} = 9$)



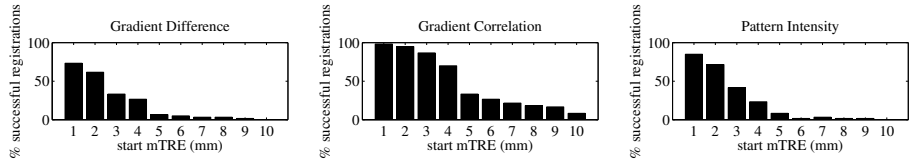
(f) Simultaneous perturbation stochastic approximation ($o_{GD} = 0, o_{GC} = 0, o_{PI} = 3$)



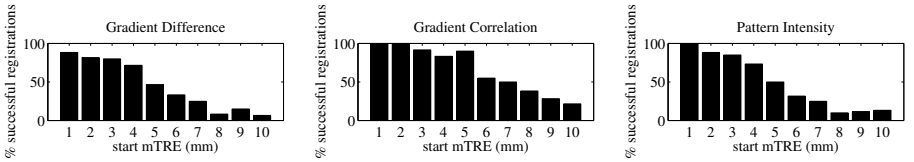
(g) Evolution strategy (outliers: $o_{GD} = 129, o_{GC} = 11, o_{PI} = 100$)

Figure 4: Results of the registration experiments. Final mTRE is given as a function of start mTRE. The threshold for success (= 2 mm mTRE) and the line of no improvement are indicated by the solid and dashed line, respectively. For each row the number of outliers is indicated by o_{GD}, o_{GC}, and o_{PI}.

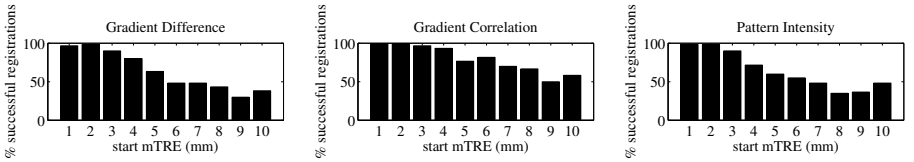
Final registration errors below this threshold are considered successful. The dashed line indicates where final mTRE is equal to start mTRE. For data points on or above this line registration did not lead to improvement. Experiments with final mTRE > 20 mm were considered outliers and were left out for visibility. For each graph, the numbers of outliers for gradient difference, gradient correlation, and pattern intensity are indicated by o_{GD}, o_{GC}, and o_{PI}, respectively, and are given in the caption. The graphs show that different combinations of com-



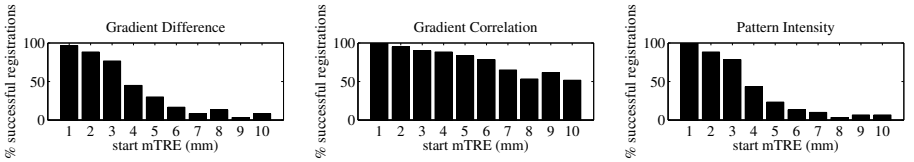
(a) Gradient descent



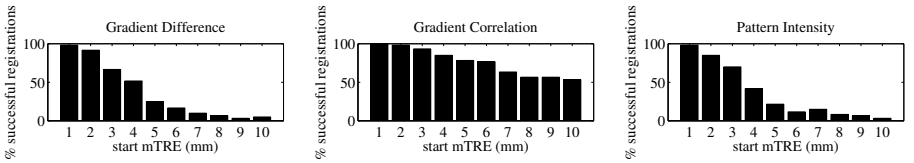
(b) Nelder-Mead



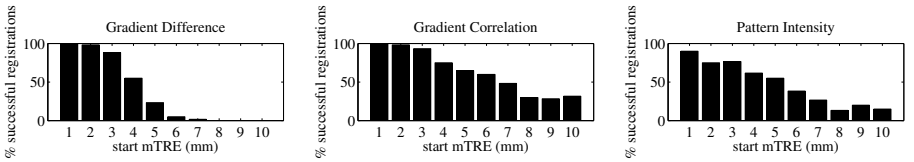
(c) Powell-Brent



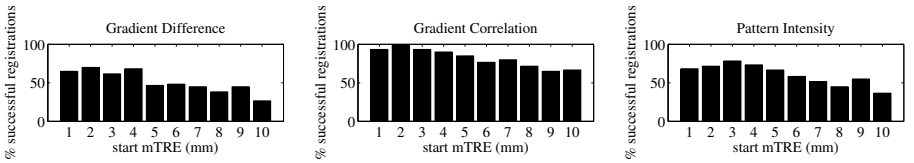
(d) Quasi-Newton



(e) Nonlinear conjugate gradient



(f) Simultaneous perturbation stochastic approximation



(g) Evolution strategy

Figure 5: Results of the registration experiments. Percentage of successful registrations (final mTRE < 2 mm) is given as a function of start mTRE.

ponents lead to different registration performances. The registration results for all component combinations are summarized in Table 1. Mean registration error ($\bar{\epsilon}$) in mm, calculated for all successful registrations, and capture

ranges (cr) are given. Standard deviations (σ) of registration errors and capture ranges obtained with different data sets were calculated. Large values of σ indicates large variance in performance. The capture range of a similarity measure, strongly depends on the search strategy that is used. Gradient difference in combination with gradient descent, Nelder-Mead, and evolution strategy resulted in capture ranges of 0.72 mm, 0.77 mm, and 0.49 mm, respectively, whereas Powell-Brent and simultaneous perturbation stochastic approximation gave capture ranges of 3.57 mm and 3.32 mm, respectively, using the same similarity measure. Similarly, gradient correlation and pattern intensity in combination with gradient descent resulted in relatively small capture ranges (2.93 mm and 0.75 mm, respectively), whereas Powell-Brent led to better results (5.31 mm and 3.61 mm, respectively). The largest capture range (5.31 mm) was obtained using gradient correlation and Powell-Brent. The differences in accuracies were relatively small. Mean registration errors varied from 0.54 mm to 0.97 mm. For all similarity measures, simultaneous perturbation stochastic approximation resulted in the poorest accuracies. Overall best accuracy was obtained with pattern intensity in combination with Powell-Brent.
The percentages of successful registration as a function of start mTRE are given Figures 5. Results were divided into bins (size 1 mm) based on their start mTRE. From these graphs it can be derived that for some methods, reasonable performances are obtained outside their capture range. A good example is evolution strategy in combination with gradient correlation; this combination led to high success percentages for start mTRE up to 10 mm. The same search strategy, however, performed worse in combination with gradient difference. Overall, the best performance in terms of percentage successful registrations was obtained with evolution strategy with gradient correlation. This combination led to 87.2% successful registrations for all experiments.
In Table 2 the mean number of iterations (N_{it}), mean total registration time (t) in seconds, and mean number of DRRs required for registration (N_{DRRs}) are given for all component combinations. The mean time required to calculate the similarity between x-ray and DRR image was 506, 547, and 314 milliseconds for gradient difference, gradient correlation, and pattern intensity, respectively. For all similarity measures, Powell-Brent required the least time and DRRs for registration. Generally, simultaneous perturbation stochastic approximation, evolution strategy and nonlinear conjugate gradient required the most time and DRRs. It should be noted that for simultaneous perturbation stochastic approximation no tolerance values are defined and therefore each registration requires the maximum number of iterations that is given. A notable result is the number of DRRs required for regular step gradient descent in combination with gradient difference, which was much larger than those for the same optimization method combined with the other two similarity measures. Experiments measuring registration times were performed on an Intel Xeon E5420 Dual-Core, 2.5 GHz.

7. DISCUSSION AND CONCLUSIONS

Although the quality of similarity measures for intensity-based 2D-3D registration methods has been well evaluated and compared, their performance in combination with various search strategies had not. Klein *et al.* have shown, that for the same similarity measure, different optimization methods can lead to differences in precision.[8] We have evaluated the effect of three well evaluated similarity measures in combination with seven optimization methods on registration accuracy and capture range. Experiments were performed using CBCT and x-ray data, acquired during cerebral interventions. Registration software used for this study was developed in `elastix`, a medical image registration software package. 2D-3D registration components will be available in release 4.5 and

Table 2: Results of the experiments performed for measurements of relative registration times. Mean number of total iterations (N_{it}), mean registration time (t), and mean number of DRRs used for registration (N_{DRRs}) were calculated for 10 registration experiments with start mTRE of approximately 1 mm.

	GD			GC			PI		
	N_{it}	t	N_{DRRs}	N_{it}	t	N_{DRRs}	N_{it}	t	N_{DRRs}
RSGD	108.7	832.0	2363.1	75.2	358.2	990.6	56.9	191.7	753.7
NM	541.6	191.66	550.7	654.2	241.4	661.7	659.4	171.0	668.8
PB	4.1	118.3	342.7	4.5	122.6	336.6	6.3	132.6	521.5
QN	38.6	257.0	745.9	40.4	269.9	741.0	40.0	214.3	846.0
NCG	72.8	440.5	1283.8	104.0	651.3	1800.5	36.6	209.5	818.7
SPSA	1000	1030.9	3000	1000	1087.1	3000	1000	762.3	3000
ES	211.0	717.6	2112.0	206.8	752.5	2069.0	248.2	632.0	2484.0

higher at http://elastix.isi.uu.nl.
For gradient difference, Powell-Brent and simultaneous perturbation stochastic approximation led to the largest capture ranges (3.57 and 3.32 mm, respectively), of which the first gave the best accuracy. For gradient correlation, largest capture ranges were obtained in combination with Powell-Brent and Nelder-Mead. The largest capture range for pattern intensity was obtained with Powell-Brent. For both gradient correlation and pattern intensity, Powell-Brent optimization resulted in the highest registration accuracies. Overall component combinations, gradient correlation in combination with Powell-Brent gave the largest capture range.
An overall well-performing optimization method was Powell-Brent. Regardless of the similarity measures used its performance in terms of capture range was high. Another successful optimization method was simultaneous perturbation stochastic approximation, which had capture ranges of 3.32 and 3.87 mm in combination with gradient difference and gradient correlation, respectively. Quasi-Newton and nonlinear conjugate gradient led to moderate capture ranges in combination with all three similarity measures.
In our experiments, the best similarity measure was gradient correlation. It showed the largest capture ranges in most experiments. Both gradient difference and pattern intensity in combination with gradient descent resulted in capture ranges in the order of 1 mm mTRE, which corresponds to the results reported by Van de Kraats *et al.*[9] Our experiments have shown, however, that the performances can be improved by using a different optimization method.
Registration accuracies varied from 0.54 mm to 0.97 mm, and were influenced by both similarity measure and optimization method. Relatively poor accuracies are achieved with simultaneous perturbation stochastic approximation. Because this method does not test whether it has converged, the accuracy of this method should be improved by increasing the number of iterations. The best accuracy was achieved with pattern intensity combined with Powell-Brent.
Calculation of gradient correlation requires no additional parameters, whereas in both gradient difference and pattern intensity scaling factors need to be defined. Weese *et al.*[11] proposed to optimize this scaling factor for pattern intensity during each iteration. However, in research reported by Penney *et al.*[3] a suitable scaling factor was used. In a later study by Penney *et al.*,[4] the scaling factor for both gradient difference and pattern intensity was optimized during each iteration. During exploratory experiments we have compared the registration performance of pattern intensity with a fixed scaling factor and a scaling factor that was optimized by an exhaustive search strategy. Registrations were performed on a single data set as described in section 5, using a Powell-Brent search strategy. The accuracy and capture range of pattern intensity improved with 0.11 mm and 0.78 mm, respectively, when the scaling factor optimization during each iteration. However, the total number of successful registrations (69%) did not improve. Because the improvement in registration performance was small compared to the additional required computational load a suitable constant scaling factor was used for pattern intensity and gradient difference, as was initially proposed by Penney *et al.* The σ value for pattern intensity is based on the noise level in the fixed image. The smoothness of the cost function is very sensitive to this parameter and it should therefore be determined with care. We have noticed that pattern intensity may not converge when the fixed image contains a lot of background pixels. Large regions of background pixels may dominate the similarity measure, which could cause the registration method to diverge. This was prevented by using a fixed image mask.
Since the number of computations that is performed within one iteration varies per optimization method, direct comparison of the total number of iterations is not possible. Comparison based on total registration time is not only machine dependent, but is also influenced by software implementation. To be able to compare the registration time of the optimization methods used in the study, total number of DRRs required for registrations were recorded, since generation of DRRs is, computationally speaking, the most expensive task in intensity-based 2D-3D registration. Registration using Powell-Brent required the least DRRs and time, irrespective of the similarity measure used.
Conventionally, gradient descent, Quasi-Newton, and nonlinear conjugate gradient use an exact calculation of the local gradient of the cost function $\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}} \mid_{\boldsymbol{\mu}_k}$ in order to determine the direction $\boldsymbol{d}_k$ of the next optimization step. Because of the complexity of this derivation, we have used a finite difference approach to calculate the derivative of the cost function. We expect that using a finite difference approach instead of a direct derivation has a significant influence on efficiency. Direct determination of the derivative is, however, outside the scope of this research.
From the results we can conclude that for the image data used in this study, the best 2D-3D registration re-

sults are obtained by a Powell-Brent search strategy. Because of its good performance with all three similarity measures, we recommend Powell-Brent as a reliable optimization method for intensity-based 2D-3D registration in cerebral interventions. It required the least registration time for convergence and only a few user-specified parameters need to be determined, which makes this method relatively easy to tune to a specific application.

REFERENCES

[1] Birkfellner, W., Wirth, J., Burgstaller, W., Baumann, B., Staedele, H., Hammer, B., Gellrich, N. C., Jacob, A. L., Regazzoni, P., and Messmer, P., "A faster method for 3d/2d medical image registration - a simulation study," *Physics in medicine and biology* **48**(16), 2665–2679 (2003).

[2] Sarrut, D. and Clippe, S., "Geometrical transformation approximation for 2D/3D intensity-based registration of portal images and CT scan," 532–540 (2001). Medical Image Computing and Computer-Assisted Intervention - MICCAI'01.

[3] Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., and Hawkes, D., "A comparison of similarity measures for use in 2D-3D medical image registration," *Medical Image Computing and Computer-Assisted Interventation - MICCAI'98* , 1153–1161 (1998).

[4] Penney, G. P., Batchelor, P. G., Hill, D. L. G., Hawkes, D. J., and Weese, J., "Validation of a two-to three-dimensional registration algorithm for aligning preoperative ct images and intraoperative fluoroscopy images," *Medical physics* **28**, 1024 (2001).

[5] Hipwell, J. H., Penney, G. P., McLaughlin, R. A., Rhode, K., Summers, P., Cox, T. C., Byrne, J. V., Noble, J. A., and Hawkes, D. J., "Intensity-based 2-d-3-d registration of cerebral angiograms," *IEEE Transactions on Medical Imaging* **22**(11), 1417–1426 (2003).

[6] Khamene, A., Bloch, P., Wein, W., Svatos, M., and Sauer, F., "Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy," *Medical Image Analysis* **10**(1), 96–112 (2006).

[7] Maes, F., Vandermeulen, D., and Suetens, P., "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Medical Image Analysis* **3**(4), 373–386 (1999).

[8] Klein, S., Staring, M., and Pluim, J. P. W., "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Transactions on Image Processing* **16**(12), 2879–2890 (2007).

[9] van de Kraats, E. B., Penney, G. P., Tomaževič, D., van Walsum, T., and Niessen, W. J., "Standardized evaluation methodology for 2-D-3-D registration," *IEEE Transactions on Medical Imaging* **24**(9), 1177–1189 (2005).

[10] Lemieux, L., Jagoe, R., Fish, D. R., Kitchen, N. D., and Thomas, D. G. T., "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs," *Medical physics* **21**(11).

[11] Weese, J., Penney, G. P., Desmedt, P., Buzug, T. M., Hill, D. L. G., and Hawkes, D. J., "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE transactions on information technology in biomedicine* **1**(4), 284–293 (1997).

[12] Nocedal, J. and Wright, S. J., [*Numerical optimization*], Springer verlag (1999).

[13] Nelder, J. A. and Mead, R., "A simplex method for function minimization," *The computer journal* **7**(4), 308–313 (1965).

[14] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., [*Numerical recipes*], Cambridge university press, Cambridge (2007).

[15] Powell, M. J. D., "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The computer journal* **7**(2), 155–162 (1964).

[16] Nocedal, J., "Updating quasi-newton matrices with limited storage," *Mathematics of computation* **35**(151), 773–782 (1980).

[17] Moré, J. J. and Thuente, D. J., "Line search algorithms with guaranteed sufficient decrease," *ACM Transactions on Mathematical Software (TOMS)* **20**(3), 286–307 (1994).

[18] Potra, F. A. and Shi, Y., "Efficient line search algorithm for unconstrained optimization," *Journal of Optimization Theory and Applications* **85**(3), 677–704 (1995).

[19] Hestenes, M. R. and Stiefel, E., "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards* **49**(6), 409–436 (1952).

[20] Dai, Y. H. and Yuan, Y., "An efficient hybrid conjugate gradient method for unconstrained optimization," *Annals of Operations Research* **103**(1), 33–47 (2001).

[21] Dai, Y. H., "A family of hybrid conjugate gradient methods for unconstrained optimization," *Mathematics of computation* **72**(243), 1317–1328 (2003).

[22] Spall, J. C., "Multivariate stochastic approximation using a simultaneousperturbation gradient approximation," *IEEE Transactions on Automatic Control* **37**(3), 332–341 (1992).

[23] Beyer, H. G. and Schwefel, H. P., "Evolution strategies. A comprehensive introduction," *Natural computing* **1**(1), 3–52 (2002).

[24] Hansen, N. and Ostermeier, A., "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation* **9**(2), 159–195 (2001).

[25] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. W., "`elastix`: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging* **29**(1), 196–205 (2010).