Towards Real-Time Multi-Modality 3-D Medical Image Registration

Thomas Netsch¹, Peter Rösch¹, Arianne van Muiswinkel², Jürgen Weese¹

¹Philips Research Laboratories, Hamburg, Germany
²Philips Medical Systems, MR–CPP, Best, The Netherlands
thomas.netsch@philips.com

Abstract

Intensity value-based registration is a widely used technique for the spatial alignment of medical images. Generally, the registration transformation is determined by iteratively optimizing a similarity measure calculated from the grey values of both images. However, such algorithms may have high computational costs, especially in the case of multi-modality registration, which makes their integration into systems difficult. At present, registration based on mutual information (MI) still requires computation times of the order of several minutes. In this contribution we focus on a new similarity measure based on local correlation (LC) which is well-suited for numerical optimization. We show that LC can be formulated as a least-squares criterion which allows the use of dedicated methods. Thus, it is possible to register MR neuro perfusion time-series (128² × 30 voxel, 40 images) on a moderate workstation in real-time: the registration of an image takes about 500 ms and is therefore several times faster than image acquisition time. For the registration of CT–MR images (51² × 87 CT, 256² × 128 MR) a multiresolution framework is used. On top of the decomposition, which requires 47 s of computation time, the optimization with an algorithm based on MI previously described in the literature takes 97 s. In contrast, the proposed approach only takes 13 s, corresponding to a speed-up about a factor of 7. Furthermore, we demonstrate that the superior computational performance of LC is not gained at the expense of accuracy. In particular, experiments with dual contrast MR images providing ground truth for the registration show a comparable sub-voxel accuracy of LC and MI similarity.

1. Introduction

The spatial alignment of images based on their intensity values has recently become the most widely used type of registration for several important medical applications [7]. Grey value-based methods offer many advantages when compared to point-based and surface-based methods: they do not require user-interaction and therefore offer the possibility to operate automatically and to align images with sub-voxel accuracy. Generally, grey value-based methods determine the registration transformation by iteratively optimizing a similarity measure which is calculated from the grey values of both images. However, the evaluation of the similarity measure during optimization involves many image reformatting steps, which results in high a computational cost.

Nevertheless, in single-modality registration, where two images show identical or similar contrast variation, real-time performance is feasible by using dedicated optimization schemes. For example, Cox et al. [1] show that in functional MRI, 3-D time-series of 80 images of the brain (128³ × 30 matrix, 37.5 Mvoxel) can be spatially aligned by rigid transformation within about one minute. Thus, the registration of an individual image is 2–3 times faster than its acquisition which justifies the term real-time. As an application, registration can be used, not only for motion compensation after image acquisition, but also to detect large head movements, which limit the post-processing of the functional data, while the experiment is still in progress. The reason for fast single-modality algorithms is that similarity measures based on image subtraction are well-understood in terms of numerical optimization. They can be very efficiently optimized by least-squares (Gauss-Newton) methods.

In the related application of MR neuro perfusion, uptake times of an administered contrast agent are determined to characterize the vitality of brain tissue. Due to the contrast agent the similarity measure must also be robust against the dissimilar tissue-to-grey-value transfer during bolus passage. So far, only mutual information (MI), an information-theoretic measure, demonstrates robust and accurate performance for a large class of multi-modality images in practice [5, 15]. However, it is very difficult to find suitable efficient optimization strategies for MI, as recently demon-
strated [6, 13]. At present, CT–MR registrations (512^2 × 48 CT, 256^2 × 128 MR, 20 Mvoxel) with sub-voxel accuracy still require of the order of several minutes CPU time on current workstations.

To further improve computational performance for multi-modality registration we focus on a new, recently proposed similarity measure based on local correlation (LC) [10, 14]. Similar to measures used in single-modality registration, we show that LC can be formulated as a least-squares criterion, and therefore allows the application of efficient optimizers. Furthermore, only a fraction of 5–20% of image voxels actually contribute to the calculation of similarity which makes the method computationally very attractive especially for volumes with large image matrix. For example, image resampling and multiresolution representations only need to be computed locally for relevant voxels and not for the entire images as in the case of MI.

The algorithm presented in this contribution achieves real-time performance for perfusion time-series and a significant speed-up for multi-modality matching such as CT–MR images, while the accuracy of the registration is comparable to MI. In the next sections, LC and its optimization is described in detail. Then, we summarize the performance of the method in the case of 3-D time-series and pairs of large image volumes. Particular attention is paid to a comparison of the registration accuracy yielded by LC and MI.

2. Similarity based on correlation

In image registration a target image t_0 has to be spatially aligned to a reference or base image b according to a chosen class of transformations T(p) with parameters p = (p_1, . . . , p_m). In the case of rigid transformations and 3-D images, m = 6, and the first three parameters p_1, p_2 and p_3 define image translation whereas the remaining three parameters p_4, p_5 and p_6 determine rotation around the coordinate axes. During the registration process, t_0 is reformatted with respect to the actual parameters p resulting in an intermediate image

\[ t = T(p, t_0). \]

After reformatting, the images b and t are compared to each other by using a suitable similarity measure M(b, t).

2.1. Local correlation

Apart from the difference image, maximum squared cross correlation is frequently used for single-modality registration, where a linear relationship between the grey values of the images can be assumed [7]. The squared cross correlation coefficient is defined as

\[ CC(b, t) = \frac{\sum_i (b_i - \bar{b})(t_i - \bar{t})^2}{\sum_i (b_i - \bar{b})^2 \sum_i (t_i - \bar{t})^2} \]

where \( b_i \) and \( t_i \) are the grey values of voxel i of b and t and \( \bar{b} \) and \( \bar{t} \) are the mean grey value of the images, respectively. For multi-modality images no global linear relationship can be assumed. However, a predominately linear transfer of image intensities can still be presumed in small neighborhoods of the images. To compute the similarity value for a number N of voxels, a neighborhood \( n(j) \) around voxel \( j \) in the base b is defined and the cross correlation coefficient of grey values \( b_i \) with \( i \in n(j) \) and the corresponding grey values \( t_i \) of the target is calculated and accumulated [10, 14]. Hence, local correlation is defined by

\[
LC(b, t) = \frac{1}{N} \sum_j \frac{\sum_{i \in n(j)} (b_i - \bar{b}_j)(t_i - \bar{t}_j)^2}{\sum_{i \in n(j)} (b_i - \bar{b}_j)^2 \sum_{i \in n(j)} (t_i - \bar{t}_j)^2}.
\]

Note, \( \bar{b}_j \) and \( \bar{t}_j \) are the average grey values of \( b \) and \( t \) in the neighborhood \( n(j) \). If one of the images has a constant grey value within \( n(j) \) the contribution to LC at \( j \) is defined to be zero. If both images show only two dominant grey values within \( n(j) \), the grey values can be transferred into each other by a linear transformation. Most contributions of LC therefore originate from edges in the images. The better the structures in the neighborhood match the larger their contribution, in particular if images are registered optimally. The definition of LC ensures

\[ 0 \leq LC(b, t) \leq 1. \]

Well-aligned images have values close to one while values are small in the case of misalignment.

Before we continue, a concise notation for the sums and averages is introduced. With

\[ [b_i]_j \overset{\text{def}}{=} \sum_{i \in n(j)} (b_i - \bar{b}_j), \]

LC defined by Eq. (1) can be compactly written as

\[
LC(b, t) = \frac{1}{N} \sum_j \frac{[b_i]_j [t_i]_j^2}{[b_i]_j^2 [t_i]_j^2}.
\]

More complicated expressions are abbreviated in the same way, for example

\[ [b_i^2 - t_i]_j \overset{\text{def}}{=} \sum_{i \in n(j)} ((b_i - \bar{b}_j)^2 - (t_i - \bar{t}_j)). \]

2.2. Voxel selection

Although local correlation is a spatial filtering technique, it is computationally attractive since only small neighborhoods of 6 adjacent voxels, and a small fraction of image voxels, are sufficient for correct alignment [8]. These
 voxels are chosen among those which have the largest variance of image grey values in the corresponding 6-neighborhoods. Therefore, mostly—with respect to the registration—anatomical relevant voxels, are considered for spatial correspondence. For example, voxels carrying contrast agent in perfusion time-series or voxels in the neighborhood of bone/soft tissue boundaries in CT-MR images.

3. Gauss-Newton optimization

Optimization iteratively determines the parameters \( \mathbf{p}^* \) minimizing (or maximizing) a given similarity measure \( M(b, t) \),

\[
\mathbf{p}^* = \arg \min_{\mathbf{p}} M(b, \mathbf{T}(\mathbf{p}, t_0)).
\]

At iteration \( k \) with actual parameters \( \mathbf{p}_k \), the result of the optimization algorithm is an update \( \mathbf{s}_k \)—a so-called Newton step—which determines the direction towards \( \mathbf{p}^* \) and width of the update. Hence, in the next iteration \( k + 1 \) the parameters \( \mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{s}_k \) are used. The iteration terminates when the similarity values no longer change significantly or when the update step \( \mathbf{s}_k \) has reached some small predefined threshold. To improve readability we omit the index \( k \) of \( \mathbf{p} \) and \( \mathbf{s} \) in the following.

The application of Gauss-Newton optimization requires a least-squares form of the similarity measure

\[
M(\mathbf{p}) = \sum_i \omega_i^2 \mathbf{m}_i^2(\mathbf{p})
\]

where \( \omega_i^2 \) are weight factors. Furthermore, the similarity \( M(\mathbf{p}) \) must be close to zero in the vicinity of the optimum corresponding to the parameters \( \mathbf{p}^* \). With these conditions the Newton step \( \mathbf{s} \) is approximated as follows. Given \( \mathbf{p} \) of the actual iteration each contribution of the measure \( M \) is expanded by its Taylor series up to first order around \( \mathbf{p} \).

\[
M(\mathbf{p} + \mathbf{s}) = \sum_i \omega_i^2 \mathbf{m}_i^2(\mathbf{p} + \mathbf{s})
\approx \sum_i \omega_i^2 \left( \mathbf{m}_i(\mathbf{p}) + \nabla^\top \mathbf{m}_i(\mathbf{p}) \cdot \mathbf{s} \right)^2.
\]

Necessary conditions for optimality are \( \frac{\partial}{\partial \mathbf{s}} M((\mathbf{p} + \mathbf{s}) = 0 \)

for \( l = 1, \ldots, m \) which results in the \( m \) equations

\[
0 = \sum_i \omega_i^2 \left( \frac{\partial}{\partial \mathbf{p}_l} \mathbf{m}_i(\mathbf{p}) \right) (\mathbf{m}_i(\mathbf{p}) + \nabla^\top \mathbf{m}_i(\mathbf{p}) \cdot \mathbf{s}).
\]

Writing the gradient as sum over \( h = 1, \ldots, m \) and interchanging the order of summation with respect to \( i \) and \( h \), the equations build a \( m \times m \) system of linear equations

\[
\sum_h \sum_i \omega_i^2 \left( \frac{\partial}{\partial \mathbf{p}_l} \mathbf{m}_i(\mathbf{p}) \right) \left( \frac{\partial}{\partial \mathbf{p}_h} \mathbf{m}_i(\mathbf{p}) \right) s_h \\
= - \sum_i \omega_i^2 \left( \frac{\partial}{\partial \mathbf{p}_l} \mathbf{m}_i(\mathbf{p}) \right) \mathbf{m}_i(\mathbf{p}) \quad (3)
\]

with unknowns \( s_h \). The matrix of the linear system is symmetric and positive definite [2] and therefore has a unique solution which can be robustly determined, for example, by a Cholesky decomposition of the corresponding matrix [9]. In the next section, LC is formulated in the least-squares form of Eq. (2) and in Sec. 3.2 the partial derivatives \( \frac{\partial}{\partial \mathbf{p}_l} \mathbf{m}_i(\mathbf{p}) \) are calculated.

3.1. Least-squares form of LC

Apart from the least-squares form of the objective function, Gauss-Newton optimization requires small similarity values in the neighborhood of the desired solution \( \mathbf{p}^* \). When the images are perfectly aligned, LC is not close to zero. However, the local correlation coefficients all tend to one. Therefore, we do not consider LC itself but \( 1 - LC \), thus

\[
1 - \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} = \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} - 2 \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} + \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j}
\]

The expression within the sum is further transformed

\[
1 - \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} = 1 - \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} + \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j}
\]

and finally it follows

\[
M(\mathbf{p}) = \frac{1}{N} \sum_j \frac{1}{[\bar{t}_i]^2 j} \left( b_i - \frac{[b_i t_i]_j^2}{[\bar{t}_i]^2 j} \right)^2
\]

yielding the required expression in least-squares form. It should be noted that the sum over the squares in Eq. (2) is represented by the \( \ldots \) \( j \). The sum over \( j \) yields an additional summation of terms \( \mathbf{m}_i(\mathbf{p}) \) where the remaining constant terms in Eq. (4) contribute to the weights \( \omega_i^2 \).
3.2. Partial derivatives of LC

The partial derivatives $\frac{\partial}{\partial p_l} m_{ij}(p)$ are determined by calculus. The image $b$ does not depend on $p_l$, only $t$ depends on the actual parameters $p$. By our convention and by abbreviating $\frac{\partial}{\partial p_l} t$ with $t'$,

$$\frac{\partial}{\partial p_l}[t_{ij}] = \left[ \frac{\partial}{\partial p_l} t'_i \right]_j.$$  

Thus, it follows

$$\frac{\partial}{\partial p_l} m_{ij}(p) = \frac{\partial}{\partial p_l} \left( (b_i - \bar{b}_j) - \frac{[b_i t_{ij}]_j (t_i - t_j)}{[t'_{ij}]_j} \right) = \frac{2[b_i t'_{ij}]_j (t_i - t_j)}{[t'_{ij}]_j^2}.$$  

Finally, the remaining question is, how the image derivatives $\frac{\partial}{\partial p_l}$ are calculated. This is described in the next section in more detail.

4. Calculation of image derivatives

The calculation of $\frac{\partial}{\partial p_l} t$ needs particular attention since $m$ derivatives have to be determined in each iteration, which requires additional, expensive image reformating steps. During optimization this is the most time-consuming element. If the derivatives are determined, however, then the remaining calculation of the matrix of the linear systems in Eq. (3) is fast: for each $j$ the grey values of voxels $i \in n(j)$ have to be processed according to Eq. (5). We describe in the following three alternatives which have significant impact on the computational cost of the optimization.

4.1. Standard method

At iteration step $k$, the image $t_0$ is first reformatted according to the actual optimization parameters $p$, yielding the image $t$. The derivative of $t$ with respect to $p$ is calculated by a central-difference approximation [2]

$$t' = \frac{\partial}{\partial p_l} t = \frac{T(p + \epsilon_l, t_0) - T(p - \epsilon_l, t_0)}{2\epsilon_l}$$

where $\epsilon_l$ is a small number. In all experiments, $\epsilon_l$ is set to 0.1 voxel for translation and to 0.5 deg for rotation. It is observed that a change of these values does not affect the results within the limit of the registration accuracy. The central-difference approach requires two resampled images for each derivative. In the case of rigid transformations there are 12 additional images to be calculated which makes this method rather expensive.

4.2. Application of chain rule

The number of reformating steps can be reduced by application of the chain rule which shifts the finite-difference approximation for $p_l$ to an approximation with respect to the 3-D image space. Of course, the central differences have to be calculated, but only with respect to $x, y$ and $z$ which are the usual grey value gradients in direction of the coordinate axes. Therefore, only 6 resampled images are required, in particular, this number is independent of the number of parameters $m$. By the chain rule

$$\frac{\partial}{\partial p_l} t = \left[ \frac{\partial}{\partial x^t} \cdot \frac{\partial}{\partial y^t} \cdot \frac{\partial}{\partial z^t} \right] \cdot \frac{\partial}{\partial p_l} T(p)$$

the derivative of $t$ with respect to $p_l$ is calculated as the scalar product of image gradients with respect to $x, y$ and $z$ and the partial derivative of the transformation $T$ with respect to $p_l$ at position $(x, y, z)^T$ in the 3-D image space. Note that the image gradients are identical to the partial derivatives with respect to the translation parameters $p_1, p_2$ and $p_3$ which are calculated first. Rotational derivatives are then determined according to Eq. (6).

4.3. Change of base and target image

A Gauss-Newton iteration $k$ according to Eq. (3) is based on the image $b$, the reformatted image $t = T(p_k, t_0)$ and the $m$ derivative images, which are calculated using the images $T(p_k \pm \epsilon_l, t_0)$ as shown in Fig. 1 (top). The Newton
The Newton step $s_k$ can also be determined in a different way relating the calculation of the derivatives to the base image, which is not changed throughout the optimization. As indicated in Fig. 1 (bottom), the Newton step at iteration $k$ can be determined as well after the base and the target image are interchanged: we assume that $t$—after reformating according to $p_k$—is the fixed (base) image and that $b$ is the reformatted (target) image. Since $b$ has not been reformatted at all the actual parameters are zero and the derivatives are calculated according to $T(\pm e, b)$. The solution of the corresponding equations is the Newton step $r_k$, which is the update of the base $b$ for the next iteration. However, the calculated Newton step $r_k$ can be used for the original optimization with respect to $t$: obviously, the target $t$ can be reformatted according to the inverse translation and rotation parameters $r^{-1}$. It should be noted, however, that the new parameters $p_{k+1}$ are not just the sum $p_k + r^{-1}$, but must be calculated by the concatenation of the transformations

$$
T(p_{k+1}, t_0) = T(r^{-1}, t)
= T(r^{-1}, T(p_k, t_0))
= T(r^{-1}) \circ T(p_k) |_{t_0}.
$$

As an advantage of the image change, derivatives must be only computed once in a pre-processing step prior to the optimization. This drastically reduces the number of image reformattings, in particular for time-series registration. If the number of target images is large, the amount of CPU time for the pre-calculation of derivatives becomes marginal when compared to the time for the entire registration.

5. Results

In this section we summarize CPU times of the proposed method for the registration of time-series, as well as for image pairs. The CPU times are given for ANSI C code with highest level of compiler optimization running on a moderate 400 MHz Sun workstation with 512 MByte RAM. Furthermore, we compare the registration accuracy of LC and MI and show that the superior computation time of the LC optimization is not achieved at the expense of registration quality.

5.1. MR Neuro perfusion time-series

In MR neuro perfusion the uptake time of a contrast agent provides quantitative parameters to characterize the vitality of brain tissue [4]. The administration of the contrast agent results in a visible decrease of image intensity as illustrated in Fig. 2 (top) which is proportional to the blood volume in each voxel. In our experiments we use a multislice FFE–EPI time-series of 40 images with an image matrix of $128^2 \times 30$ (19 Mvortex).

Higher computational performance may be always achieved by relaxation of the requirements on registration accuracy. We therefore intend to relate the outcome of our experiments with respect to CPU time and registration accuracy. Specifically, we plot accuracy as a function of elapsed CPU time for different LC fractions, MI, and optimizations. Since the true registration transformations of the time-series are not known we address registration capability of an algorithm by the statistical consistency of the registration results [3, 8]. For three images $t_j, t_k, t_l$ of the time-series the cyclic registration transformations $T_{jk}, T_{kl}, T_{lj}$ are determined. Their composition $T_{jk} \circ T_{kl} \circ T_{lj}$ yields some residual rotational and translational errors which are small when the three images are well-registered. For each possible combination of three images of the time-series the residual errors are calculated and accumulated. The resulting rotational and translational mean errors characterize the con-
Fig. 3 shows residual mean errors as a function of the average CPU time per image. The plots depict rotational (top) and translational (bottom) consistency, respectively. First, we evaluate the Gauss-Newton registration for several voxel fractions and interchange of base and target (Sec. 4.3). All intermediate images are resampled with linear interpolation. The graphs show that the proposed method is more consistent for larger voxel fractions, which can be explained by the fact that more voxels contribute to the registration (refer to labels “5%”, “10%”, ...). With 10% voxel fraction, for example, one image of the time-series can be registered within 500 ms. This is much faster than the acquisition of an image which is of the order of 1–3 s [4].

To relate these results to MI we implemented a second optimization method based on a simple hill-climbing scheme which does not require derivative information. Given an actual parameter \( p \), a current step size is added and subtracted to each of the parameters \( p_1, \ldots, p_m \) followed by image resampling and calculation of the similarity measure under investigation. If the similarity value increases or decreases, \( p \) is immediately updated appropriately, otherwise the step size is reduced until a final accuracy is reached. In practice, the algorithm proves to be as accurate as the popular downhill simplex algorithm [9] but has better computational performance. Since the hill-climbing only evaluates comparisons of the objective function it can be easily used for an alternative implementation of LC as well. The labels “LC 10%” and “MI” in Fig. 3 show the results for the hill-climbing optimization. LC registration is more consistent than MI and more than twice as fast. Note also that the derivative-based Gauss-Newton algorithm is more consistent than LC with hill-climbing and about 20 times faster: the entire time-series is registered in 21 s.

Finally, Tab. 1 shows the performance of the Gauss-Newton optimizer in more detail. The CPU times for the standard calculation of derivatives (4.1), the application of the chain rule (4.2) and the change of base and target (4.3) are considered. Obviously, a reduction of the number of evaluated voxels (LC fraction) by a factor of 2 decreases the CPU time for optimization by a factor of 1.5–2. The application of the chain rule only slightly speeds up computation, while the change of base and target gives a significant gain of a factor of 4–6 when compared to the standard calculation of the derivatives.

### 5.2. Dual contrast MR images

To further demonstrate the registration properties of the proposed method two dual contrast MR images of the brain are investigated. The images are acquired simultaneously with a multi-slice dual contrast TFE sequence on an image matrix of size \( 8 \times M \) voxel. As an advantage of the dedicated imaging protocol, the spatial correspondence of the images is maintained providing ground truth for the registration. Slices from each image are shown in Fig. 2 (middle). To test the accuracy of the registration algorithm the image on the left in Fig. 2 is reformatted with the parameters \( p = (10.3 \text{ mm}, 12.7 \text{ mm}, -3.5 \text{ mm}, 9.4 \text{ deg}, 11.1 \text{ deg}, -4.2 \text{ deg}) \). Afterwards, the reformatted image is registered to the dual one.

To overcome the computational burden of large image matrices, multiresolution image decompositions are often suggested [7]. Multiresolution pyramids allow optimization using a coarse-to-fine strategy. First, images at lower resolution are registered where convergence is fast because the
corresponding image matrices are small. Then, finer levels are registered with the result of the previous scale used for initialization. Generally, each optimization method can be enhanced by a multiresolution representation. Speed-ups of a factor of 2–5 are reported in the literature, which is in agreement with our own experiences. The LC optimization is therefore complemented by a multiresolution strategy proposed by Studholme et al. for MI [11].

Tab. 2 summarizes the CPU time for the registration and the absolute differences of the reported registration results from the initial transformation. CPU times in the table do not include the calculation of the multiresolution representation of the images which takes 7 s for each experiment. First we evaluate the Gauss-Newton registration for different voxel fractions, application of the chain rule (Sec. 4.2) and resampling with linear interpolation. Experiments show that the interchange of base and target (4.3) has only minor influence on the CPU time. Although derivatives are calculated only once, the additional overhead for storing and retrieving the data during registration is almost of the same order as the computation of derivatives by the chain rule on the fly. Similar to the time-series experiments, the use of larger voxel fractions—given in the first four rows of the table—yields more accurate registrations but also slower computation. The translation inaccuracy remains to be well below the voxel size of 0.92 × 3.0 mm² and the rotation inaccuracy is very small.

In the following, the result obtained for LC is related to MI. Therefore, the MI optimization by Studholme et al. [11]—we already use their multiresolution decomposition—is applied to the dual contrast images and complemented by an implementation of LC similarity. The algorithm is based on a steepest-gradient-descent approach [9]. The rows with label “MI” and “LC 10%” in Tab. 2 summarize the results. The dedicated Gauss-Newton optimization with 10% of image voxels is 4 times faster than the gradient-descent implementation of LC using the same voxel fraction and 7 times faster when compared to MI. In the case of steepest-gradient-descent optimization both MI and LC recover the initial registration transform with high accuracy, although MI appears to be slightly superior. The accuracy of the Gauss-Newton approach with 10% LC fraction and the gradient-descent of MI, however, have comparable accuracy.

### 5.3. CT–MR images

We conclude this section by the application of the proposed LC optimization to a clinical CT–MR image pair of the brain which is a prominent application for multimodality registration. Slices from each image are shown in Fig. 2 (bottom). The size of the image matrices is 512² × 87 and 256² × 128 voxels, almost 30 Mvoxel are processed during registration.

With 10% LC fraction the images are registered in about one minute. The optimization requires only 13 s, while the remaining CPU time of 48 s is used by the multiresolution decomposition which has so far not been optimized. The MI algorithm by Studholme et al. [11] takes 97 s for the optimization which is about 7 times slower than the LC approach. The same factor is also found for the registration of the dual contrast images in the previous section. The true registration transformation of the images is not known, but the difference of the reported translations of LC and MI is below half of the voxel size for both images and the rotation differences are smaller than 0.5 deg which again indicates good agreement of both registrations.

### 6. Summary and conclusions

Although mutual information (MI) shows robust and accurate results for a large class of multi-modality registrations, efficient optimizations of the similarity measure are not suggested so far—mainly, because MI cannot be formulated as least-squares criterion. Registrations are still of the order of several minutes, which limits its use for clinical applications and system integration. Due to these difficulties, this contribution focuses on a new type of multi-modality measure which is based on local cross correlation coefficients (LC), and a very fast Gauss-Newton optimizer is proposed.

To demonstrate the performance of LC, the method is first tested with a clinical MR neuro perfusion time-series (19 Mvoxel) since the passage of an administered contrast agent results in intensity variation, demanding the use of multi-modality similarity measures. Consistency testing is introduced to evaluate the registration results. In particular, registration consistency is related to the CPU time to indicate a possible trade-off between accuracy and computational performance. To relate the results to MI, a second optimization based only on comparisons of the objective function is implemented. These experiments show that
sub-voxel accuracy is achieved and that the accuracy of the proposed LC optimization is comparable or slightly better as MI. A speed-up by a factor of more than 20 is achieved by the new optimizer. Individual images of the perfusion time-series are registered in real-time within 500 ms on a moderate workstation, which is at least twice as fast than their acquisition.

In a second evaluation, we applied the LC approach to multi-modality matching. Such registrations, for example of CT–MR images, are generally more time-consuming. The burden of large image matrices is usually overcome by a multiresolution representation of the images allowing to perform the optimization in a coarse-to-fine strategy. In a recently published paper comparing different optimization strategies for MI it is concluded that CPU times of about 5 minutes can be achieved on advanced workstations for comparable data [6]. These results, however, are difficult to relate to the proposed LC method because a different number of scales and a different resolution of the finest level may have significant impact on the CPU time.

We therefore compared the LC optimization to a MI algorithm based on a steepest-gradient-descent for a CT–MR image pair (30 Mvoxel) [11]. The registration is performed on top of the same pyramid decomposition of the data which requires 47 s computation time. In the case of MI the optimization requires 97 s while the proposed approach only needs 13 s corresponding to a speed-up about a factor of 7. In particular, the optimization itself is now much faster than the decomposition part, which needs further attention in future. Additionally, the registration accuracy is investigated with dual contrast MR images, providing ground truth for the registration transformation. As in the case of the time-series and CT–MR matching, both LC and MI show comparable accuracy results.

So far, we have not fully exploited the local properties of LC. Intermediate images during optimization, as well as the pyramid images, are completely reformatted although only a fraction of 5–20% of image voxels are actually considered by the measure. Furthermore, the voxel selection strategy and the size of the neighborhood, in particular with respect to the tradeoff between efficiency and accuracy, needs further attention. By the use of faster multiresolution decompositions [12] and by restricting image reformatting steps to the voxels of interest, CPU times can be further reduced, which makes LC also very promising for fast non-rigid image registration.

Acknowledgments

We would like to thank EasyVision AD (Philips Medical Systems, Best), C. Studholme (University School of Medicine, Yale), D. L. Hill and D. J. Hawkes (Guy’s Hospital, London) for providing the MI registration algorithm.

References