Presented at the Int. Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'00)
Hilton Head Island, South Carolina, USA – June 2000

3D Non-Rigid Registration by Gradient Descent on a Gaussian-Windowed Similarity Measure using Convolutions

Pascal Cachier, Xavier Pennec
INRIA - EPIDAURE Project, 2004 route des Lucioles, F-06902 Sophia Antipolis
{Pascal.Cachier, Xavier.Pennec}@sophia.inria.fr

Abstract

Non-rigid registration of medical images is usually presented as a physical model driven by forces deriving from a measure of similarity of the images. These forces can be computed using a gradient-descent scheme for simple intensity-based similarity measures. However, for more complex similarity measures, using for instance local statistics, the forces are usually found using a block matching scheme. In this article, we introduce a **Gaussian window** *scheme, where the local statistics (here the sum of local correlation coefficients) are weighted with Gaussian kernels. We show that the criterion can be deducted easily to obtain forces to guide the registration. Moreover, these forces can be computed very efficiently by* **global convolutions** *inside the real image of the Gaussian window in a time independent of the size of the Gaussian window. We also present two minimization strategies by* **gradient descent** *to optimize the similarity measure: a linear search and a Gauss-Newton-like scheme. Experiments on synthetic and real 3D data show that the sum of local correlation coefficients optimized using a Gauss-Newton scheme is a fast and accurate method to register images corrupted by a non-uniform bias.*

1 Introduction

In the context of this paper, the goal of non-rigid registration is the retrieval of the motion that has occurred between two volumetric images of the same organ. Because we are working on two images I and J, we have to rely on features of these images. Then, the registration process consists in finding a transformation T that matches well the features, with the hope that these features are significant enough to justify the assumption "features are matched hence homologous physical points are matched." Numerous features are possible, such as points, curves, or intensity. In this paper, we focus on intensity-based methods.

When the intensity is conserved during the motion, an appropriate similarity energy[1] is the sum of square differences (SSD). Its gradient is easy to compute and has been used as the guiding force of numerous non-rigid registration techniques [1, 2, 3, 4]. When this assumption is invalid, for example with magnetic resonance images (MRI) corrupted by a non-uniform bias, we have to use more complex criterions, such as the sum of local correlation coefficients (LCC). In this case, the minimization is usually done by a block matching technique using an exhaustive search [5, 6, 7, 8] or a 0th-order minimization strategy [9]. Both the computation of statistics in each hard block and the search of the minimum are computationally intensive for volumetric images. As a consequence, windows are usually sparse and the minimization is done for translations only.

In this paper, we propose to use an **isotropic Gaussian window** instead of a hard block for the computation of local statistics. This way, we can compute the criterion and its derivatives at each voxel by **global convolutions**, in a time independent of the extension of the Gaussian window, thus obtaining dense force fields. Section 2 presents the method used to compute local statistics with Gaussian convolutions. Section 3 presents our two-step non-rigid **gradient descent** based registration algorithm. Finally, we present results of the algorithm on synthetic and real images in Section 4 and we compare different optimization strategies.

2 Computation of Local Statistics

When the intensity conservation hypothesis that drives numerous mono-modal non-rigid registration algorithms is violated in medical images, this is mainly due to a non-uniform bias that effects the imaging modality. Therefore, the use of local statistics, and

[1] For convenience, we shall talk about similarity energies instead of similarity measures. This terminology supposes that the energy should be positive and takes zero as a minimum. A similarity measure S with an upper bound S^* can be transformed into a similarity energy E by setting $E = S^* - S$.

particularly the sum of local correlation coefficients, has been increasingly popular. It has even been suggested that it could drive a multi-modal registration [10].

Most of the time, these statistics are computed in a window. The image is divided in some number of square windows that may, or may not, overlap. In its simplest form, the block matching scheme consists in moving each window into several positions, computing a local similarity energy inside the window, and keeping the position of the window leading to the best match. However, it is costly to compute the similarity energy for every position of the window. Hence there is a trade-off between precision and computation time. Generally, one chooses fewer windows than voxels. Also, the motion of the window is often restricted to translation, because adding rotation will multiply the number of estimations, especially in 3D images.

Local Gaussian Correlation Coefficients

In the method we propose, we compute local similarity measures around each voxel. The similarity measure we use in this paper is the correlation coefficient.

Let $G_{\mathbf{p}}$ be the window function centered around a voxel $\mathbf{p}$ obtained by translation of a discrete, normalized, symmetric kernel G, and I and $J \circ T$ be the two images defined on $\mathbf{N}^n$ to register with a transformation T. We define the local mean of I, the local correlation between I and $J \circ T$, the local variance of I, and the local correlation coefficient between I and $J \circ T$ (all of them centered around $\mathbf{p}$) respectively by:

$$\begin{aligned}
\overline{I}_{\mathbf{p}} &= \sum_{\mathbf{q}\in\mathbf{N}^n} G_{\mathbf{p}}[\mathbf{q}].I[\mathbf{q}] \\
< I, J \circ T >_{\mathbf{p}} &= \sum_{\mathbf{q}\in\mathbf{N}^n} G_{\mathbf{p}}[\mathbf{q}].\left(I[\mathbf{q}] - \overline{I}_{\mathbf{p}}\right)\left(J \circ T[\mathbf{q}] - \overline{J \circ T}_{\mathbf{p}}\right) \\
\sigma^2_{\mathbf{p}}(I) &= < I, I >_{\mathbf{p}} \\
CC_{\mathbf{p}}\left(I, J \circ T\right) &= \frac{< I, J \circ T >_{\mathbf{p}}}{\sigma_{\mathbf{p}}(I)\,.\,\sigma_{\mathbf{p}}(J \circ T)}
\end{aligned}$$

The sum of local correlation coefficients criterion is then defined by:

$$LCC(I, J \circ T) = \sum\nolimits_{\mathbf{p}\in\mathbf{N}^n} CC_{\mathbf{p}}(I, J \circ T) \qquad (1)$$

There are multiple choices for the kernel G (hard block, hard sphere, Gaussian, etc.). Here, we have chosen a Gaussian window. The choice of the sum of local correlation coefficients as a registration criterion, and of the discrete, normalized Gaussian as the window function, has been done in the light of the following considerations.

The use of the correlation coefficient enables to compute a distance between two images that is invariant by an affine rescaling of the intensity range on one of the images. It is therefore insensitive to a uniform linear bias.

Because we use local correlation coefficients, the estimation of the bias is done locally. The assumption implicitly done is that the bias should not vary too much inside the window in which the local correlation coefficient is computed. However, the bias is allowed to vary with the position in the image. Hence, we are able to tackle non-uniform biases. We will show later that the LCC criterion (eq. 1) can actually tackle a bias with a variation up to 50% of the intensity range.

The locality of the correlation coefficients is obtained by introducing a Gaussian weight (or Gaussian window) in the formulas. The Gaussian window fulfills two theoretical needs. The first one is the isotropy[2] of the window. Hence, if we compute some statistics around a point $\mathbf{p}$, the influence of another point $\mathbf{q}$ on the results depends only on its distance to $\mathbf{p}$. The second one is the gradual decrease of the influence of remote points. Remote points should have less influence than points located near the center of the window. This is similar to the implementation of Bajcsy [11], where the correlation criterion is not locally weighted but the images are decomposed onto an Hermite orthogonal polynomial basis.

Finally, we compute local correlation coefficients around every voxel, and not only a subset of the voxels. Thus, we do not have any privileged point or location in our images, and our measure is invariant by any integer translation.

Practical Computation using Convolutions

Fortunately, we are able to compute the local statistics and their derivatives at a relatively low computational cost, because the windows are not treated separately but at once, using convolutions. Using the notations previously introduced, local statistics around every voxels can be computed altogether with one or two Gaussian convolutions (denoted by $G*$):

$$\begin{aligned}
\overline{I}_{\mathbf{p}} &= (G * I)[\mathbf{p}] \\
\sigma^2_{\mathbf{p}}(I) &= \overline{I^2}_{\mathbf{p}} - \overline{I}^2_{\mathbf{p}} = G * (I^2)[\mathbf{p}] - (G * I)^2[\mathbf{p}] \\
<I, J \circ T>_{\mathbf{p}} &= \overline{I.J \circ T}_{\mathbf{p}} - \overline{I}_{\mathbf{p}}.\overline{J}_{\mathbf{p}} \\
&= G * (I.J \circ T)[\mathbf{p}] - (G * I)[\mathbf{p}].(G * J \circ T)[\mathbf{p}]
\end{aligned}$$

Note that this can actually be done for any symmetric window function G. However, the Gaussian win-

[2] The Gaussian should be isotropic in the *physical space*. If the voxels themselves are not isotropic, then the Gaussian window follows this anisotropy in the *image space*.

dow, in addition to its theoretical properties presented above, has two practical advantages:

1. The convolution by a Gaussian is separable, i.e., we only have to smooth the image successively with three 1D Gaussian kernels to smooth a 3D image with a 3D Gaussian kernel.

2. The 1D Gaussian convolution can be approximated by a recursive filter, yielding a fast convolution in a time *independent of the standard deviation of the Gaussian window* [12].

Optimization of the Criterion

To minimize the similarity energy, we compute its derivatives with respect to the displacement field, and use a gradient descent minimization. Different optimization strategies will be presented in the next section. If necessary, we use a multi-resolution scheme to avoid being trapped in local minima.

Because we use Gaussian windows, the derivatives of the criterion can be easily computed using some more Gaussian convolutions. For instance, we put in Table 1 the derivative of the sum of local correlation coefficients criterion with respect to the transformation $T[\mathbf{p}]$ of a point $\mathbf{p}$ (Eq. 2). We also derived an approximation of the previous formula, where all final convolutions are dropped, in order to be faster (Eq. 3 in Table 1).

In the following, the gradient descent using formula (2) is called *Local Correlation Coefficients* (LCC). The one using formula (3) is called *Simplified Local Correlation Coefficients* (SLCC).

We compute a gradient with respect to the displacement of each voxel. Hence, we have a dense vector field. The first advantage of doing so is that we can handle local, subtle deformations, independently of the spread of the Gaussian used to compute local statistics. A second advantage is that the local similarity of the two images is computed using the *real* image of the window[3] and not its rigid or translational approximation (Fig. 1).

Numerical Precision Issues

A mathematical equality may hide numerical problems. For example, consider this equality for the local variance: $\overline{(I-\overline{I}_{\mathbf{p}})^2}_{\mathbf{p}} = \overline{I^2}_{\mathbf{p}} - \overline{I}^2_{\mathbf{p}}$. Whereas the left-hand side will only sum positive and relatively small numbers, the right-hand side will compute two big yet close numbers, $\overline{I^2}_{\mathbf{p}}$ and $\overline{I}^2_{\mathbf{p}}$, and then subtract them: we lose several bits of precision. To quantify the loss

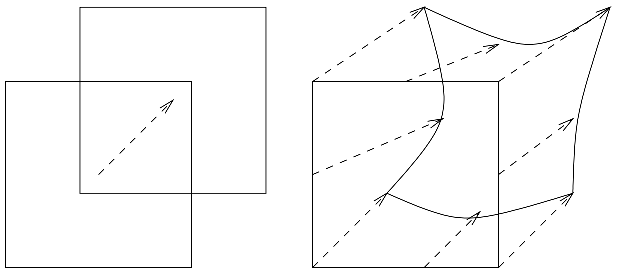


Figure 1: *On the left, the common block matching scheme assumes that motion is uniform on the entire window and equal to the motion of its center. On the right, using a dense displacement field allows us to use the real image of the window to compute the similarity energy.*

of precision, we have compared the results computing the local variances of the image 2(a) using the left-hand side and using the right-hand side. We have used real numbers coded on 4 bytes, and as a preprocessing we have subtracted the mean of the whole image. It appears that the error in the computation of the local variance using the difference of mean relatively to the accurate method has an average of 0.14%±0.83% and a median error of 0.00033%. This means that we have a good numerical precision everywhere, except for strong outliers which appear to lie on the borders. They are due to the conjugate effects of the very low variance on these plain black regions, and the boundary effects of the filter. They are mostly less than 1%, but reach up to 40% in the 4 corner pixels.

In our experiments, these errors never had a meaningful incidence on the results, but we should be aware that it can potentially induce some significant distortions in the deformation field at the image borders.

3 A Two-Step Registration Algorithm

Non-rigid registration using free form transformations is an ill-posed problem that needs regularization. Within the medical image registration community, regularization is often done by choosing a smoothing energy in addition to the similarity energy, and by minimizing both together. Of course, there is a trade-off between the similarity energy, reflected by the visual quality of the registration, and the smoothing energy, reflected by the regularity of the transformation (the term "regularity" should be taken in its broadest sense, since the smoothing energy may allow occasional discontinuities in the displacement field [13]). Therefore, we show the result of a non-rigid registration using the deformed image *and* the transformation itself.

In the regularization theory framework, one min-

[3]If we use a recursive implementation for the Gaussian, the window actually fills the entire image.

$$\begin{aligned}\frac{\partial LCC(I, J\circ T)}{\partial T[\mathbf{p}]} &= \left[I[\mathbf{p}].G * \frac{1}{\sigma_{\mathbf{q}}(I).\sigma_{\mathbf{q}}(J\circ T)}[\mathbf{p}] - J\circ T[\mathbf{p}].G * \frac{< I, J\circ T >_{\mathbf{q}}}{\sigma_{\mathbf{q}}(I).\sigma^3_{\mathbf{q}}(J\circ T)}[\mathbf{p}] \right.\\ &\quad \left. + G * \left(\frac{< I, J\circ T >_{\mathbf{q}} \, \overline{J\circ T}_{\mathbf{q}}}{\sigma_{\mathbf{q}}(I).\sigma^3_{\mathbf{q}}(J\circ T)} - \frac{\overline{I}_{\mathbf{q}}}{\sigma_{\mathbf{q}}(I).\sigma_{\mathbf{q}}(J\circ T)} \right)[\mathbf{p}] \right] \nabla J(T[\mathbf{p}]) \qquad (2)\\ &\approx \frac{1}{\sigma_{\mathbf{p}}(I).\sigma_{\mathbf{p}}(J\circ T)} \left(\left(I[\mathbf{p}] - \overline{I}_{\mathbf{p}}\right) - \left(J\circ T[\mathbf{p}] - \overline{J\circ T}_{\mathbf{p}}\right) \frac{< I, J\circ T >_{\mathbf{p}}}{\sigma^2_{\mathbf{p}}(J\circ T)}[\mathbf{p}] \right) \nabla J(T[\mathbf{p}]) \qquad (3)\end{aligned}$$

Table 1: *Derivative of the sum of local correlation coefficients criterion with respect to the transformation $T[\mathbf{p}]$ of a point $\mathbf{p}$ (equ. 2) and an approximation without Gaussian convolutions to speed up the computations (equ. 3).*

imizes a weighted sum of both energies $E_1 + \lambda.E_2$, E_1 and E_2 being respectively the similarity and the smoothing energies, and λ being a trade-off coefficient. This formulation has proven to be successful for data approximation, and has been used for various approaches of non-rigid registration algorithms [13, 14].

Another widely used method attempts to separate the image measure from the transformation measure, and could be compared with the approach of game theory. It consists in alternatively decreasing the similarity energy and the smoothing energy. This includes in particular most of block-matching algorithms [5, 6, 7, 8, 9] and some optical-flow-based techniques [15, 16]. Note that a regularization-theoretic approach may lead in practice to a two-step algorithm [16].

In this paper, we use the alternated-minimization method. We now detail in turn minimization of the two energies.

Minimization of the Similarity Energy

The image registration step consists in minimizing the similarity energy using a gradient descent [17]. We have tested two strategies: a Gauss-Newton-like scheme, and a linear search.

Gauss-Newton When we want to minimize a non-linear least squares function $\mathbf{z}(\mathbf{x})^T.\mathbf{z}(\mathbf{x}) = \sum_i (z_i(\mathbf{x}))^2$, the Gauss-Newton scheme realizes a second order gradient descent using an approximation of the Hessian matrix from the gradient of this criterion to minimize the computations. The evolution equation is:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - (J(\mathbf{x}_t)^T J(\mathbf{x}_t) + \lambda.Id)^{-1} J(\mathbf{x}_t)^T \mathbf{z}(\mathbf{x}_t)$$

where $J = (\partial z_i / \partial x_j)$ is the Jacobian matrix of $\mathbf{z}$ (the matrix of the first order derivatives). In the case of the SSD criterion on two images I and J, this formula adds at each iteration the following vector field to the current displacement field [4]:

$$\mathbf{u}[\mathbf{p}] = \frac{(I[\mathbf{p}] - J\circ T[\mathbf{p}])}{\|\nabla J\circ T[\mathbf{p}]\|^2 + \lambda} \nabla J\circ T[\mathbf{p}]$$

which can also be written:

$$\mathbf{u}[\mathbf{p}] = -\frac{2E_1[\mathbf{p}]}{\|\nabla E_1[\mathbf{p}]\|^2 + 4\lambda E_1} \nabla E_1[\mathbf{p}] \qquad (4)$$

where $E_1[\mathbf{p}] = (I[\mathbf{p}] - J\circ T[\mathbf{p}])^2$ is our local similarity energy and $\nabla E_1[\mathbf{p}] = 2(J\circ T[\mathbf{p}] - I[\mathbf{p}])\nabla J\circ T[\mathbf{p}]$ its gradient with respect to the transformation $T[\mathbf{p}]$. To extend this scheme to our non quadratic criterion (the sum of local correlation coefficients), we simply use formula (4) with the similarity energy $LCC(I, J\circ T)$ described in Equation 1. Even if the Gauss-Newton scheme was developed for non-linear least squares problems, we will see in the experiments section that our adaptation still performs very well.

Linear search We have also tried a *linear search* to compare with our Gauss-Newton adaptation. The search is done along the descent direction given by (4), which can be seen as a pre-conditioning of ∇E_1. The use of a linear search is computationally more expensive since it requires several estimations (typically 8) of the criterion at each iteration.

Minimization of the Smoothing Energy

The regularization of the transformation is done here by decreasing by a certain amount the following stretch energy of the displacement field $\mathbf{U} = Id - T$:

$$E_2(\mathbf{U}) = \int \left\| \frac{\partial \mathbf{U}}{\partial \mathbf{x}} \right\|^2 = \int \left\| \frac{\partial \mathbf{U}}{\partial x} \right\|^2 + \left\| \frac{\partial \mathbf{U}}{\partial y} \right\|^2 + \left\| \frac{\partial \mathbf{U}}{\partial z} \right\|^2$$

This is done by Gaussian smoothing [18], the spread of the kernel growing with the amount of reduction done.

4 Experiments

To evaluate the different methods previously presented, we first ran a series of experiments on 2D data with synthetic deformations. This allows us to exhibit some key features about each method and criterion while evaluating the accuracy of the non-rigid registration. Then, we show that the algorithms perform similarly on real 3D data. All experiments were performed on a 450 MHz Pentium II and used a Gaussian window with a standard deviation of 4 voxels to compute the local statistics.

Synthetic 2D experiments

In the first experiment, we took a "noise-free" 211×228 MRI slice (obtained by anisotropic diffusion), and we deformed it with a known synthetic transformation meant to be highly non-rigid yet local. Then, we added a white Gaussian noise of standard deviation 3 to both the deformed (Fig. 2(b)) and non-deformed image (Fig. 2(a)).

We tried the Gauss-Newton scheme with the standard SSD, the LCC, and the SLCC formula. We also ran a linear search using the LCC criterion, to compare with the Gauss-Newton-like minimization. 20 iterations were allowed, which is sufficient for the convergence of the Gauss-Newton scheme; the linear search scheme converged around 60 iterations.

In the second experiment, we also added a strong non-uniform bias to the deformed image, ranging linearly from 0 (at the top left corner) to 130 (at the bottom right corner), while maintaining the intensities in the range [0; 255] (Fig. 2(c)).

We have used a Gaussian kernel with a standard deviation of 1.4 for the smoothing. A rough preliminary study has shown that best results were obtained with this standard deviation for these couple of image.

The result of the registration is evaluated by the mean Euclidean distance between the synthetic transformation T^* used to deform the image and the transformation T found by the registration algorithm:

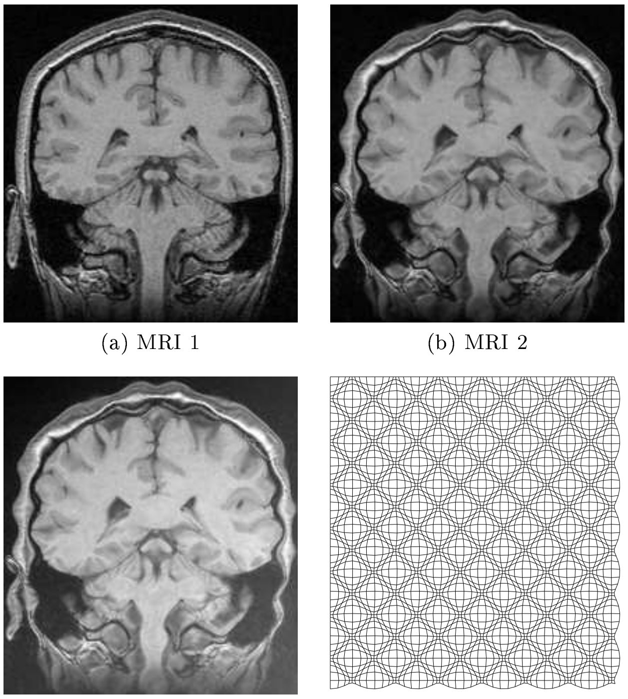
$$D1(T) = \frac{1}{\#I} \sum_{\mathbf{p} \in I} \|T[\mathbf{p}] - T^*[\mathbf{p}]\|$$

where $\#I$ is the number of pixels of the image. Since it is impossible to recover the transformation in the black regions outside the head, we have also computed their mean distance inside the head:

$$D2(T) = \frac{1}{\#\Omega} \sum_{\mathbf{p} \in \Omega} \|T[\mathbf{p}] - T^*[\mathbf{p}]\|$$

where Ω is a mask obtained by manual segmentation.

Influence of the minimization scheme First, we observe that the Gauss-Newton-like scheme, which is a special adaptation of a Gauss-Newton for non quadratic criterions, has a proper minimization behavior. The accuracy results are more or less equivalent with the linear search for LCC and SLCC (see Table 2), but the linear search is slower because it needs several computation of the criterion (typically, 8) at each iteration.



(a) MRI 1 (b) MRI 2

(c) MRI 3 (d) T^*

Figure 2: *Synthetic images and transformation.* ***From left to right:*** *Image 2(a) has been transformed into image 2(b) using a deformation depicted on image 2(d), and white Gaussian noise ($\sigma = 3$) has been added afterwards. Image 2(c) has been obtained by adding a bias varying linearly from 0 at the top left corner to 130 at the bottom right corner to image 2(c) (the original image intensity ranges from 0 to 250).*

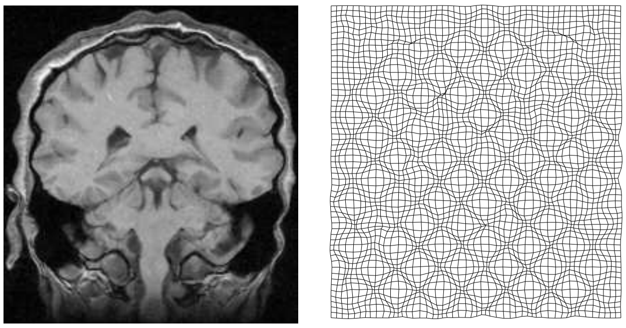


Figure 3: *Result of non-rigid registration of image 2(a) toward image 2(c) using the SLCC and Gauss-Newton strategy.* ***Left image*** *should be compared to the target image 2(c). Structures are almost perfectly matched.* ***Right image*** *should be compared to the synthetic transformation 2(d). The transformation is recovered in regions carrying information.*

Criterion	Minimization	Without bias (MRI 2) D1	D2	#It	Time	With bias (MRI 3) D1	D2	#It	Time
Initial		2.04	2.04			2.04	2.04		
SSD	Gauss-Newton	1.10	0.93	20	14s	6.84	7.60	20	15s
LCC	Gauss-Newton	1.16	1.00	20	37s	1.20	1.05	20	36s
SLCC	Gauss-Newton	1.13	0.96	20	24s	1.15	0.97	20	24s
LCC	Linear Search	1.27	1.12	20	77s	1.37	1.23	20	89s
LCC	Linear Search	1.13	0.96	60	225s	1.21	1.06	60	219s

Table 2: *Summary of synthetic experiment results: comparison of accuracy and computation times of algorithms with respect to the criterion and the minimization method used.*

Influence of the criterion The sum of local correlation coefficient performs almost as well as the classical SSD on unbiased images. However, as expected, the SSD criterion completely fails for biased images while LCC and SLCC still perform equivalently, showing their robustness to non-uniform bias. The additional computation time due to the use of the local correlation coefficients criterion instead of the SSD criterion is reasonable: a factor 1.75 for SLCC to a factor of 2.50 for LCC. These ratios should actually be smaller when more complex physical models are used for smoothing the deformation field.

Influence of the derivative simplification (SLCC vs LCC) Surprisingly, the algorithm performs slightly better with the derivative simplification (SLCC) than with the exact derivative (LCC), while we were expecting the contrary. We have no explanation for this yet. This result may be due to the fact that we are alternating minimizations and not minimizing a sum of energies. As far as the computation times are concerned, SLCC runs in two thirds of the computation time of LCC.

Analysis of the deformation field errors Errors are mainly due to the inability to totally recover the motion in homogeneous areas and the component of the motion that is tangential to the boundaries of the image and makes them slide. Errors on the component of the motion normal to boundaries are generally much smaller (Fig. 4). Therefore, the result image, whose aspect is somewhat insensitive to the tangential and homogeneous-area errors, looks very much like the target image (Fig. 3).

Real 3D experiment

In this experiment, we have registered two $52 \times 89 \times 45$ MRI centered on the ventricles of a patient who is suffering from focal aphasia. This disease is seen by a dramatic increase of the size of the ventricles between the first MRI 5(a) and the second MRI 5(b) taken a

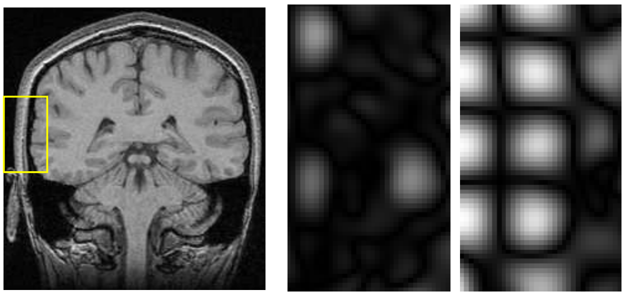


Figure 4: *Errors on the displacement field for the region delimited on the left image.* ***Middle image*** *shows the errors along the x-component, normal to the boundary of the skull. Errors are around 0.1 pixel.* ***Right image*** *shows the errors along the y-component, tangential to the boundary of the skull. Local maxima of the error are between 2.6 and 3.0 pixel (local maxima of the initial displacement is 3.0 pixel). The tangential component of the displacement is not recovered and constitutes the main source of errors.*

year later. This pair of images presents a bias as well as a contrast change.

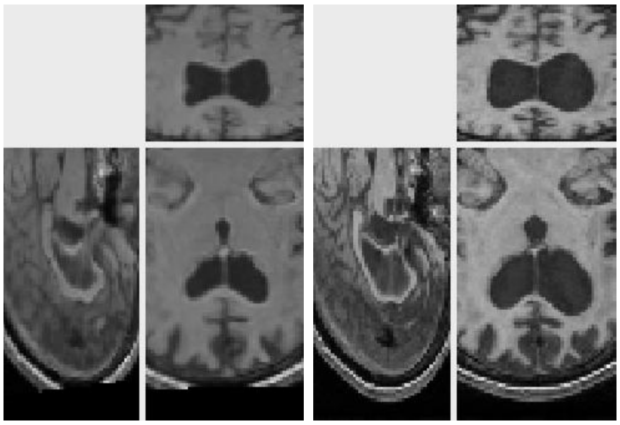
The original images were previously rigidly registered with the robust algorithm of [19]. One of the original image does not show the top of the head, which explains why a part of the skull is missing at the bottom of the rigidly registered image 5(a).

We used a Gauss-Newton-like gradient descent with the SLCC formula to register the images, and a Gaussian with a standard deviation of 1.8 to smooth the transformation.

The results presented in Fig. 6 have been obtained in 103 seconds on a 450MHz Pentium II. They show a good match of the structures, as well as a smooth transformation, with deformations localized near the ventricles and the missing part of the skull. This later movement is due to the lack of specific treatment of occlusions in our algorithm. Results using the LCC formula are almost the same, but run in 127 seconds.

A remarkable fact is that an isovalue of the Jacobian of the recovered transformation almost perfectly segments the ventricles which have grown between both acquisitions. The dilatation ratio recovered by the transformation is the same for big structures as well as small structures.

We have also registered these images using the SSD criterion, and a uniform correction of the bias and contrast change, i.e. we have set $I' = (I - \overline{I}_{\mathbf{p}}) * \sigma(J)/\sigma(I) + \overline{J}_{\mathbf{p}}$ and registered I' and J, I being image



(a) 3D MRI 1 (b) 3D MRI 2

Figure 5: *Rigidly-registered 3D MR images of the ventricles. Notice the bias and the deformation in relation with focal aphasia.*

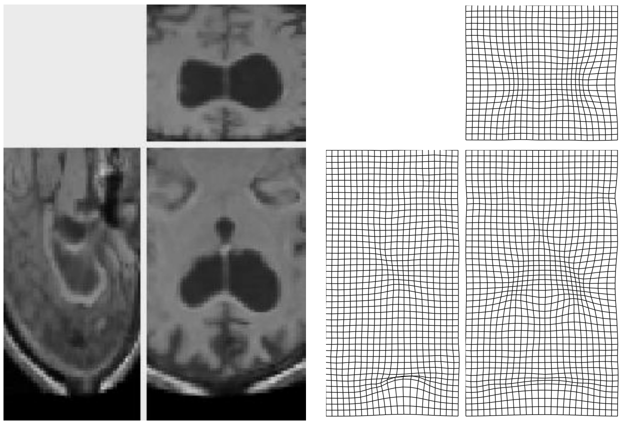


Figure 6: *Result of non-rigid registration of the images 5 using the SLCC+Gauss-Newton algorithm.* ***Left Image*** *shows the image result, which should be compared to the image 5(b). Ventricles are good, but grey matter is difficult to match, and the skull occlusion perturbs the deformation.* ***Right Image*** *shows the transformation result. It is smooth, and the biggest deformations are inside the ventricles, and near the missing part of the skull.*

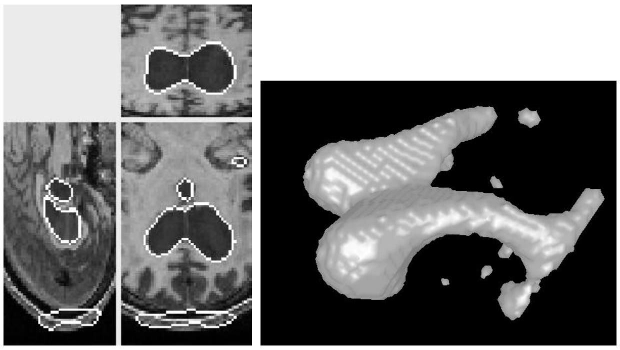


Figure 7: *An isovalue of the recovered transformation segments the ventricles.*

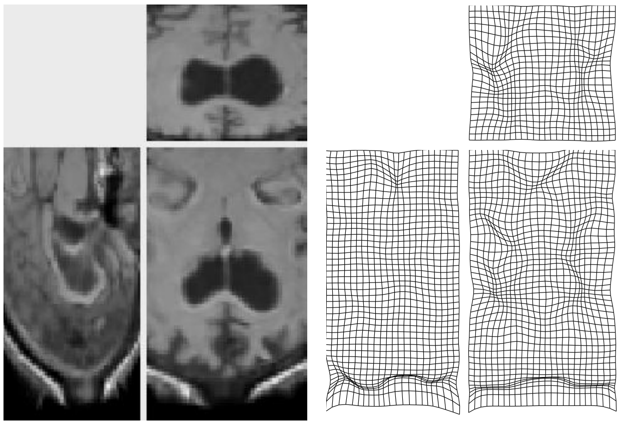


Figure 8: *Results using the SSD + Gauss-Newton algorithm with uniform intensity correction. Due to the growth of the ventricles, the uniform intensity correction underestimates the intensities. This is enough for the SSD to fail.*

5(a). The registration ran in 64 seconds. The results are poor: this is due to the fact that the uniform intensity correction is not efficient. For example, the background intensity is 0 in J but -9 in I', which explains the problems near the skull. The intensity value of white matter is also underestimated (100 instead of 120). The uniform intensity correction does not work because the histogram is altered by changes (the skull occlusion and mainly the growth of the ventricles) other than the bias and contrast change.

5 Conclusion

In this paper, we showed how to use the derivatives of the sum of local correlation coefficients (LCC) everywhere in the image to drive a non-rigid registration algorithm. This is computationally feasible thanks to the Gaussian window because we can compute the derivatives very efficiently using separable Gaussian filters. We also compared two gradient descent techniques, an adaptation of the Gauss-Newton method and a linear search, to optimize this non quadratic criterion.

We found that the LCC and its approximation SLCC are almost as accurate as the SSD on images without bias. They work almost as well with biased images, where the SSD fails, showing the robustness of this criterion with respect to non-uniform bias. The LCC and the SLCC methods demonstrate equivalent accuracies, but SLCC is faster than LCC: they were respectively 1.75 and 2.50 times slower than the SSD in our experiments.

A few points still need to be investigated to fully control and automate the algorithm. Firstly, we have to study the impact of the boundary conditions of the filters on the computations of the similarity measure and the gradient. The spread of the Gaussian window used to compute the local statistics could be chosen automatically. Last but not least, we should study and compare the behavior of our technique using a minimization of the sum of the similarity and the smoothing energies.

We believe that numerous non-rigid registration algorithms that use dense (i.e., at each voxel) forces to guide the registration could be extended from the SSD to the LCC criterion using the approach presented here. Moreover, even if we presented the formulas and the experiments with the sum of local correlation coefficients, the method is intrinsically generic and could be applied to many other local criterions, the only limitation being the ability to compute the derivative of the criterion.

Acknowledgments: We thank S. Ourselin for his proofreading and fruitful discussions.

References

[1] H. Lester *et al.* Non-Linear Registration with the Variable Viscosity Fluid Algorithm. In *Proc. of IPMI'99*, pages 238 – 251, 1999.

[2] G. E. Christensen, S. C. Joshi, and M. I. Miller. Volumetric Transformation of Brain Anatomy. *IEEE Trans. on Medical Imaging*, 16(6):864–877, 1997.

[3] M. Bro-Nielsen and C. Gramkow. Fast Fluid Registration of Medical Images. In *Proc. of VBC'96*, pages 267–276, 1996.

[4] X. Pennec, P. Cachier, and N. Ayache. Understanding the "Demons" Algorithm: 3D Non-Rigid Registration by Gradient Descent. In *Proc. of MICCAI'99*, pages 597 – 605, 1999.

[5] Y. H. Lau, M. Braun, and B. F. Hutton. Non-Rigid 3D Image Registration Using Regionally Constrained Matching and the Correlation Ratio. In *Proc. of WBIR'99*, pages 137 – 148, 1999.

[6] J. B. A. Maintz, E. H. W. Meijering, and M. A. Viergever. General Multimodal Elastic Registration based on Mutual Information. *Image Processing*, 1998.

[7] M. G. Strintzis and I. Kokkinidis. Maximum Likelihood Motion Estimation in Ultrasound Image Sequences. *IEEE Signal Processing Letters*, 4(6), 1997.

[8] J. C. Gee, L. Le Briquer, C. Barillot, D. R. Haynor, and R. Bajcsy. Bayesian Approach to the Brain Image Matching Problem. In *SPIE Medical Imaging*, 1995.

[9] D. L. Collins and A. C. Evans. ANIMAL Validation and Applications of nonlinear registration based segmentation. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 11(8):1271 – 1294, 1997.

[10] J. Weese, P. Rösch, T. Netsch, T. Blaffert, and M. Quist. Gray-Value Based Registration of CT and MR Images by Maximization of Local Correlation. In *Proc. of MICCAI'99*, pages 656 – 663, 1999.

[11] R. Bajcsy and S. Kovačič. Multiresolution Elastic Matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, 1989.

[12] R. Deriche. Recursively Implementing the Gaussian and Its Derivatives. In *Proc. of 2nd Int. Conf. on Image Processing*, pages 263–267, 1992.

[13] P. Hellier, C. Barillot, E. Mémin, and P. Pérez. Medical Image Registration with Robust Multigrid Techniques. In *Proc. of MICCAI'99*, pages 680 – 687, 1999.

[14] M. Ferrant *et al.* 3D Image Matching using a Finite Element Based Elastic Deformation Model. In *Proc. of MICCAI'99*, pages 202 – 209, 1999.

[15] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3), 1998.

[16] B. K. P. Horn and B. G. Schunk. Determining Optical Flow. Technical Report A.I. Memo 572, MIT Artificial Intelligence Laboratory, April 1980.

[17] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, 1983.

[18] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser, 1995.

[19] S. Ourselin, A. Roche, and X. Pennec. Automatic Alignment of Histological Sections. In *Proc. of WBIR'99*, 1999.