AN EFFICIENT MUTUAL INFORMATION OPTIMIZER FOR MULTIRESOLUTION IMAGE REGISTRATION

Philippe Thévenaz and Michael Unser
Swiss Federal Institute of Technology (EPFL)
mailto: philippe.thevenaz@epfl.ch

Abstract

We propose a new optimizer in the context of multimodal image registration. The optimized criterion is the mutual information between the images to align. This criterion requires that their joint histogram be available. For its computation, we introduce differentiable and separable Parzen windows that satisfy the partition of unity. Along with a continuous model of the images based on splines, this allows us to derive exact and tractable expressions for the gradient and the Hessian of the criterion. Then, we develop an optimizer based on the Marquardt-Levenberg strategy. Our new optimizer is specific to mutual information, in the same sense that Marquardt-Levenberg is specific to least-squares.

We show that our optimizer is particularly well-adapted to an iterative coarse-to-fine approach. We validate its accuracy by comparing its performance to that of several results available in the literature.

1. Introduction

The mutual information between two images can be regarded as a statistical tool to measure the degree to which an image can be predicted from the other. It has been introduced independently by [1] and [2] in the context of multimodal registration, along with a stochastic optimizer. In a previous ICIP paper [3], we used this registration criterion in a multiresolution context and presented an approach based on Parzen estimates of the joint histogram of both images. It was characterized by a consistent model based on splines for the computation of the resolution pyramid and of the interpolation model. Because we did no attempt to use the gradient available through the Parzen strategy, our exploratory investigations were suffering from an inefficient optimizer. In this paper, we develop a new optimizer that is well adapted to multiresolution multimodal registration. We validate our approach by applying it to a benchmark data base.

In the specific context of multiresolution, the optimizer has to be particularly efficient when the initial solution is already close from the desired solution. If this requirement is not satisfied, the multiresolution approach loses its attractiveness because one has to iterate at the finest level anyway. Few optimizers satisfy this requirement; for example, some algorithms are potentially super-linear only after they have explored in some detail the surroundings of the initial condition (*e.g.*, Fletcher [4], Powell [5]). This exploration can be costly in terms of number of function evaluations required. On the other hand, an algorithm of the Marquardt-Levenberg type enjoys no such restriction [6]; unfortunately, the class of problems that can be solved by the original formulation of Marquardt-Levenberg is limited to least-squares, even though they might be non-linear.

Mutual information is not a least-squares criterion. For this reason, we develop a new optimizer by modifying the Marquardt-Levenberg strategy to accommodate for that fact. In addition, we show that the selection of a particular type of Parzen windows allows for a tremendous reduction of the computational burden. This reduction leads to tractable forms for the gradient and the Hessian of the criterion.

The Marquardt-Levenberg approach requires that gradients be computed. In turn, those require a continuous image model. By using an image model based on splines, we satisfy these requirements and achieve the computation of exact derivatives (as opposed to only their estimates, as proposed elsewhere [1, 7]). We keep the same spline model for generating the least-squares pyramid necessary for a legitimate multiresolution implementation—this approach incorporates anti-aliasing, as opposed to subsampling. In addition, the spline model is also used for performing the geometric transformations involved in searching for the best registration parameters. This results in an algorithm that is entirely consistent, fast (because of the efficiency of our modified Marquardt-Levenberg algorithm), robust (because of the multiresolution), and accurate (because the consistent use of the continuous spline model introduces no need for approximation).

2. Image Registration

One out of many applications of image registration is to help physicians in the precise delineation of lesions or tumors, which is a necessary step in the planning of radiation therapy. In brain imaging, it is common to acquire both a computed tomography scan (CT) and a magnetic resonance image (MRI). The former typically provides precise anatomical characteristics that help in the accurate aiming of the radiation beams, while the latter is less accurate with respect to anatomy but offers much better opportunities for the delineation of tumors—which are usually invisible in the CT image. A third modality is also sometimes used: positron emission tomography images (PET) provide a measure of the local activity of the brain. In return, this last modality offers no clue to the main anatomical features; for example, the skull is indistinguishable from the background. It is nonetheless of clinical interest to align PET and MRI in order to associate structure and function.

-8186-8821-1/98 $10.00 Copyright 1998 IEEE

The combination of complementary information from separate imaging studies into a single coherent study can be prospective (in which case the radiation oncologist must plan for, and setup corresponding fiducial markers in both modalities before scanning) or retrospective (in which case the registration is based on data without the help of external fiducial markers). Because prospective registration is not applicable to data acquired without prior or coherent planning, only retrospective registration is available in the majority of situations.

Routine multimodal registration algorithms are based on markers [8], either prior or inferred from data, and surface matching (*e.g.*, chamfer registration [9], iterated closest point [10]). Since the accuracy of these routine registration algorithms cannot be better than the accuracy of the position of markers or the accuracy of the segmented surfaces, we propose a new algorithm that dispenses with these limitations. The principle is to measure, in a well-defined statistical sense, the degree to which the first image can predict the second one. Registration is achieved when the images are oriented in a way that maximizes this prediction measure.

This solution relies on recovering a direct relation between the intensities of the images, for which several similarity measures are available [11]. One way to take advantage of them is to restrict their use to regions of interest where we know a priori that they are relevant [12]. Another way is to perform registration based on a specific metric that is general enough to cope well with parts of the image where the correspondence is not apparent. Often, such measures rely on an adequate intermediate feature, namely, the joint histogram of the two images to register. Intuitively, this histogram is clustered when the images are aligned, and spread out otherwise. Several measures of clustering have been proposed [13, 14], but the most successful approach is to give up clustering in favor of another information theoretic concept: mutual information [1, 2, 7]. This latter has the interesting property that it can be applied without regard to the specific pair of modalities to register.

3. Mutual Information

In this paper, we specifically deal with the Kullback-Leibler measure of mutual information, for which an essential element is the joint histogram of the test and reference image. The test image is also sometimes called the reslice image because it undergoes a reslicing operation while the transformation is performed. By contrast, the reference image—which is sometimes called the target image—is not transformed. Thus, we expect the histogram of the transformed test image to depend slightly on the transformation parameters (because of the interpolation process), and the histogram of the reference image to be constant, exactly. The most important dependence with respect to the geometric transformation parameters is to be found in their joint histogram.

A partial-volume solution to its computation has been proposed recently [7]. Its specific advantage is that the marginal probabilities of the joint histogram of the test and reference image will never increment a bin that should otherwise stay empty. Its disadvantage is that the marginal histogram that corresponds to the histogram of the reference image depends on the transformation parameters, which is clearly unsatisfying, for indeed the reference image is never transformed.

We propose instead to compute—and minimize—the negative of the mutual information S between a test image f_T and a reference image f_R as follows:

$$S(\mu) = -\sum_{\iota \in L_T} \sum_{\kappa \in L_R} p(\iota,\kappa;\mu) \log_2 \left(\frac{p(\iota,\kappa;\mu)}{p_T(\iota;\mu)\, p_R(\kappa;\mu)} \right), \tag{1}$$

where the marginal discrete probabilities are given by

$$p_T(\iota;\mu) = \alpha(\mu)\, h_T(\iota;\mu) = \sum_{\kappa \in L_R} p(\iota,\kappa;\mu) \tag{2}$$

and

$$p_R(\kappa;\mu) = \alpha(\mu)\, h_R(\kappa;\mu) = \sum_{\iota \in L_T} p(\iota,\kappa;\mu), \tag{3}$$

and where the joint probability p is proportional to the joint histogram h by a transformation-dependent factor α

$$p(\iota,\kappa;\mu) = \alpha(\mu)\, h(\iota,\kappa;\mu). \tag{4}$$

The joint histogram itself is computed through the use of Parzen windows w

$$h(\iota,\kappa;\mu) = \sum_{\mathbf{x}_i \in V} w(\iota - f_T(\mathbf{g}(\mathbf{x}_i;\mu)))\, w(\kappa - f_R(\mathbf{x}_i)), \tag{5}$$

where $\mathbf{g}$ is a geometric transformation parametrized by μ, V is a volume of interest in which we select as many discrete coordinates $\mathbf{x}_i$ as we care to, and $\iota \in L_T$ and $\kappa \in L_R$ are discrete intensity values. Under these conditions, the normalization factor α is given by

$$\alpha(\mu) = \frac{1}{\sum_{\zeta \in L_T} \sum_{\eta \in L_R} h(\zeta,\eta;\mu)}. \tag{6}$$

The argument of the logarithm function in (1) might take a zero value. For example, when f_T and f_R are perfectly aligned and come from the same modality, we expect that $p(\iota,\kappa;\mu) = 0$ if $\iota \neq \kappa$. However, this difficulty is irrelevant because these terms can be discarded in reason of the well-known relation $\lim_{x \to 0^+} x \log(x) = 0$.

4. Parzen Window

To yield a consistent, unbiased density estimator when its width vanishes, the most important conditions that a Parzen window w must satisfy are integrability, absolute integrability, boundedness, and sufficient decay [15]. For the purpose of this paper, we impose an additional restriction by asking that w satisfies a partition of unity condition for every given real value f

$$1 = \sum_{\iota \in L} w(\iota - f) \qquad \forall f \in \mathbb{R}. \tag{7}$$

When this constraint is satisfied, it is easy to verify that the normalization factor α is independent of the transformation

$$\alpha(\mu) = \frac{1}{\sum_{\mathbf{x}_i \in V} \underbrace{\sum_{\zeta \in L_T} w(\zeta - f_T(\mathbf{g}(\mathbf{x}_i;\mu)))}_{1} \underbrace{\sum_{\eta \in L_R} w(\eta - f_R(\mathbf{x}_i))}_{1}} = \frac{1}{(\#V)}, \tag{8}$$

where $(\#V)$ represents the number of discrete coordinates $\mathbf{x}_i$ considered in the volume of interest V. In addition, the marginal discrete probability p_R of the reference image f_R is also independent[†] of μ, as shown in Expression (9)

[†] The Parzen window used in [1] is Gaussian—thus separable, like ours. However, it does not satisfy the partition of unity constraint. For the approach [1] to be consistent, we would expect the entropy of the reference image to depend on the transformation, a fact that is by-stepped in [1].

-8186-8821-1/98 $10.00 Copyright 1998 IEEE

$$\frac{\partial S(\mu)}{\partial \mu} = -\alpha(\mu)^2 \sum_{\iota}\sum_{\kappa}\left[\left(\left(\frac{1}{\alpha(\mu)}\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu}\right)-\left(h(\iota,\kappa;\mu)\sum_{\zeta}\sum_{\eta}\frac{\partial h(\zeta,\eta;\mu)}{\partial \mu}\right)\right)\log_2\left(\frac{h(\iota,\kappa;\mu)}{\alpha(\mu)h_T(\iota;\mu)h_R(\kappa;\mu)}\right)\right.$$
$$\left.+\log_2(e)\left(\left(\frac{1}{\alpha(\mu)}\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu}\right)+\left(h(\iota,\kappa;\mu)\sum_{\zeta}\sum_{\eta}\frac{\partial h(\zeta,\eta;\mu)}{\partial \mu}\right)\right)-\frac{\log_2(e)}{\alpha(\mu)}\frac{h(\iota,\kappa;\mu)}{h_T(\iota;\mu)h_R(\kappa;\mu)}\left(\frac{\partial h_T(\iota;\mu)}{\partial \mu}h_R(\kappa;\mu)+h_T(\iota;\mu)\frac{\partial h_R(\kappa;\mu)}{\partial \mu}\right)\right] \quad (10)$$

$$\frac{\partial^2 S(\mu)}{\partial \mu_1 \partial \mu_2} = -\alpha\left(\left(\sum_{\iota}\sum_{\kappa}\frac{\partial^2 h(\iota,\kappa;\mu)}{\partial \mu_1 \partial \mu_2}\log_2\left(\frac{h(\iota,\kappa;\mu)}{h_T(\iota;\mu)}\right)\right)+\left(\sum_{\iota}\sum_{\kappa}\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu_1}\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu_2}\frac{\log_2(e)}{h(\iota,\kappa;\mu)}\right)-\left(\sum_{\iota}\frac{\partial h_T(\iota;\mu)}{\partial \mu_1}\frac{\partial h_T(\iota;\mu)}{\partial \mu_2}\frac{\log_2(e)}{h_T(\iota;\mu)}\right)\right) \quad (12)$$

$$\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu} = \sum_{\mathbf{x}_i \in V} w(\kappa - f_R(\mathbf{x}_i))\left.\frac{\partial w(\xi)}{\partial \xi}\right|_{\xi=\iota-f_T(\mathbf{g}(\mathbf{x}_i;\mu))}\left(\left.\frac{-df_T(t)}{dt}\right|_{t=\mathbf{g}(\mathbf{x}_i;\mu)}\right)^T\frac{\mathbf{dg}(\mathbf{x}_i;\mu)}{\partial \mu} \quad (14)$$

$$p_R(\kappa;\mu) = \alpha(\mu)\sum_{\mathbf{x}_i \in V} w(\kappa - f_R(\mathbf{x}_i))\underbrace{\sum_{\iota \in L_T} w(\iota - f_T(\mathbf{g}(\mathbf{x}_i;\mu)))}_{1}. \quad (9)$$

We also ask that w be positive. This is obviously necessary when the width of the Parzen window is not asymptotically small, in order to forbid the histogram (5) to take negative values. Finally, we ask that w be differentiable, which will allow us to develop an optimization algorithm that takes advantage of an explicit form for the gradient of the criterion S with respect to a component of the transformation parametrized by μ. These requirements are all satisfied when we select a B-spline in the role of the Parzen window.

5. Gradient and Hessian Computation

The disadvantage of our Parzen window approach is that some entry (ι_0,κ_0) of the joint histogram might display a non-zero probability even if there is no datum corresponding to that particular entry, because the contributions will have been spread over potentially empty bins. In counterpart, the marginal histogram that corresponds to the fixed reference image does not depend on the transformation parameters. This last benefit is essential to allow the removal of dependencies between the gradient of the criterion and the gradient of the reference image.

This results in a tremendous overall simplification of the expression for the gradient of the criterion with respect to transformation parameters. We show that the lengthy gradient of the criterion S with respect to a transformation parameter μ given by (10) can be simplified into

$$\frac{\partial S(\mu)}{\partial \mu} = -\alpha\sum_{\iota}\sum_{\kappa}\frac{\partial h(\iota,\kappa;\mu)}{\partial \mu}\log_2\left(\frac{h(\iota,\kappa;\mu)}{h_T(\iota;\mu)}\right). \quad (11)$$

We emphasize here that (11) is not an approximation of (10). In fact, both are identical, provided the Parzen window used in the computation of the joint histogram satisfies the partition of unity condition. The expression for the Hessian (second derivative) of the criterion enjoys an even greater reduction in the number of terms needed. Starting from (11) and taking an additional derivative, we get (12). Had we started from (10), the number of terms would have been prohibitively large.

6. Optimizer

Marquardt and Levenberg [6] have introduced an iterative optimizer that takes advantage of both the robustness of the gradient approach and the efficiency of the Newton algorithm, while avoiding their respective drawbacks: slow convergence for the gradient, and lack of robustness for Newton. This result is achieved by combining those two solutions in an adaptive way: The Newton solution predominates when the previous iterations have been successful—thus accelerating the algorithm—while the gradient solution predominates when the previous iterations have failed to improve the criterion, at the same time reducing the step-size in the gradient direction—thus gaining in robustness. Their algorithm is restricted to the minimization of a least-squares criterion; in this case, it is not only possible, but even beneficial to simplify the Hessian matrix needed for computing the Newton contribution by ignoring second-order derivatives.

Unfortunately, the mutual information criterion (1) is not least-squares. We nevertheless follow the same strategy than Marquardt-Levenberg. In particular, we too ignore the second-order terms in (12). We justify this simplification by the following argument: when the images are perfectly aligned, and when they are totally dependent, we have that $h(\iota,\kappa) = h_T(\iota)\,h_R(\kappa)$. Then, our partition of unity condition implies that

$$\sum_{\iota}\sum_{\kappa}\frac{\partial^2 h(\iota,\kappa)}{\partial \mu_1 \partial \mu_2}\log_2\frac{h(\iota,\kappa)}{h_T(\iota)} = \sum_{\kappa}\log_2\left(h_R(\kappa)\right)\underbrace{\frac{\partial^2 h_R(\kappa)}{\partial \mu_1 \partial \mu_2}}_{0} = 0. \quad (13)$$

Since the main contribution of the Hessian is limited to iterations in the vicinity of the desired solution, second-order terms can be safely ignored. We note however that this simplification is not as beneficial than for the least-squares criterion: in the latter case, the Hessian becomes positive-definite, while in our case this is not necessarily true.

7. Image Model

The Hessian and the gradient all depend on the $\partial h/\partial \mu$ term given in (14). This term introduces the spatial gradient of the test image, which thus requires to be given through a continuous and differentiable model. In addition, the geometric transformation $\mathbf{g}$ also requires a continuous model for describing the image in-between samples. We select this model to be

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in V} c(\mathbf{x}_i)\,\beta^{(n)}(\mathbf{x} - \mathbf{x}_i), \quad (15)$$

where $\beta^{(n)}(\mathbf{x})$ is a separable convolution kernel given by a tensor-product of B-splines $\beta^{(n)}(x)$ of degree n, and where $c(\mathbf{x}_i)$ is a series of coefficients obtained by recursive digital filtering of the samples $f(\mathbf{x}_i)$ (see [16, 17]). The spatial gradient of the image is thus given by Expression (16)

-8186-8821-1/98 $10.00 Copyright 1998 IEEE

TABLE I
RESIDUAL MEDIAN DIFFERENCE IN MM BETWEEN THE PROSPECTIVE GOLD-STANDARD AND SEVERAL RETROSPECTIVE REGISTRATION ALGORITHMS. THE ALGORITHM PROPOSED IN THIS PAPER IS LABELED TH.

	BA	CO	HA	HI	MAI	MAL	NO	PE	RO3	RO4	TH
CT-T1	1.6	1.5	3.4	*1.2*	5.1	4.3	3.3	2.7	5.7	5.4	**1.5**
CT-PD	1.9	*1.5*	3.1	1.9	4.1	4.0	7.8	1.9	4.9	4.8	**1.7**
CT-T2	2.5	1.5	4.2	1.5	3.9	5.0	3.9	2.5	5.4	4.7	***1.1***
CT-T1r	1.4	*0.7*	3.3	*0.7*	4.9	5.4	3.4	2.2	6.3	5.9	**0.9**
CT-PDr	1.7	0.8	3.0	*0.7*	3.0	4.0	4.6	2.1	5.5	5.5	**1.0**
CT-T2r	2.1	*0.8*	3.5	*0.8*	4.3	5.3	4.2	2.9	5.3	5.3	**0.9**
PET-T1	4.6	3.6	*2.8*	3.2	3.5	4.2	3.6	2.9	4.0	3.4	**3.0**
PET-PD	5.2	2.9	4.2	3.1	4.7	4.0	4.1	3.3	4.3	3.3	***2.7***
PET-T2	4.7	2.8	2.7	*2.4*	5.3	4.9	4.6	3.3	4.0	3.6	**2.6**
PET-T1r	3.2	2.8	3.6	2.5	3.9	3.6	3.9	2.8	3.8	3.6	***1.9***
PET-PDr	4.5	3.0	3.2	3.0	4.7	3.6	4.4	2.8	3.6	4.1	***2.0***
PET-T2r	3.9	2.0	3.3	2.2	4.0	3.6	5.2	2.9	3.8	3.4	***1.9***

$$\frac{\partial f(\mathbf{x})}{\partial x} = \sum_{\mathbf{x}_i \in V} c(\mathbf{x}_i) \frac{\partial \beta^{(n)}(\mathbf{x}-\mathbf{x}_i)}{\partial x}. \quad (16)$$

This gradient is exact provided $\beta^{(n)}(x)$ is differentiable, which is true for all arguments x as soon as $n>1$ (this excludes linear interpolation). In practice, we select cubic splines ($n=3$) because this model yields better performance than quadratic splines, while their computational cost is essentially the same. For the same reasons, we also select a cubic B-spline in the role of the Parzen window w.

8. Multiresolution

The model proposed in Section 7 allows the computation of an image pyramid that is optimal in a least-squares sense [18]. We use this pyramid for performing optimization in a coarse-to-fine strategy. Registration is first performed at the coarsest scale, where convergence is fast because there are few data, which largely compensates for a potentially higher number of iterations needed to reach convergence (the initial condition at the coarsest scale is arbitrary). Moreover, it is likely that the criterion to optimize has a reduced number of local optima; this is due to a loss of image details and results in enhanced robustness. In return, the solution is only approximate.

Registration is then performed on a finer scale with the result of the previous scale as initial condition. This process is iterated until the finest scale is reached. Especially in 3D, there is a great difference in the amount of data between pyramid levels. This means that only the number of iterations performed at the finest level is relevant for the computational cost of the whole optimization. Thus, it is very important that the initial condition for this last level be the best possible in order to reduce the amount of refinement necessary to reach convergence. This implies that intermediate pyramid levels are relevant in the optimization process.

A coarse-to-fine strategy is successful when the optimizer takes particular advantage from good starting conditions. Very few optimizers satisfy this constraint. Some claim super-linear convergence (*e.g.*, Powell, Fletcher, Polak), but reach it only after at least N criterion evaluations have been performed, where N is the number of dimensions of the parameter μ. By contrast, the optimizer we proposed in Section 6 is able to converge in just one step when the conditions are favorable.

One could naively assume that the quality of the coarse levels is only moderately relevant because, after all, the final accuracy depends on the finest level only. Along these lines, a simple pyramid (*e.g.*, subsampling) and a simple interpolation scheme (*e.g.*, linear) would seem good enough to prepare the initial condition for optimizing at the finest level. Reality is dramatically different, though. To get optimal starting conditions, it is crucial that the coarse levels of the pyramid most faithfully represent the finest level. In addition, only the use of high-quality interpolation models will ensure convergence to an accurate solution. Since the final iterations are so costly, it is very advantageous to pay special attention to the work performed at coarser resolution, because every spared fine-scale iteration is worth many coarser-scale ones. Moreover, robustness depends almost exclusively on high-accuracy registration and faithfulness at the coarse levels.

9. Experiments

We applied our algorithm to the registration of volumes acquired by CT (Computed Tomography) and three MRI (Magnetic Resonance Imaging) modalities: PD (Proton Density), T1 (T1 Relaxation Time) and T2 (T2 Relaxation Time). The goal was to align the CT volumes with the MRI ones, which represent two very different measurements since the former use X-rays while the latter deal with the interaction between spins and magnetic field. The MRI volumes were available in two versions: raw (PD, T1, T2), and corrected (rectified) for geometric distortion (PDr, T1r, T2r). There were seven patients in each case.

We compare the results of our multimodal brain image registration algorithm to those of several other approaches published in the literature. The comparison is based on a methodology proposed by West and Fitzpatrick [19], who let selected researchers access a standard set of volumes to be registered. They also act as a repository for the ideal registration transformations (gold-standard) acquired by a prospective method using physical markers. These markers are erased before the volumes are disclosed to the investigators, who then face a retrospective blind registration task. After registration,

-8186-8821-1/98 $10.00 Copyright 1998 IEEE

they report back a set of transformation parameters that are compared to the gold-standard. This results in a geometric error measured in mm, and allows for a simple ranking of the competing algorithms—from an accuracy point of view.

Table I shows in mm the median accuracy reached by the investigators taking part in that study (for a complete discussion of the experimental protocol, see [19]). We observe that our algorithm shows very promising results, given that it is entirely automatic (which is not true for several of the competing algorithms), and given that it performed its task without being specifically tuned to any pair of modalities.

10. Conclusions

We have developed a multimodal registration algorithm that takes advantage of a consistent image model based on cubic splines. We have advocated the use of a coarse-to-fine optimization strategy, and explained why the quality of the model is particularly important at coarse resolutions for speed issues, robustness and accuracy.

The mutual information between two images is our registration criterion, which requires the computation of their joint histogram. We have proposed to perform this computation through Parzen windows that satisfy the partition of unity condition. This allowed the development of a tractable expression for the gradient and the Hessian, and led to the development of an optimizer based on the Marquardt-Levenberg strategy. We justified the use of a simplified form for the Hessian.

We have validated our algorithm with biomedical volumetric brain images and compared its performance to that of several other approaches available in the literature. In many instances, our algorithm shows a better accuracy, thanks to the consistency of our approach.

Acknowledgments

The images and the gold-standard transformations of Section 9 were provided as part of the project, "Evaluation of Retrospective Image Registration", National Institutes of Health, 1 R01 NS33926-02, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN., U.S.A.

Bibliography

[1] P. Viola and W.M. Wells III, "Alignment by Maximization of Mutual Information," in Proc. *International Conference on Computer Vision*, Boston, MA, USA, June, 1995, pp. 16–23.

[2] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens and G. Marchal, "Automated Multi-Modality Image Registration Based on Information Theory," in *Information Processing in Medical Imaging*, Y.J.C. Bizais, Ed., pp. 263–274, 1995.

[3] P. Thévenaz and M. Unser, "A Pyramid Approach to Sub-Pixel Image Fusion Based on Mutual Information," in Proc. *IEEE International Conference on Image Processing*, Lausanne, Switzerland, September 16-19, 1996, vol. I, of 3, pp. 265–268.

[4] R. Fletcher and C.M. Reeves, "Function Minimization by Conjugate Gradients," *The Computer Journal*, vol. 7, pp. 149–154, July 1964.

[5] M.J.D. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," *The Computer Journal*, vol. 7, pp. 155–162, July 1964.

[6] D.W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, June 1963.

[7] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, April 1997.

[8] L.R. Schad, R. Boesecke, W. Schlegel, G.H. Hartmann, V. Sturm, L.G. Strauss and W.J. Lorenz, "Three Dimensional Image Correlation of CT, MR and PET Studies in Radiotherapy Treatment Planning of Brain Tumors," *J. Cat.*, vol. 11, pp. 948–954, 1987.

[9] H. Jiang, R.A. Robb and K.S. Holton, "A New Approach to 3-D Registration of Multimodality Medical Images by Surface Matching," in Proc. *Visualization in Biomedical Computing*, 1992, vol. Proc. SPIE 1808, pp. 196–213.

[10] L.G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, December 1992.

[11] D.L.G. Hill, C. Studholme and D.J. Hawkes, "Voxel Similarity Measures for Automated Image Registration," in Proc. *SPIE Proceedings of the Third Conference on Visualization in Biomedical Computing*, 1994, pp. 205–216.

[12] P. Gerlot-Chiron and Y.J.C. Bizais, "Registration of Multimodality Medical Images Using a Region Overlap Criterion," *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing*, vol. 54, no. 5, pp. 396–406, September 1992.

[13] A. Collignon, D. Vandermeulen, P. Suetens and G. Marchal, "3D Multi-Modality Medical Image Registration Using Feature Space Clustering," in Proc. *Computer Vision, Virtual Reality, and Robotics in Medicine*, Nice, France, April, 1995, pp. 195–204.

[14] D.L.G. Hill, D.J. Hawkes, N.A. Harrison and C.F. Ruff, "A Strategy for Automated Multimodality Image Registration Incorporating Anatomical Knowledge and Imager Characteristics," in *Information Processing in Medical Imaging*, Springer-Verlag, pp. 182–196, 1993.

[15] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, San Francisco, London: Academic Press, 1972.

[16] M. Unser, A. Aldroubi and M. Eden, "B-Spline Signal Processing: Part II—Efficient Design and Applications," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 834–848, February 1993.

[17] M. Unser, A. Aldroubi and M. Eden, "B-Spline Signal Processing: Part I—Theory," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–832, February 1993.

[18] M. Unser, A. Aldroubi and M. Eden, "The L2 Polynomial Spline Pyramid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 364–379, April 1993.

[19] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer, R.M. Kessler, R.J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P.A. van den Elsen, S. Napel, T.S. Sumanaweera, B. Harkness, P.F. Hemler, D.L.G. Hill, D.J. Hawkes, C. Studholme, J.B.A. Maintz, M.A. Viergever, G. Malandin, X. Pennec, M.E. Noz, G.Q. Maguire Jr., M. Pollack, C.A. Pelizzari, R.A. Robb, D. Hanson and R.P. Woods, "Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques," *Journal of Computed Assisted Tomography*, vol. 21, no. 4, pp. 554–566, 1997.

-8186-8821-1/98 $10.00 Copyright 1998 IEEE