

# Affine Registration with Feature Space Mutual Information

Torsten Butz and Jean-Philippe Thiran

Swiss Federal Institute of Technology (EPFL), Signal Processing Laboratory (LTS),  
CH-1015 Lausanne, Switzerland  
{torsten.butz, jp.thiran}@epfl.ch,  
<http://ltswww.epfl.ch/~brain>, FAX: +41-21-693-7600

**Abstract.** This paper introduces two important issues of image registration. At first we want to recall the very general definition of mutual information that allows the choice of various feature spaces to perform image registration. Second we discuss the problem of finding the global maximum in an arbitrary feature space. We used a very general parallel, distributed memory, genetic optimization which turned out to be very robust. We restrict the examples to the context of multi-modal medical image registration but we want to point out that the approach is very general and therefore applicable to a wide range of other applications. The registration algorithm was analysed on a LINUX cluster.

## 1 Introduction

In the last years mutual information (MI) has had a large impact on multi-modal signal processing in general and on medical image processing in particular. Since the initial work of Viola et al. [1] and Maes et al. [2] several groups have analysed the modification of the optimization objective MI itself. The work is so rich, that a comprehensive overview would over-charge this paper. We still want to mention the work of Studholme et al. [3] who introduced normalized mutual information to rigidly register multi-modal images with different fields of view, Plum et al. [4] who added a gradient-based term to the MI in order to decrease the number of local maxima and Rueckert et al. [5] who used second order entropy estimation to model the dependency of a voxel's gray value on the intensities of a local neighborhood around that voxel.

On the other hand, several papers discuss image registration with image features. Maintz et al. [6] studied different ridge detectors and used correlation for multi-modal image registration. In [7], Rangarajan et al. discussed feature point registration with MI. This direct combination of image features with MI is very interesting as it fuses two approaches that have mostly been treated independently in the literature. Our own work continues along this line and merges most of the approaches mentioned above in one single formalism. For this we simply used the fact that MI is defined on an arbitrary probability space, i.e. an arbitrary feature space (Figure 1).

Unfortunately the behavior of MI in an arbitrary feature space is very hard to predict. As a consequence, most optimization schemes would get stuck in local maxima. Viola et al. [1] proposed stochastic gradient descent in order to avoid local maxima. Nevertheless this approach avoids mainly maxima that are related to imaging noise but is still just locally convergent. For example the symmetry in brain images can cause stochastic gradient descent to fail (e.g. front/back and left/right symmetry). This might be a minor problem in the context of brain images, as a rough interactive registration can easily be performed in order to avoid these local optima. When passing to more complex feature spaces than the widely used intensity space, the number and positions of local maxima may make this approach impossible.

Therefore we used genetic optimization for maximization of MI in a general feature space. We propose a multi-scale genetic optimization and implemented a master/slave parallelization scheme [8]. The genetic optimization itself is based on the open source library written by Matthew Wall [9]. We added the communication utilities for distributed memory architectures using the MPICH implementation of the “Message Passing Interface” (MPI) [10], [11].

## 2 Methods

### 2.1 Feature Space Mutual Information

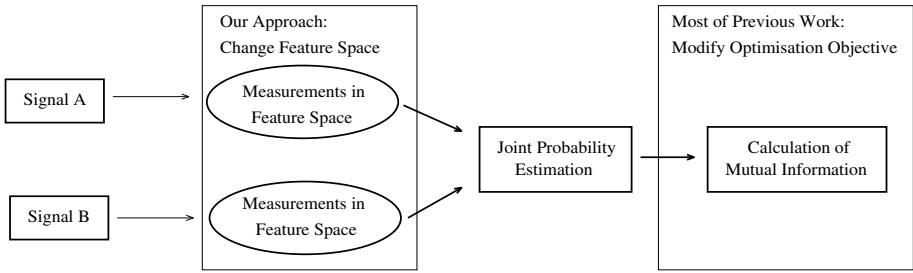
MI is a widely used information theoretical distance measure between probability densities [12]. Let’s shortly recall its definition: Let  $X$  and  $Y$  be two random variables with marginal probability distributions  $p(x)$  and  $p(y)$ , and joint probability distribution  $p(x, y)$ , then the MI between  $X$  and  $Y$  is:

$$I(X; Y) = H(Y) - H(Y|X) \quad (1)$$

$$\text{(evt. discretisation)} = \sum_{x,y} p(x, y) \cdot \log\left(\frac{p(x, y)}{p(x) \cdot p(y)}\right) \quad (2)$$

where  $H(\cdot)$  stands for the Shannon-Wiener entropy of a continuous or discrete random variable.  $X$  and  $Y$  are arbitrary random variables and can therefore stand for discrete, continuous, single-variate or multi-variate variables. In order to calculate the MI between two signals, a representative observation of the signal has to be taken (choice of an adequate feature space). Then a probability estimation will create a probabilistic model of the signal from the observation. The sampling space is therefore arbitrary, but should be chosen to model the signal as accurately as possible. This is summarized in figure 1.

In [1], the measured features are simply the voxel intensities, in [5] the sampling space is defined by the intensities of two neighbored voxels (i.e. a two-dimensional feature space) and in [7] it is defined by the voxel being or not being a particular feature point (i.e. a discrete feature space). For the latter, the maximum entropy principle is used for density estimation [13].



**Fig. 1.** We show the general pipeline to estimate the MI of two signals. Instead of modifying the MI itself, we analyse the choice of different sampling or feature spaces.

In this paper, we restrict the discussion to a feature space that can be represented by continuous variables, called edgeness. We never conclude from this measurement on whether a specific voxel is an edge or not. This fact allows us to use joint-histogramming to estimate the joint probability of the two images. The approach is very general and can easily be extended to other image features or to the combination of different feature types.

### 2.2 Edgeness Measure

We have tested two different edgeness measures. The first is simply the norm of the gradients. The second is slightly more complex, as it considers the intensity variance within a variable distance from the voxel. Let’s call  $d$  a fixed radius and  $\mathbf{d}_0$  the coordinate vector of a voxel. It’s edgeness is defined as:

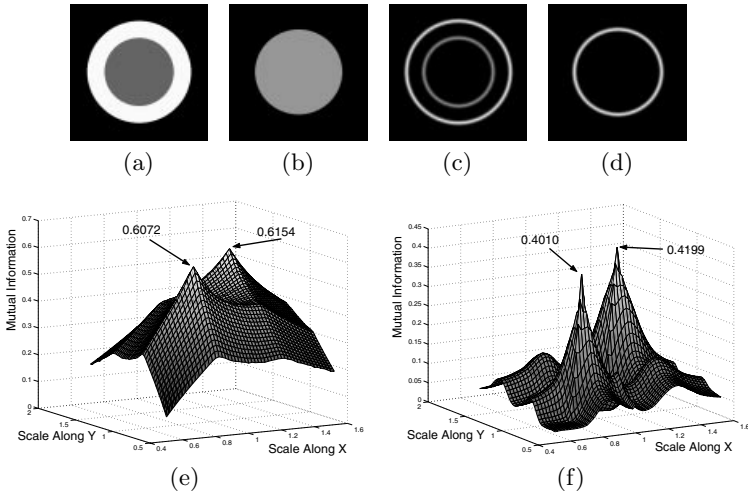
$$c(\mathbf{d}_0) = \sum_{|\mathbf{d}_i - \mathbf{d}_0| < d} |g(\mathbf{d}_i) - g(\mathbf{d}_0)| \tag{3}$$

where  $g(\cdot)$  stands for the image intensities. Figures 2 a) through d) show an output from this operator. Other edgeness operators are possible.

### 2.3 Parallel Genetic Optimization

So far, the presented approach is very similar to the well known maximization of MI, except that we recalled that MI is by far not restricted to the intensity space. The problem with other spaces is the presence of local maxima which will cause local optimization algorithms to fail. Let’s underline the fact that stochastic gradient descent does not solve this problem, as the local maxima are effectively present in the feature space MI and are not due to imaging noise. Figure 2 should clarify the problem.

In order to find the global maximum, we have to use a globally convergent algorithm. We employed Matthew Wall’s genetic algorithm library [9], but won’t discuss genetic optimization in this text. [14] is a good introductory reference.



**Fig. 2.** To show the problem of feature space MI for affine registration, we present two template images of concentric circles (a and b) as well as their edginess images (c and d) defined by equation 3 respectively. We see that both, the intensity based MI (e) as well as the edginess MI (f) have local maxima. But the edginess based measure has more and very pronounced local maxima, which underlines the necessity of a globally convergent optimization scheme.

We will neither discuss the parallelization scheme which we added on top of the genetic optimization. Let’s just mention that it’s known as a master/slave parallelization [8] with a SPMD (Single Program Multiple Data) model. The parallelization is independent of the objective function to be optimized and can therefore be used for any chosen feature space. For further speed-up we used a multi-scale genetic optimization.

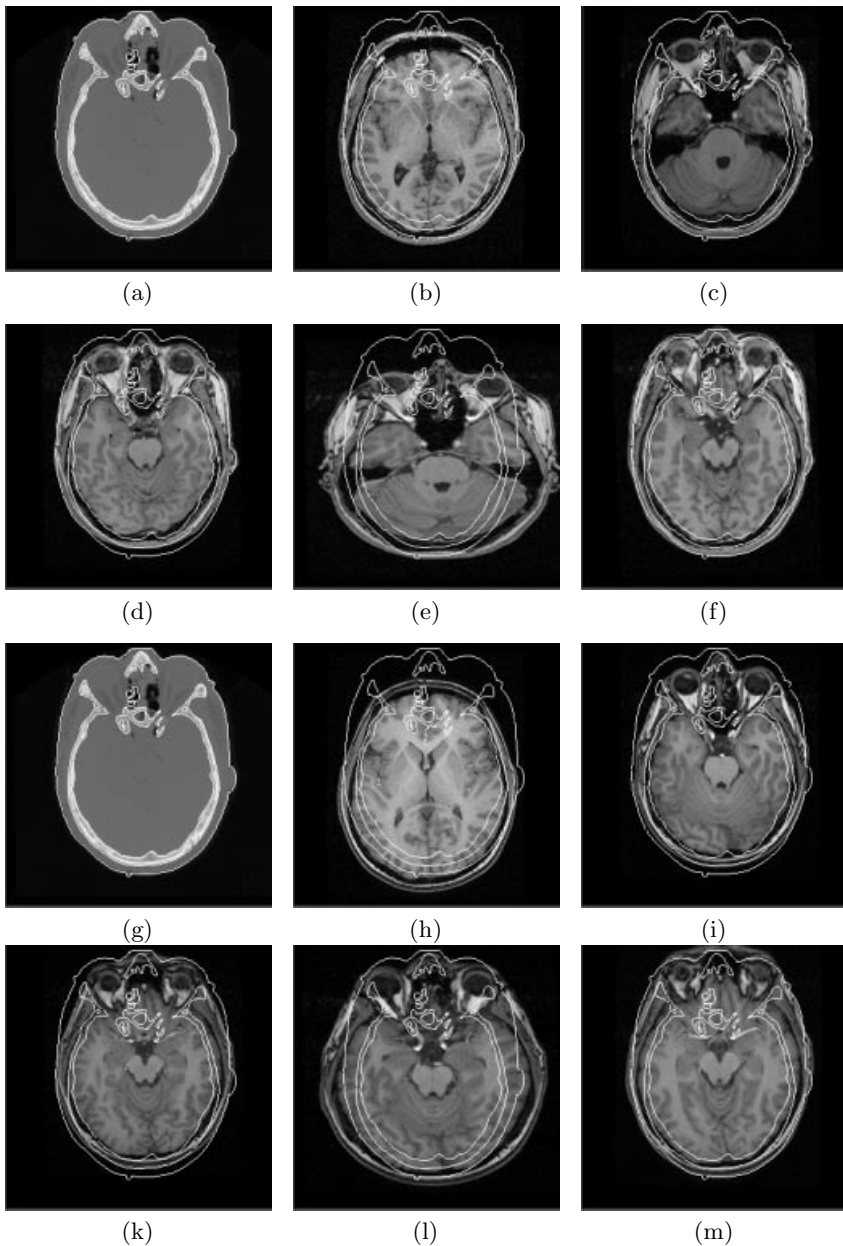
### 3 Results

#### 3.1 Affine, Multi-modal, Inter-patient Image Registration

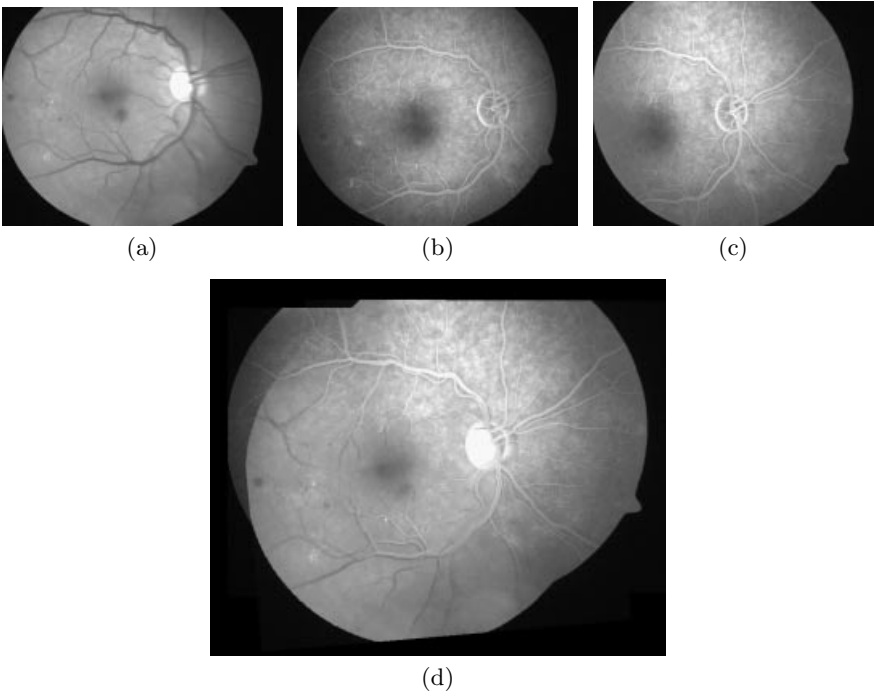
Multi-modal affine registration is a very important task for image registration as it gives a good initialization for several non-rigid registration algorithms [15], [16]. In this study, we registered MR-scans onto a CT reference image of another patient (inter-patient) and compared the gradient based MI with the intensity based MI. The results for two MR-scans are shown in figure 3. Their interpretation, in particular of the images e/f) and l/m) resp., is presented in 4.

#### 3.2 Angiograms of the Retinal Blood Vessels

In this study, we show that the presented feature space MI combined with the globally convergent genetic optimization can extend the MI based image registration to other medical applications.



**Fig. 3.** a) is the CT-target image. In b), the contours of the target image are superposed on the floating MR-scan. In c) and d) we see the results after a rigid optimization, when using resp. the intensity based MI and the edginess MI. In e) and f) we show the corresponding results for affine registration. Figures g) through m) show the results for a second MR-scan. In e) and f) (resp. l) and m)) we recognize a significant improvement with the edginess based MI; resp. that the global maximum of intensity based MI doesn't correspond to good registration.



**Fig. 4.** The figures a), b) and c) had first to be registered in order to reconstruct the extended view shown in d).

In figure 4 we show how three partial views of the retinal vascular system can be combined to provide a virtually extended view. The different intensity distributions in the images are caused by an injected contrast agent which enables the study of the retinal blood flow for diabetic retinopathy. The feature space of choice was the edgeness defined in eq. 3 with an atom radius  $d$  of 5 pixels.

### 3.3 Genetic Optimization

We ran the parallel, genetic optimization algorithm on a LINUX cluster (10 bi-processor PCs: 2x PIII 550MHz each. Memory: 7x 512MByte, 2x 256MByte and 1x 384MByte). For the results of section 3.1, the execution time for all 20 processors was about 30 minutes (image size: 256x256x123) and for the results of section 3.2 about 90 seconds (image size: 451x367) wall clock time.

## 4 Discussion

In section 3.1 we have compared the gradient based MI with intensity based MI. For rigid registration the quality of the results is comparable, while for affine transformations the global maximum of intensity based MI might not correspond

to good registration and the presented feature space defines a much better result. The edginess defined by the gradient emphasizes contours in the medical images while the intensity based MI over-emphasizes the volumetric information in the scans and therefore risks to neglect finer but important features in the images. An example is the skull and the brain: The brain covers lots of volume while the human skull is a relatively fine but anatomically important structure. Therefore the intensity based registration favors the statistical matching of the brain. On the other hand the gradient based MI reflects the statistical presence of surfaces. As a result, the skull and the brain have about the same importance and a compromise for their fitting is obtained. Figure 3, in particular the images e/f) and l/m) resp., shows a significant improvement with this approach.

In section 3.2 we have shown an additional application of image registration where the chosen feature space has been adapted to the image contents. The example also underlines the importance of a globally convergent optimization scheme as the fine blood vessels in the retina give rise to local optima due to "accidental" partial matching of non-corresponding short vessel segments. The global optimization forces convergence to maximum matching and therefore towards correct registration of the entire vessels and not just vessel segments.

Finally the timing results showed that the genetic optimization can be competitive when a parallel implementation is used. From a practical point of view, optimal results can be obtained when a local optimization refines the output of the genetic optimization.

## 5 Conclusion

This paper shows that specifically designed feature space MI out-performs the widely used intensity based MI optimization for several medical registration tasks. The mathematical expression of MI itself doesn't constrain the choice of feature spaces and therefore incorporates the described approach.

The described drawback of local maxima in the optimization objective can be solved by globally convergent genetic optimization. A parallel implementation provides a powerful algorithm for a wide range of optimization tasks.

Future work will study and compare additional feature spaces (e.g. multi-dimensional spaces) and its applications to other registration tasks. It's important to note that the curse of dimensionality in multi-variate density estimation [17] limits the maximum possible dimension of the chosen feature space.

## 6 Acknowledgements

We want to thank Conor Heneghan, Ph.D., (Digital Signal Processing Laboratory, University College Dublin) for providing the fundus images of section 3.2.

## References

- [1] W.M. Wells III, P. Viola, H. Atsumi S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, March 1996.
- [2] 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.
- [3] C. Studholme, D.J. Hawkes, and D.L.G. Hill, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, pp. 71–86, 1999.
- [4] Josien P.W. Pluim, J.B. Antoine Maintz, and Max A. Viergever, "Image registration by maximization of combined mutual information and gradient information," October 2000, vol. 1935, pp. 452–461.
- [5] D. Rueckert, M.J. Clarkson, D.L.G. Hill, and D.J. Hawkes, "Non-rigid registration using higher-order mutual information," in *Proceedings of SPIE 2000*, February 2000, pp. 438–447.
- [6] J.B. Maintz, Petra A. van den Elsen, and Max A. Viergever, "Evaluation of ridge seeking operators for multimodality medical image matching," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 353–365, April 1996.
- [7] Anand Rangarajan, Haili Chui, and James S. Duncan, "Rigid point feature registration using mutual information," *Medical Image Analysis*, vol. 3, no. 4, pp. 425–440, 1999.
- [8] Erick Cantú-Paz, "A survey of parallel genetic algorithms," Tech. Rep., The University of Illinois, 1997, IlliGAL Report No. 97003, <ftp://ftp-illigal.ge.uiuc.edu/pub/papers/IlliGALs/97003.ps.Z>.
- [9] Matthew Wall, *GAlib 2.4.5: A C++ Library of Genetic Algorithm Components*, Massachusetts Institute of Technology, <http://lancet.mit.edu/ga/>.
- [10] W. Gropp and E. Lusk, *User's Guide for mpich, a Portable Implementation of MPI Version 1.2.1*, <http://www-unix.mcs.anl.gov/mpi/mpich/>.
- [11] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming With the Message-Passing Interface*, MIT Press, second edition, 1999.
- [12] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
- [13] E.T. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939–952, 1982.
- [14] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Company, Inc., 1989.
- [15] A. Guimond, A. Roche, A. Ayache, and J. Meunier, "Multimodal brain warping using the demons algorithm and adaptative intensity corrections," Tech. Rep., Inst. National de Recherche en Informatique et en Automatique, Sophia Antipolis, 1999.
- [16] Jean-Philippe Thiran and Torsten Butz, "Fast non-rigid registration and model-based segmentation of 3d images using mutual information," in *Medical Imaging*, 2000, pp. 1504–1515.
- [17] L. Devroye, *A Course in Density Estimation*, Birkhäuser, 1987.