

# Towards Application-specific Multi-modal Similarity Measures: a Regression Approach.\*

Olivier Pauly<sup>1</sup>, Nicolas Padoy<sup>1,2</sup>, Holger Poppert<sup>3</sup>, Lorena Esposito<sup>3</sup>, Hans-Henning Eckstein<sup>4</sup> and Nassir Navab<sup>1</sup>

<sup>1</sup> Computed Assisted Medical Procedures, Technische Universität München, Germany

<sup>2</sup> LORIA-INRIA Lorraine, France

<sup>3</sup> Neuro-Kopf Zentrum, Klinikum Rechts der Isar, Germany

<sup>4</sup> Klinik für Gefäßchirurgie, Klinikum Rechts der Isar, Germany  
{*pauly, padoy, navab*}@cs.tum.edu

**Abstract.** In multi-modal registration, similarity measures based on intensity statistics are the current standard for aligning medical images acquired with different imaging systems. In fact, the statistical relationship relating the intensities of two multi-modal images is constrained by the application, defined in terms of anatomy and imaging modalities. In this paper, we present the benefits of exploiting application-specific prior information contained in one *single* pair of registered images. By varying the relative transformation parameters of registered images around the ground truth position, we explore the manifold described by their joint intensity distributions. An adapted measure is fitted using support vector regression on the training set formed by points on the manifold and their respective geometric errors. Experiments are conducted on two different pairs of modalities, MR-T1/MR-TOF and MR-T1/SPECT. We compare the results with those obtained using mutual information and Kullback-Leibler distance. Experimental results show that the proposed method presents a promising alternative for multi-modal registration.

## 1 Introduction

Image registration is a crucial processing step in all image analysis tasks in which information from various imaging sources needs to be combined. Establishing correspondences between images acquired with different medical imaging modalities is a challenging task known as multi-modal registration. Objective functions that evaluate the quality of alignment, known as similarity measures, are optimized to identify the geometric transformation that maps the coordinate system of one modality to the other [8]. The choice of the appropriate measure is not straightforward, because it implicitly models the relationship between the different images to register [4]. Classical measures such as sum of square differences (SSD) or correlation coefficient (CC) make the assumption of a linear functional mapping between the intensities of the images to align. But this hypothesis is far from being realistic according to the physics of different

---

\* This research is partially funded by the Munich Center for Advanced Photonics and by the Gefäßchirurgie Department, Klinikum Rechts der Isar, Germany

imaging systems. Modeling the real relationship between different imaging modalities is very difficult and this explains why statistical measures have become more and more popular. Since its introduction by Viola and Wells [6] and Collignon et al [2], mutual information remains the state of the art of multi-modal registration of medical images.

Even though the statistics relating intensities of two multi-modal images is modality-specific, there were only few attempts to incorporate prior knowledge in such similarity measures. Chung et al. [1] proposed to use as prior information a reference joint probability distribution of registered images from different modalities. Images are then aligned by minimizing the Kullback-Leibler distance between an observed and the expected joint histogram. Leventon et al. [3] compared two methods to model this reference histogram from a training set of registered images, namely a mixture of Gaussians and Parzen windowing. The distance to this expected histogram is then estimated by using log likelihood. In these works however, the use of prior information remains limited to one reference joint distribution.

Zhou et al.[7] propose an approach based on Adaboost to learn local similarity measures for anatomic landmarks detection in echocardiatic images. It uses an atlas of the left ventricle containing pairs of local patches with their relative displacements. In a mono-modal scenario, the method shows that incorporating prior information can improve the detection results. This approach requires however extensive initial supervision.

Joint histograms of multi-modal images warped with different relative transformations describe a manifold embedded in the joint distribution space. Our contribution is to define a similarity measure relating the topology of such manifolds to the registration error. This yields an application-specific similarity measure, which requires one single pair of registered images as prior information. Using a set of relative transformations between the two images, we generate a training set of data points from the corresponding joint histograms and their associated geometric error values defined in section 2. The similarity measure is then learned by performing a support vector regression on this data.

The remainder of the paper is organized as follows: Section 2 presents our regression approach to define an application-specific similarity measure. Section 3 reports experiments performed in two different and challenging applications in comparison to classical methods such as mutual information and Kullback-Leibler distance. Results show that our approach presents a promising alternative for multi-modal registration. Section 4 concludes the paper and gives an outlook on future work.

## 2 Methods

### 2.1 Problem statement

The goal of multi-modal image registration is to identify the geometric transformation that maps the coordinate system of one modality to the other. Let us consider two 2D images defined on the domains  $\Omega_1$  and  $\Omega_2$  with intensity functions  $I_1 : \Omega_1 \subset R^2 \rightarrow R$  and  $I_2 : \Omega_2 \subset R^2 \rightarrow R$ . The two dimensional case is discussed for better readability, the extension to three dimensions being straightforward. The registration task can be defined as a maximization problem, in which we want to estimate the best transformation  $T$  according to a chosen similarity measure  $S$  computed on the discrete overlap

domain  $\Omega = \Omega_1 \cap T(\Omega_2)$ :

$$T = \operatorname{argmax}_T S_\Omega(I_1, T(I_2)). \quad (1)$$

The joint intensity distribution  $p(I_1, I_2)$  of both images can be evaluated by histogramming or parzen windowing. In most of statistical measures, the similarity  $S_\Omega$  is a mapping from the joint distribution space  $\mathcal{J}$  into  $\mathbb{R}$ . While Mutual Information (MI) gives a measure of the distance between the joint histogram of both images and what it would be if their intensity distributions were independent, the Kullback-Leibler distance (KL) [1] evaluates the distance between an observed  $p_o$  and an expected  $p_e$  joint histogram:

$$MI(I_1, I_2) = D(p(I_1, I_2) || p(I_1)p(I_2)) \quad (2)$$

$$KL(I_1, I_2) = D(p_o(I_1, I_2) || p_e(I_1, I_2)), \quad (3)$$

where  $D$  in its general form is defined on two histograms  $p$  and  $q$  as:

$$D(p||q) = \sum_x p(x) \log \left( \frac{p(x)}{q(x)} \right). \quad (4)$$

The statistical relationship relating the intensities of two different multi-modal images is constrained by the application. With “application”, we mean the combination of the modalities to relate and the different tissues appearing in the imaged anatomy, e.g. blood, bones or muscles. Joint histograms of images warped with different relative transformations describe a manifold  $\mathcal{M}$  embedded in  $\mathcal{J}$  which is application-specific. In [1], Chung et al. makes use of one expected joint histogram, corresponding to one single reference point on such a manifold. The used Kullback-Leibler divergence is however not adapted to its topology.

Instead, we propose to model an application-specific similarity  $\Psi$  taking into account how the topology of  $\mathcal{M}$  relates to the registration error. By using a set of relative geometric transformations  $\{T_i\}_{i \in \mathbb{N}}$  between a source and a target image, we sample  $\mathcal{M}$  by the joint histograms  $J_{I_1, T_i(I_2)}$ . Each of these “points” is then associated to a geometric error derived from the corresponding transformation parameters, generating thereby a set of data points. Finally, the similarity  $\Psi$  is defined by performing a regression on these points. The following section presents how to generate data points to relate this manifold  $\mathcal{M}$  to the geometric error.

## 2.2 Data points generation

Our objective is to model a similarity  $\Psi$  learned on the full manifold  $\mathcal{M}$ :

$$\Psi : \mathcal{M} \rightarrow \mathbb{R}, \quad (5)$$

which has favorable characteristics for registration purposes, namely convexity, smoothness and the ability to estimate the geometric error. To model an accurate mapping  $\Psi$ ,

the manifold  $\mathcal{M}$  must be sampled thoroughly as a function of the transformation  $T$ , whose space is parameterized as follows:

$$T(t_x, t_y, \theta) \text{ where } \begin{cases} t_x \in [-M, +M] \\ t_y \in [-N, +N] \\ \theta \in [-\phi, +\phi] \end{cases} \quad (6)$$

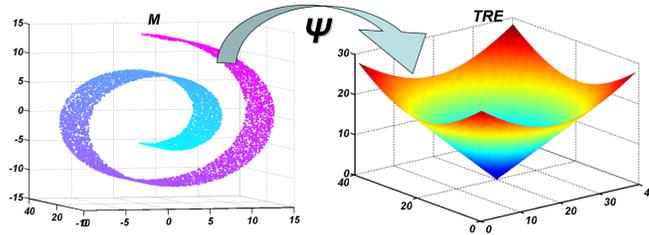
By sampling the space of transformations, a set  $\{T_i\}_{1 \leq i \leq Q}$  of  $Q$  transformations is generated. Then, by using a pair of registered images from different modalities, joint histograms are computed according to these  $\{T_i\}_{1 \leq i \leq Q}$ . As illustrated by Fig. 1, each joint histogram is then associated to a geometric error value. In medical image registration, the *target registration error* (TRE) permits the evaluation of error in translation and orientation between corresponding structures or organs appearing in both modalities. The TRE is computed by comparing the positions of a set of points  $\{p_i, 1 \leq i \leq P\}$  after being mapped by the estimated transformation  $T$  and by the ground truth transform  $G$ :

$$\mathcal{E}(T) = \frac{1}{P} \sum_{i=1}^{i=P} \|T(p_i) - G(p_i)\|. \quad (7)$$

This procedure permits us to generate following couples:

$$\{(J_{I_1, T_i(I_2)}, \mathcal{E}(T_i))\}_{1 \leq i \leq Q}, \quad (8)$$

which we denote  $\{(J_i, \mathcal{E}_i)\}_{1 \leq i \leq Q}$  for better readability.



**Fig. 1.** Our regression approach: learn a similarity  $\Psi$  mapping each point of the manifold  $\mathcal{M}$  (abstract representation on the left) to a value of the geometric error (on the right).

### 2.3 Fitting the similarity model through support vector regression

We propose to learn the similarity by approximating the function  $\Psi$  with the previously generated data points. Since this function is a high dimensional non-linear mapping, we use support vector regression for its ability of modeling complex non-linear functions. We consider the problem of fitting a similarity function on the set of  $Q$  data points  $\{(J_i, \mathcal{E}_i)\}_{1 \leq i \leq Q}$ . The  $\{J_i\}$ , as discrete approximations of the joint intensity distributions, consist of  $B \times B$  bins. They are linearized into a vector of dimensionality  $B^2$ . Let  $\varphi$  be a non-linear mapping from  $\mathcal{M}$  into a hidden feature space  $H$  with dimensionality  $\dim(H) > B^2$  used to model non-linear relationships between joint histograms

and their corresponding geometric error values. The mapping  $\Psi$  is modeled by the following function:

$$\Psi(J) = w \cdot \varphi(J) + b, \quad (9)$$

where  $w$  is a linear separator of dimensionality  $\dim(H)$  and  $b$  a bias. The optimal regression function is then given by the minimum of the following functional [5]:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^Q (\xi_i^+ + \xi_i^-), \quad (10)$$

where  $C$  controls the flexibility of the model. This functional aims at minimizing the norm of  $w$  and the regression errors on the data points, characterized by the slack variables  $\xi_i^+$  and  $\xi_i^-$ . The optimal vector  $w_0$  can be written as a linear combination of the training vectors in  $H$  with weights  $\{\alpha_i\}_{1 \leq i \leq Q}$ :

$$w_0 = \sum_{i=1}^Q \alpha_i \varphi(J_i). \quad (11)$$

The regression function becomes then:

$$\Psi(J) = \sum_{i=1}^Q \alpha_i \varphi(J_i) \cdot \varphi(J) + b = \sum_{i=1}^Q \alpha_i K(J_i, J) + b, \quad (12)$$

where  $K$  is the kernel associated to  $\varphi$  in  $H$ . To handle non-linear relations between the manifold  $\mathcal{M}$  and the TRE,  $K$  is chosen as a RBF kernel, giving thus the following similarity model:

$$\Psi(J) = \sum_{i=1}^Q \alpha_i \exp\left(\frac{-|J_i - J|^2}{\sigma^2}\right) + b. \quad (13)$$

### 3 Experiments and Results

Our regressed similarity measure is evaluated on two challenging applications for multi-modal registration: rigid registration of MR-T1 and MR-TOF (Angiography) images of the carotid artery, and of MR-T1 and SPECT images of the brain. In this paper, we focus on 2D rigid-body experiments to prove the concept of our novel approach.

This permits in particular to show that the approach is not limited to pairs of images with a tissue distribution similar to the image pair used for training. Indeed, in the following experiments, a pair of corresponding images from a 3D dataset is used for training. The obtained similarity measure is then evaluated on pairs of images taken from the 3D datasets of the other patients. For statistical relevance, the pairs are chosen randomly and the tests are repeated. It must be noted that the tissue distribution varies depending on the randomly chosen slices, which can originate from the neck or from the head.

### 3.1 Experimental Setup

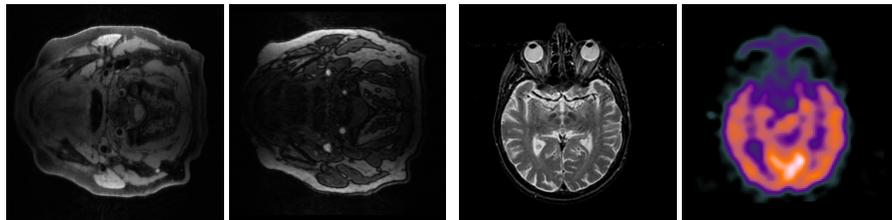
Our similarity measure will be compared to normalized mutual information (NMI) and Kullback-Leibler distance (KL) in terms of *success rate*, *accuracy* and *capture range*.

We consider a registration experiment as *successful* when the final target registration error is inferior to a given threshold  $t_e$ . In fact, this permits to quantify the ability of an approach to converge in the neighborhood of the right solution. We then define the *accuracy* as the mean target registration error on all registered images after the removal of such outliers. *Capture range* is evaluated by assessing the success rate as function of an increasing initial TRE. Knowing the ground truth position of each dataset, an initial random perturbation is applied to each pair of images according to a given value of TRE. Experiments are then repeated with an increasing initial target registration error.

The objective of our experiments is to highlight the benefits of a similarity measure taking advantage of prior information. Since the convergence to the right solution depends on the topography of the search space offered by a similarity measure, we use a Downhill-Simplex optimizer, that does not require any gradient information. For fair comparison, all measures have the same number of joint histogram bins ( $32 \times 32$ ) and are tested in the same conditions.

In both experimental setups T1/TOF and T1/SPECT, a cross-validation of  $N$  tests is performed on a set of  $P$  patients. A test consists of one *regression step* performed on a random pair of slices from a given patient and one *validation step* consisting of  $P - 1$  evaluations performed on the other  $P - 1$  patients. During the regression step, our similarity measure and the expected joint histogram needed by KL are computed on the same pair of images. During an evaluation, all measures are tested in the same conditions on a random pair of slices taken from another patient with the same initial perturbation. By using 10 initializations with an increasing TRE per evaluation, we can investigate the ability of each measure to converge towards the right solution and thereby assess their capture range.

The transformation space is sampled as follows:  $-40 \leq t_x \leq +40$  (in pixels),  $-40 \leq t_y \leq +40$  and  $-40 \leq \theta \leq +40$  (in degrees) with a step of 4 for each parameter, generating thereby 9261 data points. For the choice of the hyperparameters  $\sigma$  and  $C$ , a grid-search has been performed. All experiments are performed with the Spider environment for MATLAB on an Intel Core 2 Duo CPU 2.40 GHz.



**Fig. 2.** From left to right: T1 and TOF MR Angiography of the neck of the same patient. Then T1 and SPECT-Tc images of the brain of a healthy patient

**MR-T1 and MR-TOF Angiography images:** experiments are conducted on images (refer to Fig. 2) taken from  $P = 8$  patients (48 pairs of images) with different staging of

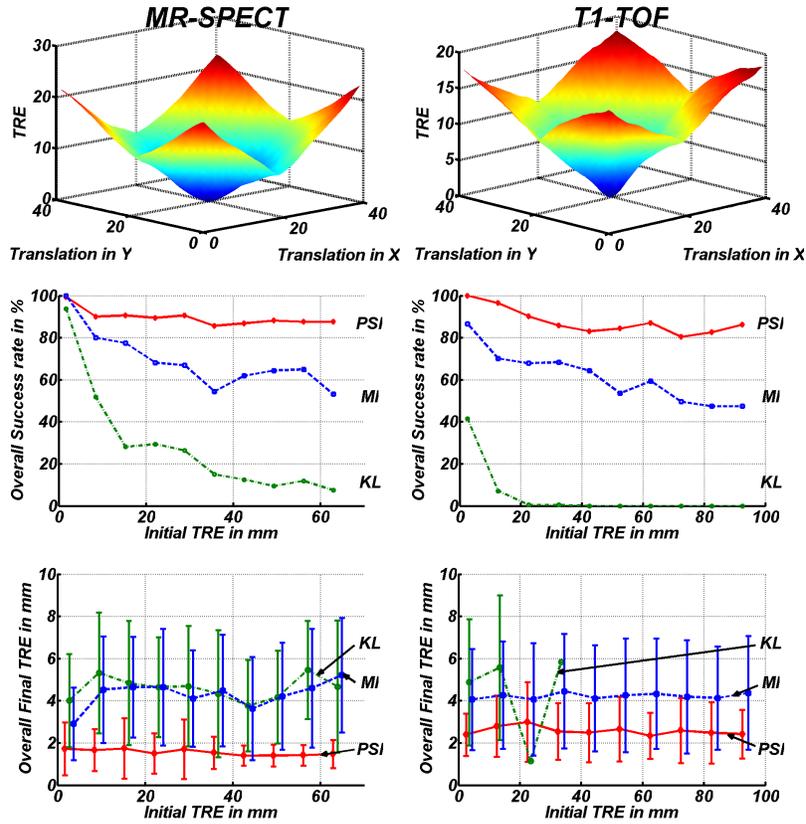
atherosclerosis. Both sequences were consecutively acquired, patients were positioned on a vacuum pillow and the acquisition was ECG gated to ensure perfect alignment. Images have a resolution of 128x128 with a pixel size of 2.5mm x 2.5mm. The threshold  $t_e$  is set to 1cm which corresponds to 4 pixels. A cross-validation of  $N = 32$  tests has been performed, which then corresponds to  $N \times (P - 1) \times 10 = 2240$  registration experiments.

**MR-T1 and SPECT-Tc images:** experiments are conducted on images (refer to Fig.2) taken from  $P = 5$  patients (73 pairs of images): a healthy patient, one with a glioma, one with a carcinoma, one with a stroke and finally one with an encephalopathy. These already registered datasets are taken from the publicly available Whole Brain Atlas database. Images have a resolution of 128x128 with a pixel size of 1.67mm x 1.67mm. The threshold  $t_e$  is set to 1cm which corresponds to 6 pixels. A cross-validation of  $N = 40$  tests has been performed, which then corresponds to  $N \times (P - 1) \times 10 = 1600$  registration experiments.

### 3.2 Results

The objective of our experiments is to show the benefits of a similarity measure taking full advantage of prior information. As shown on Fig. 3, the optimal regression model provides a smooth and convex search space, which is very close to the original TRE surface to approximate. Moreover, the global optimum has been preserved at the right position. In fact, smoothness and convexity are crucial characteristics to prevent the optimizer of being stuck in a local optimum and to ensure its convergence to the global one. The great advantage of our approach is its ability to model the convexity, the smoothness and the capture range of the similarity measure. Indeed, its convexity can be changed by choosing another function of the geometric error. The choice of hyperparameters  $C$  and  $\sigma$  influences the flexibility of the regression and thus the smoothness of the resulting function. During the regression process, increasing the sampling range of the transformation space permits to increase the capture range of the trained similarity. A high capture range is crucial when no good initialization parameters are available. Results presented in Fig. 3 shows the overall success rate and the final TRE as functions of the initial TRE. While the success rate of other measures sinks with an increasing initial TRE, our regressed similarity measure shows a good behaviour. This highlights its greater capture range and this, for a better accuracy. In the T1-TOF experiments, KL provided once a better accuracy for an initial TRE of 22.5 mm. This comes from the fact that KL was only successful on three registration experiments:  $\Psi$  and MI were actually better than KL in these specific experiments, but in the displayed results their accuracy is averaged on many more experiments as they have much higher success rates.

Our method was robust face to different tissue distributions, e.g. coming from patients affected by different kinds of disease or from different locations of the head that were not learned during the regression phase. For example, while slices from the top of the skull contain mostly skin, bone, cerebrospinal fluid, grey and white matter, slices in the middle of the head also consists of muscles and eyes. This could suggest that the manifold on which the similarity was learned is not strictly dependent on the anatomy. This needs however to be extensively studied with further experimentations.



**Fig. 3.** From top to bottom: Plot of the similarity  $\Psi$  (PSI) for variations in translation in x and y between  $-20$  and  $+20$  pixels. Plot of the success rate (in percent) and final TRE (mean and standard deviation in mm) according to an increasing initial TRE. Left MR-SPECT, right T1-TOF

## 4 Discussion and Conclusion

In this work, we propose to take advantage of prior information, namely a registered pair of images, in order to improve results in multi-modal registration. Our contribution is to define, with a regression approach, a new similarity measure relating the manifold described by joint histograms of two different modalities to the registration error. Experiments conducted on MR-T1/MR-TOF and MR-T1/SPECT images show that the presented method is a promising alternative for multi-modal registration. We empirically demonstrated that these manifolds are not dependant on the choice of the particular training pair within the dataset. This means that such an adapted application-specific measure can be defined by using a single pair of manually registered images from the specific application. Moreover, its robustness to different or new tissue distributions suggests that such manifolds could be modality-specific. In future work, we will further study their dependence to the variations of tissue distribution within the images.

## References

1. A. Chung, W. Wells, A. Norbash, and E. Grimson. Multi-modal image registration by minimising kullback-leibler distance. *Proc. of MICCAI Conf.*, pp:525–532, 2002. [2](#), [3](#)
2. A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3d multi-modality medical image registration using feature space clustering. *Proc. of Computer Vision, Virtual Reality and Robotics in Medicine Conf.*, pp:195–204, 1995. [2](#)
3. M. E. Leventon and E. Grimson. Multi-modal volume registration using joint intensity distributions. *Proc. of MICCAI Conf.*, pp:1057–1066, 1998. [2](#)
4. A. Roche, G. Malandain, N. Ayache, and S. Prima. Towards a better comprehension of similarity measures used in medical image registration. *Proc. of MICCAI Conf.*, pp:555–567, 1999. [1](#)
5. A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004. [5](#)
6. P. Viola and W. Wells. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, 1997. [2](#)
7. S. K. Zhou, J. Zhou, and D. Comaniciu. A boosting regression approach to medical anatomy detection. *Proc. of CVPR Conf.*, pp:1–8, 2007. [2](#)
8. L. Zollei, J. Fisher, and W. Wells. A unified statistical and information theoretic framework for multi-modal image registration. *Proc. of IPMI Conf.*, pp:366–377, 2003. [1](#)