

# Maximum a Posteriori Local Histogram Estimation for Image Registration

Matthew Toews<sup>1</sup>, D. Louis Collins<sup>2</sup>, and Tal Arbel<sup>1</sup>

<sup>1</sup> Center for Intelligent Machines, McGill University, Montréal, Canada  
{mtoews, arbel}@cim.mcgill.ca

<sup>2</sup> McConnell Brain Imaging Center, Montreal Neurological Institute,  
McGill University, Montréal, Canada  
louis@bic.mni.mcgill.ca

**Abstract.** Image similarity measures for registration can be considered within the general context of joint intensity histograms, which consist of bin count parameters estimated from image intensity samples. Many approaches to estimation are ML (maximum likelihood), which tends to be unstable in the presence sparse data, resulting in registration that is driven by spurious noisy matches instead of valid intensity relationships. We propose instead a method of MAP (maximum a posteriori) estimation, which is well-defined for sparse data, or even in the absence of data. This estimator can incorporate a variety of prior assumptions, such as global histogram characteristics, or use a maximum entropy prior when no such assumptions exist. We apply our estimation method to deformable registration of MR (magnetic resonance) and US (ultrasound) images for an IGNS (image-guided guided neurosurgery) application, where our MAP estimation method results in more stable and accurate registration than a traditional ML approach.

## 1 Introduction

The evaluation of intensity similarity between two images for the task of image registration can be framed in the general context of joint intensity histograms, for a wide variety of similarity measures including correlation, correlation ratio, mutual information and others [12]. In this context, the task of similarity evaluation is to calculate the relevant similarity measure based on *an estimate* of the joint intensity histogram. Thus, the quality of image similarity evaluation is dependent on the quality of joint histogram estimation.

Histogram estimation can be considered in the context of statistical parameter estimation, where the parameters to be estimated are the bin counts of the joint histogram. A variety of estimation techniques have been presented in the literature, the majority of which are variants of ML (maximum likelihood) estimation [6]. The hallmark of ML estimators is that they become unstable in the presence of sparse image data, and are undefined in the absence of image data. As such, ML estimators perform poorly in the presence of sparse data, tending to latch onto spurious, noisy matches despite the variety of techniques designed to improve their performance such as Parzen windowing, partial volume interpolation, robust ML estimation, etc.

In this article, we present a method of joint histogram estimation based on MAP (maximum a posteriori) parameter estimation. In contrast to ML estimation, MAP histogram estimates are well-defined in sparse data or even in the complete absence of data by a prior distribution over histogram bins. As such, MAP histogram estimates are not subject to instability as are ML estimates. Furthermore, MAP estimation provides a principled way incorporating prior information in terms of the number of samples required to obtain a valid histogram estimate. As a result, one can bias histogram estimates such that if the number of samples is insufficient, the MAP estimate will favor a benign prior unlikely to lead to spurious matches.<sup>1</sup>

In practical terms, our method involves pre-populating joint histogram bin counts before the arrival of data, according to an estimate of number of samples  $M$  required for valid joint histogram estimation. This pre-population constitutes our prior belief as to the joint histogram, which is gradually outweighed as data samples are added. This general approach is known to stabilize probability estimates of infrequently observed events [10], but it is not commonly used to stabilizing image similarity estimates for registration.

Our MAP estimation method was developed for the purpose of on-line non-linear registration of MR (magnetic resonance) and US (ultrasound) imagery in the context of an IGNS (image-guided neurosurgery) application. This is a demanding registration task for the following reason: the joint intensity relationship between MR and US imagery is noisy, multi-modal and highly non-stationary (i.e. varying with spatial location), thus computing image similarity for registration requires a sophisticated similarity measure such as MI (mutual information) with a high degree of parameterization. At the same time, precise nonlinear registration requires evaluating similarity within small local image windows, meaning a relatively large number of histogram parameters must be estimated from a relatively small number of intensity samples. A principled method of histogram estimation is therefore needed to overcome the problem of sparsity. Preliminary results based on stereotactic ground truth are promising, indicating a greater degree of registration stability and accuracy when based on MAP as opposed to ML histogram estimation.

## 2 Methods for Dealing with Sparsity

Sparsity in histogram estimation is a problem touched on by many authors, although rarely in the framework of statistical parameter estimation. Here we present some common approaches histogram estimation, particularly for the purpose of dealing with data sparsity:

**Parzen windowing:** Parzen windowing [15], involves smoothing histograms with a Parzen window kernel, often a Gaussian. Parzen windowing has the drawback of populating histogram estimates with fictitious samples, and does not generally converge to the true histogram estimate with increased sample size. To

---

<sup>1</sup> We stress here that we refer not to MAP estimation for registration, but the estimation of the joint histogram used to evaluate similarity given a fixed image alignment.

see this, consider the case in which the joint intensity histogram is completely contained in a single histogram bin. Using Parzen windowing, this histogram will always contain a smoothed peak, regardless of the amount of image data.

**Coarse histogram binning:** Coarse histogram binning reduces the number of parameters to be estimated from a limited amount of data [16]. Coarse histogram binning in itself is not sufficient to deal with sparsity, however, as the coarser the histogram quantization, the more impoverished the histograms become for the purpose of registration. In addition, even highly quantized histograms can suffer from undersampling given small local window sizes. Finally, the question of an optimal quantization scheme is difficult and task dependent.

**Probabilistic Segmentation of Before Registration:** A new approach which serves to reduce problems with sparse intensity histogram estimation is to compute histograms based on probabilistically labeled images [2]. Here, histograms are based on low-dimensional probability distributions, resulting in bins that are non-zero. Although seemingly well suited to its purpose, such an approach requires the extra step of probabilistic data segmentation, which in our experience tends to blur out local structures in favor of global segmentation.

**Weighting Local Histograms with Global Intensity Relationships:** Weighting local histograms with global intensity relationships has been touched on by several authors [9,7,4,8,14], suggesting agreement that a means of dealing with sparsity is to populate histograms based on other sources of relevant information. The drawback with considering global histograms is that they do not generally reflect local intensity relationships, as in the case of MR and US modalities (See Figure 1), and can result in inappropriately biased estimates. That being said, our method of MAP estimation can be used to bias local histogram estimates using a variety of prior information within the principled context of statistical parameter estimation, allowing incorporation of global histogram information when relevant and providing other options when it is not.

### 3 MAP Histogram Estimation

The majority of techniques have difficulty with local histogram estimation because they are based on ML estimation, which becomes increasingly unstable in the presence of sparse data. In practical terms, this results in registration that tends latch onto spurious incorrect matches as data sparsity increases. To deal with this instability, we propose MAP estimation with a maximum entropy prior, which tends to produce strong matches only when justified by sufficient intensity samples.

#### 3.1 Maximum Likelihood vs. Maximum a Posteriori

The task of estimating a local histogram for the purpose of registration is one of estimating a set of  $K$  discrete bin frequency counts  $\theta = \{\theta^1, \dots, \theta^K\}$ . The goal in histogram estimation is to determine the values of  $\theta$  that maximize  $p(\theta|I)$ , the conditional probability of  $\theta$  given the image data  $I$ . By Bayes rule, the following equality holds:

$$p(\theta|I) = \frac{p(I|\theta)p(\theta)}{p(I)}. \quad (1)$$

For the purpose of statistical parameter estimation,  $p(I)$  is constant as image data  $I$  is constant,  $p(I|\theta)$  naturally takes the form of a multinomial distribution, and  $p(\theta)$  and  $p(\theta|I)$  are Dirichlet distributions [6].

There are two significantly different methods of estimating  $p(\theta|I)$ : ML and MAP. We advocate the MAP strategy, as it provides a principled mechanism for explicitly incorporating prior information in the form of the number of intensity samples  $M$  required for valid estimation.

**ML Estimation:** ML estimation is based on the assumption that  $p(\theta)$  is constant. Under this assumption, we seek an estimate  $\theta_{ML}$  such that:

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \{ p(I|\theta) \}. \quad (2)$$

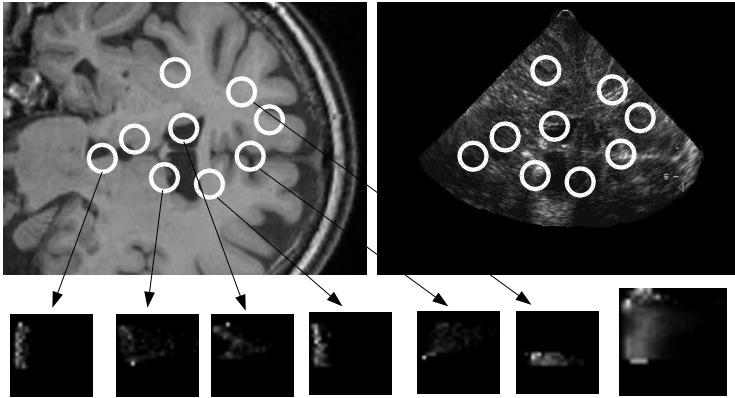
Here, to maximize  $p(\theta|I)$ , it suffices to maximize the term  $p(I|\theta)$ , which is known as the *likelihood*, hence the name maximum likelihood estimation. In ML estimation, histogram bin counts  $\theta$  are simply set to counts of intensity data  $I$ , optionally processed with Parzen windowing, etc.

**MAP Estimation:** MAP estimation does not treat all histograms as equally probable, and  $p(\theta)$  is not constant. In particular, certain histograms are more probable than others, based on prior assumptions we may have regarding  $\theta$ , and we seek an estimate  $\theta_{MAP}$  such that:

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} \{ p(I|\theta)p(\theta) \}. \quad (3)$$

It can be shown that as the number of intensity samples  $I$  approaches infinity, both ML and MAP estimation converge to the same optimal histogram estimate of  $\theta$  [6]. In intuitive terms, this is because the prior distribution  $p(\theta)$  of the MAP approach becomes swamped by  $p(\theta|I)$ . Being swamped with intensity data is hardly a problem in histogram estimation however, particularly when attempting to calculate similarity for the purpose of localized deformation [11]. It is precisely in the case of sparse data samples  $I$  that the difference between ML and MAP estimation is most telling. MAP estimation allows the incorporation of a prior distribution  $p(\theta)$  as to the value of  $\theta$  in the absence of data  $I$ . In the presence of sparse data, i.e. an image window  $I$  with insufficient samples to estimate the histogram, we would like estimation to default to our expectation as to the true histogram values, or at least a benign histogram  $\theta$  that is not likely to result in spurious, noise-driven matches.

In general,  $p(\theta)$  could be based on a variety of prior assumptions, i.e. the global histogram approach, although the global histogram may not be representative of local intensity relationships, as in the case of MR/US registration. In the absence of constraints, we follow the rule of maximum entropy [5] and suggest a uniform prior - in the case of undersampling, the uniformly-weighted histogram is adverse to making a strong decision regarding registration, and the effect of spatial neighborhood constraints will dominate. In the case where the number of intensity sam-



**Fig. 1.** The image intensity relationship between MR and US image modalities is multi-modal and non-stationary, i.e. varying spatial position. The upper left image is a 2D slice from T1-weighted MRI brain volume, and the upper right image is a corresponding US image slice. Correspondence was determined via a stereotactic positioning of the US probe relative to the volume coordinate system. The white circles overlaying the image indicate local regions within which local likelihood histograms  $p(US|MR)$  are calculated. The smaller images along the bottom are likelihood histograms, where the vertical axis is MR intensity and the horizontal axis is US intensity. The histograms corresponding to local regions are indicated by arrows, and the larger histogram in the bottom right is the global histogram. It can be seen that the statistical likelihood pixel relationship is non-stationary, as it varies significantly with spatial location. In addition, local likelihood relationships are significantly different from the global likelihood.

ples is significant, the uniform prior will be outweighed by evidence and valid registration will occur. The intuition is that  $M$  should be set according to the number intensity samples required to obtain a valid histogram estimate, which is related to a number of factors such as noise, the number of histogram bins to be estimated, etc. Our MAP estimation approach can be summarized as follows:

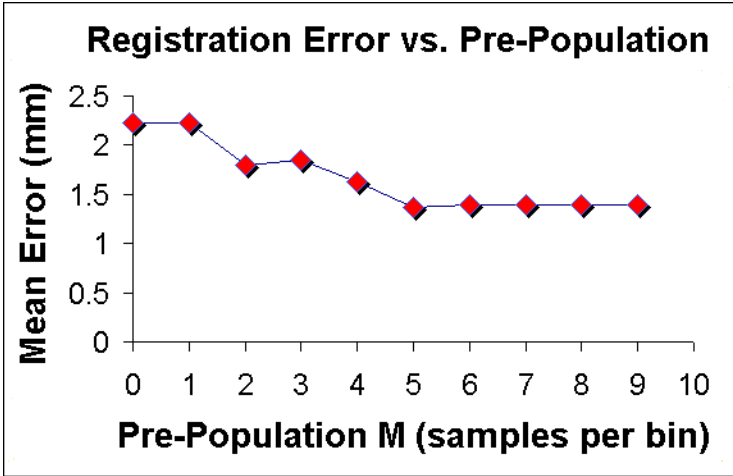
#### MAP Histogram Estimation:

- 1) Generate an estimate as to the number of intensity samples  $M$  required to obtain a valid estimate of the histogram bins.

- 2) Pre-populate histogram counts according to  $M$ , either uniformly or according to other sources of prior information.

## 4 Deformable MR to US Registration

We developed our MAP estimation method for the purpose of deformable registration of MR and US imagery in the context of an IGNS application, where the goal is to update a detailed pre-operative 3D MR volume of the brain used in intervention planning with real-time US gathered inter-operatively, in order to reflect brain shift that occurs once the dura lining has been entered. With



**Fig. 2.** A plot of mean displacement vector registration error vs. the degree of uniform prior incorporated MAP estimation. The error for a 9-point displacement field drops by a factor of 1.6 as the joint histogram pre-population is increased from  $M=0$  to  $M=5$  samples per bin. Note that when  $M=0$ , MAP and ML estimation are equivalent. The error is high for low  $M$ , as the posterior is dominated by spurious local MI maxima. Error reaches a minimum at  $M = 5$  samples per histogram bin, after which point it rises slightly as the elastic prior begins to dominate the posterior.

both the US probe and the patient’s head registered rigidly via a stereotactic tracking apparatus, the task becomes one of updating the 3D MR volume based on non-linear registration with 2D US slices.

For the purpose of validation, we have compiled a database of US slices taken inter-operatively. To test registration, we focus on recapturing the deformation relative to stereotactic ground truth. US image structure is due to the reflection of acoustic waves at the boundaries of structures of differing density, and is an inherently noisy and difficult image modality to work with, due to speckle noise, signal attenuation artifacts such as acoustic shadows, etc. Figure 1 illustrates how the joint intensity relationship between MR and US varies to a large degree with spatial position.

In order to model nonlinear brain shift, we are interested in recovering a field of displacement vectors  $\mathbf{T} = \{\mathbf{t}_i\}$  mapping fixed points in the MR image to their displacements in the US image. Adopting a Bayesian strategy as in [3], we formulate registration as a posterior probability over  $\mathbf{T}$  given the images to be matched:

$$p(\mathbf{T}|US, MR) \propto p(US|\mathbf{T}, MR)p(\mathbf{T}). \quad (4)$$

The Bayesian formulation requires specification of the terms  $p(US|\mathbf{T}, MR)$  and  $p(\mathbf{T})$  which are referred to as the likelihood and prior, respectively. In registration, the likelihood is the data term serving to evaluate similarity between images  $US$  and  $MR$  given  $\mathbf{T}$ , and the prior serves to incorporate regularization

constraints on  $\mathbf{T}$  independent of image information, such smoothness, elasticity, etc. The strength of Bayesian registration is that the likelihood and prior terms can be changed to suit the registration task at hand without altering the overall formulation.

For our purposes, we choose to model  $p(US|\mathbf{T}, MR)$  using the MI (mutual information) of intensities. The MI is a widely-used measure of statistical intensity similarity based on information theory [15,1]. Many excellent references exist regarding the details of MI calculation [11]. Alternatively, we could have adopted the correlation ratio approach of [13] based on intensity and gradient images or a learning based similarity measure as in [7]. In order to model nonlinear deformation, we choose to model  $p(\mathbf{T})$  using an elastic prior between pairs of neighboring deformation vectors in  $\mathbf{T}$ . The final Bayesian posterior is of the form:

$$p(\mathbf{T}|US, MR) \propto \exp\left\{\alpha \sum_i^N (MI(US|\mathbf{t}_i, MR) - MI_{max}) - \beta \sum_{i,j}^{N,N} d(\mathbf{t}_i, \mathbf{t}_j)\right\}, \quad (5)$$

where  $MI_{max}$  is the maximum MI achievable using a 25x25 bin joint histogram,  $N$  is the number of vectors in  $\mathbf{T}$ ,  $d(\mathbf{t}_i, \mathbf{t}_j)$  represents an elastic prior energy between  $\mathbf{t}_i$  and  $\mathbf{t}_j$ , and  $\alpha = 5$  and  $\beta = 2$  are empirically determined parameters that balance the relative strengths of the likelihood and prior terms.

For experimentation, we attempt to recover the known transform  $\mathbf{T}$  between  $US$  and  $MR$  image pairs. Given a fixed local window size, we compare the result of registration  $\mathbf{T}$  using ML and MAP estimation. The number of joint histogram bins used is 25x25=625 and the local window size is 43x43=1849 pixels. Optimization of the posterior in (5) was achieved via gradient ascent from 100 random seeds, and the best solution, i.e. the deformation field maximizing equation (5) was used in the result analysis. Random seeds were generated by perturbing vectors  $\mathbf{t}_i$  to random displacements within a 25 pixel radius of the known transform. The resolution of registration was 1mm per pixel.

Figure 2 shows the impact of MAP estimation on registration error, as the degree of prior pre-population in the joint histogram is increased. Here we see that with little or no histogram pre-population, i.e. ML estimation, poor histogram estimates result in unstable, inaccurate registration.

## 5 Conclusion

In this article, we presented a principled means of estimating joint intensity histograms for the purpose of similarity calculation in the presence of sparse image data. Given the number of intensity samples  $M$  required to reliably estimate intensity histograms, we proposed a MAP estimation method based on a uniform prior histogram. The advantage of this method is that histogram estimates only result in strong matches when sufficient evidence exists to justify them, i.e. when the number of intensity samples is sufficiently high. In the case of undersampling, histogram estimates will default to a benign prior histogram that is unlikely to result in strong false matches. Traditional ML estimation, on the other hand, tends to produce strong false matches in the case of undersampling, which can

throw off registration. Practically, our method is simple to implement, where histogram bin counts are pre-populated by  $M$  uniformly distributed samples prior to estimation, representing the prior assumption of MAP estimation. We expect that MAP histogram estimation will result in improved registration in other methods of similarity calculation based on histograms, such as correlation, correlation ratio, etc. Future work will involve further clinical validation, testing with other similarity measures, and determining of an optimal degree of prior information  $M$  to incorporate given the images to be registered.

## References

1. A. Collignon, *Multi-modality medical image registration by maximization of mutual information*, Ph.D. thesis, Catholic University of Leuven, Leuven, Belgium, 1998.
2. Emiliano D'Agostino, Frederik Maes, Dirk Vandermeulen, and Paul Suetens, *An information theoretic approach for non-rigid image registration using voxel class probabilities*, MICCAI2003, 2003, pp. 812–820.
3. J.C. Gee, L. LeBriquer, and C. Barillot, *Probabilistic matching of brain images*, IPMI95, Kluwer, July 1995.
4. A. Guimond, A. Roche, N. Ayache, and J. Meunier, *Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections*, IEEE Transactions on Medical Imaging **20** (2001), no. 1, 58–69.
5. E.T. Jaynes, *Prior probabilities*, I.E.E.E Transactions of systems, science, and cybernetics **SSC-4** (1968), no. 3, 227–241.
6. Michael I. Jordan, *An introduction to probabilistic graphical models*, In preparation, 2003.
7. Michael E. Leventon and W. Eric L. Grimson, *Multi-modal volume registration using joint intensity distributions*, Lecture Notes in Computer Science **1496** (1998), 1057–1066.
8. B. Likar and F. Pernus, *A heirarchical approach to elastic registration based on mutual information*, Image and Vision Computing **19** (2001), no. 1-2, 33–44.
9. J.B.A. Maintz, H.W. Meijering, and M.A. Viergever, *General multimodal elastic registration based on mutual information*, Medical Imaging (1998), 144–154.
10. Alon Orlitsky, Narayana P. Santhanam, and Junan Zhang, *Always good turing: Asymptotically optimal probability estimation*, Science **302** (2003), 427–431.
11. J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, *Mutual-information-based registration of medical images: a survey*, MedImg **22** (2003), no. 8, 986–1004.
12. A. Roche, N. Malandain, Ayache, and S. Prima, *Toward a better comprehension of similarity measures used in medical image registration*, MICCAI99, 1999, pp. 555–566.
13. A. Roche, X. Pennec, G. Malandain, and N. Ayache, *Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information*, IEEE Transactions on Medical Imaging **20** (2001), no. 10, 1038–1049.
14. P. Rogelj, S. Kovacic, and J. C. Gee, *Point similarity measures for non-rigid registration of multi-modal data.*, Computer Vision and Image Understanding **92** (2003), no. 1, 112–140.
15. P.A. Viola and W.M. Wells, III, *Alignment by maximization of mutual information*, International Journal of Computer Vision **24** (1997), no. 2, 137–154.
16. L. Zollei, W.E.L. Grimson, A. Norbash, and W.M. Wells, III, *2d-3d rigid registration of x-ray fluoroscopy and ct images using mutual information and sparsely sampled histogram estimators*, CVPR01, 2001, pp. II:696–703.