

# BAYESIAN REGISTRATION FOR ANATOMICAL LANDMARK DETECTION

Camille Izard<sup>1,2</sup> \*

Bruno Jedyak<sup>1,2</sup> †

<sup>1</sup>Johns Hopkins University  
Center for Imaging Science  
3400 N Charles St, Baltimore MD, USA

<sup>2</sup>Université des Sciences et Technologies de Lille  
Laboratoire Paul Painlevé  
Cité Scientifique, Villeneuve d'Ascq, France

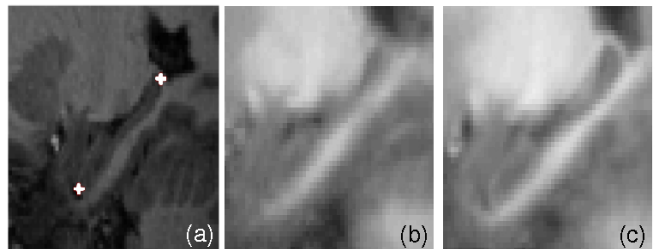
## ABSTRACT

In order to perform medical image registration, landmarks are used to settle correspondences between images. A landmark is a voxel in the image that corresponds to a well-defined point in the anatomy. Manual landmarking is a difficult, tedious and time-consuming procedure that would gain to be automated. We propose a bayesian approach for automatic landmarking. Using training data, we learn the geometry through a probabilistic template. Landmarking consists then in estimating an affine transformation mapping the image onto the template. We use gradient ascent in the likelihood function to perform this task. Experiments validate the methodology for landmarking the temporal lobe in MR brain images.

## 1. INTRODUCTION

If landmarks by themselves can provide a first analysis of the geometry of brain structures, they are overall useful for image registration, see Figure 1. They can be used to initialize correspondences between structures or to give control points in the alignment process, [1]. We define a landmark, as in [2], as the voxel in the image which corresponds to a specific point of the anatomy. They can be points like the apex of the Head of the Hippocampus (HoH), the Tail of the Hippocampus (HT) and the Splenium of the Corpus Callosum (SCC). Figure 1 shows some examples of landmarks. The manual detection of these points is a tedious and time consuming task. Although some methods have been developed to attempt detecting some types of landmarks automatically, based on 3D filters [3] or by fitting a local parametric model on the intensities [4], the detection of landmarks remains a manual task. In both cases indeed one needs prior information on the underlying structure and a detectable contour, but landmarks that neuroscientists define do not necessarily fulfill the latter requirement.

We propose a generic approach for automatic landmark detection. We assume that an image results from the deformation of a template. Because deformations in the brain can



**Fig. 1.** (a) the bottom left cross represents the Head of the Hippocampus landmark, the right top cross is the Tail of the Hippocampus, (b) shows the mean image of the hippocampus, obtained by averaging the intensities of 14 images without registration, (c) is the average image after rigid registration using the landmarks HoH and HT.

be extremely complicated, we use a probabilistic template. We present a Bayesian approach to estimate the transformation  $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  which maps the landmark location in the image onto the landmark location in the template. Given a training set of images  $(X^{(k)}, \phi^{(k)})$ ,  $X^{(k)}$  the vector of intensities of the image  $k$ , we define a generative model  $P(X|\phi)$  and estimate the best transformation for a new image using the maximum a posteriori (MAP) estimator:  $\max_{\phi} P(\phi|x)$ . In section 2 we detail the image model. Parameter estimation is in section 3. Section 4 presents the gradient ascent algorithm. Finally we demonstrate the algorithm on brain MRI to detect the location of the Head and the Tail of the Hippocampus and also the Splenium of the Corpus Callosum.

## 2. BAYESIAN INTENSITY MODEL

### 2.1. The transformation set

Let  $\Phi$  be a set of transformations from  $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ , such that there is a unique  $\phi \in \Phi$  that maps the landmarks in the image onto a given configuration in the template. We will consider transformations not more complicated than affine transformations or compositions of transformations inspired from the procedure specialists use to define landmarks. Firstly, HoH is located using a translation, secondly, the hippocampus is rotated around the sagittal axis to fix its orientation and then it is scaled to overlay the tail. We will present experiments on translation in this paper while the more complex transformations will appear elsewhere.

\*supported by the Doctoral Fellowship of the Université des Sciences et Technologies de Lille

†funded by ARO/DAAD19/-02-1-0337, general funds of the Center for Imaging Science, JHU and NIH, ADRC Pilot Project Award, 2005 . The authors are deeply thankful to Craig Stark for kindly providing the images and for locating the landmarks on them.

## 2.2. The generative model

Let us describe the generative model,  $P(\mathbf{X}|\phi)$ , where  $\mathbf{X}$  is the gray-level image and  $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a transformation. We assume that the voxel intensities are independent given the landmarks location, i.e. given the transformation, so that we can write the conditional joint probability as a product:  $P(\mathbf{X}|\phi) = \prod_{v \in V} P(X_v|\phi)$ , with  $V$  the set of voxels. We introduce a new variable  $Z_v$  representing the matter at voxel  $v$ . The intensity distribution is indeed slightly different from one image to the other, because anatomically equivalent voxels appear with different intensities in two images. Let us assume that the brain is composed of 6 matters: CSF, GM and WM to which we add the mixed voxels composed of CSF+GM and GM+WM. We also identify a sixth class corresponding to very high intensities due to blood vessels or to the skull. The discrete random variable  $Z_v$  takes values on  $\{1, \dots, 6\}$ . Conditioning by the matter at each voxel, the generative model can be written as

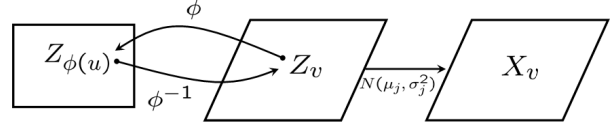
$$P(X_v = x_v|\phi) = \sum_{j=1}^6 P(X_v = x_v|Z_v = j, \phi)P(Z_v = j|\phi).$$

We assume that the intensity at the voxel  $v$ , given the matter at the same voxel is independent of the transformation  $\phi$ . This assumption allows us to model the photometry independently of the geometry. The first term characterizes the photometry of the matter  $j$  in the image, which is modeled with a Gaussian distribution. The second term corresponds to the probability of observing the matter  $j$  at the voxel  $v$  given the transformation  $\phi$  which maps the image onto the template. This is the probability distribution of the matter at the transformed voxel  $\phi(v)$ . Under the preceding assumptions the generative model can be rewritten as

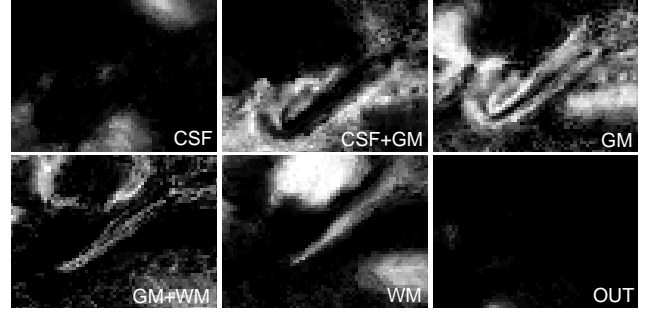
$$P(\mathbf{X}|\phi) = \prod_{v \in V} \sum_{j=1}^6 g_j(x_v)P(Z_{\phi(v)} = j).$$

## 2.3. The probabilistic template

In this Bayesian setting, we define the template as the probability distribution of the matter when the landmarks lie in a standardized configuration. Figure 2 represents the model with which new images can be generated by drawing a matter at each voxel according to the distribution in the template, then drawing a random transformation. Applying the inverse transformation  $\phi^{-1}$  one creates the geometry of the new image and finally using a set of Gaussian distributions we can assign an intensity to each voxel. The template gives the probability to observe each matter at each voxel. If the probability at one voxel to observe one of the matters is high, it means that most of the images of the registered training set contain this matter at this point. The probabilistic template allows us to capture the matter variability in addition of the global geometry. We conjecture that using such a template



**Fig. 2.** To generate a new image (rightmost), draw a random segmentation (leftmost) based on the distribution of matters contained in the template. Apply a random deformation  $\phi^{-1}$  to find the new image segmentation (middle). Finally assign an intensity chosen in the corresponding Gaussian distribution.



**Fig. 3.** Template obtained when  $\phi$  is a translation with HoH as landmark. Each image corresponds to one matter. The white voxels have a high probability to belong to the corresponding matter. The represented section of the template corresponds to the sagittal slice containing HoH. Notice that the hippocampus appear both in CSF+GM and in GM.

will allow us to deal with much simpler transformations and achieve results comparable to classical methods for registration.

## 2.4. Bayesian versus classical registration methods

Considering a new image, one wants to maximize the likelihood of the observation over the parameters  $\theta$  of the transformation  $\phi$ . The likelihood of the model can be expressed with the generative model and the prior distribution on the transformation:  $P(\phi|\mathbf{X}) \propto P(\mathbf{X}|\phi)P(\phi)$ . The prior is a distribution over the parameters of  $\phi$ . In the case of a translation for example, we use a Gaussian model on the parameter  $[a \ b \ c]$  so that we compute from the training set a mean vector of parameters in  $\mathbb{R}^3$  and an empirical covariance matrix in  $\mathbb{R}^3 \times \mathbb{R}^3$ . The log-likelihood of the model is:

$$l(\theta) = \sum_{v \in V} \ln \sum_{j=1}^6 g_j(x_v)P(Z_{\phi(v)} = j) + \ln P(\theta).$$

As in the classical cost function used for image registration, the likelihood contains two terms. A data term measuring the way the model fits to the data and a regularization term. Using a Bayesian approach allows us to deduce naturally the cost function from the model and the template gives a variable weight to the voxels of the image. When the entropy of the matter distribution is low, the corresponding voxel has potentially a large weight: it increases the likelihood if the

observation and the model match and penalizes it in case of mismatch.

Algorithm 1 describes the outline of the procedures whose steps are detailed in the following section.

---

### Algorithm 1 (Bayesian Registration)

---

**Learning step**,  $n$  images in the training set:

1. *Photometry estimation*: Using the EM-algorithm, learn the intensity distribution of each matter  $j$  in each image  $k$ ,  $\{(\mu_{k,j}, \sigma_{k,j}^2)_{j \in \{1, \dots, 6\}}\}_{k \in \{1, \dots, n\}}$ .
2. *Superimposition of the training set*: Given a set of transformations  $\Phi$ , find  $\{\phi_1, \dots, \phi_n\}$  such that for each image the transformed landmarks are mapped to the same location.
3. *Prior distribution estimation*: Estimate the prior distribution based on the set of transformations estimated on the training set.
4. *Template estimation*: Based on the transformed images  $\{(X_{\phi(v)}^{(k)})_{v \in V}\}_{k \in \{1, \dots, n\}}$ , use the EM-algorithm to estimate the matter proportions at each voxel  $\phi(v)$ .

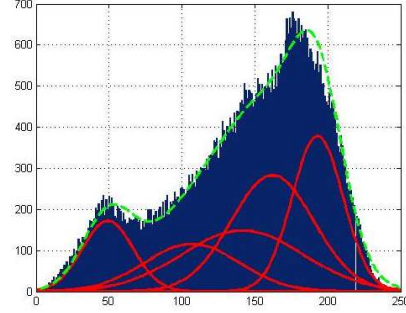
**For a new image**

1. *Photometry estimation*: Similarly to the learning step, learn the matter distributions  $(\mu_j, \sigma_j^2)_{j \in \{1, \dots, 6\}}$ .
  2. *Transformation estimation*: Find  $\phi$  which maximizes the likelihood of the observations, using a gradient ascent over the transformation parameters,  $\theta$ .
- 

## 3. LEARNING ALGORITHM

### 3.1. Photometry estimation

As outlined in the previous section, we model the intensity of each matter in the brain with a Gaussian distribution, so that the intensity at a voxel results from a mixture of Gaussian distributions. This model has been commonly used in MRI segmentation with a variable number of components [5]. Given the matter segmentation of the image, it would be easy to estimate the mixture parameters, it would also be straightforward to find the segmentation, if the model parameters were known. However, since the matter at each voxel is unknown, we use the EM algorithm to alternatively compute the classification of the voxels and estimate the model parameters. The EM-algorithm [6] maximizes the log-likelihood of the mixture model over the parameters  $(\alpha_j, \mu_j, \sigma_j^2)$  with  $j \in \{1, \dots, 6\}$ . In the case of Gaussian mixture, both the E-step and the M-step can be written in closed form and convergence to a local minimum has been proved. We initialize the EM with Kmeans.



**Fig. 4.** Example of Gaussian mixture parameter estimation on a 41-by-41 voxel subpart of the image. x-axis: intensities, y-axis: count. The plain lines are the 5 estimated Gaussian densities and the dot line represents the estimated histogram. The outliers distribution does not appear on this figure as the mean is usually around 300.

### 3.2. Template estimation

Using a training set, composed of  $n$  images, typically between 10 and 100, on which the landmarks have been located, one computes for each image  $k$  the transformation  $\phi_k \in \Phi$  which maps the voxels coordinates onto the template. The superimposition procedure provides us with a collection of transformations  $\{\phi_1, \dots, \phi_n\}$ . We use the set of parameters of these transformations to estimate the prior distribution on  $\theta$ . At each location  $\phi(v)$  in the template corresponds a set of intensities  $(x_{\phi(v)}^{(1)}, \dots, x_{\phi(v)}^{(n)})$  resulting from the registration of the training set. Because the transformation does not map perfectly the image onto the template, the matters are not perfectly superimposed in the transformed images, also at  $\phi(v)$  the intensities correspond to different matters of the brain. Since the matters and their distribution are unknown we need to use the EM-algorithm to estimate the mixture parameters. The likelihood maximized by EM, is written as

$$\max_{P(Z_{\phi(v)=j})} \sum_k \log \sum_{j=1}^6 \frac{P(Z_{\phi(v)} = j)}{\sqrt{2\pi\sigma_{k,j}^2}} \exp\left(-\frac{(x_{\phi(v)}^{(k)} - \mu_{k,j})^2}{2\sigma_{k,j}^2}\right).$$

We use the photometric parameters  $(\mu_{k,j}, \sigma_{k,j}^2)$  estimated during the previous step. Figure 3 shows one slice of the template obtained if  $\phi$  is assumed to be a translation based on the location of HoH.

## 4. TRANSFORMATION ESTIMATION BY GRADIENT METHOD

Given a new image, recovering the landmarks location is equivalent to finding the transformation  $\phi$  mapping the image onto the template.  $\phi$  is a transformation from  $\mathbb{R}^3 \rightarrow \mathbb{R}^3$  of parameter  $\theta$ . Given the generative model, one wants to maximize the likelihood of the observations over the transformation parameters, the likelihood is a mapping from  $\mathbb{R}^d \rightarrow \mathbb{R}$ , where  $d$  is the number of parameters of the transformation. Let us

simplify the notation  $\forall j, P(Z_{\phi(v)} = j) = P_j(Z_{\phi(v)})$ . We compute the gradient of the log-likelihood:

$$l(\theta) = \sum_{v \in V} \ln \sum_{j=1}^m g_j(x_v) P_j(Z_{\phi(v)}) + \ln P(\theta),$$

$$\nabla_{\theta} l(\theta) = \sum_{v \in V} \frac{\sum_{j=1}^6 g_j(x_v) \nabla_{\theta} P_j(Z_{\phi(v)})}{\sum_{j=1}^6 g_j(x_v) P_j(Z_{\phi(v)})} + \nabla_{\theta} \ln P(\theta),$$

with  $\forall j, \nabla_{\theta} P_j(Z_{\phi(v)}) = \left\langle \begin{array}{c} \frac{\partial P_j(Z_{\phi(v)})}{\partial x} \\ \frac{\partial P_j(Z_{\phi(v)})}{\partial y} \\ \frac{\partial P_j(Z_{\phi(v)})}{\partial z} \end{array}, \begin{array}{c} \frac{\partial \phi_1}{\partial \theta} \\ \frac{\partial \phi_2}{\partial \theta} \\ \frac{\partial \phi_3}{\partial \theta} \end{array} \right\rangle$ .

The gradient of the log-likelihood over the parameters of the transformation  $\theta$  can be written as a function of the template derivative. Consequently it is possible to compute the cartesian derivatives of the template offline that makes the optimization algorithm much faster. The estimation of the transformation parameters is equivalent to a  $d$ -dimensional optimization problem with  $d$  small in this approach, up to 12 in the case of affine transformation.

## 5. EXPERIMENTS ON BRAIN MRI

The training set is composed of 38 T1-weighted MR brain images acquired on a Philips-Intera 3-Tesla scanner, with resolution  $1\text{mm}^3$ . Brains were first transformed into standardized Talairach space using Analysis of Functional Neuroimages (AFNI) to provide a canonical orientation (anterior and posterior commissures (AC and PC) made co-linear) and approximate alignment. All the images have the same size after the transformation:  $161 \times 191 \times 151$  voxels or mm. An expert located the splenium of the corpus callosum (SCC), the apex of the Head of the Hippocampus (HoH) and define on the same sagittal slice as HoH the Tail of the Hippocampus (HT). We apply the algorithm to the detection of HoH, SCC and HT.

In the case of HoH and SCC, the considered set of transformations is the group of translations in  $\mathbb{R}^3$ . The derivatives of the template are exactly the cartesian derivatives of each matter. We use a gradient ascent to find the maximum of the likelihood. The results are presented on both the training and the testing set, 9 MR brain images from the same scanner and landmarked by the same specialist. We compare the results of the algorithm to the expert landmark by computing the Euclidian distance between the two points. The detection of HT is conditional to the position of the HoH landmark, hence it reduces to a 2D problem, assuming that the location of HoH is given. Once again we consider the group of translations but now in  $\mathbb{R}^2$ . Table 1 presents the results obtained on the training and the testing set. The specialist's intra-variability is  $0.71\text{mm}$  ( $\sigma = 0.61\text{mm}$ ) for SCC and  $1.22\text{mm}$  ( $\sigma = 0.92\text{mm}$ ) for HoH, while the non-expert intra-variability on the same images for HoH is  $3.58\text{mm}$  ( $\sigma = 0.98\text{mm}$ ). The resulting inter-observer variability for HoH is  $3.26\text{mm}$  ( $\sigma = 0.98\text{mm}$ ).

**Table 1.** Mean prediction error obtained with our algorithm for SCC, HoH and HT using the training set (38 images) and the testing set (9 images).

	error (training set)	error (testing set)
SCC	1.81mm ( $\sigma = 1.42\text{mm}$ )	2.46mm ( $\sigma = 1.92\text{mm}$ )
HoH	2.75mm ( $\sigma = 1.94\text{mm}$ )	3.70mm ( $\sigma = 1.48\text{mm}$ )
HT	0.26mm ( $\sigma = 0.51\text{mm}$ )	2.19mm ( $\sigma = 1.11\text{mm}$ )

The algorithm obtains performance comparable to the precision of the expert on the training set. As for the performance on the testing set, it is closer to the precision a non-specialist would reach.

## 6. CONCLUSION

We have presented a generic approach to address the issue of automatic landmarking medical images. We develop a Bayesian approach which results in the construction of a probabilistic template for the matter of the brain. Using a training set of images, it is easy to estimate the parameter of the generative model thanks to the EM-algorithm. Finally we showed that the gradient ascent of the log-likelihood can be computed efficiently to estimate the transformation mapping the landmarks in the template. The method we developed can be used either for specialists training or as a starting point for large deformation registration.

## 7. REFERENCES

- [1] S. Joshi and M. Miller, "Landmark matching via large deformation diffeomorphisms," in *IEEE Trans. in Image Processing*, 2000, vol. 9, pp. 1357–1370.
- [2] J. Talairach and P. Tournoux, *Co-planar stereotaxic Atlas of the Human Brain*, Thieme Medical Publishers, 1988.
- [3] J.-P. Thirion, "New feature points based on geometric invariants for 3d image registration," *Int. J. of Computer Vision*, vol. 18:2, pp. 121–137, 1996.
- [4] K. Rohr, *Landmark-based Image Analysis using Geometric and Intensity Models*, Kluwer Academic, Dordrecht, 2001.
- [5] Carey E. Priebe, Michael I. Miller, and J. Tilak Ratananather, "Segmenting magnetic resonance images via hierarchical mixture modelling," *Computational Statistics and Data Analysis*, 2004.
- [6] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Stat. Soc.*, vol. 39, pp. 1–38, 1977.