Robust Computation of Mutual Information Using Spatially Adaptive Meshes

Hari Sundar[1,2], Dinggang Shen[1], George Biros[1], Chenyang Xu[2], and Christos Davatzikos[1]

[1] Section for Biomedical Image Analysis, Department of Radiology, University of Pennsylvania
[2] Imaging and Visualization Department, Siemens Corporate Research, Princeton, NJ

Abstract. We present a new method for the fast and robust computation of information theoretic similarity measures for alignment of multi-modality medical images. The proposed method defines a non-uniform, adaptive sampling scheme for estimating the entropies of the images, which is less vulnerable to local maxima as compared to uniform and random sampling. The sampling is defined using an octree partition of the template image, and is preferable over other proposed methods of non-uniform sampling since it respects the underlying data distribution. It also extends naturally to a multi-resolution registration approach, which is commonly employed in the alignment of medical images. The effectiveness of the proposed method is demonstrated using both simulated MR images obtained from the BrainWeb database and clinical CT and SPECT images.

1 Introduction

Inter-modality image alignment is a fundamental step in medical image analysis. It is required to bring image data from different modalities to a common coordinate frame in order to accumulate information. It is usually presented as an optimization problem requiring the minimization of a certain objective function. Objective functions, or similarity measures based on information theoretic principles have been successful in aligning images from differing modalities. Mutual Information (MI) was proposed as an image similarity measure by Collignon [1], Viola [2] and Wells [3] and is widely used for rigid inter-modality registration. Several modifications have been proposed to make MI more robust and increase its capture range, including Normalized Mutual Information [4]. However, MI-based methods are very sensitive to the way the probability distributions are estimated and the accuracy of the estimated probability distributions have a great influence in the accuracy and robustness of the registration results [5].

A common assumption made in estimating the probability distribution is that each voxel is an i.i.d. realization of a random variable. Therefore, the probability distributions (including the joint distribution) can be computed by using all voxels in the reference image and the corresponding voxels in the transformed subject image. In general this can be quite expensive to compute and several multi-resolution and sampling techniques have been proposed for faster estimation of the probability distribution. Downsampling, both uniform and random, have been used quite commonly to speed up the estimation of the distributions [5]. Nonlinear sampling techniques where the local sampling rate is proportional to the gradient magnitude have also been proposed [6].

N. Ayache, S. Ourselin, A. Maeder (Eds.): MICCAI 2007, Part I, LNCS 4791, pp. 950–958, 2007.

In general, although these methods have performed quite well for registering different structural modalities (like CT and MR), they have been less successful in being able to register structural modalities to functional modalities,which is important for diagnosis and treatment planning applications. Functional modalities do not image all tissues and are therefore more sensitive to errors in the estimation of probability distributions. We shall use the example of registering Single Photon Emission Computed Tomography (SPECT) with CT images of the same patient to demonstrate this problem.

In this paper we present a new method for rigid alignment of multi-modality images using an octree based partitioning of the reference image. Octrees allow us to partition the image into spatially adaptive regions (octants) such that homogeneous regions produce large octants. The octree allows us to define a non-linear sampling of the image that is proportional to the underlying image complexity. The samples thus obtained are closer to the usual i.i.d. assumption in that they tend to be less statistically interdependent, which in turn help us obtain more accurate estimates of entropy. The MI is the sum of the entropy of the subject image and the entropy of subject conditional on the target. Consequently, improved entropy estimates provide better estimates of MI.

The rest of the paper is organized as follows. In Section 2 we present a brief introduction to estimating entropy in images, and lay the foundation for our arguments in favor of the octree-based estimation of distributions, which is described in Section 3. Section 4 discusses the incorporation of the octree-based mutual information metric into a registration framework for inter-modality alignment of images, within a multi-resolution framework. Experimental results and comparisons with other methods are provided in Section 5.

2 Estimating the Entropy of Images

Shannon's entropy [7] for a discrete random variable X with a probability distribution $\mathbf{p}(X) = (p_1, \cdots, p_n)$, is defined as,

$$H(X) \stackrel{\Delta}{=} -\sum_{i=0}^{n} p_i \log p_i. \tag{1}$$

The definition can be extended to images by assuming the image intensities to be samples from a high dimensional signal. A common assumption made while defining entropy for images is that each voxel is an i.i.d. realization and that the underlying probability of pixel intensities can be estimated via the normalized histogram. The probability of a pixel having an intensity y_i, $p(y_i) = \text{hist}_Y(y_i)/d$, where $\text{hist}_Y(y_i)$ is the number of voxels in image Y with intensity y_i and d is the total number of voxels. Equation 1 can be then used to estimate the entropy of the image. This however does not capture the spatial complexity of the underlying image. For example, if we shuffle the pixels within the image, we will lose structure and be left with an image which is spatially random. However since the histogram is unchanged, the entropy is the same. This is demonstrated in Figure 1, where Figure 1(a) is highly structured and the red and blue blocks are placed at regular intervals. It can be seen that although the image in Figure 1(b) is produced from that in Figure 1(a) by adding random noise to the position of the

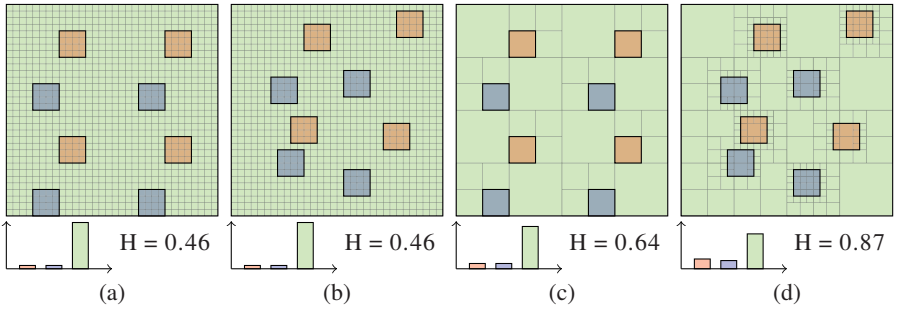


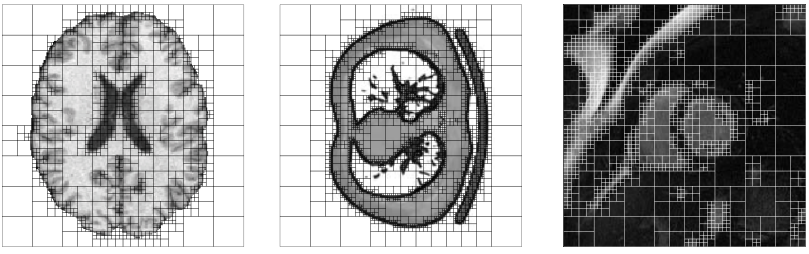
Fig. 1. The difficulty in estimating the entropy of images. **(a)** An image with the corresponding histogram and entropy estimate, **(b)** An image with objects moved within the image to make it appear more *random* than (a). The standard estimate for the entropy does not change, since the histogram is still the same. **(c)** Octree-based entropy estimation for the image shown in (a), and **(d)** octree-based entropy estimation for the image shown in (b). Note that in this case the increase in entropy was captured by the octree.

objects in the scene, the entropy is unaffected. This happens because of our assumption that each voxel intensity is an independent sample. This is not true since voxel intensities depend on the underlying objects, and samples from the same object cannot be assumed to be independent. Also observe that the gradient based approaches will not be able to capture this difference either, because it is not affected by the spatial variation between the two configurations shown in Figures 1(a) and (b).

It is widely accepted that images can be successfully modeled as Markov random fields [8,9]. Now, assuming that instead of an i.i.d. assumption, the samples form a Markov random field, then the error in the estimation of the entropy is lowered if the samples are largely independent of each other. Sabuncu et al. [6] suggest two non-linear sampling strategies based on the magnitude of the image gradient, in order to make the samples more independent. These suggested sampling methods, however, are for 2D images, expensive to compute and not easily extensible for 3D images. Algorithms have been proposed to make mutual information more robust by incorporating geometric information using gradients [10] and image salience [11]. However, the gradient captures only the local image complexity and it is not straightforward to extend this within a scale independent framework. This suggests using an adaptive sampling strategy, and adaptive image representations have been commonly used in non-linear image registration for representing the transformation [12] and for estimating local image similarities. However, to the best of our knowledge, it has not been used to estimate the global similarity of images.

3 Octrees Based Estimation of MI

If we can define a partitioning on the image that is spatially adaptive, and creates largely independent regions, we will be able to use the partition to define a better sampling strategy. Binary space partitions (BSP) [13] allow us to generate such spatially adaptive



(a) (b) (c)

Fig. 2. (a) An example of a quadtree generated from a 2D slice of a MR brain image demonstrating how the quadtree adapts to the underlying data, **(b)** an example of a quadtree generated from a 2D slice of a thoracic CT image, and **(c)** the quadtree generated from a cardiac short axis MR image

partitions of data by recursively subdividing a space into convex sets by hyperplanes. BSP trees can be used in spaces with any number of dimensions, but *quadtrees* [14] and *octrees* [15] are most useful in partitioning 2D and 3D domains, respectively. Quadtrees and octrees use axis aligned lines and planes, respectively, instead of arbitrary hyperplanes and are efficient for the specific domains they are intended to work on.

The octree structure introduces non-stationarity in space [16]. This is important since the resulting octree is not shift-invariant and can create block artifacts. Such artifacts have been reported before [17,18]. Approaches like generating octrees with overlapping leaves [19] could possibly be applied for estimating shift-invariant octrees. Making the samples shift-invariant is left for future work.

We use a standard top-down approach to construct the octree for a given image. Starting with the entire domain, we test each block to see if it meets some criterion of homogeneity. If a block meets the criterion, it is not divided any further. If it does not meet the criterion, it is subdivided and the test criterion is applied to those blocks. This process is repeated iteratively until each block meets the criterion. In the experiments reported in this paper, we used a simple intensity based test criterion of homogeneity, wherein a block is split if all the voxels within it are not within a specified threshold, which was based on the number of bins of the histogram used to estimate the probability distribution. An example of a quadtree constructed from a 2D slice of a MR image of the human brain, a 2D slice of a Cardiac CT image, and a cardiac short axis MR image are shown in Figure 2.

In our sampling method, the same number of samples are used per octant. In other words the density of octants specifies the sampling frequency. The octree-based entropy can be computed by using the following estimate of the probability distribution in (1),

$$p_i = \frac{\sum_x (T(x) \in \texttt{bin}(i))}{\sum_x 1}, \quad \forall x \in \text{Oct}(T), \tag{2}$$

where $\texttt{bin}(\cdot)$ defines the histogram binning operator and $\text{Oct}(T)$ is the octree computed for the template image T. In order to understand why an octree-based sampling appears to be better than uniform or gradient based sampling, consider the example discussed

earlier, shown in Figure 1. We showed earlier that traditional sampling methods (both uniform and gradient based) estimate the same entropy for the images shown in Figures 1(a) and (b). However, the octree-based sampling is able to capture this difference, as seen in Figures 1(c) and (d). This is because the spatial complexity of the octree increases as a result of the change in the randomness of the scene. The octree captures the spatial complexity of the scene and consequently the entropy of a complex scene is higher as it has a denser octree. It is important to point out that the octree is not necessarily the best estimate of spatial complexity, since it is partial towards objects that are axis aligned and will not be able to capture such variations. A better estimate of the scene complexity would be a connectivity graph of segmented objects. This however would increase the complexity of both computing such a graph and also in using it during the registration. It would be difficult and computationally expensive to define transformations and interpolation functions on a connectivity graph of segmented tissues. The octree is a compromise that provides a sufficiently accurate and robust estimate of the entropy (see Section 5) and is also easy to compute and manipulate. Interpolation and transformations (linear and non-linear) can be easily defined on the octree. Importantly, octree-representations are amenable to parallel computing, which can dramatically expedite the performance of algorithms that use it. In all the experiments reported in this paper, the octree was computed only for the template image and each octant was sampled at the center of each octant in the template image. Although, it would be better to average over the entire octant instead of sampling at the center, we opted for the latter to improve the computational efficiency of the method.

4 Rigid Inter-modality Registration Using Adaptive MI

Given a template image, $T : \Omega \rightarrow \mathbb{R}^n$ and a subject image $S : \Omega \rightarrow \mathbb{R}^n$, where $\Omega \in \mathbb{R}^d$, the goal is to determine a rigid transformation χ that aligns the two images. The similarity measure is the metric that quantifies the degree of alignment and allows us to present the alignment problem as an optimization problem. We use the octree-based similarity measures as described in Section 3. We formulate the determination of the optimal transformation, $\hat{\chi}$, as an optimization problem:

$$\hat{\chi} = \arg\max_{\chi} I(T(x), S(\chi(x))), \quad \forall x \in \text{Oct}(T), \tag{3}$$

where, $I(.;.)$ is the Mutual Information. Powell's multidimensional set method [20] is used to iteratively search for the maxima along each parameter using Brent's method [20]. In order to increase the robustness and to improve speed, we use a multi-resolution framework defined on the octree. The octree at lower (coarser) resolutions are generated by skipping all octants at the finer levels.

5 Results

In this section we describe experiments that were carried out to test the effectiveness of octree-based MI in the rigid registration of inter-modality images. We first describe

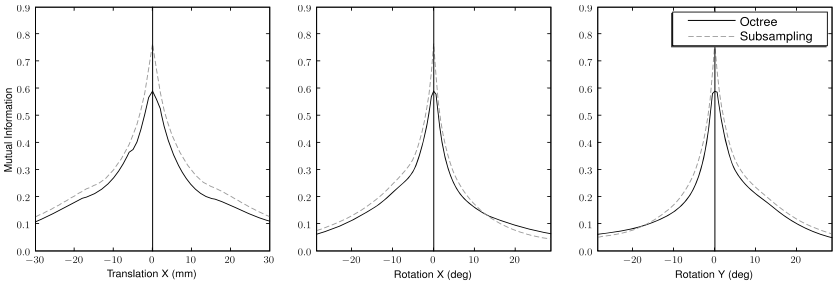


Fig. 3. Comparison of the mutual information computed via uniform sampling (dotted lines) and using the proposed octree-based sampling (solid lines), on BrainWeb datasets. The plots shown are for a comparison between a T1-weighted (T1) image and a proton density (PD) image with 9% noise and 40% intensity non-uniformity.

the similarity profiles when an artificial transformation is introduced between two registered images. We compared the octree-based method with uniform sampling based estimation of mutual information. The first experiment was performed using simulated MR datasets obtained from the BrainWeb database [21]. The second experiment was performed with 13 CT datasets with corresponding SPECT images. These images were all acquired using a Siemens Symbia™ T, a TruePoint SPECT-CT system and are assumed self registered. We analyzed the mutual information profiles while varying the transformation. The transformation parameters were varied one at a time, and the similarity profiles were plotted. The plots for translation along the x-axis, and for rotation about the x and y axes are shown in Figures 3 and 4, for T1-PD MR images and CT-SPECT images, respectively[1]. The profiles for translation and rotation along the other axes were similar. In all cases we compare the octree-based sampling with uniform sampling, where the total number of samples are similar. The octree reduced the number of samples by a factor of 8 on an average, therefore we subsampled by a factor of 2, along each direction, for the uniform sampling strategy, to have the same number of samples in both cases. As can be seen from Figure 3, both methods perform equally well on the BrainWeb datasets. Both sampling techniques have smooth curves with sharp peaks and very good capture ranges. However, when we look at the results from the CT-SPECT comparison, shown in Figure 4, we observe that the octree-based sampling performs much better. Although, both approaches have good profiles subject to translation, for the profiles subject to rotation, the uniform sampling approach exhibits a weak maxima at the optimal value with a very small capture range. In contrast the octree-based approach exhibits a strong maximum at the optimal value and also has a much larger capture range. The fact that the neighboring maxima in the vicinity of the optimum are lower further implies that it is likely that a multi-resolution approach can potentially be used to increase the capture range. The uniform sampling approach will in most cases converge to the wrong result since the neighboring maxima are much larger in that case.

[1] The profiles generated from using all the voxels in the image were almost identical to those obtained by uniform subsampling, and are not presented here for clarity.

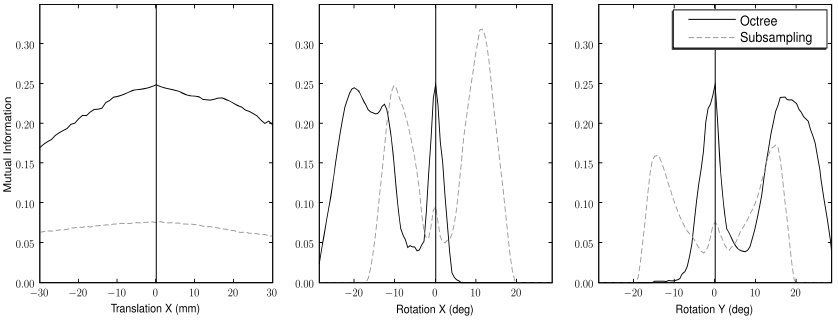


Fig. 4. Comparison of the mutual information computed via uniform sampling (dotted lines) and using the proposed octree-based sampling (solid lines), with CT-SPECT datasets. The plots shown are for a comparison between a CT cardiac image ($512 \times 512 \times 25$) and a SPECT image ($128 \times 128 \times 128$).

Table 1. Means and standard deviations of the registration errors for the different test cases

	Uniform sampling			Octree-based		
Dataset	Success	Trans. error (mm)	Rot. error (deg)	Success	Trans. error (mm)	Rot. error (deg)
T1 - T2	82.4%	0.48 ± 0.63	0.17 ± 0.24	86.1%	0.53 ± 0.59	0.21 ± 0.19
T1 - PD	79.7%	0.57 ± 0.66	0.2 ± 0.33	81.3%	0.59 ± 0.62	0.22 ± 0.23
CT - SPECT	31.1%	0.73 ± 0.69	0.23 ± 0.28	68.5%	0.64 ± 0.57	0.21 ± 0.31

Registration was performed on a number of datasets to quantitatively assess the performance of the octree-based mutual information within the registration framework. We selected a T1-weighted image with no noise and uniform intensity as the template image. T2-weighted and Proton Density (PD) images with varying levels of noise ($0-9\%$) and intensity non-uniformity ($0 - 40\%$) were registered to the template image, with a pseudo-random initial transform applied. The random transform was selected such that the initial translation was at most half the size of the image (to ensure overlap) and the rotational components were less than $60°$. The same set of pseudo-random transformations were used for both methods. The registration was considered successful if the final error was less than 2mm for the translational parameters and less than $2°$ for the rotational parameters. Similar experiments were performed for the CT-SPECT dataset and are summarized in Table 1. The error in the estimation of the translation and rotation parameters is calculated using only the cases in which a successful registration was performed. We can see from Table 1, that the octree-based sampling performs slightly better than uniform sampling in case of T1-T2 and T1-PD registration. We would like to emphasize that the success rate for the CT-SPECT registration is much better using the octree-based sampling as compared to uniform sampling, owing mainly to the broader capture range of the octree-based method. The average time to perform the registration was 13 seconds using the octree-based approach, 14 seconds for uniformly sampled approach and 85 seconds when using all voxels. All experiments were performed on an Intel Xeon 2.8GHz with 2GB of RAM. The time for the octree-based method includes the time to compute the octree.

6 Conclusion

We have presented a spatially adaptive sampling method for the estimation of image entropies and mutual information. We also demonstrate the improvements in the rigid registration of CT and SPECT images using the new sampling method. The proposed sampling offers flexibility between robustness and speed. When used at full resolution, octree-based sampling provides better estimates of image entropy and mutual information between images. In addition, it is better to use octree-based sampling in order to speed up the estimation of the similarity measure as opposed to using uniform sampling approaches. When compared with uniform sampling approaches, the octree-based sampling was more accurate for comparable computational speeds, and appears to be a better method for speeding up the estimation of MI.

Acknowledgments

The authors would like to thank Dr. Parmeshwar Khurd for useful discussions. This research was partially funded by a research grant from Siemens Corporate Research, Princeton, NJ.

References

1. Collignon, A., et al.: Automated multimodality medical image registration using information theory. Proc. Information Processing in Medical Imaging 3, 263–274 (1995)
2. Viola, P.: Alignment by Maximization of Mutual Information. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts (1995)
3. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1), 35–51 (1996)
4. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 32(1), 71–86 (1999)
5. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE Trans. on Medical Imaging 22(8), 986–1004 (2003)
6. Sabuncu, M.R., Ramadge, P.J.: Gradient based nonuniform sampling for information theoretic alignment methods. In: Proc. Int. Conf. of the IEEE EMBS, San Francisco, CA, IEEE Computer Society Press, Los Alamitos (2004)
7. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27 (1948)
8. Li, S.Z.: Markov random field modeling in computer vision. Springer, Heidelberg (1995)
9. Prez, P.: Markov random fields and images. CWI Quarterly 11(4), 413–437 (1998)
10. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Image registration by maximization of combined mutual information and gradient information. IEEE Trans. on Medical Imaging 19(8), 809–814 (2000)
11. Luan, H., Qi, F., Shen, D.: Multi-modal image registration by quantitative-qualitative measure of mutual information Q-MI. In: Proc. CVBIA, pp. 378–387 (2005)
12. Timoner, S.: Compact Representations for Fast Nonrigid Registration of Medical Images. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts (2003)
13. Fuchs, H., Kedem, Z.M., Naylor, B.F.: On visible surface generation by a priori tree structures. In: Proc. SIGGRAPH, pp. 124–133 (1980)

14. Finkel, R.A., Bentley, J.L.: Quad trees: A data structure for retrieval on composite keys. Acta. Inf. 4, 1–9 (1974)
15. Meagher, D.: Geometric modeling using octree encoding. Computer Graphics and Image Processing 19, 129–147 (1982)
16. Laferté, J., Perez, P., Heitz, F.: Discrete Markov image modeling and inference on the quadtree. IEEE Trans. on Image Processing 9(3), 390–404 (2000)
17. Bouman, C.A., Shapiro, M.: A multiscale random field model for Bayesian image segmentation. IEEE Trans. on Image Processing 3(2), 162–177 (1994)
18. Luettgen, M.R., Karl, W.C., Willsky, A.S.: Efficient multiscale regularization with applications to the computation of optical flow. IEEE Trans. on Image Processing 3(1), 41–64 (1994)
19. Irving, W.W., et al.: An overlapping tree approach to multiscale stochastic modeling and estimation. IEEE Trans. on Image Processing 6(11), 1517–1529 (1997)
20. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, New York (1992)
21. Collins, D.L., et al.: Design and construction of a realistic digital brain phantom. IEEE Trans. on Medical Imaging 17(3), 463–468 (1998)