TEXTURAL MUTUAL INFORMATION BASED ON CLUSTER TREES FOR MULTIMODAL DEFORMABLE REGISTRATION

Mattias P. Heinrich[1,2]*, Mark Jenkinson*[2]*, Sir Michael Brady*[3]*, Julia A. Schnabel*[1]

[1]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK
[2]Oxford University Centre for Functional MRI of the Brain, UK
[3]Department of Oncology, University of Oxford, UK

ABSTRACT

Mutual information (MI) has been widely used in image analysis tasks such as feature selection and image registration. In particular, it is the most widely used similarity measure for intensity based registration of multimodal images. However, a major drawback of MI is that it does not take the spatial neighbourhood into account. An effective way of incorporating spatial information could be of great benefit in a number of challenging applications. We propose the use of cluster trees to efficiently incorporate textural information from the local neighbourhood of a voxel into the computation of MI, while at the same time limiting the number of bins used to represent this higher-order information. This new similarity metric is optimised using a Markov random field (MRF). We apply our new method to the registration of dynamic lung CT volumes with simulated contrast. Experimental results show the advantages of this technique compared to standard mutual information.

Index Terms— multimodal image registration, mutual information, cluster trees

1. INTRODUCTION

Mutual information (MI) was first introduced as a similarity measure for the rigid alignment of medical images [1][2]. It is derived from information theory and measures the statistical dependency of the intensity distributions of two images. MI is defined between two images I and J as the difference between both marginal entropies and their joint entropy based on their image intensity distributions:

$$\mathrm{MI}(I,J) = \sum_{i\in I}\sum_{j\in J} p(\mathbf{i}) \log \frac{p(\mathbf{i})}{p(i)p(j)} \qquad (1)$$

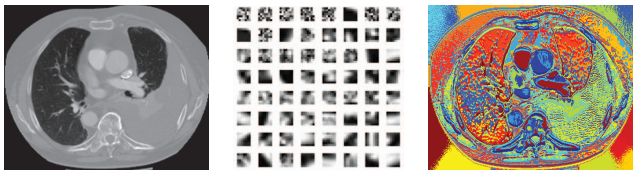
where $p(\mathbf{i})$ is the joint probability of the co-occurrence of an intensity pair $\mathbf{i} = (i,j)^T$ in the two images I and J and the two marginal probabilities are $p(i)$ and $p(j)$. MI has been applied to a range of applications for medical image registration [3]. Its robustness and registration accuracy has been demonstrated in particular for rigid body registration of multimodal images.

We aim to align multimodal lung images and compensate for respiratory motion. However, adapting MI for deformable registration is not straightforward and remains an active area of research. Several weaknesses have been identified when MI is used for non-rigid registration [4]. It is especially sensitive to the initial estimation of the joint histogram and is often susceptible to local minima during optimisation. An inherent limitation of the original formulation is the absence of local information for an image intensity pair. This leads to a negative influence of image noise, non-uniform bias fields, and missing data. As a result, incorporating spatial information has attracted much interest. Rueckert et al. [5] first introduced additional spatial information into the joint histogram estimation, by taking the intensity of one spatial neighbour per voxel into account. However, this increases the dimensionality of the histogram and limits the approach to a very small spatial neighbourhood, since one would otherwise be faced with the curse of dimensionality. In [4], an extension to MI for deformable registration is proposed by introducing a third channel representing a regional label (based on the local position) and improved accuracy in the presence of non-uniform intensity variations are demonstrated. Very recently, Yi and Soatta [6] presented a method which organises image patches into orbits under the action of Euclidean transformations and in this way introduces spatial context into MI. However, the approach is computationally very complex and so far limited to rigid or 2D registrations.

2. METHOD

In this work we propose a novel image similarity measure: *textural mutual information* (tMI), which incorporates intensity information from local neighbourhoods. Our approach can effectively include both small local neighbourhoods and the local position of a corresponding intensity pair into the

We would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre. JAS also acknowledges funding from EPSRC EP/H050892/1.



(a) CT lung scan (b) Textons (c) Clustered image

Fig. 1. Overview of our method (shown on an example CT slice). A representative texton dictionary is learnt from the image using hierarchical tree clustering. Each pixel is then assigned to the closest texton using nearest neighbour search (pixels with the same colour belong to the same cluster).

estimation of the joint histogram.

In order to achieve this, we propose a novel approach that makes use of cluster trees to reduce the dimensionality of the histogram. A metric tree is built for each image based on the distance of patches, which effectively divides the image into a number of clusters. Each leaf node of the tree then contains similar textural patches. The representative textons form a dictionary, where each texton directly corresponds to a bin in the histogram. The similarity between two images is then measured as the overlap of similar textures.

2.1. Textons

Mutual information, as a multimodal similarity measure, is based on the assumption that voxels of corresponding anatomical structures are represented by a common intensity pair. However, due to degradations of medical images this is not fulfilled in real life. Imaging-related artefacts can cause a complex intensity distribution within the same tissue. Our approach is motivated by the fact that although single intensities might provide only a limited representation of the underlying anatomical structure, a patch including several neighbouring voxels can more effectively capture the texture and thus better describe the actual situation.

Textural representation and classification has previously mainly employed the responses of high-level filter banks (e.g. Gabor filters [7]). A filter bank approach has the advantages of representing a large support area while still having a low-dimensional feature vector. However, the choice of filters is crucial and has to be learnt for each specific image. In [8], it is shown that an excellent representation of texture can be directly obtained by using small image patches (as small as 3x3). In their application of texture classification, so called textons are learnt based on segmented supervised training data and clustered forming a texton dictionary. In the next section, we present our approach, which employs hierarchical tree clustering to obtain a representative texton dictionary.

2.2. Cluster trees

Finding similar patches using a tree structure has been extensively studied and can be usually performed with $\mathcal{O}(n \log n)$ complexity. We use the vantage-point tree, which was introduced by Yianilos [9] and achieved the best results in a recent comparison of clustering methods [10] in terms of computational complexity for both the clustering and the retrieval of image patches. First, a pivot is chosen (an optimized pivot is selected using the element, which results in the largest spread for a random subset) and the distances to all remaining elements is calculated. The elements are then split into two equal-sized branches based on the median distance. These steps are repeated recursively until a fixed minimum leaf size is reached. The distance metric used in our experiments between two patches $\mathcal{P}$ within the same image I located at $\mathbf{x}$ and $\mathbf{y}$ is defined as their sum of squared differences, and a spatial distance can be added using a weighting term λ:

$$d(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{x}' \in \mathcal{P}} \left(I(\mathbf{x}-\mathbf{x}') - I(\mathbf{y}-\mathbf{x}')\right)^2 + \lambda ||\mathbf{x}-\mathbf{y}|| \quad (2)$$

where $||\mathbf{x}-\mathbf{y}||$ defines the Euclidean distance between the two voxels $\mathbf{x}$ and $\mathbf{y}$.

The median element within a leaf node is selected as a representative texton and stored in the texton dictionary T. We then assign the texton label $t(\mathbf{x}) \in T$ to each voxel $\mathbf{x}$, which has closest distance d (see example in Fig. 1 (b,c)). As Eq. 2 describes a metric, the triangle inequality is used to accelerate the search for the closest textons. Additionally, we use the second closest textons for the calculation of the joint histogram (we also use the two closest intensity bins to obtain the conventional intensity histogram).

2.3. Local estimation of mutual information

We need to evaluate the similarity function at each location for which the local derivation of mutual information is used. Rogelj et al. [11] describe a straightforward estimation of a point-wise measure based on a global joint distribution using the simplifying assumption that the real marginal and joint distributions can be assumed to be very similar to the initial estimation based on the non-registered images.

In Eq. 1 the global MI is obtained by summing up all contributions over the intensity variables. For a local point-wise evaluation of standard MI (sMI), the order of summation is reversed, yielding:

$$\mathrm{sMI}(I,J) = \sum_{\mathbf{x}\in\Omega} \log \frac{p\left(I(\mathbf{x}), J(\mathbf{x})\right)}{p\left(I(\mathbf{x})\right) p\left(J(\mathbf{x})\right)} \cdot \frac{1}{c} \quad (3)$$

$$c = \sum_{\mathbf{x}\in\Omega} p(I(\mathbf{x})) \log(p(I(\mathbf{x}))) \quad (4)$$

the global entropy c is used to obtain normalised MI (nMI)[1].

[1] In our MRF registration framework, where the global histogram is only calculated once, c is only a constant, so nMI and sMI are equivalent.

For the case of textural MI (tMI), the similarity between two images is now defined by the statistical dependency of their texton representations t^I and t^J.

$$\text{tMI}(I,J) = \sum_{\mathbf{x}\in\Omega} \log \frac{p\left(t^I(\mathbf{x}), t^J(\mathbf{x})\right)}{p\left(t^I(\mathbf{x})\right) p\left(t^J(\mathbf{x})\right)} \tag{5}$$

2.4. Non-rigid MRF registration framework

We adapt our previously-developed, efficient, discrete optimisation framework for non-rigid registration[12]. It has the advantages that no derivatives of the cost function are necessary, and the images have neither to be resampled nor warped. It is based on an MRF energy minimisation scheme for stereo reconstruction using hierarchical belief propagation [13], together with a recent extension called *constant space belief propagation* (CSBP), which additionally reduces complexity by reducing the search space hierarchically [14].

The corresponding graph consists of a set of nodes (which correspond to a set of pixels $p \in P$ in an image). Each random variable corresponds to a node and takes values from a set of labels $f_p \in \mathcal{L}$, (which correspond to displacements). We aim to find the label with maximum posterior probability (MAP). The energy cost is given by:

$$E(f) = \sum_{p\in P} S_p(f_p) + \sum_{(p,q)\in N} R(f_p, f_q) \tag{6}$$

$S_p(f_p)$ is the cost of the similarity term (see Eq. 4 and 5) for assigning label f_p to pixel p. $R(f_p, f_q)$ measures the pairwise cost of assigning labels f_p and f_q to two neighbouring pixels, and is equivalent to a regularization term. To perform the inference on the MRF, we use the max-product belief propagation (BP) message passing. The update at time t is found by calculating:

$$m^t_{p\to q}(f_q) = \min_{f_p}(S_p(f_p) + R(f_p, f_q) + \sum_{s\in N(p)\setminus q} m^{t-1}_{s\to p}(f_p)) \tag{7}$$

After T iterations, the label that minimizes the final belief vector $b_q(f_q) = S_q(f_q) + \sum_{p\in N(q)} m^T_{p\to q}(f_q)$ is selected.

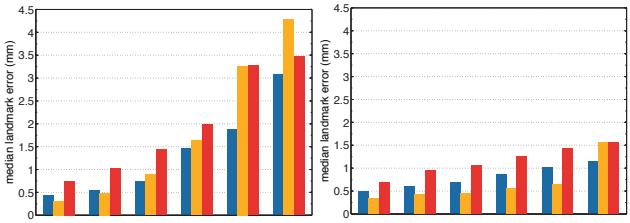
3. EXPERIMENTS

3.1. Local minima detection

For the first experiment, we automatically select around 100 landmarks in an axial slice of a CTPA scan (see Fig. 1 (a)). A multimodal scenario with non-linear intensity relation is simulated by generating a second image for which the image intensities are transformed by the square root of their inverse.

We then calculate the proposed pointwise similarity measures in the second image within a search region of a radius of 12 mm around each landmark (which ensures the displaced landmark is within the search region). We average the cost function over a patch of 7x7 pixels, as we do not apply any regularization in this experiment. The minimum of the cost function is selected and compared with the known displacement of the landmark. The median Euclidean distance of all landmarks and their gold standard positions are evaluated for three scenarios: increasing non-rigid deformations (using random displacements of B-spline control points), additive Gaussian noise and non-uniform bias fields. We tested a range of values for the histogram computation for both approaches, varying the number of bins between 64, 128, 256 and selected the best parameter for each degradation. A patch size of 3x3 voxels was found to be best for the texton computation (the weighting for the spatial coordinate λ was set to 0).

The results are displayed in Fig. 2. We obtain very similar results with both methods for the highest image quality. It is observable that sMI deteriorates for stronger image degradations in all three test cases, while our proposed method maintains very high accuracy even for the lowest image quality. This demonstrates how well our approach can model the underlying image structures when more informative patches are used, compared to individual intensities.



(a) standard MI (sMI) (b) textural MI (tMI)

Fig. 2. Resulting median landmark localisation error (in mm). x-axis defines increasing ■ Gaussian noise (PSNR: $34-18$ dB), ■ non-uniform multiplicative bias field (90^{th}: percentile $1-1.5$) and ■ non-rigid deformations ($\mu = 0-5$ mm).

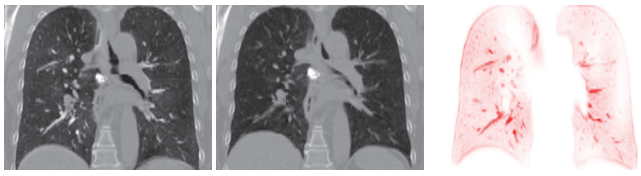


Fig. 3. Example of dynamic 3D registration experiment. Left: Coronal plane of lung CT scan with locally varying contrast uptake (maximum inhale). Centre: Coronal plane without contrast (maximum exhale). Right: Simulated contrast.

3.2. Registration of 3D CT with simulated contrast

To validate our approach on real clinical data, we perform registrations (using the MRF framework described in Section 2.4) between the two extreme breathing phases in dynamic

lung CT volumes. The data set consists of 5 dynamic lung CT image volume sequences acquired for a standard treatment planning process and has 300 expert annotated anatomical landmarks [15]. Additionally, we simulate a contrast agent with locally varying uptake in the pulmonary vessels. Contrast enhanced scans are commonly performed in cancer imaging to examine the heterogeneity of tumors and lung nodules and pose a multimodal registration problem. An example of the simulated contrast images is shown in Figure 3. Results for both standard and textural MI are given in Table 1. We used 128 histogram bins for both methods and a Parzen window kernel of size 5x5 with a standard deviation of 0.5 for the joint entropy estimation in sMI. The registration error at landmarks was on average 0.3 mm smaller for our method than for standard MI.

Table 1. Target registration error (for 300 manual landmarks) for 3D CT scans with simulated contrast (all results in mm). Voxel dimensions are $\approx$1x1x2.5 mm. Average computation times are 60 minutes for sMI and 70 minutes for tMI.

Dataset	**before reg.**	**sMI**	**tMI**
Case 1	3.89±2.78	2.62±3.71	1.70±1.84
Case 2	4.34±3.90	1.91±3.32	1.95±3.64
Case 3	6.94±4.05	1.94±2.49	2.07±2.19
Case 4	9.83±4.86	3.45±4.03	3.29±4.20
Case 5	7.48±5.51	2.78±3.79	2.28±2.58
Average	**6.50 mm**	**2.54 mm**	**2.26 mm**

4. CONCLUSION

We have presented a novel way to incorporate information from a local neighbourhood into the computation of mutual information. This approach enables inclusion of both neighbouring intensities and the spatial location of a voxel, while limiting the number of histogram bins by using a hierarchical tree clustering. A representative texton dictionary is obtained for each image using vantage-point trees. The joint histogram is calculated using a fast nearest neighbour search for all image patches. Synthetic experiments show that our approach is able to locate landmarks in a local search region more accurately than standard MI and is less susceptible to non-rigid deformations, additive noise and non-uniform bias fields. Registrations of 3D dynamic CT scans using a discrete MRF optimization framework show in average a lower target registration error for manual expert landmarks using our proposed textural mutual information. In the future we will further evaluate the benefits of our method on more challenging real multimodal datasets.

5. REFERENCES

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, 1997.

[2] P. Viola and W.M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[3] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, 2003.

[4] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IPMI*, pp. 725–737, 2007.

[5] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes, "Non-rigid registration using higher-order mutual information," 2000, vol. 3979, pp. 438–447, SPIE.

[6] Z. Yi and S. Soatto, "Multimodal registration via spatial-context mutual information," in *IPMI*, 2011.

[7] Y. Ou and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," in *Proc. IPMI*, 2009, pp. 50–62.

[8] A. Varma and A. Zisserman, "Texture classication: Are filter banks necessary?," in *Proc. CVPR*, 2003.

[9] P.N. Yianilos, "Data structures and algorithms for nearest neighbour search in general metric spaces," in *Proc. ACM-SIAM*, 2003.

[10] N. Kumar, L. Zhang, and S. Nayar, "What is a good nearest neighbors algorithm for finding similar patches in images?," in *Proc. ECCV*, 2008, pp. 364–378.

[11] P. Rogelj, S. Kovacic, and J.C. Gee, "Point similarity measures for non-rigid registration of multi-modal data," *Comput. Vis. Image Und.*, vol. 92, no. 1, pp. 112 – 140, 2003.

[12] M.P. Heinrich, M. Jenkinson, J.M. Brady, and J.A. Schnabel, "Non-rigid image registration through efficient discrete optimization," in *MIUA*, 2011, pp. 1–5.

[13] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, pp. 41–54, 2006.

[14] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," in *Proc. CVPR*, 2010, pp. 1458–1465.

[15] R. Castillo, E. Castillo, R. Guerra, V.E. Johnson, T. McPhail, A.K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.*, vol. 54, no. 7, pp. 1849, 2009.