Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 90 (2016) 194 – 199

Procedia Computer Science

International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK

FASTR: Using Local Structure Tensors as a Similarity Metric

James M. Sloan[a,b,*], Keith Goatman[a], Paul Siebert[b]

[a] *Toshiba Medical Visualisation Services Europe, Ltd., Leith, Edinburgh, EH6 5NP, Scotland*
[b] *University of Glasgow, Kelvingrove, Glasgow, G12 8QQ, Scotland*

Abstract

We describe a novel structural image descriptor for image registration called the Fractionally Anisotropic Structural Tensor Representation (FASTR), calculated from the local structural tensor (LST). The metric has several characteristics that are advantageous for multi-modality registration, such as not depending on absolute voxel intensities, and being insensitive to slowly varying intensity inhomogeneities across the image. This latter property is very useful, since many imaging modalities suffer from such artefacts. Registration accuracy is tested on both computed tomography (CT) to cone-beam CT (CBCT) rigid registration, and CT to magnetic resonance (MR) rigid registration. The performance is compared with Mutual Information (MI) metric and the Self Similarity Context (SSC) descriptor. The results show that, for images with significant intensity inhomogeneity, FASTR produced more accurate results than MI, and faster results than SSC. The results suggest FASTR gives similar benefits in images with intensity inhomogeneity, but at a fraction of the computation and memory demand.

Peer-review under responsibility of the Organizing Committee of MIUA 2016

Keywords: Medical; image; registration; FASTR; Structure; Tensors ; Similarity; Metric

1. Introduction

Medical image registration is concerned with the automatic alignment of multiple datasets to a common space. It is an essential component in a diverse array of applications, including diagnosis, treatment planning, atlas construction and augmented reality.

One categorisation of medical image registration concerns whether the datasets were acquired using the same imaging modality (i.e. *mono-modality*), or using different imaging modalities (i.e. *multi-modality*). In general, multi-modality registration is a harder problem, since different tissue types can have vastly different appearances (intensity, contrast, noise properties etc.) in each modality. Indeed, when aligning a functional and a non-functional imaging modality — such as Positron Emission Tomography (PET) and Computed Tomography (CT) — there may be no visible correlation between many parts of the images.

* Corresponding author.
E-mail address: JSloan@tmvse.com

Peer-review under responsibility of the Organizing Committee of MIUA 2016
doi:10.1016/j.procs.2016.07.015

A key component of any registration algorithm is some means for determining a figure of merit for how well aligned the images are. In general this is achieved using a similarity metric, which assesses the similarity between a reference image, R, and a test image, T. Typical similarity metrics include the sum of squared differences (SSD) between corresponding voxels, or the Mutual Information (MI)[1,2] between R and T. More recently, various multi-dimensional descriptors have been proposed to quantify the relationship between R and T. Examples of such descriptors include Normalised Gradient Fields (NGF)[3], Modality Independent Numerical Descriptors (MIND)[4], and Self Similarity Context (SSC)[5]. In this paper we propose a new descriptor named *Fractionally Anisotropic Structural Tensor Representations* (*FASTR*), and a similarity metric based upon it, which has a number of properties advantageous to multi-modality registration. Specifically:

- Similar to the NGF, since it is based on local gradient orientations the metric does not rely on absolute intensities.
- Furthermore, since it aligns parallel and anti-parallel gradients, it handles cases where the gradient is in the opposite direction in the two images (i.e. where a boundary goes from light to dark in one modality, but dark to light in the other).
- Since the vector field is estimated locally, the metric is robust to global illumination inhomogeneity, e.g. bias-field artefacts in magnetic resonance (MR) or cone-beam CT (CBCT) images.

2. Method

FASTR is based on *Local Structure Tensors*[6] (LST). These are positive semi-definite matrices which describe the distribution of gradients within a given image neighbourhood, $\mathcal{N}$. The LST is calculated at voxel $\mathbf{x}$, in an image I, within a neighbourhood $\mathcal{N}$ using

$$LST(I, \mathbf{x}, \sigma_{\mathrm{LST}}) = \sum_{i \in \mathcal{N}} w(i, \sigma_{\mathrm{LST}}) \begin{pmatrix} \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial x} & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial z} \\ \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial y} & \frac{\partial I}{\partial y} \cdot \frac{\partial I}{\partial z} \\ \frac{\partial I}{\partial z} \cdot \frac{\partial I}{\partial x} & \frac{\partial I}{\partial z} \cdot \frac{\partial I}{\partial y} & \frac{\partial I}{\partial z} \cdot \frac{\partial I}{\partial z} \end{pmatrix} \tag{1}$$

where $w(i, \sigma_{\mathrm{LST}})$ is a weighting function that decreases monotonically with distance from the centre of $\mathcal{N}$. In the following experiments $w(i, \sigma_{\mathrm{LST}})$ is a Gaussian function with standard deviation σ_{LST}. The size of the neighbourhood region, $\mathcal{N}$, is chosen such that $w(i, \sigma_{\mathrm{LST}}) \approx 0$ at the edge of $\mathcal{N}$.

2.1. Computing the FASTR descriptor

The fractional anisotropy (FA) may be calculated from the eigenvalues of the LST: $\lambda_1, \lambda_2, \lambda_3$ (where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$). The principal eigenvector, $\mathbf{v}_1$, points in the dominant direction of the gradient vectors, ∇I, of the local patch $\mathcal{N}$. $\mathbf{v}_2$ and $\mathbf{v}_3$ are discarded as they are — by definition — perpendicular to the principal eigenvector.

The strength of the dominant direction (i.e. how dominant it is) may be calculated from the coherence for 2D images, and the FA for 3D volumes. Both coherence and FA share the following common set of properties[6]:

1. They are scalar, and are bounded between 0 and 1.
2. If they equal 0, there is no dominant direction within the local neighbourhood $\mathcal{N}$.
3. If they equal 1, there is an absolute dominant direction within the local neighbourhood $\mathcal{N}$, i.e. all non-zero second order gradient vectors are parallel.

The FA may be computed straightforwardly from the eigenvalues λ_1, λ_2 and λ_3 of the LST:

$$\mathrm{FA} = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{{\lambda_1}^2 + {\lambda_2}^2 + {\lambda_3}^2}}. \tag{2}$$

By multiplying the FA and principal eigenvector, $\mathbf{v}_1$, the FASTR descriptor can be computed for the voxel $\mathbf{x}$:

$$\text{FASTR}(I, \mathbf{x}) = FA \cdot \mathbf{v_1} \,. \tag{3}$$

The contribution to the overall metric for each corresponding voxel location $\mathbf{x}$ in R and T is given by the square of the dot product of the FASTR descriptors of R and T, i.e.

$$\text{Cost}(R, T, \mathbf{x}) = \langle \text{FASTR}(R, \mathbf{x}), \text{FASTR}(T, \mathbf{x}) \rangle^2 \tag{4}$$

The squared dot product is a metric that increases as the images become more similar, since as the FASTR vectors in the two images become more parallel/anti-parallel, the square of the dot product increases. Note that it is common to treat parallel and anti-parallel vectors in the same way to assess similarity, such as performed by Haber et al.[3] with their NGF descriptor and its numerous variants.

2.1.1. Influence of σ_{LST}

Similarity metrics often have one or more parameters which control their behaviour. For example, calculating mutual information from joint histograms requires decisions about the number of histogram bins; the normalised gradient fields have a critical parameter in the denominator to account for image noise; MIND and SSC require a neighbourhood size and displacement. FASTR has just a single parameter to consider, σ_{LST}, which intuitively matches the scale of features being aligned. Figure 1 shows the effect of changing σ_{LST}, shown in the FA images computed using different σ_{LST}. These images show the different scales of features captured using different values for σ_{LST}. Note that in this implementation, the size of $\mathcal{N}$ is solely dependent on σ_{LST}, and is therefore not an independent parameter.

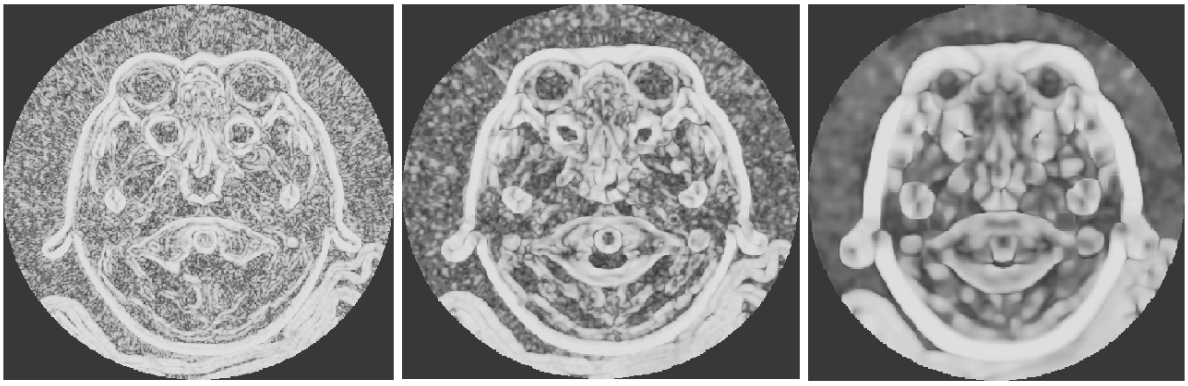

Fig. 1. Example fractional anisotropy images from a MR-T2 weighted head dataset, calculated using $\sigma_{\text{LST}} = 1$ (left), $\sigma_{\text{LST}} = 2$ (centre) and $\sigma_{\text{LST}} = 5$ (right). Example image data courtesy of Dr Subash Thapa, Radiopaedia.org, rID: 40310.

3. Experiments

To demonstrate the performance of FASTR in multi-modality, rigid registration, the method was applied to a series of head and neck images: first to CT and CBCT datasets; and secondly to MR and CT datasets. Although CT and CBCT scanners are both based on the same physical principle, i.e. X-ray attenuation, the appearance of the images is sufficiently different to consider this a multi-modality registration task. Two important differences between CT and CBCT scanners lead to a greater number, and severity, of artefacts in the CBCT images. Specifically, less collimation in the CBCT detector and a smaller field-of-view (truncating some of the projections) result in much higher scatter fraction (and therefore lower contrast) and shadow artefacts.

The FASTR method was tested with varying values for σ_{LST}: 1, 2.5 and 5, and compared with two state-of-art and widely used similarity metrics, MI[2] and SSC[5]. For the SSC, a Gaussian kernel variety was used (see section 3.1 of Heinrich et al.[5]) to aggregate neighbourhood responses, with $\sigma_{\text{SSC}} = 0.8$ and displacement $\Delta = 2$. The SSC similarity metric was calculated as the sum of squared differences between SSC(R) and SSC(T).

All registrations were done in a multi-scale fashion (1/8, 1/4 and 1/2 scales) with Newtonian optimisation of the cost function. For MI, pre-determined optimal sizes of the joint histogram were used: 160×160 bins at the finest scale of the image pyramid, and 32×32 bins otherwise.

Registration accuracy was assessed using the target registration error (TRE), calculated from a set of manually placed corresponding landmarks. There was an average of 4 landmarks per dataset for the CT $\rightarrow$ CBCT registration experiments, and an average of 17 landmarks per dataset for the CT $\rightarrow$ MR registration experiments.

3.1. Bilateral filter

To investigate whether the robustness of the FASTR metric can be improved by removing smaller gradients, a bilateral filter[7] was applied to the original image data before registration. The bilateral filter was of the form

$$I'(\mathbf{x}_c) = \sum_{\mathbf{x}_i \in \mathcal{N}} w_{i,\text{bilateral}} \cdot \exp\left(-\frac{(I(\mathbf{x}_i) - I(\mathbf{x}_c))^2}{\sigma_{\text{noise}}}\right) \cdot I(\mathbf{x}_i). \tag{5}$$

where $w_{i,\text{bilateral}}$ is a Gaussian kernel with standard deviation $\sigma_{\text{bilateral}}$. For these experiments we set $\sigma_{\text{bilateral}} = 10$, $\sigma_{\text{noise}} = 300$, and set the neighbourhood region to a $21 \times 21 \times 21$ voxel patch with $\mathbf{x}_c$ at the centre of the patch. This allows modest smoothing while preserving strong edge gradients.

Figure 2 shows an example slice from a head CT dataset, the bilateral filtered result, and the difference image between the unprocessed and processed images. Note that the difference image is close to zero in regions of strong structures, but areas of weak structure have been smoothed and are thus non-zero in the difference image.

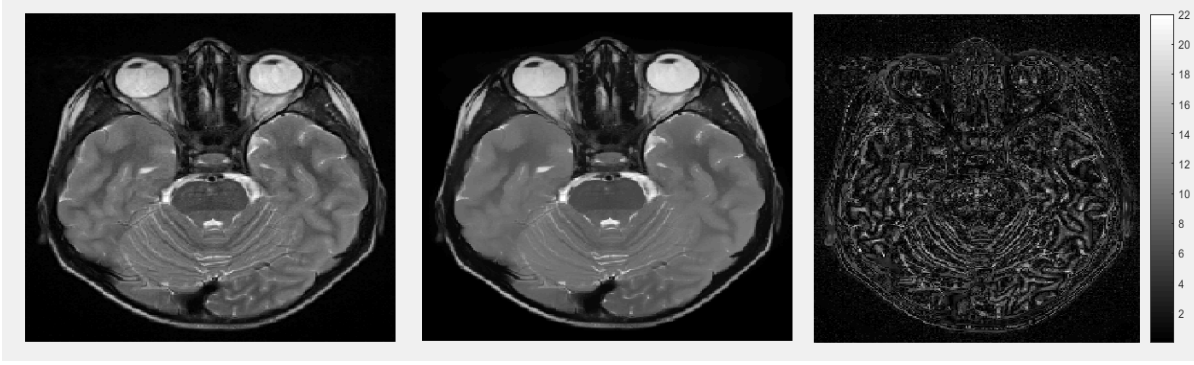

Fig. 2. Original MR T2-weighted slice (left), same slice processed with bilateral filter $\sigma_{\text{noise}} = 300, \mathcal{N} = 21 \times 21 \times 21$ (centre) and absolute difference image between original and bilateral filtered slice (right).

3.2. Results

As demonstrated in tables 1 and 2, FASTR aligns these images with lower TRE than MI. One reason for the difference is that MI is a global metric, and as such can be confounded by inhomogeneous "illumination" across an image, such as is often seen in MR and CBCT images. Interestingly, FASTR gives comparable results to the state-of-the-art SSC, while running significantly faster (see table 3). Furthermore, FASTR is less memory intensive than the SSC, as it uses 3 feature channels compared with the 12 feature channels required for SSC.

Registration ID	**1**	**2**	**3**	**Mean TRE**
Initial error	13.6	13.2	13.1	13.3 ± 0.3
MI	2.0	6.2	8.1	5.4 ± 3.1
SSC	2.1	2.3	3.0	2.4 ± 0.5
FASTR, $\sigma_{\mathrm{LST}} = 1$ (filtered)	2.0	2.4	3.0	2.5 ± 0.5
FASTR, $\sigma_{\mathrm{LST}} = 1$ (not filtered)	2.0	2.4	3.0	2.5 ± 0.5
FASTR, $\sigma_{\mathrm{LST}} = 2.5$ (not filtered)	2.1	2.5	3.2	2.6 ± 0.6
FASTR, $\sigma_{\mathrm{LST}} = 5$ (not filtered)	3.2	3.9	4.3	3.8 ± 0.6

Table 1. Target registration errors (TRE) for CT → CBCT registration using Mutual Information (MI), Self-Similarity Context (SSC), and FASTR. All results are in millimetres. 'Filtered' and 'not filtered' refers to whether the datasets were processed with a bilateral filter before FASTR computation.

Registration ID	**1**	**2**	**3**	**Mean TRE**
Initial error	14.3	27.7	17.5	19.9 ± 7.0
MI	4.7	7.3	6.2	6.1 ± 1.3
SSC	3.6	3.6	5.9	4.4 ± 1.3
FASTR ($\sigma_{\mathrm{LST}} = 1$)	3.9	3.8	5.7	4.4 ± 1.1

Table 2. Target registration errors (TRE) for CT → MR registration using Mutual Information (MI), Self-Similarity Context (SSC), and FASTR. All results are in millimetres.

Registration ID	**1**	**2**	**3**
MI	8.1s	3.7s	3.8s
SSC	453.3s	438.0s	497.9s
FASTR ($\sigma_{\mathrm{LST}} = 1$)	76.0s	91.7s	91.4s

Table 3. Runtime (seconds) of the similarity metrics for each CT → CBCT registration experiment.

As can be seen from table 1, the results from FASTR $\sigma = 1$, with and without a bilateral filter applied, are identical. This implies that FASTR is insensitive to smaller image gradients, suggesting it should be robust to image noise.

4. Conclusion

We have described a novel structural image descriptor for image registration called the Fractionally Anisotropic Structural Tensor Representation, or FASTR. Calculated from the LST, it does not depend on absolute voxel intensities, and is insensitive to slowly varying intensity inhomogeneity across the image. This is a very useful property, since many imaging modalities suffer from such artefacts, i.e. MR bias-field artefacts; CBCT truncation artefacts; ultrasound attenuation inhomogeneity.

Experiments demonstrated that for rigid registration of CT → CBCT and CT → MR datasets, FASTR produced more accurate results than MI, and faster results than SSC. Similar accuracy between SSC and FASTR was anticipated, since they both encode the directionality of gradients within a local patch.

Future work will perform similar experiments with a larger and more varied cohort of datasets, and evaluate FASTR for non-rigid registration.

The results suggest that, compared with SSC, FASTR gives similar benefits in images with intensity inhomogeneity, but at a fraction of the computation and memory demand.

References

1. Maes, F., Vandermeulen, D., Marchal, G., Suetens, P.. Multimodality Image Registration by Maximization of Mutual Information. *IEEE transactions on medical imaging* 1997;**16**(2):187–198.
2. Viola, P., Wells, W.M.. Alignment by Maximization of Mutual Information. *International Journsal of Computer Vision* 1997;**24**(2):137–154.
3. Haber, E., Modersitzki, J.. Intensity gradient based registration and fusion of multi-modal images. *Medical image Computing and Computer-assisted Intervention* 2006;**4191**(Mi):726–733.
4. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, S.M., et al. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis* 2012;**16**(7):1423–1435.
5. Heinrich, M.P., Jenkinson, M., Papiez, B.W., Brady, S.M., Schnabel, J.A.. Towards realtime multimodal fusion for image-guided interventions using self-similarities. *Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention* 2013;**16**(1):187–94. URL: `http://www.ncbi.nlm.nih.gov/pubmed/24505665`.
6. Haussecker, H., Jähne, B.. A tensor approach for local structure analysis in multidimensional images. *3D Image Analysis and Synthesis* 1996;:171–178doi:10.5281/zenodo.16444.
7. Tomasi, C., Manduchi, R.. Bilateral Filtering for Gray and Color Images. In: *1998 Sixth International conference on Computer Vision*. 1998, p. 839–846.