ABSTRACT

Recently, different approaches have emerged for voxel based nonrigid image registration using local instead of global similarity measures. Benefits are more accurate registration or the ability to subdivide the global similarity in local contributions. Within this article, we provide a general method to localise similarity measures using overlapping regions. Moreover, we extend the concept of partial volume estimation, introduced for mutual information (MI), to other similarity measures. We compare local and global sum of squared differences (SSD), cross correlation (CC) and MI for different sizes of the local regions. In general, local MI gives the highest accuracy, even for image pairs of the same modality. Partial volume estimation slightly improves the accuracy for local measures; the improvement is more pronounced for label images.

Index Terms— voxel-based similarity measure, nonrigid registration, mutual information

1. INTRODUCTION

Recently, different approaches have emerged evaluating similarity measures for voxel based nonrigid image registration no longer global but – depending on the source – local, regional or conditional. The motivation behind this is that the local transformation should be guided by the local similarity. Studholme et al. [1] proposed regional mutual information (MI). Loeckx et al. [2, 3] have shown that a similar concept – named conditional MI – can outperform global MI for multimodality image registration or images contorted with a bias field. Glockel et al. [4] have a different motivation, requiring local similarity to enable discrete optimisation. But already in 1999, Weese et al. [5] proposed to use local instead of global cross correlation (CC) applied to affine multimodal image registration. They argued that in a sufficiently small region of the image rarely more than two structures are visible, and the CC hypothesis of a linear relationship between corresponding intensities holds when only two intensities are visible in each region.

Within this article, we first discuss localisation of similarity measures. For clarity, we use the term local similarity for the total image similarity calculated as the sum of the localised similarities. The localised similarity is the similarity calculated over a region or part of the image. We distinguish between spatially separable and non spatially separable similarity measures. For the former, the similarity can be expressed as a sum over all corresponding voxels, independent of their neighbourhood. The latter measures somehow take the intensities of the neighbouring voxels into account as well. A spatially separable similarity measure, such as sum of squared differences (SSD), will be equal to the sum of the localised measures. This does not hold for measures that take neighbourhood information into account, such as CC (using the average and standard deviation of the image) and MI (using the joint histogram).

We also extend the concept of partial volume (PV) estimation, as proposed by Maes et al. [6] for MI, to other similarity measures. PV requires no interpolation in the deformed image, yet the similarity is evaluated and weighted over a neighbourhood of voxels around the transformed position.

We validate the proposed similarity measures on intensity and label images obtained from the BrainWeb database [7], using known transformations.

2. METHODS

2.1. Local Similarity

Assume reference and floating images R and F, and a transformation g (xR; μ), governed by a set of parameters μ, that maps each position xR in the reference image to its corresponding position xF = g(xR; μ) in the floating image. For each instantiation of μ, the similarity S(R, F; μ) between the reference and deformed floating image can be calculated.

To localise a similarity measure S(R, F; μ), we subdivide the reference image domain in overlapping regions x ∈ X. Assume a spatially separable similarity measure S = 1/nX ∑xR s (R(xR), F(g(xR))), with e.g., s(i, j) = (i − j)2 for SSD. With p(x|xR) the contribution of voxel xR to region x, the localised similarity is given by

\[ s_x(R, F) = \frac{1}{n_x} \sum_{xR} p(x|xR) s \left( R(xR), F(g(xR)) \right) \] (1)
and the local similarity (over the whole image) is then \( S_X = \sum_{x} p(x|x_R) \), the weighted sum of the localised similarities over all regions \( x \). Note that in this case, due to the spatial separability, \( S_X = S \).

This can be generalised for most other similarity measures, as long as they are based on sums over the reference and floating intensities. The localised similarity for each region is obtained by inserting \( p(x|x_R) \) inside the summation over all voxels.

Localisation of non separable similarity measures is illustrated using squared normalised cross correlation \( CC = \frac{1}{\sigma_R \sigma_F} \left( \sum_{x} (I_R - \bar{I}_R)(I_F - \bar{I}_F) \right)^2 \). Similar to (1), local CC becomes \( CC_X = \sum_{x} \frac{p(x|x_R)}{\sigma_R \sigma_F} (\sum_{x} p(x|x_R) (I_R - \bar{I}_R)(I_F - \bar{I}_F))^2 \). As also the average and standard deviation are calculated locally, \( CC_X \neq CC \).

2.2. Partial volume estimation
The essence of partial volume (PV) estimation [6] is that the interpolation in the floating image is replaced by a weighted average over the voxels neighbouring the target location. This has two main advantages. First of all, no new intensities are introduced. Secondly, the derivative of PV measures does not require image derivatives. Most similarity measures are based on sums over the reference and deformed floating intensities \( R(x_R) \) and \( F(g(x_R; \mu)) \). The PV counterpart is defined as

\[
S(R, F; \mu) = \sum_{x_R, x_F} w_v(x_F; g(x_R; \mu)) s(R(x_R), F(x_F))
\]

with \( w_v(x_F; g(x_R; \mu)) \) a vicinity kernel modelling the relation between the transformed reference position \( g(x_R; \mu) \) and the floating position \( x_F \).

For both the PV and the overlapping regions, 3D second degree B-spline kernels were used. FFD [8] was chosen as transformation model.

3. EXPERIMENTS AND RESULTS
To compare the proposed similarity measures, we created several sets of a priori registered 2D image pairs, based upon the images shown in figure 1. The intensity images show the central slices of the t1-weighted (t1) and proton density (pd) images from the BrainWeb database (noise level: 3%) [9, 7]. The label images show, for the same slices, the different tissue labels as available in the database. Each slice measures 181 \times 217 voxels with a 1 mm voxel size in each dimension. Intensities range from 0 to 4000 for the intensity images. The label images contain 10 discrete labels; for Label B the label indexes are randomly permuted compared to Label A. As the images are calculated from a single phantom, they are a priori in perfect alignment. We have used both undistorted BrainWeb images, and images distorted by a 40% bias field.

Fig. 1. Sample original (top) and deformed (bottom) images used for the experiments.

The experiments were repeated for different final mesh spacings \( \Delta_f = \{4, 8, 16, 32, 64, 128, 256\} \). For each experiment, the mesh spacing started at \( \Delta=256 \) voxels and was repeatedly halved until the desired final mesh spacing was reached. The deformed images are generated by applying a random transformation to the reference image. This transformation is generated as the sum of random B-spline transformations obtained over the different mesh spacings \( \Delta \) from 4 to 256. For each mesh spacing, \( \mu \) was chosen from a uniform distribution with a maximum amplitude of \( \Delta/4 \).

For each image pair and final mesh spacing, 50 registrations were performed, each using a different random deformation. First, the reference image was deformed with the known random transformation. Next, the floating image was registered to the deformed reference image, yielding the transformation to be validated. The applied random transformation was compared to the transformation to be validated obtained by nonrigid registration. The quality was measured over all voxels within the brain using the warping index, which is the root mean square of the transformation difference in each voxel. Calculation of the global measures takes 1 - 20 seconds, depending on the final mesh spacing, with local measures being 5 - 10 times slower.

Figure 2 plots the warping index for different final mesh spacings without using a PV kernel. For intensity images, results using a partial volume kernel manifested the same trends. The label images showed more discrepancies between both approaches: for figure 2(c) the optimal \( \Delta \) remained the same, while for figure 2(d) the PV curves showed an optimum for \( \Delta_f = 8 \). Figure 3 provides a boxplot of the optimal final mesh space, including the results with and without PV.

For monomodal registration in the absence of a bias field, all proposed similarity measures perform more or less similar, although the detail provided in figure 3(a) shows that even
Fig. 2. Initial warping error and results for the different image pairs, similarity measures (without partial volume) and final mesh spacings. The curve shows the median of 50 registrations, whiskers indicate the 15% and 85% quantile. Subscript $X$ and thick lines indicate local measures.

Fig. 3. Boxplots comparing normal and partial volume (PV) similarity measures. The optimal final mesh spacing $\Delta_f$ was chosen from figure 2. Subscript $X$ indicates local measures, the prime symbol $'$ indicates PV measures.
in this most simple case, MI and local CC are slightly better than global CC or SSD. When a bias field is introduced, the balance shifts in favour of local MI and local CC, with a slight advantage of PV local MI over the others. The optimal registration results are obtained at a final mesh spacing of 8 voxels.

In multimodal registration, not only SSD but also both local and global CC perform badly, and only MI survives. Again, in the absence of a bias field, local and global MI are comparable, while the bias field increases the gap between both measures in favour of local MI, again with a slight preference for the PV approach. The optimal final mesh spacing increases from 8 voxels to 16 voxels.

The label images show a more pronounced difference between results with and without PV, for local and global MI and for local CC. Note that, contrary to intensity images, global PV MI significantly outperforms global normal MI for label images.

4. DISCUSSION

Validation of nonrigid registration is a tricky business, as the performance of each method depends on the nature of the images to be registered. Moreover, for real clinical images it is nearly impossible to obtain a ground truth. To illustrate the proposed methodology, we rely on realistic simulated images and known artificial deformations. This allows evaluating the quality of the similarity measure independent of the transformation model. The results are still confounded with the optimisation method, although it is the same for all experiments.

The bad performance of local CC for multimodal registration is in contrast with the results of Weese et al. [5]. In their article, Weese et al. have used the method for affine registration and find results comparable to MI. Nonrigid registration is more complicated. For affine registration, errors can average out over the whole image field to create the global optimum. Our results indicate that for multimodal registration the reasoning of Weese et al. is locally violated. This allows the nonrigid transformation to adapt to local errors.

SSD only achieves reasonably for mono-modal registration, although still worse than MI or local CC. Looking at the registration results, we see that while in most regions the SSD accuracy is excellent, in some regions it seems to be stuck in a local optimum. This could be caused by a higher sensitivity of SSD to noise in the neighbourhood of the optimum.

For intensity images, the differences between PV and standard similarity measures are small, except for some cases of MI and MI\_X. The results for MI are in accordance with previous work [2, 3], indicating that mutual information is more accurate when no PV estimation is used. However, this changes for local measures as well as for label images. Especially for the latter, the advantage of PV measures is clearly visible. Interpolating measures assume that voxel intensities are a continuum, while PV measures consider intensities as discrete labels. Some advantages of using PV estimation for label images can also be seen for local CC. As always, the best choice depends on the actual application to solve.

5. CONCLUSION

We have proposed a general method to localise similarity measures, and compared local and global SSD, CC and MI. Local MI generally outperforms all other methods, even for monomodal image registration. We have also evaluated PV estimation. PV improves the accuracy for local measures; the improvement is more pronounced for label images.

6. REFERENCES


