# Dense Multi-modal Registration with Structural Integrity using Non-local Gradients

Sheshadri Thiruvenkadam
*MIAL, GE Global Research, Bangalore, India*

Keywords: Multi-modal, Non-rigid Registration, Non-local Gradients.

Abstract: In this work, the challenging problem of dense non-rigid registration [NRR] for multi-modal data is addressed. We look at a class of differentiable metrics based on weighted L2 distance of non-local image gradients. For intensity dependent choice of weights, the metric is seen to give enhanced multi-modal capability than using just gradients. In a variational dense deformation setting, the metric is coupled with non-local regularization to make the framework feature based. The above combination maintains the visual quality of the registered image, and gives a good correspondence for features of similar geometry under the challenges of noise, large motion, and presence of small structures. We also address computational speed ups of the energy minimization using an approximation scheme. The proposed approach is demonstrated on synthetic and medical data, and results are quantitatively compared with MI based, diffeomorphic NRR.

## 1 INTRODUCTION

Multi-modality and large non-rigid motion (relative to scale of structures) are the primary challenges of accurate image registration. Mutual information [MI] based approaches and their extensions have been quite successful for coarse non-rigid registration of multi-modal data, see survey (Pluim et al., 2003). Recent works have also extended MI to recover local deformations (Likar and Pernus, 2001; Hellier and Barillot, 2000; Loeckx et al., 2010; Lu et al., 2010). Though MI has been demonstrated and widely used as a popular multi-modal metric, the non-convexity of the metric due to interpolation induced artifacts makes it computationally challenging to use for dense NRR. The above reason has motivated works to look at alternative multi-modal metrics that are well behaved e.g Normalized gradients (Haber and Modersitzki., 2007). Image gradients have also been widely addressed in the vision literature e.g. (Lowe, 2004; Mikolajczyk and Schmid, 2005) for many applications. Intuitively, match of gradients should give some multi-modal capability (to spatially linear intensity variations) since differences in relative intensities are only penalized. While the above image gradient based works capture only local intensity interactions, we are motivated to look at non-local interactions between intensities which should give better robustness to noise and intensity variations. Specifically we model intensity differences between pixel pairs in a large neighborhood through non-local gradients. We wish to note that a recent work (Heinrich et al., 2011) has also looked at modelling non-local intensity relationships to achieve promising results on multi-modal data.

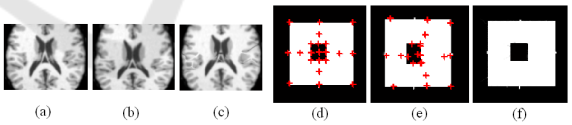


Figure 1: Structural integrity and point correspondences.

The other challenge of large deformations has been addressed using multi-resolution strategies, fluid flow regularization, and diffeomorphic frameworks; for a recent survey see (Holden, 2008). These works address the problem of *spatial integrity* (preserving the topology of structures) under large deformations. Another aspect which has not been carefully addressed in the registration community is *structural integrity*, i.e. the visual quality of the registered image, which is challenged by noise, large motion, and presence of small structures. To illustrate, in Fig. 1, the registered T1 MR image (c) looks very similar in intensity to the fixed image (a), but has lost definitive features of the moving image (b). Although it seems difficult to directly model and subsequently evaluate the registered image's visual quality, a key driver is


Thiruvenkadam S..
Dense Multi-modal Registration with Structural Integrity using Non-local Gradients.
DOI: 10.5220/0004211702580263
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 258-263
ISBN: 978-989-8565-47-1

the quality of correspondences between key feature points. Also in many applications such as motion tracking, it is desirable that the computed map gives a good correspondence of points of similar geometry. In Fig. 1 (synthetic example), the registered image (f) looks exactly like the fixed image (d), but key feature points (Red '+') are incorrectly mapped on the moving image (e). Feature based methods (Zitova and Flusser, 2003) as against intensity based methods have been successful in achieving structural integrity atleast in neighborhoods of strong features. In both intensity and feature based methods, the computed maps follow the data near strong feature locations and are driven to homogeneous regions by local regularization. As a result, one notices un-realistic motion in homogeneous regions (e.g. in Fig. 1 (e), the quality of correspondences are worse in homogeneous regions). One way to handle this issue is through the use of patch based metrics, e.g. (Bruhn et al., 2005), where the data term is effective in a neighborhood of strong features as defined by the patch size. Another possibility is to use non-local regularization (Sun, 2010; Werlberger, 2010) of the deformation fields which gives increased robustness to large motion/noise, compared to local approaches.

In this work, we contribute towards a variational framework capable of dense NRR for Multi-modal data. The framework also addresses the aspects of structural integrity of the registered image and quality of correspondences. A class of differentiable metrics based on weighted L2 distance of non-local gradients is proposed. Intensity dependent weights within the metric give enhanced multi modal capability similar to mutual information based approaches. Since minimizing the above energy is expensive due to non-linearities, an approximation scheme for speed up is proposed. Next, we couple our metric with non-local regularization (Sun, 2010; Werlberger, 2010) to make the framework feature based. The above combination maintains the visual quality of structures in the registered image and also gives a correspondence of similar features under challenges of noise, large motion, and presence of small structures. The approach is demonstrated using experiments on synthetic and medical data, and quantitative comparisons are shown with MI based dense, diffeomorphic NRR.

## 2 FORMULATION

Given template(fixed) and target(moving) images $f, m : \Omega \to \Re$, we want to recover transform $t$ between the image domains. As mentioned previously, the motivation is to look at intensity-relationships between pixel pairs in a larger neighborhood, thus improving robustness to noise and intensity variations.

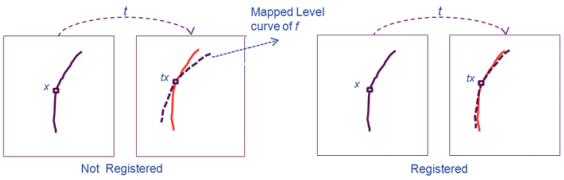


Figure 2: Map of level curves of $f$(Brown), $m$(Red) before and after registration

### 2.1 Metric using Non Local Gradients

Although there are many ways to define intensity-relationships between pixels, in this work, we want to preserve *intensity differences* between mapped pixel pairs in the template and target domains. Let $tx$ denote the mapped location for pixel $x$. We look at a general class of metrics defined through non-local gradients:

$$E[t] = \int_\Omega \int_\Omega w(x,y)\big((m(ty)-m(tx)) - (f(y)-f(x))\big)^2 dydx \quad (1)$$

Here $w$ is a weight function that defines the importance of the pixel-pair $(x,y)$. For simplicity, assuming that the weights do not depend on $t$, we arrive at the following Euler Lagrange equation [EL] for $E$. Denoting $\bar{w}(x,y) = \frac{w(x,y)}{\int_\Omega w(x,y)\,dy}$, we have,

$$\Big(\int_\Omega \bar{w}(x,y)m(ty)\,dy - m(tx) - \Big(\int_\Omega \bar{w}(x,y)f(y)\,dy - f(x)\Big)\Big)\nabla m(tx) = 0 \quad (2)$$

The choice of weights impacts multi-modal capability and computational expense. Obvious choices of weight functions are $w(x,y) = B_R(|x-y|)$, where $B_R$ is the indicator function for $[0,R]$, and $w(x,y) = G_\sigma(|x-y|)$, $G_\sigma$ is a Gaussian $(0,\sigma)$. For these weight functions, it is interesting to see that the descent equation of (2) is similar to that of SSD-intensity, after intensity normalization of $f$ and $m$ in local neighborhoods around each pixel. This gives some robustness to intensity variations. Further, the integrals in (2) are convolutions and hence fast to compute.

For better multimodal capability, we consider intensity dependent weights, $w(x,y) = \delta(|f(x) - f(y)|)G_\sigma(|x-y|)$, $\delta$ is the Dirac delta function. Now the energy (1) becomes:

$$\tilde{E}[t] = \int_\Omega \int_\Omega w(x,y)\big(m(ty)-m(tx)\big)^2 dydx \quad (3)$$

With EL:

$$\Big(\int_{\Omega} \bar{w}(x,y)m(ty)\,dy - m(tx)\Big)\nabla m(tx) = 0 \qquad (4)$$

Intuitively, the metric now maps level curves of $f$ to level curves of $m$ locally at each pixel $x$, Fig. 2. In fact, the descent of (4) would converge when for each $x$, the level curve passing through $f(x)$ maps to a level curve through $m(tx)$. Since the actual intensity of the iso-contours of $f$ and $m$ is not taken into account while matching, we get better flexibility to multimodal data as compared to the gaussian choice of weights. Here, the descent will not involve convolutions, hence approximations for computational speed up are needed, which would be dealt with shortly.

## 2.2 Non Local Regularization

In many registration applications, the visual quality of the registered image and correspondence between features that the computed map achieves are important. Feature based methods (Zitova and Flusser, 2003; Haber and Modersitzki., 2007) are partly successful in preserving structure around features and also drive the registration based on feature correspondence.

The proposed metric (1) is not yet feature based; e.g. in a monomodal scenario, once the registered image $m \cdot t$ ($\cdot$ denotes composition) matches the fixed image $f$, the geometry (e.g. gradients, hessian) would match as well. Analogous to matching image gradients, we need to match $\nabla f(x)$ and $\nabla m(tx)$ for feature based capability, and not $\nabla f(x)$ and $\nabla(m \cdot t)(x)$. Representing $tx = x + u(x)$, we want to match $f(y) - f(x)$ with $m(y+u(x)) - m(x+u(x))$, instead of with $m(y+u(y)) - m(x+u(x))$. One possibility would be to directly use a patch based metric of Non-Local gradients taking the form $\int_{\Omega}\int_{\Omega} w(x,y)\big(m(y+u(x)) - m(x+u(x)) - (f(y)-f(x))\big)^2 dydx$. But this would be expensive even for Gaussian weights, since the EL would not involve convolutions. Secondly, the above patch based metric intrinsically assumes rigid misalignments, thus proving very restrictive while recovering non-rigid motion. These issues are highlighted in the Results section.

With this understanding, to make our energy (1) feature based, we consider a penalty of the form $\int_{\Omega}\int_{\Omega} \tilde{w}(x,y)|u(y)-u(x)|^2\,dydx$ with $\tilde{w}(x,y)$ as a weight function. Adding the above penalty to (1) gives:

$$E_{reg}[u] = \int_{\Omega}\int_{\Omega} w(x,y) \Big((m(y+u(y)) - m(x+u(x)) - (f(y)-f(x))\Big)^2 dydx + \lambda_{reg}\int_{\Omega}\int_{\Omega} \tilde{w}(x,y)|u(y)-u(x)|^2\,dydx \qquad (5)$$

$\lambda_{reg}$ is a parameter balancing the two terms. The regularization term penalizes non-local gradients of the deformation field and has found recent interests in optic flow methods. For computational simplicity, we have used a $L2$ metric for the regularization term and a straight forward choice for the weight term $\tilde{w}(x,y) = G_{\tilde{\sigma}}(|x-y|)$. Alternatively, the regularization term can be replaced with NL-TV and intensity dependent weights for better results, e.g. (Sun, 2010; Werlberger, 2010). The distinct advantage over patch based methods is that the rigidity of matching can be controlled through $\lambda_{reg}$ thus allowing better capture of local deformations. Our approach also shows the benefits of patch based methods namely robustness to noise, structural integrity, and match of features.

# 3 NUMERICAL IMPLEMENTATION

To minimize (5), we use time descent given by the EL and discretized using explicit finite differences. For gaussian weights, the descent equation involves just convolutions and is fast to compute. When the weight is intensity dependent for better multimodal capability, the integrals in the EL are no longer convolutions and are expensive to compute. Below, an approximation scheme for speed up is proposed.

Consider the energy (3) with intensity dependent weights, and EL given by 4. Using the co-area formula for smooth functions (Jost and Li-Jost, 1998), assuming $|\nabla f| > 0$ a.e., it can be shown that (4) is solved by the solutions of

$$\nu(x,f(x))(m(tx) - \mu(x,f(x)))\nabla m(tx) = 0,$$

$$\text{where } \nu(x,\lambda) = \int_{f^{-1}(\lambda)} G_{\sigma}(|x-y|)d\tilde{s}_{\lambda},$$

$$\mu(x,\lambda) = \frac{\int_{f^{-1}(\lambda)} G_{\sigma}(|x-y|)m(ty)d\tilde{s}_{\lambda}}{\nu(x,\lambda)} \qquad (6)$$

Here, $d\tilde{s}_{\lambda} = \frac{ds}{|\nabla f(x)|}$, $d\tilde{s}_{\mu} = \frac{ds}{|\nabla f(y)|}$, and $ds$ is the arc-length differential. $f^{-1}(\lambda)$ denotes the level curve corresponding to $\lambda$. We discretize the above equations with finite differences and solve using iterative descent. The descent is analogous to that of the SSD metric where the update equations seek a transform $t$ to match intensities of the moving image $m$ and the evolving fixed image $\mu$. The above arc-length integrals can be reduced to region integrals by working in a band around the level set $\lambda$. Also $\mu$ and $\nu$ are 'binned' at discrete locations $(x_i, \lambda_i)$ and interpolated to get the value at $(x,\lambda)$. Our experiments indicate

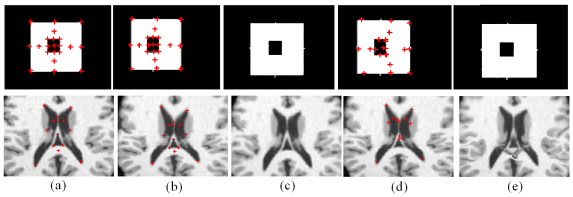


Figure 3: Quality of correspondences of key feature points (Red '+') (a) fixed image (b) Moving image with mapped points using Approach (c) Registered image using Approach (d) moving image with mapped points using SSD NRR (e) Registered image using SSD NRR

that $\mu$ and $\nu$ have to be updated only once in 10-20 iterations effectively making the cost of every iteration as number of pixels.

## 4 RESULTS

Here we present results on synthetic and medical data to illustrate the algorithm's performance. The use of non-local gradients for proposed choice of weights gives computationally tractable dense NRR capability for multi-modal data. Also, combination with non-local regularization gives structural integrity of the registered image (under noise/large motion/presence of small structures), and gives a good quality of correspondences.

### 4.1 Quality of Correspondences and Robustness

In Fig. 3, we illustrate quality of correspondences using the proposed approach. In the two examples shown in Fig. 3, the fixed image (a) is registered with moving image (b). The objective is to see if key feature points (red '+' in (a)) are mapped to respective locations on the moving image after NRR. In (d), results from SSD based NRR (SSD intensity metric + local regularization of motion fields) shows incorrect correspondences even in regions where the registered image (e) is close to the fixed image. Remarkably, our approach maps points to correct geometric locations even at homogeneous regions (b), and gives very good visual quality for the registered image (c). Next, in Fig. 4, we show capability of preserving small structures while recovering large motion, and robustness to noise. In Fig. 4, (a), (b) are noisy, fixed and moving images (First row: synthetic, Second Row: PET phantom data) with structures of different scales. (c) is the overlay showing the mismatch, (d) is the result of our approach. In both examples, the large motion of small lesion-like structures is corrected under the presence

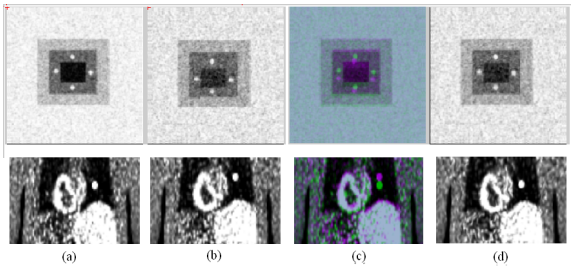


Figure 4: Quality of registered image with large motion relative to scale of structures and noise (a) Fixed image (b) Moving image (c) Overlay showing initial mismatch (d) Registered image using approach.

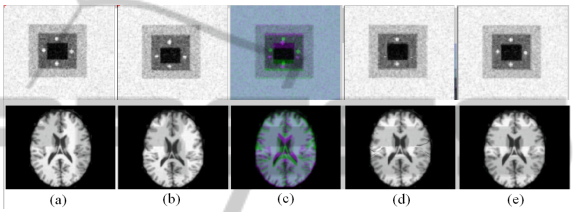


Figure 5: Comparison of Approach with Patch SSD (a) Fixed image (b) Moving image (c) Overlay showing initial mismatch (d) Registered image using Patch SSD (f) Registered Result using approach.

Table 1: Timing of Methods.

| Method | Time (sec) |
|---|---|
| SSD NRR | 27 |
| Patch SSD | 232 |
| Approach with Gaussian weights | 46 |
| Approach with Intensity based weights | 52 |

of noise. Local NRR approaches would not work here since there is no overlap of the lesions in the fixed and moving images, and these methods would collapse the lesions after NRR. Typical multi-resolution strategies would also not work for large motion of small structures, since (structure size to motion) ratio would be preserved across resolutions.

### 4.2 Comparisons

In subsequent discussion, we compare our approach with Patch based NRR (e.g. (Bruhn et al., 2005)) and MI based NRR (e.g. (Lu et al., 2010)):

a) **Patch based Methods.** Starting with the work of (Bruhn et al., 2005), there has been a lot of interest in use of patch based methods within registration mainly due their robustness to noise and ability to preserve details of structures under large motion. In this section we compare the results of our approach on some monomodal images with *Patch SSD* (patch based SSD metric + local regularization of motion). We wish to

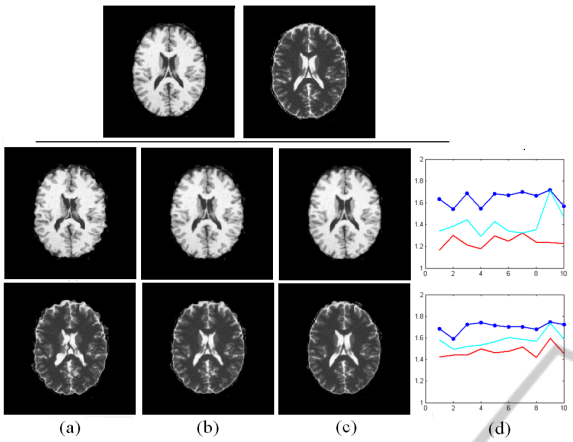


Figure 6: Synthetic MR comparisons with MI NRR. (a) Ground truth T1 and T2 images, (b) moving image (c) result using MI NRR (d) result using approach (e)Error plot (Blue - before NRR, Cyan - MI NRR, Red - Approach).

comment that the metric used in [Bruhn et al.] is a linear approximation to the patch based SSD metric, and is suitable for recovering small deformations only. In spite of the advantages patch based methods offer, these methods are restrictive in recovering local non-rigid deformations and are computationally expensive. Our method shows the advantages of patch based methods without the above drawbacks. In Fig. 5, in the two examples shown, the fixed image (a) is non-rigidly deformed to generate the moving image (b), (c) the overlay showing the mismatch. In (d), we see that patch SSD has done quite well in recovering deformations under noise. However, artifacts are seen in the registered image since the metric is too restrictive to recover local non-rigid deformations. (e) shows results using our Approach.

A note on the timing (Table. 1), for the T1 NRR experiment (250x250 images) in Fig. 5. All runs were in Matlab, on a 2.6 GHz laptop. It is noted that the timing for our approach using intensity based weights is comparable to that with gaussian weights due to the approximation scheme. The main cost of our approach over SSD NRR is the use of non-local regularization with a gaussian kernel of width (25x25). Patch SSD is seen to be expensive since the EL does not involve convolutions and also takes more iterations to converge under local deformations.

b) **Mutual Information [MI] based NRR.** There has been recent interest in extending MI based metrics for dense NRR. In (Lu et al., 2010), an MI based extension to diffeomorphic Demon's algorithm is implemented in a multi-resolution framework [MI NRR]. The main challenge of MI based methods for dense NRR is that the number of samples available to construct the joint histograms at each point could be less, resulting in interpolation artifacts. This would mean optimizing a non-convex metric leading to sub-optimal solutions, and giving un-realistic deformations due to over emphasis on regularization.

Now, quantitative comparisons with MI NRR for MR synthetic data and gated PET-CT data are shown. In Fig. 6, we consider two experiments, MR T1-T1 NRR and MR T1-T2 NRR. The first row shows the ground truth T1 and T2 images that are well aligned. We then generated 10 T1(T2) images by applying increasing ranges of random non-rigid motion to the ground truth T1(T2) images. The 10 T1(T2) images are then non-rigidly registered to the ground truth T1 image using MI NRR and our approach. A representative example is shown in (a)-(c), (a) is the moving image, (b) is the result using MI NRR, and (c) is the result using proposed approach. The plot (d) is the maximum absolute difference error between the ground truth T1(T2) image and the registered images. The Blue curve is the error before NRR, the Cyan curve is error using MI NRR, and the Red curve is using proposed approach. As clearly indicated in the error plots and in the examples, the motion recovery and the quality of the registered images is better using the proposed approach.

Next, we look at NRR for gated PET-CT. In gated PET-CT, PET and Cine CT are synchronized across breathing phases, using e.g. an external tracking device. The data consisted of 6 cases ( 3 clinical and 3 synthetic PET Phantoms). Each of the data cases consisted of 6 PET gates (128x128x50), and 6 corresponding in-phase CT gates (512x512x72). We then picked the central coronal slices for our 2D experiment and brought the CT images to PET resolution. For each of the 6 data sets, a PET image and its corresponding in-phase CT image were picked as the reference. Now, the other 5 CT images were non-rigidly registered to the reference PET image resulting in 30 NRR computations for the 6 data cases. For validation, the reference CT image is compared with the registered CT images for each of the data cases.

We illustrate on a clinical case in Fig. 7. Column (a) shows the PET reference image and the corresponding in-phase CT image (ground truth for comparison). Column (b) is the registered CT image, Column (c) is the overlay of the CT image (b) with the reference PET image, Column (d) is the difference image of (b) with the reference CT image. The first, second and third rows are the outputs before NRR, using MI NRR, and using the proposed approach. For quantitative comparison (Fig. 8), we look at the absolute difference errors between the registered image and the reference CT image around 10 key landmark

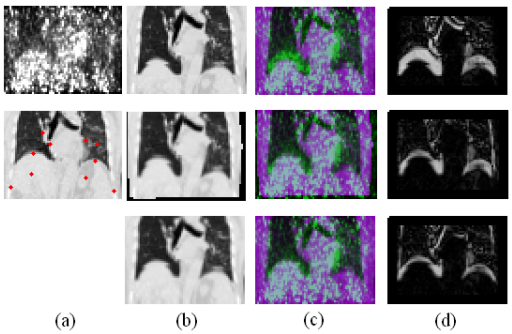


Figure 7: PET CT NRR for a clinical case. Column (a): PET reference image and the CT reference image with key landmarks (Red dots), First Row: Before NRR, Second Row: after MI NRR, Third Row: using proposed approach. Column (b): registered CT images, Column (c): Overlay of the CT image (b) with the reference PET image, Column (d): difference image of (b) with the reference CT image.

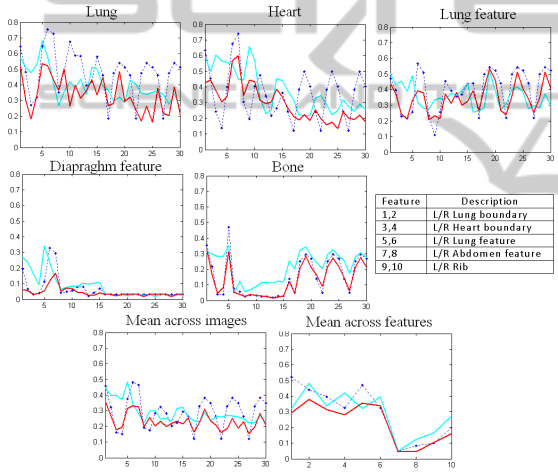


Figure 8: First and Second Rows: Error across individual feature locations (errors due to LEFT/RIGHT feature pairs have been aggregated for display), Last Row: Mean errors shown. (Dotted blue - before NRR, Cyan - using MI NRR, Red - using proposed approach).

locations (red dots shown on the reference CT image in Fig. 7 (a)). We evaluate the maximum absolute difference in a 5x5 patch around every landmark point, for each of the 30 registered images, before NRR (dotted Blue), after MI NRR (Cyan), and after proposed approach (Red). The errors at different landmark locations (for display purposes only, the errors due to left and right landmark pairs are aggregated to save space) are separately shown. The last row shows the mean error across images, and across landmark locations. For the 30 test cases, the proposed approach is clearly seen to give lesser error values than MI NRR at key landmark locations.

# 5 CONCLUSIONS

A weighted L2 match of Non-Local gradients is proposed as a metric for dense multi-modal NRR. The coupling of NL-regularization gives good correspondence between points of similar geometry in the fixed and moving images, and preserves the structural integrity of the registered image. Our method is also seen to give robustness to noise and ability to preserve small structures. The approach is demonstrated on synthetic/medical data, and comparisons are shown with MI based diffeomorphic NRR.

# REFERENCES

Bruhn, A., Weickert, J., and Schnrr, C. (2005). Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61:211–231.

Haber, E. and Modersitzki., J. (2007). Intensity gradient based registration and fusion of multimodal images. methods of information in medicine. *Methods of information in medicine*, 46(3):292–299.

Heinrich, M. P., Jenkinson, M., and et al. (2011). Non-local shape descriptor: a new similarity metric for deformable multi-modal registration. MICCAI'11, pages 541–548.

Hellier, P. and Barillot, C. (2000). Multimodal non-rigid warping for correction of distortions in functional mri. *MICCAI*, pages 512–520.

Holden, M. (2008). A review of geometric transformations fornonrigid body registration. *IEEE Transactions on Medical Imaging*, 27(1).

Jost, J. and Li-Jost, X. (1998). Calculus of variations. *Cambridge Univ. Press*.

Likar, B. and Pernus, F. (2001). A hierarchical approach to elastic registration based on mutual information. *Image Vision Comput.*, 19(1-2):33–44.

Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., and Suetens, P. (2010). Nonrigid image registration using conditional mutual information. *IEEE Transactions on Medical Imaging*, 29(1).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2).

Lu, H., Reyes, M., and et al. (2010). Multi-modal diffeomorphic demons registration based on point-wise mutual information. *ISBI*, 27(1).

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630.

Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2003). Mutual information based registration of medical images: A survey. *IEEE Trans. Med. Imaging*, 22(8):986–1004.

Sun (2010). Secrets of optical flow estimation and their principles. *CVPR*, 12(1):234–778.

Werlberger (2010). Motion estimation with non-local total variation regularization. *CVPR*, 13(1):234–778.

Zitova, B. and Flusser, J. (2003). Image registration methods:a survey. *Image Vision Comput.*, 21(11):977–1000.