

# Gradient Intensity: A New Mutual Information-Based Registration Method

Ramtin Shams, Parastoo Sadeghi and Rodney A. Kennedy  
Research School of Information Sciences and Engineering (RSISE)  
The Australian National University  
Canberra ACT 0200 Australia

{ramtin.shams}; {parastoo.sadeghi}; {rodney.kennedy}@anu.edu.au

## Abstract

*Conventional mutual information (MI)-based registration using pixel intensities is time-consuming and ignores spatial information, which can lead to misalignment. We propose a method to overcome these limitations by acquiring initial estimates of transformation parameters. We introduce the concept of ‘gradient intensity’ as a measure of spatial strength of an image in a given direction. We determine the rotation parameter by maximizing the MI between gradient intensity histograms. Calculation of the gradient intensity MI function is extremely efficient. Our method is designed to be invariant to scale and translation between the images. We then obtain estimates of scale and translation parameters using methods based on the centroids of gradient images. The estimated parameters are used to initialize an optimization algorithm which is designed to converge more quickly than the standard Powell algorithm in close proximity of the minimum. Experiments show that our method significantly improves the performance of the registration task and reduces the overall computational complexity by an order of magnitude.*

## 1. Introduction

### 1.1. Background and Motivation

The aim of the registration process is to find the optimal transformation, which best aligns two or more images by minimizing some cost function. The images may have been taken at different times, by different sensors, at different viewpoints or even contain local non-rigid deformations. One such cost function is the joint entropy of two random variables representing intensity values in each image or its more widely used utility function counterpart, mutual information (MI).

A concept borrowed from the information theory, mutual information measure in recent years has been successfully adopted for automatic registration of mono-modal,

multi-modal and multi-temporal images, with applications in medical imaging [3, 8, 14] (refer to [11] for an extensive survey), surveillance, remote sensing [2], environmental monitoring, and mapping [4].

Registration typically involves an optimization step, where at each iteration of the optimization algorithm, the moving (or floating) image  $M$  is transformed and the similarity measure is calculated against the fixed (or reference) image  $F$ . The optimal transformation  $T_{opt}$  is determined by finding the transformation parameters that maximize the similarity (e.g. MI) between the images

$$T_{opt} = \operatorname{argmax}_T S(F; M(T)), \quad (1)$$

where,  $S$  is the similarity measure,  $T$  is the transformation and  $M(T)$  is the transformed moving image.

Image transformation and MI computation are both time consuming. It is hence desirable to reduce the number of iterations required by the optimization algorithm. Several optimization algorithms have been applied or adapted to improve the registration process [9]. Optimization algorithms can be trapped by local extrema and result in mis-registration [11].

In the MI-based registration literature, pixel intensity (PI) has conventionally been the feature of choice. However, PI-based methods do not explicitly take the correlation of neighboring pixel intensities into account, which has been a known source of mis-registration. Multi-resolution methods that implicitly enter spatial information into calculations improve efficiency and robustness of registration. However, they may still introduce additional local minima at lower resolutions. The lower resolutions have been reported to be more sensitive to initial order of parameters being optimized [8].

A natural choice for incorporating spatial information is to compute MI over gradient of images. However, the utility function obtained in this way is likely to have a narrow attraction range and several ripples [10]. In [10] and [7], authors show that the shape of PI-based MI function can be

improved by introducing a correction factor based on co-occurrence of image gradients and MI of image gradients, respectively. The improved robustness of the registration task, as reported by these methods, comes at the cost of increasing the computational complexity of the cost function by introducing the image gradient calculation at each step.

## 1.2. Approach and Contribution

We propose an alternative method to improve robustness and efficiency of the registration task by acquiring an initial estimate of the registration parameters. The parameter estimation is decoupled in the sense that each parameter is estimated independently or is based on previously estimated parameters. The main contribution of the paper is the introduction of gradient intensity (GI) and GI-based maximization of MI to determine the rotation parameter between the images invariant to scale and translation. GI is loosely defined as the number of significant gradient vectors in a given direction. We provide a formal definition of GI in section 2.4. Table 1 compares the concept of GI- and PI-based registration methods.

PI-Based Method	GI-Based Method
Pixel	Direction
Intensity of a pixel	Number of gradient vectors in a direction
Intensity of a pixel assumed to be a random variable	Number of gradient vectors in each direction assumed to be a random variable
MI measures similarity of pixel intensities	MI measures directional similarity of gradient vectors
MI calculated over an $m \times n$ data-set, where $m, n$ are image height and width, respectively	MI calculated over a $1 \times d$ data-set, where $d$ is the number of directions (typically 360)

Table 1. Comparison of pixel- and gradient-based concepts.

We consider the class of proper similarity transformations (proper rotation, translation and scale). Our method determines the rotation parameter efficiently and is designed to be invariant with respect to scale and translation. Unlike PI-based methods, our method does not rotate the moving image due to the fact that an ideal rotation does not change the gradient content of an image except for a circular shift in all directions. As such, time-consuming gradient calculation is not required for each change in the rotation parameter. The MI utility function is efficiently calculated for two vectors of length 360, which removes the need for a conventional optimization step, at this stage. The

GI method reduces the dimensionality of rotational registration. For a 2D image the rotation parameter is found by maximizing MI of 1D vectors. In section 6 we briefly discuss that an extension of this method can reduce dimensionality of 3D image registration to a 2D plane. Reduction in dimensionality and applicability of our method to multi-modal images, images with large intensity variances and in the presence of isotropic scale, differentiates our work from existing approaches such as the gradient field method in [6].

We obtain an independent estimate of scale parameter using the average distance of significant gradients from the centroid of the gradient image. Given the rotation and scale parameters, the translation parameters can be derived from the relative position of the centroids of gradient images.

Finally, we use estimated parameters to initialize a guided optimization algorithm based on Powell’s direction set optimization method. Our optimization algorithm operates on a conventional PI-based MI utility function to achieve sub-pixel accuracy. Our optimization algorithm is specifically designed to converge quickly when initialized in close proximity of the optimal alignment.

We show in section 4 that our method significantly improves the robustness of the registration task, while reducing the overall computational complexity by an order of magnitude. We use 2D aerial images and 3D multi-modal MR images in our experiments. For 3D images we limit our experiments to in-plane transformations. The experiments with 3D images are only intended to demonstrate the applicability of the method for multi-modal registration and to motivate further extension for full 3D medical image registration as discussed in section 6.

## 2. Concepts

### 2.1. Entropy

Entropy of a random variable is a measure of the average or expected information content of an event, whose distribution is determined by the marginal probability of the random variable. One such measure was introduced by Shannon in 1948 [13], and is defined as

$$H(X) = \sum_{x \in X} p(x) \log \frac{1}{p(x)}, \quad (2)$$

where  $p(\cdot)$  is the probability mass function (pmf) of the random variable  $X$ . Shannon entropy measures the degree of uncertainty of a random variable by scoring less likely outcomes higher than the more likely ones. This is consistent with the notion that knowledge of an outcome that can be easily predicted is considered less valuable.

### 2.2. Mutual Information

Mutual information of two random variables is the amount of information that each carries about the other and

is defined as

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y), \end{aligned} \quad (3)$$

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (4)$$

where  $H(X|Y)$  is the information content of random variable  $X$  if  $Y$  is known,  $H(X, Y)$  is the joint entropy of the two random variables and is a measure of combined information of the two random variables.  $I(X; Y)$  can be thought of as the reduction in uncertainty of random variable  $X$  as a result of knowing  $Y$ . The uncertainty is maximally reduced, when there is a one-to-one mapping between the two random variables and is not reduced at all if the two random variables are independent and do not provide any information about one another.

### 2.3. Image Gradients

Let  $\mathbf{I}(z)$  be a 2D image or a slice of a 3D image and  $\mathbf{K}_x$  and  $\mathbf{K}_y$  be differentiating kernels in  $x$  and  $y$  directions, the gradient image is then calculated using (5)-(6).

$$\mathbf{G}_x(z) = \mathbf{K}_x \otimes \mathbf{I}(z), \quad \mathbf{G}_y(z) = \mathbf{K}_y \otimes \mathbf{I}(z), \quad (5)$$

$$g_m(x, y, z) = \sqrt{g_x(x, y, z)^2 + g_y(x, y, z)^2}, \quad (6)$$

$$g_\phi(x, y, z) = \arctan \frac{g_y(x, y, z)}{g_x(x, y, z)}, \quad (7)$$

where  $\mathbf{G}_x(z)$  and  $\mathbf{G}_y(z)$  are image gradient matrices in  $x$  and  $y$  directions for the 2D image or each slice of the 3D image,  $g_m$  and  $g_\phi$  are the gradient magnitude and phase at position  $(x, y, z)$ , respectively and  $\otimes$  denotes convolution.

### 2.4. Gradient Intensity

Gradient intensity is a measure of directional strength of an image and is loosely defined as the number of significant gradient vectors in a given direction. To calculate gradient intensity we first derive a gradient map  $G$  for the image using (5)-(6) defined as the set of vectors whose gradient magnitude is not less than a given threshold

$$G = \{\mathbf{x} \triangleq (x, y, z) | g_m(\mathbf{x}) \geq t(z)\} \quad (8)$$

where  $t(z)$  is the threshold, calculated for each slice of the gradient image and is used to remove gradient vectors that do not contribute to image shape or are more sensitive to noise. We use the root mean square (RMS) criterion as the gradient magnitude threshold

$$t(z) = \sqrt{\frac{\sum_{y=1}^m \sum_{x=1}^n g_m(x, y, z)^2}{m \times n}}, \quad (9)$$

where  $m$  and  $n$  are the number of gradient matrix rows and columns, respectively.

Let  $r$  be the angular resolution, gradient intensity  $\Gamma$  is then defined at the center of each direction bin  $\phi_i$  as

$$\begin{aligned} \Gamma(\phi_i) &= \sum_{\mathbf{x} \in G} 1 - \left| \frac{g_\phi(\mathbf{x}) - \phi_i}{r} \right|, \\ \phi_i - r &\leq g_\phi(\mathbf{x}) < \phi_i + r. \end{aligned} \quad (10)$$

Equation (10) bi-linearly distributes each gradient vector between the adjacent directions. In order to be able to compare gradient intensity of different images we use a normalized version of (10)

$$\tilde{\Gamma}(\phi_i) = \frac{\Gamma(\phi_i)}{\max_j \Gamma(\phi_j)}. \quad (11)$$

Fig. 1 shows a number of images and corresponding gradient intensities encoded on a unit disk. Brighter colors indicate higher gradient intensities in a given direction. As can be seen, the gradient intensity disks clearly identify the rotation between the images. However, a one-to-one relationship between the GI maps does not exist. For example a strong response is observed at  $\frac{k\pi}{2}$  directions due to inherent bias of rectangular gradient kernels to these directions. In the case of Fig. 1(g) and Fig. 1(h), the effect of multimodal nature of the images can be seen in GI maps, where the gradient intensities cannot be linearly related. These observations indicate that a statistical similarity measure such as MI that does not assume a linear relationship between gradient intensities will be more suitable for our purposes.

### 2.5. Entropy and Mutual Information of Gradient Intensity

Entropy of gradient intensity is a measure of directional information content of an image. The more directionally versatile an image, the higher the entropy. For example, in polygons, the entropy of gradient intensity is higher for higher order polygons with the maximum entropy observed for a circle and the minimum for a line.

Mutual information of gradient intensity of two images is a measure of directional similarity of images regardless of their relative size (scale) and position (translation) and is maximum, where images are rotationally aligned.

Fig. 2 shows the MI utility function for the GI-based algorithm and the PI-based algorithm calculated for Fig. 1(c) and Fig. 1(d). Both methods determine the rotation parameter that aligns the moving image onto the fixed image to be  $10^\circ$ . The utility function of the PI-based method is smooth, as expected, and yields well to an optimization algorithm. The utility function for the GI-based method has several local maxima, but a clear global maximum and since the MI calculation takes less than 1 millisecond on a standard PC, we can quickly find the global maximum and are not concerned with the shape of the MI function.

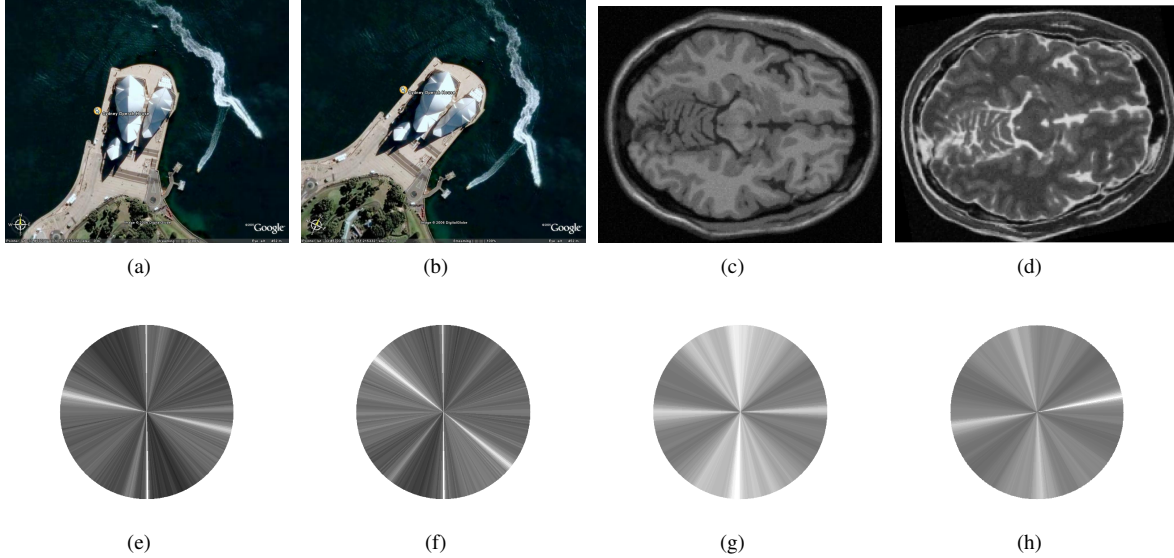


Figure 1. (a,b) Satellite images of Sydney Opera House, misaligned by  $-26.3^\circ$ . (c,d) MR-T1 and MR-T2 images of brain, misaligned by a transversal rotation of  $10^\circ$  (e-h) Corresponding gradient intensities mapped on a unit disk.

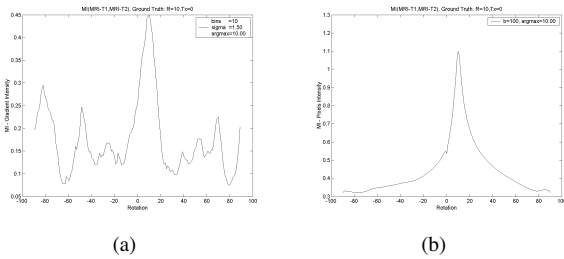


Figure 2. Sample MI functions for (a) GI-based method and (b) PI-based method; GI function is not as smooth, but this does not affect our algorithm, as we find the global maximum by searching on a small 1D space and are not affected by the MI function shape.

### 3. Method

We determine the rotation parameter of the optimal rigid transformation in (1) by maximizing the MI between gradient intensities of the images. We then use a centroid-based method to estimate the scale parameter. The translation parameters are derived using estimated scale and rotation and from the displacement of the gradient centroids. We finally use a guided optimization algorithm on a conventional PI-based MI function to achieve sub-pixel accuracy.

#### 3.1. Calculation of Image Gradients

The gradient calculations are done only once for each image using (5)-(6) and as such, we can use a more accurate filter with wider support without affecting the overall efficiency of the algorithm. A  $7 \times 7$  derivative of a Gaussian filter (which has the desired effect of differentiating the smoothed image [5]) with standard deviation  $\sigma = 1.5$  is used in our experiments.

### 3.2. Calculation of Rotation Parameter

#### 3.2.1 Gradient Intensity Histogram

The gradient phase matrix is converted into a histogram at a desired angular resolution using (10). The histogram is then normalized by the most frequent sample to create the gradient intensity vector as in (11). For 3D images, the 2D gradient of each slice is calculated separately and then all the samples are combined into a single gradient intensity vector. So regardless of the size of an image or its dimensionality, we always calculate MI for two  $1 \times d$  vectors: a fixed vector and a moving vector, where  $d$  is the number of directions. We use the angular resolution of  $1^\circ$ , which limits the gradient intensity to a vector of size  $1 \times 360$ .

#### 3.2.2 Computational Advantage of GI Method

The MI utility function is computed by iterating over the rotational range. At each step, a circular shift is applied to the moving GI vector and the MI between the resulting vector and the fixed vector is calculated using (4). This has a huge computational advantage over the PI method, where at each step one would require to rotate the moving image itself and calculate MI for the entire image. On a standard PC, MI was calculated under 1 *ms* for the GI method.

Once the MI utility function for the entire rotational range is calculated, the optimal rotation parameter is found, where the utility function assumes its global maximum.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} I(V_F; V_{M(\theta)}) \quad (12)$$

where  $V_{M(\theta)}(\phi) = V_M(\phi - \theta)$ , and  $V_F$  and  $V_M$  are the GI vectors for fixed and moving images, respectively.

### 3.2.3 Effect of Number of Bins

Increasing the number of bins used for estimating marginal and joint pmfs reduces the average number of samples per bin. For small data-sets, this results in a less accurate estimation of the pmfs of random variables. The total number of samples for the GI histograms is indeed very small and as such, we use 10 bins for pmf estimations. For PI method, where the number of samples is much larger, we use 100 bins.

### 3.3. Estimation of Scale and Translation

2D rigid transformation of a vector  $\mathbf{x}$  can be written as

$$\mathbf{x}' = s\mathbf{R}\mathbf{x} + \mathbf{t}, \quad (13)$$

where  $\mathbf{x}'$  is the transformed vector,  $s$  is the scale factor,  $\mathbf{t}$  is the translation vector, and  $\mathbf{R}$  is the rotation matrix. The distance between two points  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is defined as  $\mathbf{d}(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$ , where  $\|\cdot\|_2$  is the Euclidean  $\ell_2$  norm. It can be easily shown that

$$\mathbf{d}(\mathbf{x}'_1, \mathbf{x}'_2) = s \|\mathbf{R}(\mathbf{x}_1 - \mathbf{x}_2)\|_2 = s\mathbf{d}(\mathbf{x}_1, \mathbf{x}_2). \quad (14)$$

We need two points and their transformed locations to determine  $s$ . However since we only have the centroid and its transformed location, we estimate the scale parameter from the average distance of centroids  $\mathbf{x}_c$  and  $\mathbf{x}'_c$  from all other points in the corresponding gradient map

$$\hat{s} = \frac{\sum_{\mathbf{x}' \in G'} \mathbf{d}(\mathbf{x}', \mathbf{x}'_c)}{\sum_{\mathbf{x} \in G} \mathbf{d}(\mathbf{x}, \mathbf{x}_c)}, \quad (15)$$

where  $\hat{s}$  is the estimated scale parameter and  $G$  and  $G'$  are the gradient maps of fixed and moving images, respectively.

Given the rotation and scale parameters, the translation can be found from (13) by replacing  $\mathbf{x}$  and  $\mathbf{x}'$  with the centroid of the fixed and moving images, respectively

$$\hat{\mathbf{t}} = \mathbf{x}'_c - s\mathbf{R}\mathbf{x}_c. \quad (16)$$

### 3.4. Optimization

Powell's multi-dimensional direction set algorithm finds the minimum<sup>1</sup> of the cost function by iteratively minimizing the function along a set of  $N$  directions, where  $N$  is the number of independent parameters of the cost function. In the absence of any direction, the parameter space's origin is commonly used as the starting position for the optimization algorithm. We refer to this conventional approach as 'blind optimization'.

One problem with the Powell algorithm is that from an observer's point of view who knows where the minimum is

<sup>1</sup>In line with the optimization literature, we refer to finding minimum even though for MI we will actually maximize the function.

located, it appears to spend a lot of time, aimlessly iterating on and around the minimum. Obviously, the only way the optimization algorithm can satisfy itself that it has found the actual minimum is to spend enough time to check the surroundings. Our implementation resembles the method described in [12] with one major difference that it takes advantage of the fact that it has been initialized not far from the minimum and as such, can quickly converge by refraining from checking the perimeter excessively. We call this method 'guided optimization'. We define a minimum distance or resolution for the cost function. The cost function keeps track of each point in the  $N$ -dimensional space that it evaluates and will only evaluate a new point if falls outside all previously evaluated points by the specified minimum distance. The cost function quickly returns a previously calculated value for points inside the minimum distance of a previously evaluated point, which limits the resolution of the cost function. Obviously, this method is not suitable as a general purpose optimization tactic, and can only be used when the optimization algorithm can be properly initiated.

## 4. Experiments

We tested our method on 2D aerial images and 3D simulated MR images of brain generated by Brain Web [1]. The experiments on MR images were to demonstrate the performance of our method for the more difficult problem of registering multi-modal images and to motivate further extension of the method to support full 3D registration. We calculated the registration error for 2D images by selecting 10,000 control points uniformly distributed across the image and calculating the average distance between the transformed control points under the calculated transformation and the actual transformation. For 3D images we located the centroid of the brain volume and calculated the distance between the transformed centroids under calculated and actual transformations.

### 4.1. Registration of 2D Images

We compared GI method with the conventional PI method by conducting two sets of experiments on aerial images. For each run we randomly created 500 transformation matrices within the dynamic range of the parameter space. The reference image was transformed by each matrix and GI and PI registration was performed on the reference and the transformed images. A smaller dynamic range was used for the first set of experiments and the dynamic range was increased for the second round as shown in Table 2. The GI method was run with three resolutions high, medium and low (0.001, 0.01 and 0.1 pixels). The results are shown in Fig. 3 (darker colors are associated with the first experiment and lighter colors with the second experiment).

The robustness and performance of registration signifi-

	Round One Tests	Round Two Tests
<b>Translations</b>	$[-25 \ 25]$	$[-50 \ 50]$
<b>Rotation</b>	$[-25^\circ \ 25^\circ]$	$[-180^\circ \ 180^\circ]$
<b>Scale</b>	$[0.75 \ 1.25]$	$[0.5 \ 1.30]$

Table 2. Dynamic range of parameters used in the experiments

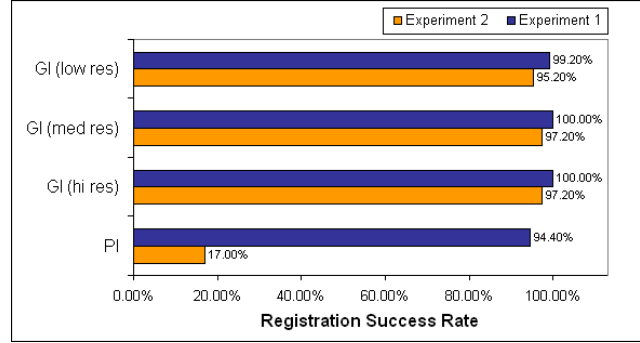
cantly improved under the GI method as shown in Fig. 3. We declared a registration failed if the average pixel alignment error was more than two pixels. As shown in Fig. 3(a), the PI method performance is reasonable in the first experiment but as we increase the dynamic range in the second experiment the performance drops below 20% since the chance of being trapped by local minima increases. The GI method, on the other hand, performs very well in both case with performance levels more than 99% and 95% for experiment one and two, respectively. Also notable is the success rate of GI method at medium resolution which successfully registered all images in the first experiment and 486 out of 500 in the second experiment. This is due to the good initial estimates of transformation parameters under the GI method. Throughout our experiments estimation of rotation and scale parameters were very close to the actual alignment with a mean error of  $0.54^\circ$  and  $0.1\%$  for rotation and scale, respectively. However, the translation estimations which are mainly based on centroids are less accurate with an average error of 4.1 pixels and deteriorate under heavy clipping of the images (centroid calculation is sensitive to partial overlap). This explains the slight performance drop in GI method in the second experiment.

In addition to improved performance, GI-method improves efficiency by allowing the optimization step to converge more quickly as shown in Fig. 3(b). Obviously the efficiency can be improved by reducing the resolution of the algorithm. Based on experiments a medium choice of resolution seems to provide a good balance between the accuracy and efficiency.

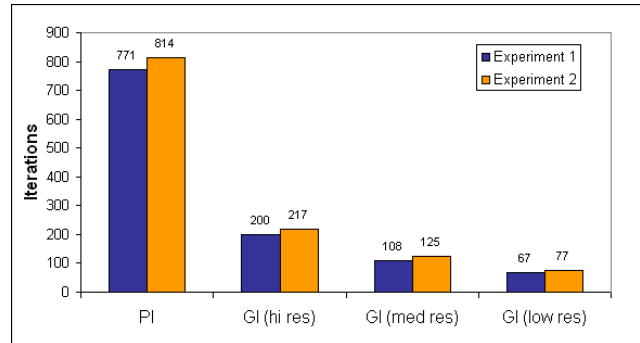
Fig. 4 shows the result of registration of aerial images of Sydney Opera House previously shown in Fig. 1(a) and Fig. 1(b) using the GI method. The GI method detected the correct transformation in 43 iterations while for the same experiment the PI method failed to register the images.

## 4.2. Registration of Multi-Modal Images

We experimented with several combinations of multi-modal MR images of brain to determine the applicability of our method in the presence of non-linear intensity variations. Fig. 6 shows a sample of T1, T2 and PD modalities used in our experiments [1]. The transformation were applied to the entire volume, however the misalignment be-

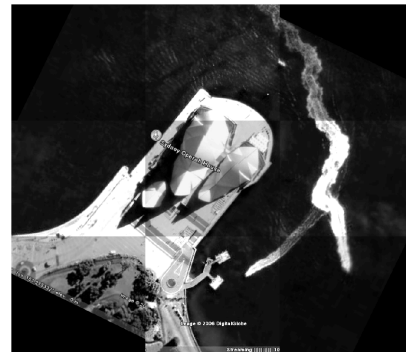


(a) Registration success rate for each method. Conventional PI method performance drops below 20% for larger dynamic ranges, while GI method sustains more than 95% success rate.



(b) Iterations required for convergence of the optimization algorithm for each method. The GI methods converge much more quickly.

Figure 3. Improved performance and efficiency of the GI method.

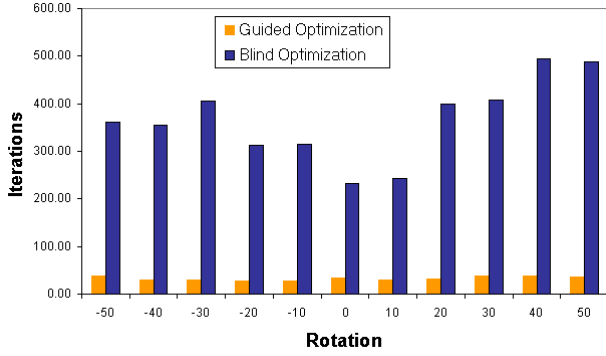


(a) GI method registers images successfully avoiding local minima

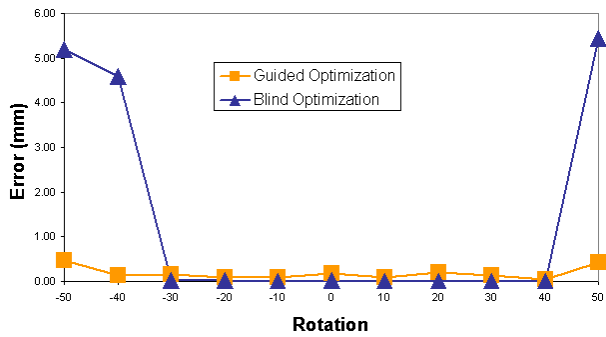
Figure 4. GI-based registration of 2D aerial images. Checkerboard overlays of images are displayed for visual comparison.

tween the images were limited to in-plane transformations.

Fig. 5 demonstrates that our optimization method improves performance and efficiency of the registration task for multi-modal images by giving an example for registration of T1 to T2 images. In this experiment, T2 was translated by  $(-20, 20)$  mm and scaled by 0.9, the rotation pa-



(a) More than an order of magnitude improvement in efficiency (execution times are proportional to iterations).



(b) The registration error for our method remains nearly constant and is well below the voxel size; the conventional method fails for rotations outside  $\pm 40^\circ$ .

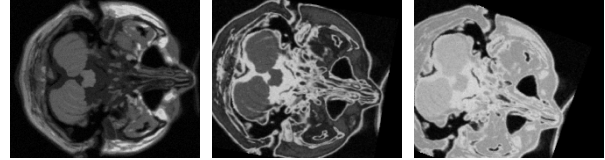
Figure 5. Superior robustness and improved efficiency of our method compared to the conventional method.

parameter was varied from  $-50^\circ$  to  $50^\circ$  in  $10^\circ$  increments. The efficiency of the registration improved on average by a factor of 12. The GI method was able to achieve accuracy well below the voxel size of 1 mm with an average error of 0.18 mm. The results shown in Fig. 5 are indicative and consistent with several other experiments, which we have not included due to space limitations. We experimented by gradually increasing other transformation parameters and tested our method on other combinations of image modalities (T1 to PD and T2 to PD).

## 5. Discussion

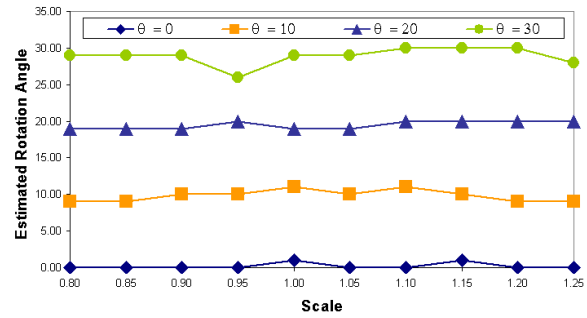
### 5.1. Accuracy of Estimating Rotation

Average error in estimating the rotation parameter for 2D experiments was  $0.54^\circ$  and for 3D experiments was below  $1^\circ$  (refer to Fig. 7), which is within the angular resolution of our experiments. This demonstrates that mutual information of gradient intensity is a strong measure of directional similarity, irrespective of translation and scaling between the images.

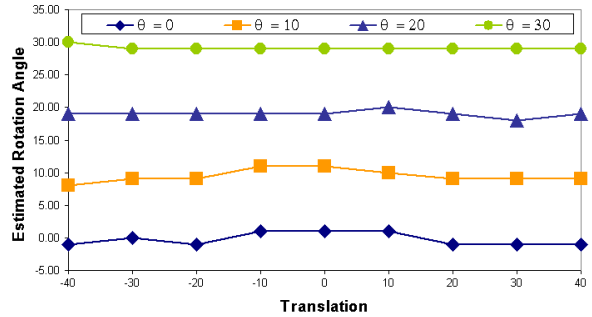


(a) MRI-T1. (b) MRI-T2. (c) MRI-PD.

Figure 6. A sample set of synthetic MR images of brain used in our experiments with voxel size of  $1 \text{ mm}^3$ .



(a) Estimated rotations for various misalignments vs. scale



(b) Estimated rotations for various misalignments vs. translation

Figure 7. Estimation of rotation parameter using GI method demonstrating invariance to scale and translation. The legend shows ground truth for each experiment.

### 5.2. Invariance to Translation and Scaling

When fixed and moving images are scaled by a factor within a reasonable range or translated, their shape remains unchanged. As such, one would expect that the directional content of image gradients be relatively maintained.

Fig. 7 underlines the invariance of the GI method to scale and translation. Fig. 7(a) shows a number of experiments, where an MRI-T2 is scaled from 0.8 to 1.25 while being rotated between  $0^\circ$  to  $30^\circ$ . The transformed image is then registered to a reference MRI-T1. The experiments show that GI method estimates the rotation parameter with an average error of  $0.6^\circ$ . Fig. 7(b) shows a number of experiments, where an MRI-T2 is translated from  $(-40, 40)$  to  $(40, 40)$  while being rotated between  $0^\circ$  to  $30^\circ$ . The transformed image is then registered to a reference MRI-T1. The exper-

iments show that GI-based method estimates the rotation parameter with an average error of  $0.9^\circ$ .

### 5.3. Robustness with Respect to Partial Overlap

In some of our experiments clipping of the transformed images has removed a significant number of important gradient vectors on the outer boundary of the image (e.g. Fig. 6(c)), however a close estimate of the rotation parameter can still be found, which demonstrates the robustness of the gradient intensity method for partially overlapping images.

### 5.4. Guided Optimization vs Blind Optimization

The guided optimization outperforms blind optimization both in terms of efficiency and accuracy. The registration error for the guided optimization remains almost constant (e.g. Fig. 5(b)). While the average registration error is lower for the guided optimization, the minimum registration error can be higher for guided optimization at lower misalignments, unless a sufficiently high resolution is used. The choice of resolution for the guided optimization is determined by minimum accuracy required by the application. For example for most medical application sub-voxel accuracy is required which in our experiments could be achieved by a medium choice of resolution.

## 6. Future Work

We are currently investigating full 3D extension of the GI method for registration of medical images with out-of-plane transformations. The rotation parameters around  $x$ ,  $y$  and  $z$  axes are being derived by maximizing the MI of spherical angles of 3D gradients, mapped on a unit sphere. This allows us to derive three rotational parameters by optimizing MI function on a 2D plane which reduces the dimensionality of the problem and significantly improves the efficiency of registration with respect to rotational parameters. The translation and scale parameters are estimated using methods similar to those presented in this paper. The results are then used to initialize a guided optimization algorithm.

## 7. Conclusion

Distribution of gradient vectors, as a measure of directional strength of an image, captured by the gradient intensity, is a robust measure of similarity between the images regardless of their relative size or location. Mutual information of gradient intensity assumes its maximum, where the images are rotationally aligned without requiring the images to be optimized with respect to scale or translation. In our experiments the MI utility function of gradient intensity could be calculated in less than 1 ms for the entire rotation range. Decoupled estimates of scale and translation,

together with rotation were used to initialize an optimization algorithm to achieve sub-pixel accuracy. We showed that our method improves robustness of the optimization algorithm and reduces the overall computational complexity of the registration task by an order of magnitude.

## References

- [1] *Brain Web*. Montreal Neurological Institute, McGill University, <http://www.bic.mni.mcgill.ca/brainweb/>, 2006.
- [2] A. Cole-Rhodes, K. Johnson, J. LeMoigne, and I. Zavorin. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE Trans. on Image Processing*, 12(12):1495–1511, Dec. 2003.
- [3] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality medical image registration using information theory. In *Proc. Int. Conf. Information Processing in Med. Imaging: Computational Imaging and Vision 3*, pages 263–274, Apr. 1995.
- [4] X. Fang, H. Rhody, and E. Saber. Automatic registration of multi-sensor airborne imagery. In *Proc. IEEE 34th Applied Imagery and Pattern Recognition Workshop (ICIP)*, 2005.
- [5] D. A. Forsyth and J. Ponce. *Computer Vision, A Modern Approach*. Printice Hall, Pearson Education, Inc., Upper Saddle River, New Jersey, 2003.
- [6] J. Gluckman. Gradient field distributions for the registration of images. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, volume 2, pages 691–694, Sept. 2003.
- [7] J. Liu, J. Tian, and Y. Dai. Multi-modal medical image registration based on adaptive combination of intensity and gradient field mutual information. In *IEEE Int. Conf. of Engineering in Medicine and Biology Society (EMBS)*, pages 1429–1432, Aug. 2006.
- [8] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging*, 16(2):187–198, Apr. 1997.
- [9] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Med. Image Anal.*, 3(4):373–386, 1999.
- [10] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. on Med. Imaging*, 19(8):809–814, Aug. 2000.
- [11] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Trans. on Med. Imaging*, 22(8):986–1004, Aug. 2003.
- [12] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, second edition, 1992.
- [13] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423/623–656, 1948.
- [14] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, 1997.