Image Registration by Minimization of Residual Complexity

Andriy Myronenko and Xubo Song
Dept. of Science and Engineering, School of Medicine
Oregon Health and Science University, Portland, OR
myron@csee.ogi.edu, xubosong@csee.ogi.edu

Abstract

Accurate definition of similarity measure is a key component in image registration. Most commonly used intensity-based similarity measures rely on the assumptions of independence and stationarity of the intensities from pixel to pixel. Such measures cannot capture the complex interactions among the pixel intensities, and often result in less satisfactory registration performances, especially in the presence of nonstationary intensity distortions. We propose a novel similarity measure that accounts for intensity non-stationarities and complex spatially-varying intensity distortions. We derive the similarity measure by analytically solving for the intensity correction field and its adaptive regularization. The final measure can be interpreted as one that favors a registration with minimum compression complexity of the residual image between the two registered images. This measure produces accurate registration results on both artificial and real-world problems that we have tested, whereas many other state-of-the-art similarity measures have failed to do so.

1. Introduction

Accurate definition of similarity measure is a key component in image registration [6]. Most commonly used intensity-based similarity measures, including Sum-of-Squared-Differences (SSD), Correlation Coefficient (CC), Correlation Ratio (CR) and Mutual Information (MI), rely on the assumption of independence and stationarity of the intensities from pixel to pixel [14]. These similarity measures are defined only between the corresponding pixels without considering their spatial dependencies. Further, the intensity relationship is assumed to be spatially stationary. As a result, such measures tend to fail when registering two images corrupted by spatially-varying intensity distortion.

Real-world images often have spatially-varying intensity distortions. For instance, brain MRI images often can be corrupted by slow-varying intensity fields; visual-band images can have illumination non-homogeneity and reflectance artifacts [6]. These complex distortions do not obey the pixel-wise independence or stationarity assumption and cannot be captured by simplistic intensity relationships. To illustrate this argument, consider aligning the two images in Figure 1. The smaller image (B) is a crop of the larger image (A) corrupted by an additive spatially-varying intensity field (simulated by the sum of random Gaussians). Figure 2 plots the values of several similarity measures, including SSD, CC, MI and our new similarity measure called Residual Complexity (RC), with respect to horizontal shift of B over A. Only RC achieves its optimum at the correct image alignment (zero translation).

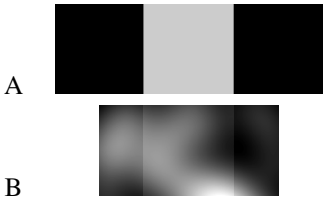


Figure 1. Gray stripe registration experiment.

To deal with complex spatially-varying intensity distortion, one has to account for non-stationary intensity relationships and spatial pixel dependencies. El-Baz et al. [2] proposed to learn a prior appearance model of the first image using a Markov-Gibbs random field with pairwise interaction, and then transform the second image to maximize its probability under the learned appearance model. Wyatt and Noble [22] proposed to use Markov random fields (MRF) to iteratively segment the images and register the class labels. Zheng and Zhang [24] proposed a MAP-MRF framework for intensity-based similarity measures. The framework includes some standard similarity measures and allows defining new ones taking advantage of local pairwise intensity interactions. The computational complexity of such methods is high, while the demonstrated performance is limited. Finally, some methods attempts to correct image intensity simultaneously with registration to remove the non-stationary distortions [3, 12, 8]. Friston et al. [3] pro-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

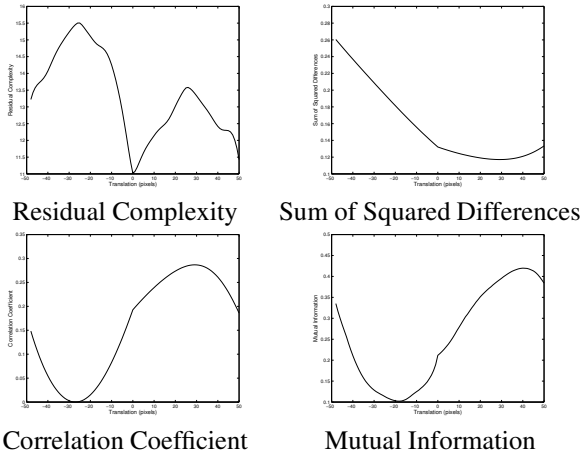


Figure 2. Plot of several similarity measures (RC, SSD, CC, MI) with respect to the translation of the smaller image over the larger one in Figure 1. Zero translation corresponds to the accurate registration. Only the Residual Complexity similarity measure gives an optimum at the correst registration with a wide convergence range.

posed to align the images using standard similarity measure (SSD), but assuming that one of the images has to be intensity corrected using a non-linear intensity transformation and a convolution filter. The intensity correction function is defined as a linear combination of some basis functions with spatially smooth-varying coefficients. The convolution filter has to be chosen manually for a specific problem or estimated from the images. Modersitzki and Wirtz [8] used a similar approach, but defined a multiplicative intensity correction function with a total variation regularizer. In more recent work, Ashburner and Friston [1] proposed a probabilistic framework for joint registration, intensity normalization and segmentation, using alternating optimization of corresponding parameters. This method demonstrates accurate performance on brain MRI images, but requires a manual choice of intensity normalization parameters.

We propose a new intensity-based similarity measure to deal with complex spatially-varying intensity distortions. We start deriving the similarity measure by introducing an intensity correction field that brings images into agreement in intensity space. We learn the adaptive regularization for the correction field. Analytically solving for the correction field and adaptive regularization allows us to derive an adaptive similarity measure that is robust to spatial intensity distortions. Interestingly, the final form of our similarity measure has many analogies in several computer vision areas, such as image compression, sparse coding and topographic learning. Our similarity measure can be interpreted as one that favors a registration with minimum compression complexity of the residual image between the two registered images. Thus, we name it the Residual Complexity (RC) similarity measure.

2. Method

Consider two images I and J to be aligned, assuming the following intensity relationship:

$$I = J(\mathcal{T}) + S + \eta \tag{1}$$

where S is an intensity correction field and η is zero mean Gaussian noise (note that for any two images there always exists a correction field S). $\mathcal{T}$ is the geometric transformation that aligns I and J. The maximum a posteriori (MAP) approach to estimate S and $\mathcal{T}$ is to maximize the probability

$$P(\mathcal{T}, S|I, J) \propto P(I, J|\mathcal{T}, S)P(\mathcal{T})P(S), \tag{2}$$

where we assume the independence of $\mathcal{T}$ and S. The first term is a joint likelihood of the images, which leads to the familiar similarity measure of SSD [14]; $P(\mathcal{T})$ is a prior used to regularize the transformation and $P(S)$ is a prior on S that reflects our assumption on spatial intensity interactions. Now, we also assume that pixel-wise probabilities are i.i.d., but only given the correction field. It is common to formulate the prior on the correction field in the form $P(S) \propto e^{-\beta\|\mathbf{P}S\|^2}$. MAP estimation in Equation 2 is equivalent to minimization of the following objective function:

$$E(\mathbf{S}) = \|\mathbf{I} - \mathbf{J} - \mathbf{S}\|^2 + \beta \|\mathbf{PS}\|^2 \tag{3}$$

where images and correction field are in column-vector form, $\|\cdot\|$ is Euclidean norm and $\mathbf{P}$ is a regularization operator for $\mathbf{S}$ (we have not yet specified the form of $\mathbf{P}$). For now we omit the transformation $\mathcal{T}$ and its regularization term, to make the derivation of $\mathbf{S}$ more clear.

Equating the gradient to zero, we solve for $\mathbf{S}$:

$$-2(\mathbf{I} - \mathbf{J} - \mathbf{S}) + 2\beta\mathbf{P}^T\mathbf{PS} = 0 \tag{4}$$

$$\mathbf{S} = (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\mathbf{r} \tag{5}$$

where $\mathbf{Id}$ is the identity matrix and $\mathbf{r} = \mathbf{I} - \mathbf{J}$ is the residual vector (difference image). The inverse always exists, because $\mathbf{P}^T\mathbf{P}$ is positive semidefinite. Substituting $\mathbf{S}$ back into the objective function (3), we get:

$$\begin{aligned} E = &\left\|\left(\mathbf{Id} - (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\right)\mathbf{r}\right\|^2 + \\ &\beta\left\|\mathbf{P}(\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\mathbf{r}\right\|^2 = \\ &\mathbf{r}^T\left(\mathbf{Id} - (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\right)\left(\mathbf{Id} - (\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\right)\mathbf{r} \\ &+ \beta\mathbf{r}^T(\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{P}(\mathbf{Id} + \beta\mathbf{P}^T\mathbf{P})^{-1}\mathbf{r} \end{aligned} \tag{6}$$

The matrix $\mathbf{P}^T\mathbf{P}$ is square, symmetric and positive semidefinite. Thus, it allows spectral decomposition:

$$\mathbf{P}^T\mathbf{P} = \mathbf{Q}\Lambda\mathbf{Q}^T, \quad \Lambda = \mathrm{d}\left[\lambda_1, .., \lambda_N\right],\ \lambda_i \geq 0 \tag{7}$$

Substituting Equation 7 in Equation 6, we can simplify the objective function greatly, because $\mathbf{Q}$ is orthogonal, and the

majority of multiplications and inversions are only with diagonal matrices within.

$$E = \mathbf{r}^T\mathbf{Q}\,\mathrm{d}(\frac{\beta^2\lambda_i^2 + \beta\lambda_i}{(1+\beta\lambda_i)^2})\mathbf{Q}^T\mathbf{r}$$
$$= \mathbf{r}^T\mathbf{Q}\,\mathrm{d}(\frac{\beta\lambda_i}{1+\beta\lambda_i})\mathbf{Q}^T\mathbf{r} = \mathbf{r}^T\mathbf{Q}\mathbf{L}\mathbf{Q}^T\mathbf{r} = \mathbf{r}^T\mathbf{A}\mathbf{r}. \quad (8)$$

where $\mathrm{d}()$ denotes a diagonal matrix. With $\mathbf{A} = \mathbf{Q}\mathbf{L}\mathbf{Q}^T$ we defined a new square, symmetric and positive semidefinite matrix with eigenvalues $\mathbf{L}$:

$$\mathbf{L} = \mathrm{d}([l_1, .., l_N]) = \mathrm{d}(\frac{\beta\lambda_i}{1+\beta\lambda_i}),\ 1 \geq l_i \geq 0 \quad (9)$$

Note that eigenvalues of $\mathbf{A}$ are all nonnegative and bounded between $[0, 1]$. Now, if we choose a particular regularization operator $\mathbf{P}$, then $\mathbf{A}$ is known and we can minimize

$$E(\mathcal{T}) = \mathbf{r}^T\mathbf{A}\mathbf{r} = (\mathbf{I} - \mathbf{J}(\mathcal{T}))^T\mathbf{A}(\mathbf{I} - \mathbf{J}(\mathcal{T})) \quad (10)$$

with respect to the transformation $\mathcal{T}$. The matrix $\mathbf{A}$ is of size N^2, where N is a number of image pixels. This equation represents the squared Mahalanobis distance between the images, where $\mathbf{A}$ represents an inverse covariance matrix.

So far we have not specified the eigenvectors and eigenvalues of the operator $\mathbf{P}^T\mathbf{P} = \mathbf{Q}\Lambda\mathbf{Q}^T$. A large class of operators $\mathbf{P}^T\mathbf{P}$ have the same eigenvector basis $\mathbf{Q}$ and differ only in eigenvalues. In our method, we will choose a particular form of $\mathbf{Q}$ (see Section 3) without specifying the eigenvalues, which are absorbed in matrix $\mathbf{L}$. Instead, we estimate $\mathbf{L}$ within the optimization framework. Such adaptive regularization allows greater flexibility with respect to the choice of $\mathbf{P}$. Consider solving for the eigenvalues matrix $\mathbf{L}$:

$$E(\mathbf{L}) = \mathbf{r}^T\mathbf{A}\mathbf{r} = (\mathbf{Q}^T\mathbf{r})^T\mathbf{L}(\mathbf{Q}^T\mathbf{r}),\ 1 \geq l_i \geq 0 \quad (11)$$

A trivial, but not interesting, minimum for $E(\mathbf{L})$ is with an all-zeros matrix $\mathbf{L}$. To avoid this, we impose a regularization on the eigenvalues $\mathbf{L}$ that prefers $\mathbf{L}$ to be flat (has relatively equal elements) and large but still bounded in $[0, 1]$. The motivation is as follows. Notice that if $\mathbf{L}$ is identity (or multiples of identity), then the inverse covariance $\mathbf{A}$ is also identity, and the objective function simplifies to regular SSD:

$$E(\mathbf{L}) = (\mathbf{Q}^T\mathbf{r})^T\mathbf{I}(\mathbf{Q}^T\mathbf{r}) = \mathbf{r}^T\mathbf{r} = \|\mathbf{r}\|^2 \quad (12)$$

which is equivalent to the assumption of no regularization on $\mathbf{S}$. SSD is a valid and optimal objective function if the images are corrupted by i.i.d. Gaussian noise. The more the eigenspectra ($\mathbf{L}$) deviate from flat, the more off-diagonal elements appear in $\mathbf{A}$, which means that the noise is more correlated. As far as the true noise covariance is unknown, we want to assume the least required correlation possible and thus a flat $\mathbf{L}$. This leads us to define a regularization term on $\mathbf{L}$ as

$$R(\mathbf{L}) = -\sum_i \log l_i + \sum_i l_i \quad (13)$$

It consists of two terms: the first term forces l_i to be nonnegative (similar to a log-barrier function) and flat (similar to ML estimation); the second term keeps eigenvalues from growing arbitrary large. Now, our objective function becomes

$$E(\mathbf{L}) = (\mathbf{Q}^T\mathbf{r})^T\mathbf{L}(\mathbf{Q}^T\mathbf{r}) + \alpha R(\mathbf{L}) \quad (14)$$

where α is a trade-off parameter. Differentiating the equation with respect to all l_i and equating the derivative to zero, we obtain

$$(\mathbf{q}_i^T\mathbf{r})^2 - \alpha\frac{1}{l_i} + \alpha = 0;\ \ l_i = \frac{1}{(\mathbf{q}_i^T\mathbf{r})^2/\alpha + 1} \quad (15)$$

where $\mathbf{q}_i$ are eigenvectors in $\mathbf{Q}$ ($\mathbf{Q} = [\mathbf{q}_1, .., \mathbf{q}_N]$). As a check we see that l_i are indeed bounded within $[0, 1]$. Substituting this result back into Equation 14, we obtain

$$E = \sum_{n=1}^{N}[\frac{(\mathbf{q}_n^T\mathbf{r})^2}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1} - \alpha\log\frac{1}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1} +$$
$$\frac{\alpha}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1}] = \alpha N - \alpha\sum_{n=1}^{N}\log\frac{1}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1} \quad (16)$$

Now, we bring back the geometric transformation $\mathcal{T}$ and ignore the terms independent of $\mathcal{T}$; we obtain a novel similarity measure

$$E(\mathcal{T}) = -\sum_{n=1}^{N}\log\frac{1}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1};\ \ \mathbf{r} = (\mathbf{I} - \mathbf{J}(\mathcal{T})) \quad (17)$$

Note that even though we have derived it starting from solving for the correction field and its regularizer, they are not explicitly present in the final similarity function form. However, we still need to define the basis eigenvectors $\mathbf{Q}$.

3. DCT basis

Now, we shall proceed to specify $\mathbf{Q}$. Recall that the functional form of basis $\mathbf{Q}$ initially comes from the eigenvectors of $\mathbf{P}^T\mathbf{P} = \mathbf{Q}\Lambda\mathbf{Q}^T$, where $\mathbf{P}$ is a regularization operator of the correction field. We choose the discrete cosine transform (DCT) basis [19] as a functional form of $\mathbf{Q}$,. The reasons are two-fold: DCT eigenvectors correspond to the discrete derivative-based regularizer $\mathbf{P}$ and they are the eigenvectors of the covariance matrix of weakly stationary stochastic process. In this case, the matrix multiplication $\mathbf{Q}^T\mathbf{r}$ is just a discrete cosine transform (DCT) of $\mathbf{r}$, which can be computed through FFT in $\mathcal{O}(N\log N)$ [19].

Derivative regularization: Consider the first order derivative regularizer (finite differences matrix) $\mathbf{P}_1$, the corresponding self-adjoint operator $\mathbf{P}_2 = \mathbf{P}_1^T\mathbf{P}_1$, which appears after taking the gradient, is a second order derivative matrix (discrete Laplacian), which decomposes as [19]:

$$\mathbf{P}_2 = \mathbf{P}_1^T\mathbf{P}_1 = \mathbf{Q}\Lambda\mathbf{Q}^T \tag{18}$$

where $\mathbf{Q}$ is the DCT basis (real and orthogonal). Depending on a choice of boundary conditions and boundary approximations, there are several DCT choices. The most common in image processing is DCT-2, which corresponds to Neumann midpoint boundary conditions [19].

If we use the second order regularizer $\mathbf{P}_2$, the corresponding self-adjoint operator, $\mathbf{P}_2^T\mathbf{P}_2$, is the fourth order discrete derivative operator $\mathbf{P}_4 = \mathbf{P}_2^T\mathbf{P}_2 = \mathbf{Q}\Lambda\mathbf{Q}^T\mathbf{Q}\Lambda\mathbf{Q}^T = \mathbf{Q}\Lambda^2\mathbf{Q}^T$, with same set of eigenvectors as $\mathbf{P}_2$. Similarly, derivative operators of higher orders and their linear combinations, lead to eigenvectors $\mathbf{Q}$ and differ only in eigenvalues. We use a DCT basis for $\mathbf{Q}$, because it corresponds to derivative-based regularization.

Weak stationarity: Another motivation to use DCT, is that it corresponds to the covariance matrix of a certain class of signals. Equation 10 can be interpreted as the Mahalanobis distance, where $\mathbf{A}$ is an inverse covariance matrix. By using DCT eigenvectors for $\mathbf{A}$, we can see that the covariance matrix is also diagonalized by $\mathbf{Q}$:

$$\Sigma = \mathbf{A}^{-1} = (\mathbf{Q}\mathbf{L}\mathbf{Q}^T)^{-1} = \mathbf{Q}\mathbf{L}^{-1}\mathbf{Q}^T \tag{19}$$

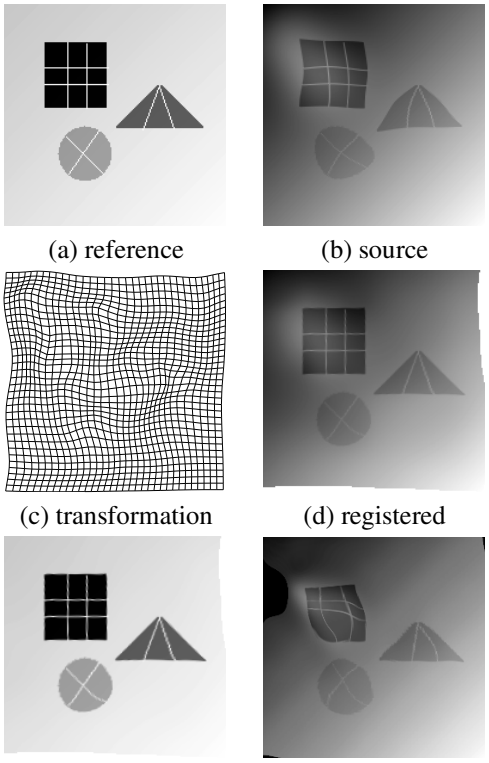
Sanchez et al. [17] showed that matrices that are diagonalized by DCT basis are symmetric Toeplitz plus scaled near-Hankel matrices. They also showed that DCT is asymptotically optimal for all finite-order Markov processes; DCT bases are the eigenvectors of the autocovariance matrix.

It is well known in signal processing theory that with the assumption of the constant mean and the Toeplitz covariance, a stochastic process is said to be weakly stationary [5]. Thus using the DCT basis $\mathbf{Q}$ is implicitly related to assumptions of a weakly stationary residual image or a finite-order Markov process.

4. Implementation

We model the transformation $\mathcal{T}$ using the Free Form Deformation (FFD) transformation with 3 hierarchical levels of B-spline control points [15]. We use the gradient descent optimization method to iteratively update the transformation parameters $\mathcal{T}$. The pseudo-code to compute the objective function and its gradient is

$$\mathbf{r} = \mathbf{I} - \mathbf{J}(\mathcal{T});\ \mathbf{c} = dctn(\mathbf{r});\ E = \sum \log(\mathbf{c}^2/\alpha + 1);$$
$$\nabla E = -idctn\left(\frac{2\mathbf{c}/\alpha}{\mathbf{c}^2/\alpha + 1}\right)\nabla\mathbf{J}(\mathcal{T})\frac{\partial\mathcal{T}}{\partial\theta}$$



(a) reference (b) source (c) transformation (d) registered (e) clean registered (f) typical NMI result

Figure 3. Synthetic experiment 1. We register the source image (b) onto the reference image (a). (d) The registered image. (c) The estimated transformation. (e) The registered image without intensity corruption. (f) Typical performance of NMI similarity measure. Other tested similarity measures also do not produce any satisfactory results, whereas, RC demonstrates accurate performance.

where $dctn$ and $idctn$ are the forward and inverse multi-dimensional DCTs, $\nabla\mathbf{J}$ is the intensity image gradient and θ represents the transformation parameters. We set the parameter α to 0.05. We implemented the algorithm in Matlab, and tested it on an AMD Opteron 2GHz Linux machine. The computation time was about 20 sec for a pair of images.

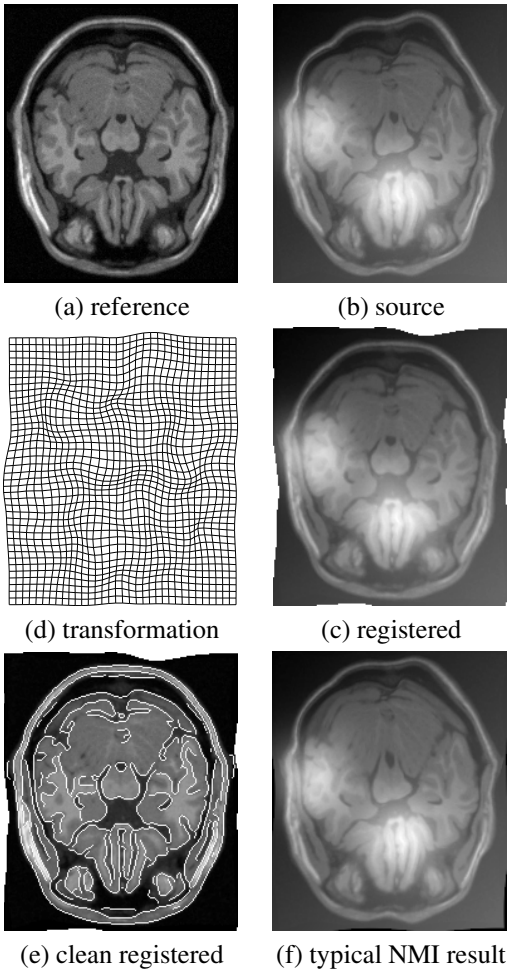
5. Results

We show the performance of the new similarity measure on several challenging synthetic and real-life examples. For the comparison, we have also tested registration with the SAD, SSD, CC, CR, MI, NMI similarity measures, implemented in Deformable Registration using Discrete Optimization (DROP) software [4] and in Image Registration Toolkit (ITK) [15].

Synthetic Examples: We generated synthetic source images by introducing both geometric and intensity distortions to the reference image, and then register source images onto the reference one. To simulate the geometric distortion, we perturb a uniform grid of points followed by the thin-plate

Table 1. The parameter values used for the distortion simulation for each of the distortion levels. The Gaussian means μ_K were selected randomly from the interval $[1, \Omega]$, where Ω is the image domain size. Values of γ are selected randomly from the listed range.

Distortion level	w	K	μ_K	σ	γ
1	0.2	1	$[1, \Omega]$	30	1
2	0.4	1	$[1, \Omega]$	30	$[0.9, 1.2]$
3	0.8	2	$[1, \Omega]$	30	$[0.8, 1.5]$
4	1.0	8	$[1, \Omega]$	30	$[0.5, 2]$
5	2.0	20	$[1, \Omega]$	30	$[0.4, 3]$



(a) reference (b) source (d) transformation (c) registered (e) clean registered (f) typical NMI result

Figure 4. Synthetic experiment 2. We register the source image (b) onto the reference image (a). (d) The registered image. (c) The estimated transformation. (e) The registered image without intensity corruption together with contours extracted from the reference image. (f) Typical performance of NMI similarity measure. Other tested similarity measures also do not produce any satisfactory results, whereas, RC demonstrates accurate performance.

spline (TPS) interpolation according to the grid deformation. The grid size was 7×7 and its random perturbation was drawn from a zero-mean Gaussian with 3 pixels std. To simulate the spatially-varying intensity distortion we corrupted one of the images according to the formula

- $I(x, y) = I^{\gamma}(x, y) + w\frac{xy}{MN} + \frac{1}{K}\sum_{k=1}^{K} e^{-\frac{\|[x;y]-\mu_k\|^2}{2\sigma^2}}$;
- Rescale I to $[0, 1]$.

where the first term represents gamma correction on I after geometric distortion, the second term models a smoothly varying global intensity field and the last term models locally-varying intensity field with mixture of K Gaussians.

We performed 100 automatic registrations for different levels of synthetic spatially-varying intensity distortions and a fixed level of known nonrigid deformation, reinitialized randomly at every run. Table 1 shows the values of the parameters used during simulation for each of the distrotion levels. To quantify the registration performance, we compute the intensity root mean square error (RMSE) between the reference and the clean registered images, where the clean registered image is obtained by applying the transformation $\mathcal{T}$ to the source image without intensity distortion. We compare the RC performance to several methods: a) Pre-processing (high-pass filtering) followed by SSD registration b) Fixed regularizer method (Eq. 10) c) Normalized MI (ITK implementation [15]). For the fixed regularizer method, we assumed a particular fixed regularization operator $\mathbf{P}$ (discrete Laplacian), and used Eq. 10 as a similarity measure.

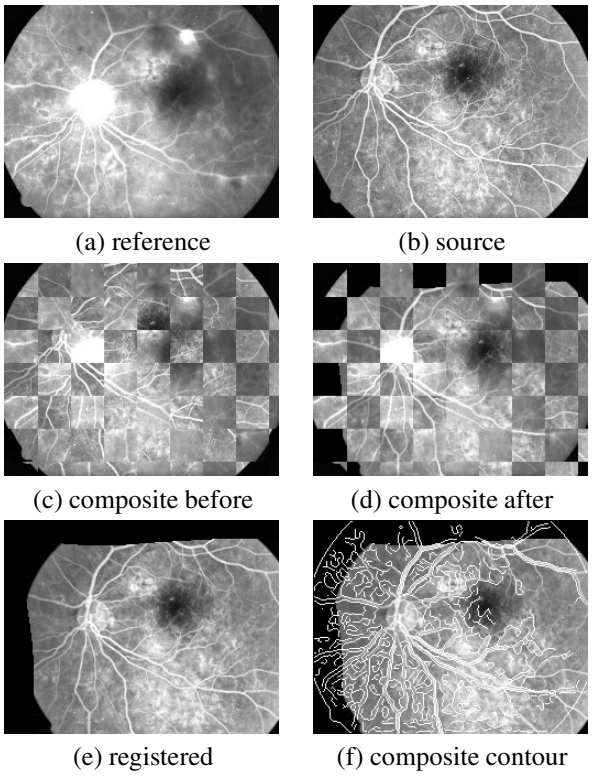
Table 2 shows the final RMSE value as a function of intensity distortion level. Figure 3 shows a typical performance on synthetic images (186×186 pixels) for the distortion level 2. Figure 4 shows a typical performance on brain MRI images (223×187 pixels) for the distortion level 3. We used brain MRI images to evaluate the performance shown in Table 2. The registration results obtained using the RC similarity measure are accurate. The intensity pre-processing approach shows poor performance. Perhaps different way of pre-processing can improve the registration performance, but the true intensity pre-processing is not known beforehand. Choosing a particular fixed regularizer $\mathbf{P}$ (we used Laplacian) demonstrates a reasonable registration performance, but still has a significant visible misalignment. Finally, normalized MI shows poor performance. Indeed, the RMSE between the images prior to intensity distortion without registration was 0.0913 on average, which is comparable to the NMI performance in Table 2. The NMI results have little or no improvement over the unregistered ones in most of the experiments.

Retina Images: We registered retina images taken 2 years apart [23]. Retinal images are used in ophthalmology to assess the evolution of illness, e.g. diabetic retinopathy [16, 10]. For this, the images have to be aligned first.

Retina images are challenging to register due to multiple intensity artifacts including non-uniform background

Table 2. Experimental results for different spatially-varying intensity distortion levels.

Distortion level	Residual Complexity RMSE, pixels	Preprocessing+SSD RMSE, pixels	Fixed regularizer (Eq. 10) RMSE, pixels	Normalized MI RMSE, pixels
1	0.0106 ± 0.0025	0.0531 ± 0.0071	0.0213 ± 0.0034	0.0416 ± 0.0065
2	0.0121 ± 0.0042	0.0585 ± 0.0088	0.0289 ± 0.0031	0.0532 ± 0.0142
3	0.0285 ± 0.0048	0.0594 ± 0.0085	0.0371 ± 0.0024	0.0786 ± 0.0121
4	0.0331 ± 0.0063	0.0693 ± 0.0092	0.0436 ± 0.0069	0.0681 ± 0.0161
5	0.0364 ± 0.0071	0.0675 ± 0.0072	0.0441 ± 0.0062	0.0924 ± 0.0189

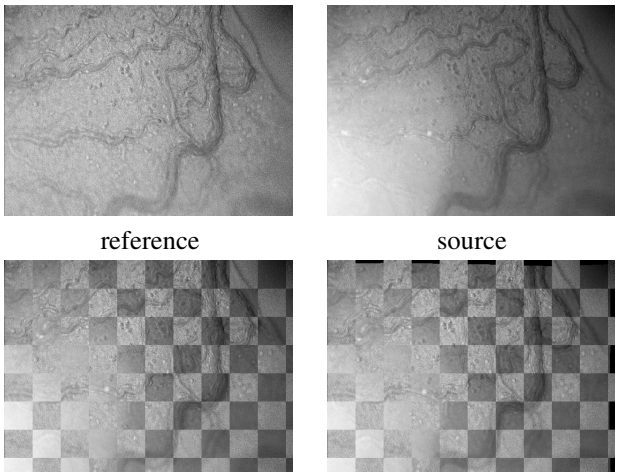


(a) reference (b) source

(c) composite before (d) composite after

(e) registered (f) composite contour

Figure 5. Registration of retina images. (a) reference image; (b) source image (taken 2 years prior); (c) composite view before registration; (d) composite view after registration; (e) registered source image; (f) composite view through contour overlap after the registration. The RC registration result is accurate; vessel structures are aligned despite intensity distortions and artifacts.



reference source

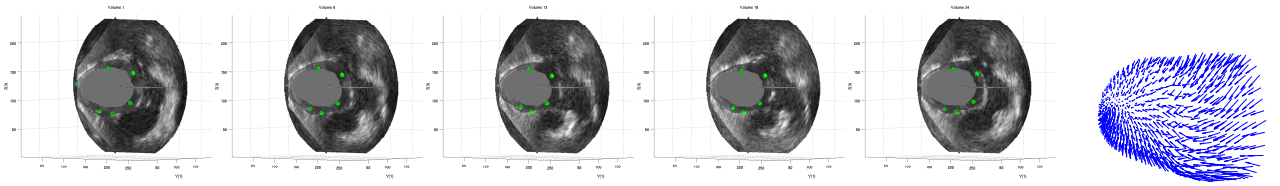
composite before composite after

Figure 6. Registration of 2 iris images. Notice the intensity variation across the images as well as intensity artifacts from moving cells, nonconsistent vessel reflections, etc. Our method accurately registers the images without any additional preprocessing.

and blood vessels (with intensity variations and changed structure) [10]. Most of the retina registration methods are feature-based [16, 23]; they extract the vascular structure or landmarks and align them. Figure 5 demonstrates the RC performance on 2 retina images (200×250 pixels). We achieved accurate registration results as demonstrated through composite views. We did not do image preprocessing or normalization and did not excluded the background outside the retina circle. We used the images as they appear in the figure. Also we did no rigid pre-registration beforehand and no regularization of the FFD transformation. None of the other tested similarity measures succeeded in registering the retina images. The CC, MI and CR methods were successful at registering the retina images only after we defined the region of interest to exclude the area outside the retina circle and after image filtering and heavy regularization of the transformation (to the extend where the transformation is almost rigid).

Iris Images: We stabilize a video sequence of microscopic iris images through frame-by-frame registration. This was necessary to remove the severe jitter and deformation across frames in order to be able to track leukocyte motion [18]. The deformation between frames is highly nonlinear. Notice the intensity variation across the images as well as intensity artifacts from moving cells, nonconsistent vessel reflections, etc. Our algorithm proved accurate and effective for these images (Figure 6). Other similarity measures, including NMI, CR, CC, SSD, showed poor performance on iris images without preprocessing, even though different combinations of parameters were tried.

3D Echocardiography: We sequentially registered a set 3D echo images (24 frames $192 \times 274 \times 248$ voxels) to find a displacement field of imaged area through time [9]. The displacement field can be used to study the myocardial



volume 1 volume 8 volume 13 volume 18 volume 24 displacement ED-ES

Figure 7. Estimated motion of the LV contour found during the cardiac cycle (24 consecutive volumes). LV achieves the maximum contraction at volume 13, then the LV dilates (diastolic phase) up to the volume 22 and starts contracting again (systolic phase). The last plot shows the displacement fields on LV endocardium between end diastole (ED) and end systole (ES) volumes. We validate the accuracy of RC registration with groundtruth motion of implanted sonomicrometers. The RC registration-based estimated motion is accurate.

deformation. We acquired the 3D echo sequences using a Philips iE33 with EKG gating from openchest piglets. Figure 7 shows the estimated motion of LV superimposed on the 3D echo images and the displacement vector field visualized at the LV endocardium position between end diastole and end systole volumes. We used a groundtruth sonomicrometer motion to validate the algorithm. We have omitted the description of this procedure in this paper, but note that the estimated motion has shown a high correlation with the groundtruth one, which shows the high accuracy of our method [9].

6. Related work

Analogy with compression and complexity: The term $\mathbf{Q}^T\mathbf{r}$ is a vector of the DCT coefficients of the difference image $\mathbf{r}$. DCT is popular in image compression, because most of the image information is represented by a few low-frequency DCT coefficients. Also DCT approximates the optimal (in the decorrelation sense) Karhunen-Loève transform with a certain Markovian assumption between the image pixels [13]. This is the idea used in earlier JPEG image compression format: roughly speaking, find DCT and ignore the smallest coefficients.

In our case, $\mathbf{Q}^T\mathbf{r}$ are the DCT coefficients of the residual image. Intuitively, the image can be highly compressed if only a few coefficient are non-zero (sparseness). Our similarity measure indeed forces sparseness of the coefficients

$$E = -\sum_{n=1}^{N} \log \frac{1}{(\mathbf{q}_n^T\mathbf{r})^2/\alpha + 1} \propto \sum_{n=1}^{N} \log((\mathbf{q}_n^T\mathbf{r})^2 + \alpha) \tag{20}$$

because, the logarithm function decreases quickly to zero compaired to its increase for larger values of the DCT coefficients. This is strongly related to minimization of so-called *zero-norm* (or cardinality) of parameters [21].

Our similarity measure implicitly enforces high compressibility of the residual image (or lower Kolmogorov complexity [7]) using the DCT coefficients (thus, the name Residual Complexity (RC)).

Analogy with sparse coding In sparse coding [11], one tries to decompose images in terms of basis functions $\mathbf{W} = [\mathbf{w}_1, .., \mathbf{w}_N]$ and sparse coefficients $\mathbf{c} = [c_1, .., c_N]^T$, by minimizing

$$E(\mathbf{W}, \mathbf{c}) = \sum_i \left\|\mathbf{x}^i - \mathbf{W}\mathbf{c}\right\|^2 + \lambda \sum_n \log(1 + (\frac{c_n}{\sigma})^2) \tag{21}$$

where $\mathbf{x}^i$ are vectorized image patches, and the bases ($\mathbf{W}$) are usually overcomplete and nonorthogonal. The last term measures the sparseness of coefficients $\mathbf{c}$ [11], which is similar to our objective function in Equation 17. Variations of this formulation lead to principal components, independent components, or more specialized filters [20].

Welling et al. [20] proposed to learn the overcomplete set of filters $\mathbf{J} = [\mathbf{J}_1, .., \mathbf{J}_N]$, using the product of experts (PoE) framework. PoE is an energy based method, which defines a probability of $\mathbf{x}$ as a normalized product of all the distributions represented by the individual experts. Unnormalized experts are chosen to be Student-t distributions, because this distribution has heavy tails, which makes it a suitable candidate for modeling constraints that are found in natural images. The energy of the PoE model is

$$E(\alpha, \mathbf{J}) = -\sum_{i=1}^{M} \alpha_i \log \frac{1}{(1 + \frac{1}{2}(\mathbf{J}_i^T\mathbf{x})^2)} + \log(Z) \tag{22}$$

which is again similar to our objective function in Equation 17, but for a different purpose. The estimated filters $\mathbf{J}_i$ are further applied for image denoising and inpainting tasks [20].

Whereas in sparse coding, the task is to search for sparse bases and coefficients to represent the image, in image registration we are given the bases and looking for the alignment where the residual image has sparse representation by the bases.

7. Discussion and Conclusion

We have derived a novel similarity measure for image registration, which accounts for spatial intensity distortions

and nonstationarities of the images. The similarity measure intuitively measures the coding complexity of the residual image. This measure produces accurate registration results on both artificial and real-world problems that we have tested, whereas many other state-of-the-art similarity measures have failed to do so.

Our similarity measure requires initialization of the coefficient α, we found its value in a range $[0.01, 0.1]$ to give accurate results with no significant difference in performance. Generally smaller values of α forced sparser coefficients, which resulted in more accurate registration, but also lead to more local minima of the similarity measure.

Our underlying assumption on the correction field was a simple additive one. Nevertheless, our similarity measure showed robust performance both during our synthetic and real experiments, where the intensity correction function can be complex. The explanation for the accurate performance of RC in such cases is as follows - even when the real residual image can be complex, its form still has to be simpler (represented sparsely by the basis functions) than the one with inaccurate registration. Thus, RC will still work with some complex intensity distortions.

References

[1] J. Ashburner and K. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.

[2] A. El-Baz, A. Farag, G. Gimel'farb, and A. Abdel-Hakim. Image alignment using learning prior appearance model. In *ICIP*, pages 341–344, Oct 2006.

[3] K. Friston, J. Ashburner, C. Frith, J. Poline, J. D. Heather, and R. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2:165–189, 1995.

[4] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Contin. deform. meet efficient opt. linear progr. In *Inf. Proc. in Med. Imag.*, July 2007.

[5] R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2006.

[6] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in medicine and biology*, 46(3):R1–R45, Mar. 2001.

[7] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitanyi. The similarity metric. *IEEE Inf. theory*, 50(12), 2004.

[8] J. Modersitzki and S. Wirtz. Combining homogenization and registr. In *WBIR06*, pages 257–263, 2006.

[9] A. Myronenko, X. Song, and D. J. Sahn. LV motion tracking from 3D echocardiography using textural and structural information. In *MICCAI*, volume 4792 of *LNCS*, pages 428–435. Springer, Oct. 2007.

[10] J. Nunes, Y. Bouaoune, E. Delechelle, and P. Bunel. A multiscale elastic registration scheme for retinal angiograms. *CVIU*, 95(2):129–149, June 2004.

[11] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: a strategy employed by v1. *Vision Research*, 37:3311–3325, 1997.

[12] K. Pohl, J. Fisher, J. Levitt, M. Shenton, R. Kikinis, W. Grimson, and W. Wells. A unifying approach to registration, segmentation, and intensity correction. In *MICCAI*, volume 3749, pages 310–318, 2005.

[13] K. R. Rao and P. Yip. *Discrete cosine transform: algorithms, advantages, applications*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.

[14] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *Int. J. of Imaging Syst. and Technology: Sp. issue on 3D imaging*, 11:71–80, 2000.

[15] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Image Processing*, 18(8):712–721, Aug. 1999.

[16] N. Ryan, C. Heneghan, and P. de Chazal. Registration of digital retinal images using landmark correspondence by expectation maximization. *Image and Vision Computing*, 22(11), Sept. 2004.

[17] V. Sanchez, P. Garcia, A. M. Peinado, J. C. Segura, and A. J. Rubio. Diagonalizing properties of the DCT. *IEEE Trans. Signal Processing*, 43(11), 1995.

[18] X. Song, A. Myronenko, S. R. Plank, and J. T. Rosenbaum. Registration of microscopic iris image sequences using probabilistic mesh. In *MICCAI*, volume 4191 of *LNCS*, pages 553–560. Springer, Oct. 2006.

[19] G. Strang. The discrete cosine transform. *SIAM Review*, 41(1):135–147, 1999.

[20] M. Welling, G. E. Hinton, and S. Osindero. Learning sparse topographic representations with products of student-t distr. In *NIPS*, pages 1359–1366, 2003.

[21] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping. Use of the zero-norm with linear models and kernel methods. *JMLR*, 3:1439–1461, 03 2003.

[22] P. P. Wyatt and J. A. Noble. MAP MRF joint segmentation and registration. In *MICCAI*, volume 2488, pages 580–587. Springer-Verlag, 2002.

[23] F. Zana, I. Meunier, and J. C. Klein. A region merging alg. using mathematical morphology: application to macula detection. In *ISMM*, pages 423 – 430, 1998.

[24] G. Zheng and X. Zhang. A unifying MAP-MRF framework for deriving new point similarity measures for intensity-based 2D-3D registration. In *ICPR*, volume 2, pages 1181–1185, 2006.