MEASURING IMAGE SIMILARITY TO SUB-PIXEL ACCURACY

G.K. Rohde [1*], *D.M. Healy, Jr.*[1], *C.A. Berenstein* [1], *A. Aldroubi* [2]

[1] Department of Mathematics, University of Maryland, College Park, MD, USA. 20742.
[2] Mathematics Department, Vanderbilt University, Nashville, TN, USA. 37235

ABSTRACT

Measuring the similarity between different images is essential for performing registration (alignment) and many related tasks. We show that popular image similarity measures such as the sum of squared differences (SSD), cross correlation (CC), and mutual information (MI), particularly in the presence of noise, may present local optima at half pixel translation intervals when low order interpolators are used to model the discrete images. Several strategies for reducing local optima artifacts during registration are discussed. Results using magnetic resonance image (MRI) data are shown.

1. INTRODUCTION

Image registration consists of finding the spatial alignment between two or more images. Registration methods represent an important technological step that enables a number of tertiary procedures of interest including data fusion, quantitative parameter extraction, as well as image based tracking. In biomedical applications image registration is often used to correct for patient motion and geometric distortions in serial MRI acquisitions such as in functional MRI as well as diffusion tensor MRI. In addition, nonlinear (or nonrigid) registration methods are used for performing so called atlas-based segmentation, or in computational anatomy studies.

Registration methods capable of measuring alignment to sub-pixel accuracy are of increasing interest. This is due to the fact that many biological structures of interest may be of similar size than the resolution of the image. Moreover, in situations in which one tries to obtain spatial transformations that contain rotations or more general nonrigid spatial transformations, one has no choice but to use a strategy that estimates image intensity values in arbitrary spatial coordinates.

Woods *et al.* [1] reported that the SSD similarity measure contains local optima with respect to image translation. Ashburner *et al.* [2] reported that these local optima oscillations can be removed by low-pass filtering the data prior to computation of the similarity measure. In parallel, several researchers reported that local optima oscillations with respect to translation can occur in MI similarity measures [3, 4, 5]. Strategies for diminishing the effects of local optima in MI curves, such as partial volume interpolation [3], image resampling [4], and others [5] have been empirically proposed and tested experimentally. Until recently [6, 7] the mechanisms causing the oscillations in SSD and MI similarity measures, were explained only through empirical observations. No relationship between the oscillations in different similarity measures was known. Moreover, misconceptions such as the intuition that higher order sinc interpolation methods had no effects on the severity of artifacts, argued with empirical observations obtained through limited experimentation [5], existed.

In [6, 7] we showed that SSD, CC, and MI similarity measures contain local optima with respect to the spatial transformation being optimized. We showed that the severity of local optima can be decreased by using higher order interpolation methods such as sinc and truncated sinc interpolation. Alternatively, the images can also be low-pass filtered prior to computations so as to lessen the variance of intensity values due to noise, also decreasing the severity of local optima artifacts. Our previous work [6] explained the oscillations in the similarity measures using concepts from uncorrelated, constant variance, random processes. In addition, in order to explain the oscillations in MI curves we also assumed that the intensity values in the images were normally distributed.

Here we relax such assumptions and show that similar analysis can be used to explain such artifacts under more general circumnstances. We show that oscillations in SSD and CC similarity measures can be explained using concepts from piece-wise stationary processes. Furthermore, the random process needs not to be white. In MI similarity measures, oscillatory behavior with respect to the spatial transformation can be explained using Gaussian-type mixture models for the probability density functions (*pdf*) of the images. Results using MRI data are shown.

2. THEORY

When registering two digital images $S(\mathbf{q})$ and $T(\mathbf{q})$, with $\mathbf{q} \in \mathbb{Z}^d$ organized in an M^d grid, we seek a spatial transformation $f : \mathbb{R}^d \to \mathbb{R}^d$ such that a given image similarity measure (cost function) $\Psi(S(f(\mathbf{q})), T(\mathbf{q}))$ is maximized (minimized). In order to compute a spatially transformed im-

*Currently a National Research Council Post Doctoral Associate at the Naval Research Laboratory, Washington, DC, USA. 20375.

age $S(f(\mathbf{q}))$ a continuous model for image $S(\mathbf{q})$ is needed. Because of computational considerations, images are usually made continuous via a linear combination of compactly supported basis functions $h(\mathbf{x})$:

$$\hat{S}(\mathbf{x}) = \sum_{\mathbf{i}\in\mathbb{Z}^d} S(\mathbf{i})h(\mathbf{x}-\mathbf{i}). \quad (1)$$

Naturally, if we wish that $\hat{S}(\mathbf{q}) = S(\mathbf{q})$ the basis function $h(\mathbf{x})$ in (1) needs to satisfy interpolation conditions.

In order to understand the oscillatory behavior of similarity measures with respect to spatial transformations it is useful to consider the stochastic properties of the interpolated image $\hat{S}(f(\mathbf{q}))$ in contrast to the stochastic properties of the original image $S(\mathbf{q})$. Consider a simple linear stochastic image model

$$S(\mathbf{q}) = \int W(\mathbf{x})\Upsilon(\mathbf{q}-\mathbf{x})d\mathbf{x} + e(\mathbf{q}) \quad (2)$$

where $S(\mathbf{q})$ is the measured image, $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{q} \in \mathbb{Z}^d$ organized in an M^d grid, $W(\mathbf{x})$ the object being imaged, $\Upsilon(\mathbf{x})$ the transfer function of the imaging system, and d the dimension of the image. $e(\mathbf{q})$ is a zero-mean random variable representing the influence of noise sources (thermal and others), and it is statistically independent from the signal part of the image.

It is easy to show that the covariance function of the interpolated image $\hat{S}(\mathbf{x})$ in (1) is given by:

$$R_{\hat{S}}(\mathbf{x}_1,\mathbf{x}_2) = \sum_{\mathbf{q}_1,\mathbf{q}_2\in\mathbb{Z}^d} h(\mathbf{x}_1-\mathbf{q}_1)R_S(\mathbf{q}_1,\mathbf{q}_2)h(\mathbf{x}_2-\mathbf{q}_2), \quad (3)$$

where $R_S(\mathbf{q}_1,\mathbf{q}_2)$ is the covariance matrix of the original data. If we are dealing with band-limited white noise, for example, the variance of an interpolated value $\hat{S}(\mathbf{x})$ is given by:

$$R_{\hat{S}}(\mathbf{x},\mathbf{x}) = \sigma^2 \sum_{\mathbf{q}\in\mathbb{Z}^d} \left[h(\mathbf{x}-\mathbf{q})\right]^2, \quad (4)$$

where $\sigma^2 = R_e(\mathbf{x},\mathbf{x})$ is the noise variance. By using simple examples, consider $d=1$ and h to be the linear 'hat' interpolating function, it is easy to see that equation (4) is not a constant function, but a quadratic one in the interval $x \in [0,1]$. If however, in the white noise example above $h(x) = \text{sinc}(x)$, $R_{\hat{S}}(x,x) = \sigma^2\ \forall x$. The random process needs not to be white in order for oscillations to occur. Consider, for example, the 2d covariance matrix given by:

$$\mathbf{R}_S = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \quad (5)$$

In Figure 1 we plot $R_{\hat{S}}(x,x)$ in equation (3) for many commonly used interpolating basis functions such as linear, cubic spline, truncated sinc, and sinc.

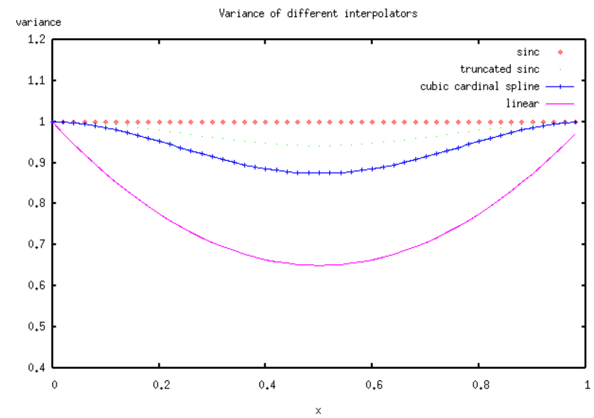

Fig. 1. Plot of equation (3), using the covariance matrix (5).

2.1. SSD-based registrationn

To demonstrate the effects of interpolation on image registration methods, we look at the following example. A two-dimensional source image $S(\mathbf{q})$, is translated with respect to a target image $T(\mathbf{q})$ by

$$\hat{S}(\mathbf{q}-\mathbf{t}) = \sum_{\mathbf{i}\in\mathbb{Z}^d} S(\mathbf{i})h(\mathbf{q}-\mathbf{t}-\mathbf{i}), \quad (6)$$

where $\mathbf{t} = \{t, 0\}$ is a simple translation vector in one dimension of the image. Note that this is a simple linear operation on the intensity values of the images. Thus, if we write $\mathbf{S}$ to represent the digital image $S(\mathbf{q})$ in vector notation, a translated version of the same image can be computed with a linear opration $\mathbf{F}_t\mathbf{S}$, with the values of $\mathbf{F}_t$ determined by the equation above. At each translation value, a similarity measure (in this case the SSD) between the target and translated source images is computed. The value t for which the discrete 2-norm is minimized is the optimal translation value. In the context of the optimization framework described earlier, this translates to:

$$\Psi(S(f(\mathbf{q})), T(\mathbf{q}), f) = \|\mathbf{F}_t\mathbf{S} - \mathbf{T}\|^2 \quad (7)$$

where $\mathbf{T}$ is the target image vector, and $\|\mathbf{S}\|^2 = \langle\mathbf{S},\mathbf{S}\rangle = \frac{1}{M^2}\sum_{i=1}^{M^2}\left(S(\mathbf{q}_i)\right)^2$, and M^2 is the size of the image.

Let $\mathbf{S} = \tilde{\mathbf{W}}_\mathbf{S} + \mathbf{e}_\mathbf{S}$ represent the vector of sampled image values (the vector notation for equation (2)). Similarly, let the target image in vector notation be $\mathbf{T} = \tilde{\mathbf{W}}_\mathbf{T} + \mathbf{e}_\mathbf{T}$. Inserting these into (7), the objective function being minimized becomes:

$$\|\mathbf{F}_t\mathbf{S}-\mathbf{T}\|^2 \sim \|\mathbf{F}_t\tilde{\mathbf{W}}_\mathbf{S} - \tilde{\mathbf{W}}_\mathbf{T}\|^2 + \|\mathbf{e}_\mathbf{T}\|^2 + \|\mathbf{F}_t\mathbf{e}_\mathbf{S}\|^2. \quad (8)$$

If the noise process is weakly stationary $\|\mathbf{F}_t\mathbf{e}_\mathbf{S}\|^2$ is equivalent to a sample variance estimate, which, as shown in Figure 1, oscillates according to equation (3) (equation (4) in the case of white noise):

$$\|\mathbf{F}_t\mathbf{e}_\mathbf{S}\|^2 \approx R_{\hat{S}}(t,t) = \sigma^2(t). \quad (9)$$

Thus computational algorithms that seek to optimize (7) are likely to converge to one of the many artificial local optima, instead of a 'true' solution.

The random process describing the stochastic properties of the measured image $S(\mathbf{q})$ is not required to be stationary but rather piece-wise stationary. Let the source image $S(\mathbf{q})$ be described by a random process which can be divided into N simply connected regions, $\Omega_1, \cdots, \Omega_i, \cdots, \Omega_N$, within which the covariance R_S^i of image samples is approximately stationary though distinct from other image regions. Now, the sum above can be rewritten as a sum of sums so that samples belonging to each region Ω_i are summed together:

$$\|\mathbf{F_t e}_S\|^2 = \frac{1}{M^2}\sum_{i=1}^{M}(\mathbf{F_t e}_S)_i^2 \tag{10}$$

$$= \frac{1}{M^2}\sum_{j=\Omega_1}^{\Omega_N}\sum_{i\in\Omega_j}(\mathbf{F_t e}_S)_i^2. \tag{11}$$

It is easy to see that the sum over each region is equivalent to a variance estimate computation, as in equation (9). Each term j in the sum above will oscillate with respect to translation parameter t, as explained by equation (3). Note that the oscillations contributions of each term j above will be in phase with each other, as long as translations are uniform.

The artifactual oscillations in the objective function can be reduced using zero degree B-spline interpolation (nearest neighbor). However, this is not advisable since sub-pixel measurements would not be possible. Oscillations can also be avoided by using a continuous 2-norm instead of the discrete one in (7). This, however, could be computationally impractical for all but the most trivial applications. Another alternative is to 'smooth' the image prior to computation of the cost function so as to minimize the oscillations in $\|\mathbf{F}_t\mathbf{e_S}\|^2$. This, however, is also not optimal since blurring can reduce edge information that can be critical for obtaining a good match. Finally, the magnitude of the oscillations can be decreased by increasing the degree of the interpolant.

2.2. Correlation based-registration

When the intensity values of the images being registered is not expected to be identical but a linear relationship between them can be assumed, it is possible to register the images using the CC similarity measure:

$$\Psi(S(f(\mathbf{q})), T(\mathbf{q}), f) = \frac{\langle \mathbf{F_t S}, \mathbf{T}\rangle}{\|\mathbf{F_t S}\|\|\mathbf{T}\|}. \tag{12}$$

Expanding the term $\|\mathbf{F_t S}\|$ one can quickly find the term $\|\mathbf{F}_t\mathbf{e_S}\|^2$ once again. Thus cost functions based on linear correlation are expected to oscillate with respect to the translation parameter causing local optima, as in the case of the SSD cost function.

2.3. MI-based registration

In the case when the distribution of intensity values of both $T(\mathbf{x})$ and $S(\mathbf{x})$ is normal, the MI between them is given by $\frac{1}{2}\log(1-\rho^2)$, where ρ is the linear correlation coefficient between them. Thus, under the normal assumption, it can be expected that the MI similarity measure will also contain local optima. In real world images, however, the intensity values of the images are hardly ever normally distributed. Under general circumstances the MI between two random variables S and T is given by: $\mathrm{MI}(S,T) = H(T) + H(S) - H(S,T)$, where $H(\cdot)$ is the Shannon entropy function.

In this section we show that the entropy of a probability density function (*pdf*) of an image, modeled here as a Gaussian mixture, that undergoes translation according to (6) is monotonic with respect to the variance of each component of the mixture.

Let the *pdf* for the intensity values in an image be defined as a mixture of Gaussians:

$$\mathrm{pr}_S(s) = \sum_{i=1}^{Q}\alpha_i \mathrm{pr}_{S_i}(s),\ \mathrm{pr}_{S_i}(s) \sim N(\mu_i, \sigma^2), \tag{13}$$

with $\sum_i^Q \alpha_i = 1$. If the basis function being used to model the image satisfies partition of unity, the mixture model of an image that undergoes such translation is modified as $\mathrm{pr}_{\hat{S}_i}(s) \sim N(\mu_i, \sigma^2(t))$, where $\sigma^2(t)$ can be computed using (3). Note that $\sigma^2(t) \leqslant \sigma^2$, with equality when h =sinc (other basis functions, such as the Haar basis function, may also apply).

Again we wish to compare the entropy of two mixture modes: $\mathrm{pr}_S(s)$ refers to the *pdf* of the original data, and $\mathrm{pr}_{\hat{S}}(s)$ refers to the *pdf* of the interpolated (translated) continuous signal. Then, for a fixed translation t we can find a^2 so that:

$$\sigma^2(t) + a^2 = \sigma^2. \tag{14}$$

The *pdf* of the original data can be written as a function of the *pdf* of the translated signal:

$$\mathrm{pr}_S(s) = \mathrm{pr}_{\hat{S}}(s) * N(0, a^2), \tag{15}$$

where $*$ refers to the one dimensional convolution operation and $N(0,a^2)$ refers to a Normal distribution with zero mean and variance a^2. Let Y_S, $Y_{\hat{S}}$, and Y_a be random variables associated with *pdf*s $\mathrm{pr}_S(s)$, $\mathrm{pr}_{\hat{S}}(s)$, and $N(0,a^2)$, respectively. By (15) $Y_{\hat{S}}$, and Y_a are independent. Then:

$$Y_S = Y_{\hat{S}} + Y_a, \tag{16}$$

and

$$H(Y_S) = H(Y_{\hat{S}} + Y_a) \geqslant H(Y_{\hat{S}}). \tag{17}$$

In words, the entropy of the *pdf* associated with the experiment of choosing intensity values from an image translated using a continuous image model is monotonic with respect to the variance of the signal: the more interpolation on image values, the lower the variance of the signal, the lower the

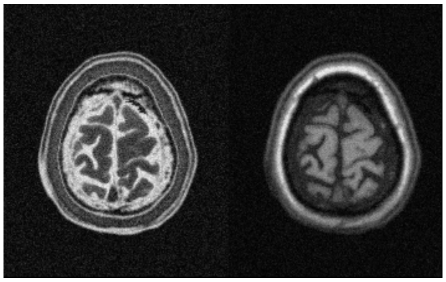

Fig. 2. Magnetic resonance images (T_1-weighted left, T_2-weighted right) used for experimental results.

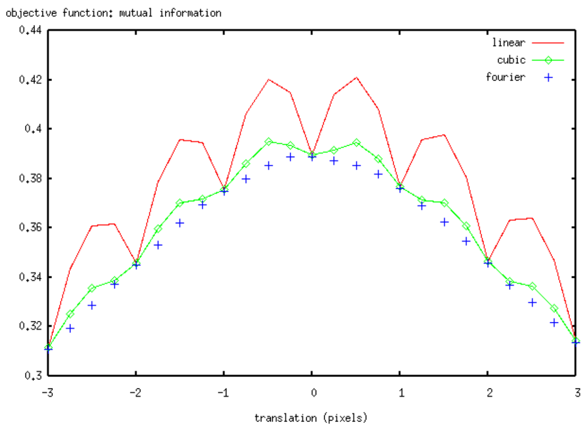

Fig. 3. MI with respect to translation.

entropy. Since the variance of any component in the mixture model defined in (13) oscillates with respect to translation parameter t according to equation (3) the entropy of a translated signal will also oscillate accordingly. The same argument can be extended to two dimensional *pdf*'s to show that the joint entropy component of the MI $H(T, \hat{S})$ also oscillates in an analogous manner [7].

3. RESULTS

The results in this section were computed based on the MRI images of the human brain shown in Figure 2. These images were taken from a realistic MRI simulator [8]. In this experiment one image was translated using linear, cubic, and the FFT algorithm. The MI between the translated image and the static one was computed, and it's value plotted in Figure 3. The MI similarity measure was computed using a Parzen window-type estimate of the joint *pdf* of the intensity values of the images. As seen from this figure, low order interpolators such as linear and cubic cause spurious local optima at (approximately) half-pixel translations, while the FFT-based algorithm produces a cost function with only one clear optima. Similar results can be obtained with the SSD and CC similarity measures [7]. For brevity, these are omitted here.

4. DISCUSSION AND CONCLUSION

We've explained the source of interpolation artifacts in SSD, CC, and MI similarity measures by tracking the variance of image intensity values using concepts from the theory of random processes. We've shown that intensity based image registration using translations is difficult when using low order basis functions because many of the commonly used similarity measures contain local optima at half-pixel translations. Thus, the solution to the registration problem will be biased towards the most 'blurred' image. The situation can be improved by performing some pre-processing of the data such as low-pass filtering, or 'regriding' as in [4]. Such data preprocessing, however, can obscure image features that can be important for obtaining accurate matches. Alternatively, continuous integration can be used to compute the SSD and CC similarity measures. This, however, may be prohibitively computationally expensive in many situations. Finally, an improvement can be obtained by using higher order interpolants (ideally sinc) for modeling the images.

5. REFERENCES

[1] R.P. Woods *et al.*, "Rapid automated algorithm for aligning and reslicing PET images", *JCAT*, vol. 16, pp. 620-633, 1992.

[2] J. Ashburner *et al.*, "Image Registration," in **Functional MRI**. Springer-Verlag, 1999.

[3] F. Maes *et al.*, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187-198, 1997.

[4] J.P.W. Pluim *et al.*, "Interpolation Artefacts in Mutual Information-Based Image Registration," *CVIU*, vol. 77, pp. 211-232, 2000.

[5] J. Tsao, "Interpolation Artifacts in Multimodality Image Registration Based on Maximization of Mutual Information," *IEEE Trans. Med. Imag.*, vol. 22, pp. 854-968, 2003.

[6] G.K. Rohde *et al.*, "Measuring image similarity in the presence of noise," . In *Proceedings of the SPIE Medical Imaging: Image Processing*, San Diego, February 2005. pp 132-143.

[7] G.K. Rohde, "Registration Methods for Quantitative Imaging," University of Maryland, College Park, Maryland, USA, 2005.

[8] *http://www.bic.mni.mcgill.ca/brainweb/*