Identification of translational displacements between N-dimensional data sets using the high order SVD and phase correlation

W. Scott Hoge,* Member, IEEE and Carl-Fredrik Westin, Member, IEEE

Abstract—This manuscript presents an extension of the phase correlation image alignment method to N-dimensional data sets. By the Fourier shift theorem, the motion model for translational shifts between N-dimensional images can be represented as a rank-one tensor. Through use of a high-order singular value decomposition, the phase correlation between two N-dimensional data sets can be decomposed to independently identify translational displacements along each dimension with subpixel resolution. Using 3D MRI data sets, we demonstrate the effectiveness of this approach relative to other N-dimensional image registration methods.

I. INTRODUCTION

The acquisition of multi-dimensional image data is fast becoming commonplace due to rapid improvement in imaging technology. This is particularly true in the author’s field of expertise of medical imaging. This includes surgical treatment planning where 3D visualization is essential, diffusion tensor imaging which is inherently multi-dimensional, and functional-MRI where analysis is improved with high resolution 3D coverage. In each of these cases, registration of multi-dimensional image sets is vital for both the comparison of images between patients and between imaging sessions of the same patient.

In response to this need, many methods have been developed to estimate the displacement and deformation between similar images [1], [2]. The most popular method in medical imaging applications today is based on the maximization of mutual information (MI), particularly for registration of images acquired from multiple modalities. A good review of MI can be found in [3]. This approach has shown to be very robust in estimating translations, rotations, and affine deformations between multiple data sets. These methods are currently based on maximization of an objective function, and often utilize gradient search approaches to find the maximum. However, in certain image registration problems, this approach can be computationally excessive.

In particular, there exist certain cases where rigid registration is sufficient. Rigid registration considers only translations and rotations between images. It is well known that these two motion models can be considered separately in the Fourier domain [4]. By estimating the rigid registration parameters in the Fourier domain, translations can be considered independently of rotations.

This independence between the two motion models allows us to consider here only translational motion between N dimensional data sets, with the expectation that this method is complimentary to other methods focused on rotation estimation, e.g. [5]. Our approach below addresses the limited case where the motion model can be completely described by translational shifts, and we demonstrate that direct global estimation methods such as ours can find a better translation estimate with greater computational efficiency than gradient search methods.

II. THEORY

For the registration of two-dimensional images, the phase correlation method (PCM) [6] is popular due to robust performance and computational simplicity. In this manuscript, we demonstrate a technique that extends the phase correlation method to identify translational displacements between multi-dimensional data sets. [1]

A. PCM registration with sub-pixel resolution

The phase correlation method [6] capitalizes on the well-known Fourier shift theorem which states that shifts in the spatial domain correspond to linear phase changes in the Fourier domain. Specifically, if a pair of two-dimensional functions, \( A \) and \( B \), are spatially related by a simple translational shift in coordinate systems, the
elements of the Fourier representation of \( \mathcal{B} \), denoted \( \mathcal{B} \), are related to \( \mathcal{A} \) by

\[
\mathcal{B}(l_1, l_2) = \mathcal{A}(l_1, l_2) \exp(-j(l_1 t_1 + l_2 t_2))
\]

(1)

where \((l_1, l_2)\) are the Fourier domain coordinates and \(t_1\) and \(t_2\) are the amount of horizontal and vertical displacement between \( \mathcal{A} \) and \( \mathcal{B} \).

In the original PCM approach, one first computes the phase correlation matrix

\[
Q(l_1, l_2) = \frac{\mathcal{B}(l_1, l_2) A^*(l_1, l_2)}{|\mathcal{B}(l_1, l_2)| A^*(l_1, l_2)} \exp(-j(l_1 t_1 + l_2 t_2)).
\]

(2)

Computing the inverse Fourier transform of \( Q \) results in a matrix where the location of the peak value identifies the translation. In ideal cases, this function could be represented as

\[
Q(x, y) = \delta(x - t_1, y - t_2).
\]

(3)

Phase correlation is widely used because estimation of motion in the phase domain is exceptionally robust in the presence of noise [3]. However, sub-pixel translations will cause the peak in \( Q(x, y) \) to spread to neighboring pixels. In addition, the resolution of the peak can degrade due to the presence of aliasing and edge-effects [9]. To address these issues, a number of extensions to PCM have been proposed, two of which are briefly reviewed here.

In [9], Stone, et. al., performed sub-pixel translation estimation via a 2D least-squares fit to the phase plane of \( Q \). They showed that two sources of spectrum distortion can degrade the estimation quality: aliasing during image acquisition, and edge effects produced by the discrete Fourier transform (DFT). To overcome these effects, they apply a spatial domain envelope to the images before comparison to ensure smooth transitions at the image edges, and use an appropriate mask during phase-plane parameter estimation. This mask allows only those phase correlation matrix elements with significant signal information into the estimation algorithm.

It was recently shown [10] that robust translation parameter estimation can also be performed in the presence of these spectrum distortions by first finding the dominant singular vectors of \( Q \). The Eigen-filtering nature of the SVD effectively suppresses noise in the phase correlation matrix and isolates the phase contribution from vertical and horizontal translation. This allows phase unwrapping along each dimension independently, rather than 2D phase unwrapping as in [9]. The SVD approach greatly simplifies the translation estimation, and can be extended to higher dimensional data sets by utilizing the high-order SVD described in the next section.

B. The high order SVD

The standard 2D matrix SVD is defined as \( A = U \Sigma V^H \). Here \( U, V \) are unitary matrices containing singular vectors, and \( \Sigma \) is a diagonal matrix of non-negative singular values, \( \sigma_i \). The SVD is typically used to identify and order the dominant column and row subspaces for two-dimensional matrices. The dominant left singular vector, \( u_1 \), is the solution to \( \max_u \| A^H u_1 \|_2 \) with \( u_1^H u_1 = 1 \), and corresponds to the largest singular value, \( \sigma_1 = \| A^H u_1 \|_2 \).

To extend this matrix SVD to data organized in a multi-dimensional coordinate system, DeLathauwer, et. al. [11], define the high-order SVD (hoSVD) using tensor operators as

\[
A = S \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)}.
\]

(4)

Here, the \( U^{(i)} \) are unitary matrices and \( S \) is a tensor of the same size as \( A \) which contains the singular values. Note that \( S \) is typically not diagonal. Instead, it is all-orthogonal, meaning that the inner-product of any two lines along a single mode is identically zero. For a 3-D example in MATLAB-style matrix notation, this would be written as \( \langle S(:, i, :), S(:, j, :) \rangle = 0 \), \( \forall \ i \neq j \) for lines along the second dimension. The \( n \)-mode product, \( \times_n \), of an \( N \)-dimensional tensor with a matrix along the \( n \)th dimension is defined as

\[
(A \times_n U)_{i_1, i_2, \ldots, i_{n-1}, j, i_{n+1}, \ldots, i_N} = \sum_{i_n} u_{i_1, i_2, \ldots, i_{n-1}, j, i_{n+1}, \ldots, i_N} u_{j, i_n}.
\]

In practice, \( n \)-mode multiplication is performed by first unfolding the tensor along a given mode to create a 2D matrix, performing the matrix multiplication along that mode, and then folding the data back to the tensor form. For a three dimensional tensor of size \( L_1 \times L_2 \times L_3 \), the unfolding operation can be written in MATLAB-style notation as

\[
\text{unfold}_1(A) = [A(:, 1, :) A(:, 2, :) \ldots A(:, L_2, :) ]
\]

\[
\text{unfold}_2(A) = [A(:, :, 1)^T A(:, :, 2)^T \ldots A(:, :, L_3)^T ]
\]

\[
\text{unfold}_3(A) = [A(1, :, :)^T A(2, :, :)^T \ldots A(L_1, :, :)^T ]
\]

The multiplication along mode \( n \) can then be implemented as

\[
A \times_n U = \text{fold}_n(U \ \text{unfold}_n(A))
\]

(5)

For reference, the familiar 2D SVD can be written in tensor notation as \( A = U \Sigma V^H = \Sigma \times_1 U \times_2 V^* = \Sigma \times_1 U^{(1)} \times_2 U^{(2)} \).
C. Extending PCM with the hoSVD

For multi-dimensional images that differ only by translational shifts, all of the relevant motion information is contained in the phase of the dominant singular vectors of the phase correlation matrix, \( Q \). To estimate the registration parameters between these two images, one forms the phase correlation matrix in the same manner as the 2D case, \( \Phi \). Masking \( Q \) to concentrate the algorithm on data closest to the DC coordinate in the Fourier domain is often practical. This prevents spectral distortion due to aliasing from impacting the translational estimates, and reduces the total number of data points to be analyzed.

The dominant singular vector along each dimension of \( Q \) is calculated by first unfolding \( Q \) along a given dimension \( n \), and solving

\[
\max_{u_1(n)} \left\| (\text{unfold}_n(Q))^{H} u_1(n) \right\|_2
\]

(6)

to identify the dominant singular vector along mode \( n \). The linear phase change along \( u_1(n) \) corresponds to the spatial translation, \( t_n \), along mode \( n \) between the two data sets. To estimate \( t_n \), the phase of \( u_1(n) \) is unwrapped, and a line is fit to this data which minimizes the least mean square error. The slope of this line, \( c \), corresponds to the shift as \( t_n = cl_n/(2\pi) \), where \( l_n \) is the length of the vector \( u_1(n) \).

To spatially align the two data sets, one can introduce a phase shift to the Fourier domain image data. Specifically, to correct for a shift \( t_n \) between images \( A \) and \( B \), one can construct a linear phase change vector

\[ \varphi_n(x) = \exp(j2\pi t_n x/l_n) \]  

(7)

and form a diagonal matrix \( \Phi_n \) with \( \varphi_n \) along the main diagonal. Once the translational shifts \( t_n \) have been calculated, they may be applied to the Fourier representation of \( B \) via

\[ B_{\text{corrected}} = B \times_1 \Phi_1 \times_2 \Phi_2 \cdots \times_n \Phi_n. \]  

(8)

Finally, the registered image can be reconstructed by performing an inverse Fourier transform on \( B_{\text{corrected}} \).

III. RESULTS

We demonstrate here the validity of this approach using three-dimensional magnetic resonance images. 3D MR images are typically acquired as a stack of multiple 2D slices. For the data presented here, the orientation of the 2D plane differs between each image set, shown in Figure 1. This produces data with voxel sizes that differ between the image sets, and introduces translational shifts between the images. The data was acquired using a production quality 3D fast spin-echo sequence on a GE Medical Systems 1.5T MR scanner. To co-register these three data sets, Set 2 and Set 3 are registered to Set 1.

Raw MR image data is typically acquired through a Fourier encoding process. Thus, the spatial domain reconstructions contain both magnitude and phase information. With the 3D-FSE protocol used to acquire the data shown, a spatial-domain phase variation was present across the reconstructed data, correlated to the low-resolution axis. Thus, to determine the registration parameters, the data was first transformed to the spatial domain, the phase information eliminated, and then the image data was transformed back to the Fourier domain.

Between the image data sets, there was a common area of overlap of consisting of a cube in the Fourier domain with 64 samples to a side. While the registration algorithm can be applied to data within this full cube, it is computationally advantageous to process a smaller cube of 32-by-32-by-32 data samples surrounding DC. This selection process is equivalent to the masking operation suggested for the 2D PCM method in \( \Phi \).

For each data set comparison, the phase correlation matrix is calculated as \( Q = BA^*/|BA^*| \). The dominant singular vector of \( Q \) is calculated for each dimension, \( n \), via solution of the maximization of

\[ \max_{u_1(n)} \left\| (\text{unfold}_n(Q))^{H} u_1(n) \right\|_2 \]

with respect to \( u_1(n) \). Figure 2 illustrates the linear character in the phase of these dominant singular vectors between Set 1-Set 2 and Set 1-Set 3. The strong linearity of this data indicates that the deviation between the data sets is dominated by linear translations. The translation along each dimension of the data is found by estimating the slope of the phase, as described in Sec. II-C.

Tables I and II compare the estimation results obtained using our hoSVD-PCM approach with results obtained using constrained, Powell-method, gradient search methods. Our hoSVD-PCM approach was compared to two alternative optimization functions: maximization of mutual information (MI) and normalized cross correlation (NCC). The implementation of these algorithms was obtained from the Statistical Parametric Mapping (SPM) toolbox \( \Phi \), which is used to analyze functional-MRI data. To present a fair comparison, results from the full SPM implementation of MI and NCC are shown, as well as a set from a modified implementation which restricted the parameter estimation to translational motion only.

Tables I and II illustrate the comparison numerically. As shown in the tables, the motion estimates are consistent across all methods. To measure the accuracy of the estimation, the registration estimates are applied to the second data set to correct the translation relative to Set 1. The mean squared error (MSE) between this corrected

accepted for publication in IEEE Transactions in Image Processing. (c) Copyright IEEE 2005.
Set 1 sized
128−by−128−by−64.
low-resolution along
Superior −Inferior axis

Set 2 sized
128−by−64−by−128.
low-resolution along
Left−Right axis

Set 3 sized
64−by−128−by−128.
low-resolution along
Anterior −Posterior axis

Fig. 1. Original data to be registered: three 3D fast-spin-echo MR acquisitions each with a unique high-resolution plane.

Fig. 2. Comparison of linear fit estimate to phase of the dominant singular vectors in the phase correlation matrix for (a) Set 1 and Set 2, and (b) Set 1 and Set 3
data set and Set 1 and is measured via
\[
E \equiv \| A - B_{corrected} \|_F^2 = \sqrt{\sum_{i,j,k} (A(i,j,k) - B_{corrected}(i,j,k))^2}
\]

For this set of data, restricting the gradient search methods to find only the translation parameters had a significant impact on the results. The rotation estimate in the non-constrained case is nearly negligible. Nonetheless, the measured MSE in the gradient search methods was reduced by restricting the parameter estimate to translational shifts, bolstering our claim that there is no rotation present between the three data sets. Restricting the gradient search methods also produced a significant decrease in computation time.

Even after restricting the MI and NCC algorithms to search for translations only, the hoSVD-PCM technique maintains the advantage by providing slightly lower error than either of the gradient search methods while being significantly faster to compute. The algorithm timing results are presented in the first row of Tables I and II and were measured using a MATLAB 13.1 implementation on a 1GHz Intel Pentium-4 based PC with 512MB of RAM.

This comparison demonstrates that the hoSVD-PCM method can be significantly faster to compute than the gradient search methods while providing more accurate estimates. The computational speed of this approach depends significantly on the size of the constructed phase correlation matrix. This matrix need be only large enough to illustrate the linear character of the phase information. Additionally, the reduced computational time needed for hoSVD-PCM is a direct result of the fact that it seeks a closed-form estimate of the motion model parameters and does not require iterative searching between the estimated parameters.

IV. DISCUSSION

A. Why this particular decomposition?

Among the choice of high-order tensor decomposition techniques [13], [14], [15], DeLathauwer’s was chosen because the method to find the dominant singular vector of a tensor along a given dimension is consistent with the 2D matrix case. In our 2D SVD-PCM technique [10], the separation of the phase correlation matrix into its dominant components depends on the maximization perspective of the SVD. DeLathauwer’s hoSVD extends this idea to multi-dimensional data sets in a natural way.

B. Applicability to general rigid registration problems

For general rigid registration problems, this approach will need to be integrated with an approach that estimates rigid rotations. As stated earlier, estimation of translation in Fourier domain is completely complementary to approaches focused solely on rotation estimation in the Fourier domain. For example, we anticipate that the combination of the proposed method with Keller’s, et. al., rotation estimation algorithm [5] will provide a computationally efficient way to perform 3-D rigid registration. Keller’s rotation estimate method is also based on phase correlation, however, it is based on the pseudo-polar Fourier transform [16] which has an numerical complexity on the order of the FFT. In contrast to the method by Reddy and Chatterji. Keller’s approach can be naturally extended to 3D image data.

C. Performance in presence of additive noise

To measure the relative performance of the technique in the presence of additive noise, we repeated the registration estimation above but added varying levels of spatially white Gaussian noise to each image before registration. Fig. 8 shows the root-mean-square-error for two such experiments relative to the level of noise added to each image. The data is shown using a relative signal to noise scale, with \( SNR = 10 \log_{10}(\| A \|_F^2 / \| n \|_F^2) \). The plots illustrate that our hoSVD-PCM technique degrades in the presence of noise at roughly the same point and at the same rate as normalized mutual information. The reduction in RMSE provided by our approach is evident at each noise level, consistent with the previous results in Tables I and II. We note that above the noise levels shown, the translation estimates of both methods rapidly deteriorate to the point where they become meaningless.

V. SUMMARY

We have presented an extension of the phase correlation image alignment method to multi-dimensional data sets. Through use of a high-order singular value decomposition, the phase correlation between two N-dimensional images can decomposed to independently identify translational displacements along each dimension with subpixel resolution. Our results show that in cases where the motion model is dominated by translational shifts, the hoSVD-PCM approach produces registration estimates with lower mean squared error and is faster to compute than gradient search techniques.

REFERENCES

<table>
<thead>
<tr>
<th>elapsed time</th>
<th>unregistered</th>
<th>MI with rot.</th>
<th>NCC with rot.</th>
<th>MI w/o rot.</th>
<th>NCC w/o rot.</th>
<th>hoSVD-PCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 est.</td>
<td>-5.985</td>
<td>-6.010</td>
<td>-6.006</td>
<td>-6.007</td>
<td>-6.016</td>
<td></td>
</tr>
<tr>
<td>Y0 est.</td>
<td>+1.450</td>
<td>+1.446</td>
<td>+1.461</td>
<td>+1.474</td>
<td>+1.542</td>
<td></td>
</tr>
<tr>
<td>Z0 est.</td>
<td>-4.184</td>
<td>-4.180</td>
<td>-4.114</td>
<td>-4.112</td>
<td>-4.115</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>-6.437 \times 10^{-3}</td>
<td>-6.765 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>-7.645 \times 10^{-3}</td>
<td>-7.133 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>-3.447 \times 10^{-3}</td>
<td>-4.057 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mean Square Error</td>
<td>1.895 \times 10^5</td>
<td>5.832 \times 10^5</td>
<td>5.880 \times 10^5</td>
<td>4.311 \times 10^5</td>
<td>4.296 \times 10^5</td>
<td>4.291 \times 10^5</td>
</tr>
</tbody>
</table>

**TABLE I**

**Numerical comparison of registration results for Set 1 vs. Set 2**

<table>
<thead>
<tr>
<th>elapsed time</th>
<th>unregistered</th>
<th>MI with rot.</th>
<th>NCC with rot.</th>
<th>MI w/o rot.</th>
<th>NCC w/o rot.</th>
<th>hoSVD-PCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 est.</td>
<td>-0.095</td>
<td>-0.091</td>
<td>-0.092</td>
<td>-0.095</td>
<td>-0.111</td>
<td></td>
</tr>
<tr>
<td>Y0 est.</td>
<td>-0.134</td>
<td>-0.144</td>
<td>-0.124</td>
<td>-0.133</td>
<td>-0.233</td>
<td></td>
</tr>
<tr>
<td>Z0 est.</td>
<td>+2.728</td>
<td>+2.723</td>
<td>+2.734</td>
<td>+2.717</td>
<td>+2.723</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>+1.023 \times 10^{-3}</td>
<td>+9.721 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>+1.334 \times 10^{-3}</td>
<td>+1.376 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>affine est.</td>
<td>-7.799 \times 10^{-4}</td>
<td>-1.757 \times 10^{-3}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mean Square Error</td>
<td>1.511 \times 10^5</td>
<td>3.576 \times 10^5</td>
<td>3.585 \times 10^5</td>
<td>3.565 \times 10^5</td>
<td>3.538 \times 10^5</td>
<td>3.457 \times 10^5</td>
</tr>
</tbody>
</table>

**TABLE II**

**Numerical comparison of registration results for Set 1 vs. Set 3**

Fig. 3. The root-mean-square-error performance of the hoSVD-PCM and normalized MI algorithms in the presence of additive noise for (a) Set 1-Set 2 and (b) Set 1-Set 3.


W. Scott Hoge (S’98-M’02) earned a B.A. degree in Physics from Colorado College, Colorado Springs, CO, in 1991, and M.S. and Ph.D. degrees in Electrical Engineering from Northeastern University, Boston, MA in 1994 and 2001, respectively. From 1994 through 1997, he was a member of the NMS Communications digital signal processing team, focused on the implementation of echo cancellation and fax modem algorithms. His Ph.D. thesis presents the application of digital communication theory to non-Fourier MR image acquisition methods. He is currently an Instructor at Brigham and Women’s Hospital and Harvard Medical School, Boston, MA. His research interest is in applying adaptive filtering, system identification, and subspace optimization concepts to magnetic resonance imaging and medical image analysis.

Carl-Fredrik Westin (M’99) received the MSc degree in Applied Physics and Electrical Engineering in 1988 from Linköping University. He joined the Computer Vision Laboratory in the Department of Electrical Engineering the same year where he did research on colour, information representation, image flow, frequency estimation, filtering of uncertain and irregularly sampled data and tensor operators in image analysis. In 1991, Dr. Westin was awarded the SAAB-SCANIA prize for his work in the field of Computer Vision. He received the Lic.Tech. degree on the topic feature extraction from a tensor image descriptions, in 1991.

In 1994 Westin graduated as Ph.D in computer vision from Linköping University. His thesis “A tensor framework for multidimensional signal processing” presents a novel method for filtering uncertain and irregularly sampled data termed normalized convolution.

In 1996 Dr. Westin joined Brigham and Women’s Hospital and Harvard Medical School. In 2001 he became Director of the Laboratory of Mathematics in Imaging (LMI) in the Department of Radiology, and a Research Affiliate of the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. Dr. Westin founded the Laboratory of Mathematics in Imaging in 2001 with the mission to improve medical image data formation and analysis with the aim of improving image quality, speed of procedures, and providing better methods for diagnosis and patient care.

His research interests are focused on medical applications of computer vision. He is currently working on analysis of Diffusion Tensor MRI data, and automated segmentation of MRI and CT data, using multidimensional signal processing techniques.