Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations

Charles R. Meyer¹*, Jennifer L. Boes¹, Boklye Kim¹, Peyton H. Bland¹, Kenneth R. Zasadny², Paul V. Kison², Kenneth Koral², Kirk A. Frey² and Richard L. Wahl¹,²

¹Department of Radiology, 3307 Kresge III Research Building, University of Michigan Medical School, Ann Arbor, MI 48109, USA
²Department of Internal Medicine, Division of Nuclear Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Abstract
This paper applies and evaluates an automatic mutual information-based registration algorithm across a broad spectrum of multimodal volume data sets. The algorithm requires little or no pre-processing, minimal user input and easily implements either affine, i.e. linear or thin-plate spline (TPS) warped registrations. We have evaluated the algorithm in phantom studies as well as in selected cases where few other algorithms could perform as well, if at all, to demonstrate the value of this new method. Pairs of multimodal gray-scale volume data sets were registered by iteratively changing registration parameters to maximize mutual information. Quantitative registration errors were assessed in registrations of a thorax phantom using PET/CT and in the National Library of Medicine’s Visible Male using MRI T2-/T1-weighted acquisitions. Registrations of diverse clinical data sets were demonstrated including rotate–translate mapping of PET/MRI brain scans with significant missing data, full affine mapping of thoracic PET/CT and rotate–translate mapping of abdominal SPECT/CT. A five-point thin-plate spline (TPS) warped registration of thoracic PET/CT is also demonstrated. The registration algorithm converged in times ranging between 3.5 and 31 min for affine clinical registrations and 57 min for TPS warping. Mean error vector lengths for rotate–translate registrations were measured to be subvoxel in phantoms. More importantly the rotate–translate algorithm performs well even with missing data. The demonstrated clinical fusions are qualitatively excellent at all levels. We conclude that such automatic, rapid, robust algorithms significantly increase the likelihood that multimodality registrations will be routinely used to aid clinical diagnoses and post-therapeutic assessment in the near future.

Keywords: automatic, fusion, joint entropy, maximization, minimization, multimodality, mutual information, optimization, registration

Received October 10, 1996; revised January 20, 1997; accepted January 23, 1997

1. INTRODUCTION
A variety of physiological imaging tests have been developed and implemented clinically, including PET and SPECT imaging. While these tests sometimes provide unique clinical information, their interpretation can be difficult without careful correlation and, ideally, registration with anatomical data sets. In addition, registration of test–retest physiological imaging studies is valuable for assessing regional physiological changes over time. We and others have used such fusion methods to assist in the diagnosis of hepatic hemangiomas.
(CT and Tc RBC SPECT), lung cancer (PET and CT), brain
tumor (fMRI and PET) and in the preoperative localization
of foci of abnormal monoclonal antibody uptake (SPECT and
CT), among others (Wahl et al., 1993, 1994). While most
would agree that image fusion is desirable and generally use-
fully clinically, the maximal clinical value of multimodality
imaging, e.g. the additional information obtained by com-
bining the sensitivity and specificity of functional imaging
with the structural resolution of anatomical imaging, can be
fully realized only when accurate, multimodality registration
becomes widely available at reasonable expense. Until then
visual comparison of unregistered images will provide only
marginal additional value. An even more valuable applica-
tion of automatic registration may be the presentation to the
diagnostician of previous comparison studies recomputed to
match the current study’s slice locations. In many cases the
previous study represents a slightly different set of acquisition
circumstances (contrast materials or choice of acquisition
parameters) which may significantly alter the appearance of
the two studies. An ideal registration algorithm should easily
handle iso- or multimodality data set pairings.

A recent review article described the state of medical im-
age and volume registration circa 1993 (van den Elsen et al.,
1993). The more useful and popular algorithms were based
on surface matching (Pelizzari et al., 1987, 1989) or mini-
mization of variance in one data set within corresponding seg-
mented regions defined in the registered data set (Woods et al.,
1993). Multimodality registration has evolved significantly
in the last few years beyond earlier stages which required
either user-biased homologous feature selection (e.g. the user-
selected perceived common points on the two different studies)
or tedious data preprocessing such as segmentation for
surface definition or tissue discrimination and stripping of
outer cranial tissue layers. Within the last few years papers
describing ‘similarity metrics’ have began to appear (Hill et al.,
1994). Such descriptors are statistical in nature, as is
mutual information. Collignon described the value of joint
entropy as a quantitative registration metric (Collignon et al.,
1995b) and perceived it to be a generalization of Woods’ ear-
lier work (Woods, 1993). Viola and Wells described the value
of the mutual information for multimodality registration only
months later (Viola and Wells, 1995) as applied to computer
vision topics. Within months other authors also described
mutual information as a criterion for medical data set registra-
tion (Collignon et al., 1995a; Studholme et al., 1995; Wells
et al., 1995). A more recent journal publication describes
multimodal volume registration and surgical applications
(Wells et al., 1996). Papers examining the resulting accuracy
and capture ranges obtained with mutual information-based
registration methods have just started to appear (Studholme et
al., 1996). Many more will undoubtedly appear soon.

Our purpose here is to briefly describe our implementation
of the algorithm and demonstrate the typical ease, accuracy
and computation times associated with the use of this algo-
rithm on a moderately high-performance workstation. We
have evaluated the algorithm in phantom studies as well as in
selected cases where few other algorithms could perform as
well. These registrations are meant to demonstrate the value
of this new method without comparison of its accuracy rela-
tive to other existing techniques. Accuracy and appropriate
case usage issues can best be evaluated in clinical trials that
compare algorithm results using diverse clinical and modality
conditions across multiple institutions.

2. METHODS

The entropy of a data set is defined as its average information
content, while joint entropy is the average information of two
data sets. The joint entropy, $H(a, b)$ of two data sets, $a$ and
$b$, is related to mutual information of $a$ and $b$, $I(a, b)$, by the
following classical relationship:

$$H(a, b) = H(a) + H(b) - I(a, b)$$

where $H(a)$ and $H(b)$ are the individual entropies of data
sets $a$ and $b$ respectively (Papoulis, 1984). As can be seen
in the classical relationship, mutual information, $I(a, b)$, is
the amount by which the sum of the individual data set en-
tropies must be reduced to account for correlations that exist
between the two individual data sets ($I$ is never negative). The
previous expression defining mutual information in terms of
entropies reduces to the following equation for mutual informa-
tion in terms of the more fundamental probability density functions for data sets $a$ and $b$,

$$I = \sum \sum p(a, b) \log_2 \left( \frac{p(a, b)/p(a)p(b)}{1} \right).$$

In implementation a geometrically transformed version of one
of the data sets, called the homologous data set, is interpolated
into the spatial frame of the other, called the reference data set.
If the initial homologous data set is similar to the reference set,
then there is little difference in performance between itera-
tively minimizing $H(a, b)$ or maximizing $I(a, b)$, since $H(a)$
and $H(b)$ are both fully visible in the reference frame and thus
are independent of the registration; either method increases
the correlation between the two data sets. However, if the
initial pose is sufficiently different, such that the homologous
data set is truncated in its transformation into the reference
space, its entropy as measured in the reference frame is no
longer constant with respect to registration variables. Under
these conditions maximizing $I(a, b)$ results in a larger cap-
ture range for the registration process, because maximizing
$I(a, b)$ encourages the homologous data sets to appear fully in
the reference frame due to its increasing contribution of $H(a)$ or $H(b)$ as it moves into the reference frame (anonymous reviewer).

Since we desire to use existing optimizer algorithms that traditionally minimize cost functions to obtain the condition of maximized $I$, we define our mutual information cost function, $MI$, to be related to $I$ by $MI = -I$. If we choose the correct registration parameters and the two data sets are highly correlated, then $I$ approaches its maximum and $MI$ approaches its minimum; if they are uncorrelated, then $MI$ approaches its upper bound of zero. Thus through registration maneuvers we can iteratively compute and minimize our cost function, $MI$. As depicted in Figure 1 our mutual information cost function, $MI$, is calculated directly from the joint probability histogram of gray-values under a given registration’s geometry mapping using the previous equation to form $-I$. The 2-D joint histogram is constructed by raster scanning through all voxels in the reference image set and incrementing bin counts corresponding to the gray-scale values from geometrically mapped voxel pairs. All registrations demonstrated here were achieved by a registration algorithm that uses the minimization loop shown in Figure 2. Note that the MI cost metric is computed using gray values from the reconstructed registered data set so that the effects of both the geometric mapping and pixel/voxel gray-value interpolants at all voxel locations are included.

The components of the control vector in our registration system are the 3-D coordinates of homologous control points;
the resultant geometric mapping, i.e. registration, is computed from their positions. The interpolated gray-scale reconstruction is computed to match the reference data set, voxel for voxel, where the position of each voxel in the reference data set is mapped into the homologous data set using the previously computed geometric mapping. The gray value of each mapped voxel in the homologous data set is computed by trilinear interpolation of its eight nearest neighbors. Note that the interpolation process always uses the original data set to compute a new, geometrically mapped, gray-scale image for each iteration, instead of using the previous interpolation to generate the next one. This approach prevents round-off and other undesired interpolation effects such as smoothing from accumulating across iterations. When the geometric transformation maps to a position outside the field of view of the homologous data set, the resulting ‘interpolated’ voxel is set to the constant value of zero.

Affine, i.e. linear, geometric mappings are computed using four point pairs for 3-D data sets. In addition affine transforms can be constrained to varying degrees of freedom by decomposing the computed full affine matrix into the geometrically closest (Nobel and Daniel, 1988) desired approximation, i.e. rotate±translate, rotate±isotropic-scale±translate and rotate±anisotropic-scale±translate. Thin-plate spline (TPS) geometric mappings, i.e. non-linear warps, are computed using at least one more point pair than used for the affine mapping (Bookstein, 1991; Bookstein and Green, 1993).

Although the user selects which input data set will be the reference and chooses the initial set of control point pairs for the reference and homologous data sets as in previous feature-based (user-biased) TPS methods, the algorithm automatically refines the position of the control points in the homologous data set under the direction of the optimizer by moving them to minimize the MI cost function.

The multivariate minimization algorithm we chose to drive the minimization of MI is the Nelder–Mead simplex method, usually referred to as ‘amoeba’ (Press et al., 1988). Since this algorithm is fully explained in many numerical analysis texts, and is widely used, we forgo further explanation of its function and refer the reader to Press et al. (1988) for further details. While certainly not optimal in terms of number of iterations to descend to the minimum, the simplex method does not require estimates of derivatives, is quite robust and thus its use is appropriate as an investigative tool.

Figure 3 illustrates a simplified 2-D example of the registration process described in the preceding paragraphs. Figures 3a and b demonstrate two ‘multimodal’ images in which the user has poorly chosen approximate homologous points. Figure 3c shows the geometric registration that maps the corresponding markers onto each other after the user’s initialization, but before automatic optimization begins. After initialization, although the markers are perfectly aligned, the images are still misregistered. Figure 3d shows the resulting final registration obtained by the optimizer driving the locations of the homologous points in Figure 3b to minimize the cost function, MI. Note that the position of the upper marker in Figure 3b has been moved by the optimization process to a new position as depicted in Figure 3e which yields the correct registration by minimizing MI.

Besides performing initialization (quick placement of initial approximate homologous markers) the user can set the stopping criterion for the optimizer and the geometric registration model to be used (affine, TPS and associated degrees of freedom). Computation times depend primarily on the size of the reference data set since the interpolator (trilinear for voxels and bilinear for pixels) computes a matching gray value for each pixel/voxel of the reference, but also depends on the geometric mapping used to achieve registration. Computation time for the TPS mapping is approximately twice as long as the affine mapping for just four control points. In addition, computation times for the TPS algorithm are proportional to the number of control points used. All timings reported here are for runs on a DEC Model 3000/500x OSF/Alpha with a clock rate of 200 MHz. All applications run under the AVS visual programming environment (Advanced Visual Systems, Waltham, MA). While good programming practices
were followed in software module coding, almost no attention has been focused on coding for increased computational speed. Standard algorithms, e.g. singular-value decomposition of matrices, were implemented using code from Numerical Recipes in C (Press et al., 1988).

3. RESULTS

3.1. Thoracic phantom study: PET/CT

A rigid, multicompartiment thorax phantom (Data Spectrum Corp., Chapel Hill, NC) was scanned on both a PET scanner and GE Genesis non-helical CT scanner. The Siemens 921 PET scanner is a 47-slice tomograph with an axial slice spacing of 3.375 mm (15 cm field of view), intrinsic in-plane resolution of \( \sim \)6 mm FWHM and an axial resolution of 4.5–5.5 FWHM. The resulting volume matrix has \( 128 \times 128 \times 47 \) voxels. CT data sets, acquired using 3 mm thick slices with no gaps or overlaps, were subsampled to \( 256 \times 256 \) voxels. CT data sets, acquired using 3 mm thick slices with no gaps or overlaps, were subsampled to \( 256 \times 256 \) image matrices before utilization in the registration algorithm to reduce memory requirements. In addition to the presence of streaks on the PET reconstruction, as to render them invisible or indistinguishable from artifact the eight total \( \ddagger \)ducials the isotope leaked from the capsules so quantitative assessment of registration accuracy. Before PET more around its inferior circumference, for use as \( \ddagger \)ducials in four around the phantom’s superior circumference and four \( \ddagger \)ducible void simulating an FDG ‘cold’ lesion. The lung \( \ddagger \)lings; the liver ‘lesion’ also included a water-lung and one liver, had higher FDG concentrations simulating thorax±abdomen scan, iodinated ionic X-ray contrast material was added to the liver cavity as well as lung and liver ‘lesions’ to increase their X-ray attenuation. Three ‘lesions’, two lung and one liver, had higher FDG concentrations simulating FDG-avid lesions; the liver ‘lesion’ also included a water-filled void simulating an FDG ‘cold’ lesion. The lung volumes were packed with small polystyrene spheres and water to mimic lung fields in CT and PET. A total of eight vitamin E capsules, 100 IU, were rigidly taped on the phantom, four around the phantom’s superior circumference and four more around its inferior circumference, for use as fiducials in quantitative assessment of registration accuracy. Before PET scanning the capsules were injected with FDG, but in two of the eight total fiducials the isotope leaked from the capsules so as to render them invisible or indistinguishable from artifact streaks on the PET reconstruction. The emission PET acquisition was 10 min in duration and collected \( 14.4 \times 10^6 \) net true coincidences. Since the PET reconstruction contained end slices with changing modulation and noise, two slices at each end of the 47-slice data set were eliminated. The remaining PET data were linearly converted to byte data. After the isotope had decayed sufficiently, the phantom was scanned with CT using 3 mm thick slices and with no skip between slices. The phantom was intentionally rotated by \( \sim \)12° about the anterior–posterior axis on the CT scanning table. The CT data were linearly converted to byte data over the range of \(-1024 \) to \(+1800 \) HU.

Initialization of the reconstruction algorithm did not use the external fiducial markers; in each modality four non-coplanar homologous points were casually picked (by ‘casual’ we mean within 2–3 cm of the true homology), two near the apices of the ‘lung’ volumes and two at the level of the superior margin of the ‘liver’ volume. Due to the rigid construction of the phantom and the known calibration of the two scanners, a six DOF affine reconstruction (rotate–translate) was selected.

The optimizer stopped when movements of \(< 0.01 \) mm were requested in the \( x \), \( y \) and \( z \) component locations of all control points. To avoid entrapment by local minima, restarts using a randomized simplex around the vector state of the previous stop are repeated until successive absolute changes in MI stopping values were \(< 0.0001 \). For each restart the simplex is initialized to the previous stopping vector plus additional vectors randomly distributed within a user-chosen, city-block ‘radius’ of the previous stopping vector, which for all cases reported here was 10 mm. The term ‘run’ refers to an initial or repeated automatic optimization that ceases when the stopping conditions are met. The joint histogram was computed using \( 256 \times 256 \) bins.

Although the PET data set was used as the reference geometry during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed here in Figure 4 (part 4) using a combination of two contrast windows to allow simultaneous visualization of the low-level ‘body’ and fiducial concentrations and the much higher ‘left-ventricle’ and ‘lesion’ concentrations. Because the clusters in the joint density histogram with the largest membership have the largest weight affecting the registration, the resulting registration is essentially insensitive to the presence of the fiducial markers whose volume is very small compared with the full thorax. The resulting locations of the fiducial vitamin E tablets were measured by computing the centroid of the resulting fiducial intensity distribution. An error analysis for the six fiducials is shown in Table 1. The mean error vector length, i.e. the average of the square root of the sum of squares of \( \Delta x \), \( \Delta y \) and \( \Delta z \), was 3.84 mm and the combined standard deviation of all error vector components was 2.20 mm, resulting in a standard error of the estimated error vector of 1.56 mm. In comparison the PET voxel size is 4.385\(^2\) by 3.375 mm\(^3\) with a mean voxel dimension of 4.05 mm. Thus these error measurements made just beyond the periphery of the PET data set (and thus beyond the region of support for the registration) are just slightly smaller than a voxel with essentially equal contributions from all components. The rotational errors are computed from the centroid of the phantom. The combined angular error mean and standard deviation for all fiducial components is \(-0.17° \) and 1.27°, respectively.
Figure 4. (Part 4) A 3-D rendering of three orthogonal slices through the thorax phantom’s PET/CT registration is shown using CT as a reference to demonstrate partial-volume registration quality. The color composite combines CT data in gray and PET data in red. For PET a combined contrast window has been applied to demonstrate both low and high concentrations without associated blooming of the high concentrations. Alternating slices of the registration can be viewed on the associated CD-ROM movie. (Part 5) A 3-D rendering of the registration of a PET cerebral blood flow study and a post-surgical coronal MRI study is shown using MRI as a reference. The color composite combines MRI in gray and PET in pseudocolor. All slices of the registration can be viewed on the associated CD-ROM movie. (Part 6) A 3-D rendering of the full affine thorax registration is shown using the downsized (256 × 256) CT as a reference. In the color composite CT is shown in gray using a lung window contrast setting and PET is shown in pseudocolor. All slices of the registration may be viewed on the associated CD-ROM movie. (Part 7) A 3-D rendering of a PET/CT thin-plate spline-warped registration of a patient’s thorax is shown using the downsized CT as a reference. In the color composite CT is displayed in gray using a combined lung and soft tissue window contrast setting and PET is shown in pseudocolor. All slices of the registration may be viewed on the associated CD-ROM movie. (Part 8) A 3-D rendering of a SPECT/CT, rotate-translate registration is shown using the downsized CT as a reference. In the color composite CT is displayed in gray using a soft tissue window contrast setting and PET is shown in pseudocolor. In addition to the expected concentrations in the liver, spleen and kidneys, note the uptake in the lymph nodes which reside near the descending aorta and the iliac arteries below the aorta’s bifurcation. All slices of the registration may be viewed on the CD-ROM.
the initial misregistration in the two data sets. More specifically, control points were chosen that resulted in the intentional 1 mm (subvoxel for the reference modalities of PET and SPECT) was requested in the starting positions, resulting in the statistics of Table 2 for the final, residual rotation and translation values of zero. This process was repeated for eight additional randomly selected starting positions, resulting in the statistics of Table 2 for the mean (\(\mu\)), population standard deviation (\(\sigma\)), two standard errors of the mean (2 SEM) and the probability (\(P\)), from Student’s \(T\)-test that the samples come from a zero mean population, all computed from nine repeated rotate-translate affine registrations of T2- and T1-weighted axial MRI data sets from NLM’s Visible Male using randomized starting poses.

<table>
<thead>
<tr>
<th>Fiducial</th>
<th>(\Delta x) (mm)</th>
<th>(\Delta y) (mm)</th>
<th>(\Delta z) (mm)</th>
<th>Error vector length (mm)</th>
<th>(\Delta \theta_x) ((^\circ))</th>
<th>(\Delta \theta_y) ((^\circ))</th>
<th>(\Delta \theta_z) ((^\circ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.46</td>
<td>-1.22</td>
<td>-1.13</td>
<td>1.73</td>
<td>0.39</td>
<td>0.91</td>
<td>-0.24</td>
</tr>
<tr>
<td>2</td>
<td>-3.83</td>
<td>0.01</td>
<td>0.79</td>
<td>3.91</td>
<td>1.24</td>
<td>0.32</td>
<td>-1.37</td>
</tr>
<tr>
<td>3</td>
<td>-2.09</td>
<td>-3.08</td>
<td>0.30</td>
<td>3.73</td>
<td>0.82</td>
<td>-3.12</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>-3.78</td>
<td>-2.91</td>
<td>-2.67</td>
<td>5.47</td>
<td>1.70</td>
<td>-1.38</td>
<td>-1.03</td>
</tr>
<tr>
<td>5</td>
<td>-0.87</td>
<td>2.22</td>
<td>-2.32</td>
<td>3.33</td>
<td>0.63</td>
<td>-0.50</td>
<td>1.64</td>
</tr>
<tr>
<td>6</td>
<td>2.08</td>
<td>-2.11</td>
<td>3.90</td>
<td>4.90</td>
<td>-0.72</td>
<td>-1.20</td>
<td>-1.29</td>
</tr>
<tr>
<td>(\times), std dev.</td>
<td>2.36</td>
<td>2.02</td>
<td>2.43</td>
<td>1.31</td>
<td>0.83</td>
<td>1.42</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Table 2. Statistics, mean (\(\mu\)), population standard deviation (\(\sigma\)), two standard errors of the mean (2 SEM) and probability (\(P\)), from Student’s \(T\)-test that the samples come from a zero mean population, all computed from nine repeated rotate-translate affine registrations of T2- and T1-weighted axial MRI data sets from NLM’s Visible Male using randomized starting poses.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>(\Delta \theta_x)</th>
<th>(\Delta \theta_y)</th>
<th>(\Delta \theta_z)</th>
<th>(\Delta x)</th>
<th>(\Delta y)</th>
<th>(\Delta z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>-0.0356</td>
<td>-0.0176</td>
<td>-0.0713</td>
<td>-0.1124</td>
<td>-0.2935</td>
<td>0.0429</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.0094</td>
<td>0.0573</td>
<td>0.0142</td>
<td>0.0067</td>
<td>0.0144</td>
<td>0.0512</td>
</tr>
<tr>
<td>2 SEM</td>
<td>0.0063</td>
<td>0.0382</td>
<td>0.0095</td>
<td>0.0045</td>
<td>0.0096</td>
<td>0.0341</td>
</tr>
<tr>
<td>(P)</td>
<td>0.0000</td>
<td>0.3845</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0362</td>
</tr>
</tbody>
</table>

3.2. Post mortem study

The National Library of Medicine’s Visible Male data was used as a source of MRI T1- and T2-weighted head acquisitions. Since the subject was motionless, we assume that the data sets are almost perfectly aligned as acquired. Data matrix sizes were \(256 \times 256 \times 33\), where each voxel had dimensions of \(1^2 \times 5\) mm\(^3\). In this registration accuracy test the T2-weighted data was used as the reference and T1-weighted data was used as the homologous set. During user initialization control points were chosen that resulted in the intentional initial misregistration of the two data sets. More specifically the initial misregistration in \(\theta_x\), \(\theta_y\), and \(\theta_z\) was \(-17.5^\circ\), \(16.5^\circ\) and \(24.4^\circ\), respectively, and the initial (mis)translation in \(x\), \(y\), and \(z\) was \(16.4\), \(-14.4\), and \(-3.4\) mm, respectively. Because the object was rigid, a six DOF affine reconstruction (rotate-translate) was selected. The optimizer was set to stop when movements of \(<0.01\) mm were requested for all component locations of all control points. Restarts using a randomized simplex around the vector state of the previous stop were repeated until successive absolute changes in MI stopping values were \(<0.0002\). Assuming the initial orientations of the two data sets were identical, a perfect registration would yield final, residual rotation and translation values of zero. This test resulted in final residual rotations of \(-0.023^\circ\), \(-0.017^\circ\), and \(-0.034^\circ\) for \(\theta_x\), \(\theta_y\), and \(\theta_z\), respectively, and translations of \(-0.12\), \(-0.32\), and \(0.046\) mm for \(x\), \(y\), and \(z\), respectively. The process was repeated for eight additional randomly selected starting positions, resulting in the statistics of Table 2 for the mean (\(\mu\)), population standard deviation (\(\sigma\)), two standard errors of the mean (2 SEM) and the probability (\(P\)), from Student’s \(T\)-test that the samples come from a zero mean population. Thus all of the resulting registration parameters are statistically different from zero at the 0.05 level except \(\Delta \theta_y\). Note however, that the means and standard deviations of all six parameters are very small; the largest translation is less than a third of a 1 mm voxel, \(\bar{\Delta y} = 0.29 \pm 0.01\) mm, mean \(\pm 2\) SEM.

3.3. Clinical studies

For all of the clinical studies which follow the optimizer stopped when movements of \(<1\) mm (subvoxel for the reference modalities of PET and SPECT) were requested in the
x, y and z component locations of all control points. To avoid entrapment by local minima, restarts using a randomized simplex around the vector state of the previous minimum are repeated until successive absolute changes in MI stopping values were <0.0002. Again, the term ‘run’ refers to an initial or repeated automatic optimization that ceases when the stopping conditions are met. The joint histograms were all computed using 256 × 256 bins.

3.4. Brain PET/MRI and missing data
In the first case we demonstrate the registration of a PET data set from a patient who had a PET cerebral blood flow study in 1990 which involved transient carotid occlusion using a balloon catheter and a post-surgical MRI study in 1992. The PET study, obtained under the occluded carotid condition resulting in a PET signal from approximately half of the ipsilateral side of the head, was used as the reference scan for registering the post-surgical MRI. The registration was constrained to a six degrees of freedom (rotate–translate) geometric mapping. An older Siemens model ECAT 931/08-12 was used for PET data acquisition. The ECAT 931 is a 15-slice tomograph (128 × 128 × 15 voxels) with axial plane spacing of 6.75 mm (10 cm field of view), intrinsic in-plane resolution of ~6 mm FWHM and an axial resolution of 7–8 mm FWHM. The MRI for the same patient was obtained from a post-surgical imaging study obtained approximately two years later. The original MRI data set was a coronal, T2-weighted, spin-echo series consisting of 16 images of 5 mm width and 1 mm separation (TR/TE: data set was a coronal, T2-weighted, spin-echo series consisting of 16 images of 5 mm width and 1 mm separation (TR/TE: approximately two years later. The original MRI data set was a coronal, T2-weighted, spin-echo series consisting of 16 images of 5 mm width and 1 mm separation (TR/TE: 3000/90, FOV = 20, 256 × 256 matrix, NEX = 1). Before registration the MRI data set was low-pass filtered in the ‘high-resolution’ coronal planes with a Gaussian point spread function of 7 mm to approximate the partial volume contributions inherent in the PET acquisition. After casual user initialization of control points but before automatic registration, the MI was ~0.9249; after two automatic runs consuming a total of 5.1 min of CPU time for 370 iterations, the final MI was minimized to ~1.0292. The final geometric mapping obtained with the low-pass filtered MRI data, was applied to the unfiltered MRI data and resulted in the registration displayed in Figure 4 (part 5). Note the lack of translation error (bias) in the resultant registration despite the missing data.

Few algorithms are capable of performing the registration demonstrated in Figure 4 (part 5). Note that the coronal head study begins anteriorly at the eyes, but stops before imaging the occipital cortex. Between the limits of both the half-head PET and coronal MRI, conservatively less than a third of the head volume is jointly available for driving the registration. Yet the resulting reconstruction shows no apparent bias. In addition, only low-pass filtering of the MRI data was required; no other preprocessing, e.g. stripping off scalp and other non-cortical structures by segmentation, was necessary. By way of comparison, most distance mapping–minimization techniques would require not only scalp and bone removal before surface definition, but would also result in a biased registration where the increased distance measures caused by the missing PET data would drive the registration to place the existing data in the center of the MRI data set. Computationally intensive cross-correlation methods could obtain an unbiased result, but in this multimodal situation cross-correlation must be applied to edge-enhanced images since gray-scale mappings are inconsistent between PET and MRI.

In contrast with the previous description of MRI low-pass filtering for registration of partial data sets, in other experiments we have found that MI-based registration of normal PET/MRI data sets with a full region of support is trivial and requires no low-pass preprocessing.

3.5. Body
3.5.1. PET/CT
The following two cases demonstrate automatic volumetric registrations in the body using PET/CT data sets. The PET acquisitions are acquired on a Siemens 921 scanner using 18FDG which traces glucose metabolism in vivo. The description of the tomograph can be found in the previous phantom registration description and methods of image acquisition are as previously outlined (Wahl et al., 1994). The CT data sets, acquired using 1 cm thick slices with no gaps or overlaps, were subsampled to 256 × 256 image matrices before utilization in the registration algorithm to reduce CPU memory requirements.

Consistent patient geometry
The first body registration case we present was acquired on the CT scanner with free, shallow respiration and arms down to mimic PET data acquisition conditions. The patient has a right lung lesion easily visible on both CT and PET. At initialization preceding automated registration, four starting control points in each data set were chosen: one approximately at the carina and three more at the level of the diaphragm: one on the anterior chest midline and one each at the posterior-left and posterior-right chest wall. The control point at the carina was intentionally misplaced ~2 cm laterally to the right of the carina in the CT data set while the carina was correctly chosen in the PET data set to demonstrate how initial control points may be inaccurately placed and still achieve registration. After user initialization and before automatic registration the MI was ~0.3387. The geometric registration chosen was a full, 12-parameter affine geometric mapping. Convergence was obtained in four runs requiring a total of 31 CPU minutes for 725 iterations yielding a final MI metric of ~0.5289. The resulting registration is shown in Figure 4 (part 6).
Although the PET data set was used as the reference during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed here in Figure 4 (part 6). Note the accurate delineation of the cardiac and vascular structures as well as that of the lesion. Although our previous experience indicates that registrations defined by surface markers represent internal organ geometries poorly, one relatively well registered sternal surface marker can be seen in one of the slices. The apparent registration accuracy using the full affine model is a tribute to positioning the patient in a consistent scanning geometry for both modalities as well as the use of an accurate registration algorithm.

**Inconsistent patient geometry**

In the next study the CT data set was acquired with the patient’s arms held up over the head and respiration held during scanning between breaths at maximum inspiration, while the PET data set was acquired under the usual arms down condition. Between these two acquisition conditions internal organs and chest wall structures have different complex deformations which cannot be modeled by an affine transformation.

The patient has a large lesion in the right chest wall that invades the lung space and is visualized on both CT and FDG PET. A five-point thin-plate spline (TPS) reconstruction was used to partially compensate for the non-affine geometric changes. At user initialization, control points were placed approximately at both lung apices, carina, anterior chest wall at the level of the carina and on the lesion. After user initialization the MI was $-0.4874$. After three automatic optimizer runs which consumed a total of 57 min of CPU time for 650 iterations, the MI was reduced to $-0.5971$. As before, although the PET data set was used as the reference geometry during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed in Figure 4 (part 7). In addition to the lesion, the lung apices, chest wall and cardiovascular structures are well registered. Note the avid $^{18}$FDG uptake seen both in the lesion and the left ventricle of the heart.

### 3.5.2. Abdominal SPECT/CT

The following case demonstrates registration of $^{131}$I anti-CD-3.5.2. Abdominal SPECT/CT left ventricle of the heart. (Kaminski et al., 1993, 1996). The CT patient data was acquired using the standard 1 cm slice thickness. The 512 $\times$ 512 CT matrix was subsampled to a 256 $\times$ 256 matrix to reduce memory requirements with little or no penalty due to the much lower voxel density SPECT data set. Initial control points in the approximate shape of an equilateral triangle with vertices located on the surface of the body wall were chosen at the level of the inferior tip of the liver. An additional fourth marker was placed near the socket of the left hip. The registration was constrained to be a rotate–translate mapping. After user initialization the MI was $-0.8793$. After two automatic runs MI was reduced to $-0.9265$. Due to the small size of the SPECT data matrix, total elapsed CPU time for the affine registration was only 3.5 min for 300 iterations. Again, although the final affine registration parameters were obtained using SPECT as the reference data set, the final registration is displayed in Figure 4 (part 8) using the CT data set as the reference.

Note that in addition to expected uptake in the spleen, kidneys and liver, there is increased uptake in the specifically targeted, enlarged abdominal lymph nodes that surround the descending aorta and iliac arteries in the abdomen and pelvis; there is little uptake in fat. External skin fiducials appropriate to each scanning modality were placed prospectively on the patient and can be partially seen in two of the slices. In general, the registration between the external fiducials is poor. Given the excellent agreement evident in other widely distributed structures, it is clear that the limited geometric model used, i.e. rotate–translate, is incapable of offsetting all the complex deformations between the two data sets. Fortunately for this application the registration algorithm is driven (weighted) by larger structures such that the most relevant internal organs of interest are well registered.

### 4. DISCUSSION

We have demonstrated and assessed the accuracy of a highly automated algorithm for 3-D image registration based on maximizing mutual information in both phantom and patient studies. Although not demonstrated here, full brain data sets such as PET and MRI are easily and rapidly registered without preliminary stripping of skull and skin. Even partial brain data sets are registered without bias by forming data sets with consistent partial volume effects using low-pass filtering and limiting registration to a rotate–translate geometry. Literally in the strict sense of the definition, it appears that as more information becomes available, restrictions can be relaxed and more DOFs can be supported. For registrations involving the comparatively low-entropy SPECT data set it was necessary to restrict the geometry model to the simplest, six DOF affine model. Attempts at allowing isotropic scaling (seven DOF) or...
anisotropic scaling (nine DOF) produced poor results. On the other hand the higher entropy PET data support registration not only with a full, 12 DOF affine model, but also with a 15 DOF TPS warping model as well. Other published neuroscience 2-D applications involving autoradiography and light imaging routinely support 18 DOF TPS warping (Kim et al., 1997).

To the extent supported by the information content of the modalities involved, the geometry deformation model should be chosen to match the physical deformation between the data sets. In the inconsistent geometry example of PET acquired with patient arms down while respiring freely and CT acquired with arms held over the head while respiration was held, the full affine registration was a poor compromise. The optimizer moved the right lung apex marker too far cranially to try to position the lesion correctly. The compromise result (not shown here) was that neither the lesion nor the lung apices and shoulders were correctly registered. When the five-point TPS warping registration was computed, the global fit was much improved as can be appreciated in Figure 4 (part 7).

Total computation times are strongly affected by the size of the reference matrix, since the homologous data set is mapped onto each voxel in the reference data set. Thus an important time-saving strategy is to use the smallest data set (measured in numbers of voxels) as the reference to compute the geometry mapping parameters. After computing the last iteration, the higher resolution data set may be used as a reference and the low-resolution data set interpolated to match the high-resolution set by applying the inverse of the previously computed, geometric mapping. The increase in computation time for thin-plate spline warping over the full affine transform is due to the increased number of control points and use of the more computationally intensive TPS reconstruction algorithm. Even so, computation times using an MRI data set \((256 \times 256 \times 23)\) as a reference and a thin-plate spline mapping of 11 control points (33 degrees of freedom) is only 15 s per iteration with convergence at 575 iterations requiring just under 2.5 h.

An important strategy to reduce further total registration computation time is to perform the initial registration runs using a decimated data set to bring the two data set geometries into close agreement, before using the complete reference data set to accommodate more subtle partial volume effects. Since the computational cost is related to the number of voxels in the reference data set, then decimating, i.e. downsizing, the reference data set by two in each of the three volumetric dimensions will result in run times shorter by nearly a factor of eight. This approach is particularly valuable when the initial poses of the two data sets are grossly different. Such an approach significantly reduces the total run times reported for the cases previously reported in this paper. Currently we routinely perform two initial runs using a downsized reference, followed by as many runs as necessary with the full size reference to satisfy the stopping condition. Although we have previously shown that increasing the number of bins in the histogram increases the sensitivity of the MI cost function (Kim et al., 1997), with downsizing it is important to reduce the number of histogram bins used in order to avoid obtaining sparsely sampled, unrepresentative distributions of the geometrically mapped data sets. For this paper we did not perform such downsizing maneuvers in order to simplify its presentation through the use of a single, simple algorithm with limited input parameter ranges.

The accuracy of the registration is affected by the signal-to-noise ratio (SNR) in the data sets. In the case of the CT/PET thorax phantom study presented in Figure 4 (part 4), the PET Poisson noise dominates. Although the accuracy of the resulting registration obtained at the periphery of the data set, mean error vector length of 3.84 mm, is slightly below the PET voxel size, other lower count, and thus more noisy, acquisitions would yield reduced accuracies, while count-rich acquisitions with less noise (volumetric acquisitions contain 6–8 times as many counts) would produce better accuracies than those reported here. Since affine registrations can be decomposed into translate, rotate, scale and shear components, rotational and scaling error components cause error vector lengths to increase at radial distances further from the centroid. Thus the errors measured just beyond the periphery of the PET data set for the phantom study are typically worst case; smaller errors are typically measured more centrally.

In line with the previous discussion of accuracy and SNR, the registration of two MRI data sets from the NLM’s Visible Male results in much smaller error estimates than those obtained with the PET/CT phantom. Using the angular and translation error variances from Table 2 for the MRI T2/T1 registration, the estimate of the mean error vector length, which has a Maxwellian distribution (Papoulis, 1984), at a radius of 10 cm from the centroid is <0.11 mm and <0.05 mm at the centroid. Also note that the non-zero means may indeed be real, i.e. the original T1- and T2-weighted acquisitions do have a fractional voxel offset. The superior-to-inferior course of three ‘strings’ around the exterior surface of the head in the two original data sets provides x and y fiducials for the 1–4 pixel inspection and visual verification of fractional displacements at nearly all axial slice levels. The most noticeable displacement is in the AP direction corresponding to the y-axis direction of Table 2. These slight displacements may be the result of object motion between scans or imperfect gradient amplifier control between the two acquisition sequences.
In summary, the inherent benefits in using the mutual information cost function, MI include:

- The registration is simple and independent of operator bias. The algorithm uses the full dynamic range of the data without edge enhancement, segmentation, disarticulation or operator-selected morphologic features.
- MI is computed from volumetric gray-scale data, not just surfaces. This facilitates geometric mappings that include warps which fit internal structures as well.

The recent work of two groups (Collignon, 1995a, b; Viola et al., 1995; Wells et al., 1995) has ushered in a new era in registration where both multi- and isomodality registrations are achieved automatically and easily by the same software algorithm with little fine tuning of control parameters. We have demonstrated several clinically interesting modality pairs registered with this robust, easily used algorithm, many of which have geometric inconsistencies between data sets that are difficult for other algorithms to register. This algorithm requires only minimal user initialization (no segmentation, surface definition or data stripping) and typically runs in minutes on currently available, high-performance workstations. Although more validation efforts are required, the apparent registration accuracy and robust convergence properties for all modality pairs demonstrated here along with ease of use suggest that algorithms of this class significantly increase the likelihood that multimodality registrations will be routinely available to aid clinical diagnosis in the near future.

ACKNOWLEDGEMENTS

The authors wish to thank an anonymous reviewer who provided very constructive comments which led to a significantly improved text. This work was in part supported by DHHS PHS NIH 1R01CA59412.

REFERENCES


