

A Practical Salient Region Feature Based 3D Multi-Modality Registration Method for Medical Images

Dieter A. Hahn^a, Gabriele Wolz^b, Yiyong Sun^c, Joachim Hornegger^a, Frank Sauer^c, Torsten Kuwert^b and Chenyang Xu^c

^aInstitute of Pattern Recognition, University of Erlangen–Nuremberg (FAU), Martensstr. 3, 91058 Erlangen, Germany

^bDepartment of Nuclear Medicine, FAU, Krankenhausstr. 12, 91054 Erlangen, Germany

^cSiemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA

ABSTRACT

We present a novel representation of 3D *salient region features* and its integration into a hybrid rigid-body registration framework. We adopt scale, translation and rotation invariance properties of those intrinsic 3D features to estimate a transform between underlying mono- or multi-modal 3D medical images. Our method combines advantageous aspects of both feature- and intensity-based approaches and consists of three steps: an automatic extraction of a set of 3D salient region features on each image, a robust estimation of correspondences and their sub-pixel accurate refinement with outliers elimination. We propose a region-growing based approach for the extraction of 3D salient region features, a solution to the problem of feature clustering and a reduction of the correspondence search space complexity. Results of the developed algorithm are presented for both mono- and multi-modal intra-patient 3D image pairs (CT, PET and SPECT) that have been acquired for change detection, tumor localization, and time based intra-person studies. The accuracy of the method is clinically evaluated by a medical expert with an approach that measures the distance between a set of selected corresponding points consisting of both anatomical and functional structures or lesion sites. This demonstrates the robustness of the proposed method to image overlap, missing information and artefacts. We conclude by discussing potential medical applications and possibilities for integration into a non-rigid registration framework.

Keywords: Hybrid Registration, Saliency, Region Features, Multi-Modality

1. INTRODUCTION

In medical image processing, registration is an important technique that computes a spatial mapping between two or more images. It can be used in a variety of applications. The fundamental and challenging requirement for the alignment transform is to overlay corresponding image content optimally. State-of-the-art registration techniques and comparisons of algorithms within this field of research can be found in several comprehensive surveys.^{1–4} Registration has become a crucial task in a variety of medical imaging applications, for example, the creation of atlases or normative databases are suitable for image or object analyses, intra- or inter-patient studies let physicians gain knowledge of the development of diseases or time based follow-up studies during cancer therapy. Using different imaging systems for the same subject can help to obtain more information but requires multi-modality registration techniques. Adding complementary information is facilitated by various medical imaging systems that can be coarsely divided into two major categories: anatomical imaging to extract morphological information (e.g. X-Ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US)) and functional imaging that allows to visualize information on the metabolism of the underlying anatomy (e.g. Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), functional MRI (fMRI)). In multi-modality image registration, the combination of different types of images is advantageous for the physician. For instance, CT images provide good spatial resolution, whereas PET images depict the functionality of the underlying tissue. The lack of functional information in the CT images can therefore be

Further author information:

D. A. Hahn: Dieter.Hahn@informatik.uni-erlangen.de, Phone: +49 9131 85 27894

compensated by a fusion with according PET images that on their part lack spatial resolution. The presented work concentrates on a registration approach that automatically extracts 3D regions, retrieves corresponding pairs and establishes a rigid transform in order to visualize the fused results for a medical application.

2. RELATED WORK

The basis for this work is given by several articles that introduce *saliency* in the context of image processing. We use it to extract distinct image regions from a reference and a template image for establishing a 3D transform. In 1998, Gilles introduced the idea of using a measure for image saliency based on the local complexity of intensities.⁵ Kadir and Brady reused that idea and developed a combined approach of image saliency, scale detection and content description.⁶ They presented a salient feature region extraction algorithm together with some example applications, e.g. content tracking across subsequent video frames. In subsequent work on this subject, the algorithm has been analysed in greater detail and an affine invariant salient feature description was published.^{7,8} Saliency operators have also been successfully applied to estimate the geometrical relationship between stereo images and to recover 3D information.^{9,10}

Various common registration approaches are based exclusively on either extrinsic, intrinsic features or voxel intensities. Hybrid techniques make use of a combination of multiple such properties instead and may be preferred especially for the registration of images from modalities, where image intensity or geometric information alone does not provide an accurate measurement basis. For instance, mutual information paired with an additional information channel that consists of region labelling information may improve the registration results of MR and PET images.¹¹ Modersitzki and Fischer presented a non-rigid hybrid image registration approach that is based on both landmark and intensity information.¹² Lately in 2005, a semi-automatic hybrid approach was introduced at the SPIE Medical Imaging that uses a combined landmark and voxel intensity based registration algorithm that produces better results, especially for low-contrast abdominal regions in CT images.¹³ Hybrid registration techniques are known, for instance, in blood plasma gel electrophoresis or protein imaging.^{14,15} Huang et al.¹⁶ were among the first to use a salient region feature approach for a 2D similarity transform registration and to show its applicability on MR, retinal and aerial images. Geometric features have also been successfully integrated into deformable registration frameworks.¹⁷ We propose an expansion of 2D salient region features to 3D and show their applicability for mono- and multi-modality medical image registration.

3. METHODS

The following section provides an overview of the methods that are incorporated in the proposed algorithm. It is described how a saliency description can be used to automatically extract features from 3D images. In order to get the most descriptive salient region features, a local maximum search based on a region growing algorithm is performed on the saliency values that are attained for each voxel. We present a straight forward and efficient solution to the problem of feature clustering that makes use of a kD-tree structure. In order to derive a transform between feature sets, the features themselves are considered as points in 3D. This reduces the problem to registering two point clouds, which can be achieved efficiently with the iterative closest point algorithm (ICP). The hybrid aspect of the proposed algorithm is included by using a form of normalized mutual information for the search of correspondences between features from different sets. As the set of feature pairs may still contain outliers, expectation maximization (EM) type optimization is used to extract the subset of pairs that yields an optimal image transform. Outliers in this context are region feature pairs that actually match neither in content nor spatial position.

3.1. Salient Region Feature Extraction

One of the key principles of salient region features is the expression of a high amount of local unpredictability or signal complexity with respect to a certain scale. In our case, scale refers to the radius of a spherical region around a voxel. The approach weights points of interest on basis of Shannon's entropy for circular regions of different scales.⁶ We assume that voxels from distinct corresponding anatomical or functional structures have similar saliency values. This indeed might not be the case if the intensity values in images from different modalities are too distinct. We observed that the positions of the local saliency maxima for the SPECT and PET images may often be locally translated within corresponding structures of interest, compared to the according

CT images. A local rigid registration step includes the sub-pixel accurate adaption of the region centers and addresses this problem. This moves the region to the corresponding location based on local intensity correlations but does not deform the image content. By introducing this step, the basic assumption of similar saliency values for corresponding features mentioned earlier is valid for multi-modality images as well. The local property of the saliency description provides a major benefit to image registration: salient region features that correspond between different images are invariant to gross spatial transforms, even if the images do not overlap. The saliency is defined for an image intensity range D as follows:

$$\mathcal{A}_D(s_p, \mathbf{x}) = \mathcal{H}_D(s_p, \mathbf{x}) \cdot \mathcal{W}_D(s_p, \mathbf{x}) \quad (1)$$

where \mathcal{H}_D denotes the entropy with respect to the image intensity values $i \in D$ within a spherical neighborhood region \mathcal{R}_s around a voxel position \mathbf{x} with scale s :

$$\mathcal{H}_D(s, \mathbf{x}) = - \int_{\mathcal{R}_s} p(i, s, \mathbf{x}) \log_2 p(i, s, \mathbf{x}) \, di \quad (2)$$

Here, $p(i, s, \mathbf{x})$ is the probability density function (PDF) of the descriptor i for the image intensity values contained in \mathcal{R}_s . $\mathcal{W}_D(s, \mathbf{x})$ is a measure for the similarity between PDFs with respect to the scale. It grows with increasing dissimilarity of the PDFs:

$$\mathcal{W}_D(s, \mathbf{x}) = s \int_{\mathcal{R}_s} \left| \frac{\partial}{\partial s} p(i, s, \mathbf{x}) \right| \, di \quad (3)$$

The scale s_p that results in a local peak of \mathcal{H}_D at \mathbf{x} is given by:

$$s_p = \left\{ s \mid \frac{\partial \mathcal{H}_D(s, \mathbf{x})}{\partial s} = 0 \wedge \frac{\partial^2 \mathcal{H}_D(s, \mathbf{x})}{\partial s^2} < 0 \right\} \quad (4)$$

After solving eq. 1 for each voxel, two temporary result images of the same size as the input image have to be analysed: one that contains the saliency values and a scale image that contains the values of eq. 4. The locally optimal and therefore most descriptive salient region points are extracted with a region-growing search approach from the saliency image. For a search space reduction, a global saliency threshold δ is used as a lower limit. An empirical setting to half the average saliency ($\delta = \frac{1}{2}\bar{\mathcal{A}}$) in the conducted experiments produced good results for the elimination of insignificant regions. An efficient local region growing approach determines the positions of local saliency maxima, which results in a list of voxel positions that are ordered according to their saliency values. One effect of the original salient feature extraction algorithm is the clustering of local maxima, which arises, for instance, if a global threshold is applied and the features are extracted only according to descending saliency values (see figure 1(a)). Feature clustering may have a negative impact on the accuracy of a subsequent registration. For example, in multi-modality feature based registration, one image may contain a dominant salient feature cluster in the upper right part, the other in the lower left part of the image but with a different underlying anatomical content due to different intensity values and local variations. This may be the case in PET-CT or SPECT-CT image pairs. A minimization of the mean square error (MSE) between the feature sets using ICP tends to align mainly the clusters. Obviously in that case, the resulting transform is biased towards the clusters, whereas if the clustering is removed, the features are spread more uniformly across various salient structures in the images (e.g. seen in figure 1(b)).

A nearest point algorithm based on a kD-tree¹⁸⁻²⁰ structuring of the region features is used to solve the clustering problem in a fast and simple manner. The kD-tree is created with the indices of the region centers of the extracted local saliency maxima as leaves. The K nearest neighbors of a specific feature f can then be efficiently found by a tree query on the region center positions, which returns the distance between f and the features in the tree. The scale parameter can be used as a minimal distance requirement: all returned features with equal or less distance than the scale of the queried feature f and with lower saliency are removed from the feature set. This restriction can be applied to the entire set in order to remove clustered regions. If a specific size of the result set is required, the list may be padded with features of lower saliency that fulfill the distance criterion. A feature is kept in the set, if its center is not situated within the region of a feature with a higher saliency value. The resulting set of 3D salient region features is therefore distributed more uniformly, which provides a better conditioned initial set for the subsequent feature correspondence search. Examples of resulting sets with the most salient regions are given in figures 1(a) – 1(e).

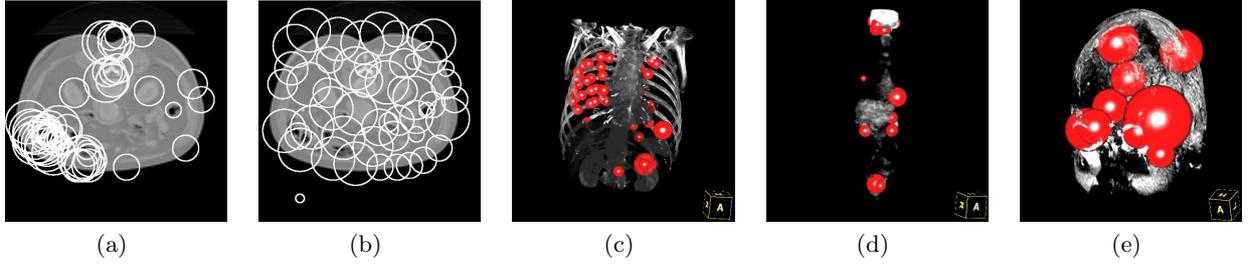


Figure 1. The set of salient feature regions before 1(a) and after 1(b) the proposed removal of the clustering. The most significant 3D salient region features are visualized after the extraction from a CT 1(c), a PET 1(d) and an MR image 1(e). The volumes are windowed with a specific transfer function to visualize the location of the features in 3D, whereas the extraction itself has been performed on the entire intensity range.

3.2. Feature Correspondence Search

In the main steps of the approach, an initial ICP algorithm aligns the entire feature sets with a locally minimal MSE. Feature pairs with a large spatial distance apart are unlikely to correspond and can be removed from the set of hypothesized correspondences. This drastically reduces the search space for joint correspondences. Let I_r be the reference and I_t the template image, the initial search step (also referred to as region component matching in literature¹⁶) estimates a set of hypothesized correspondences between features of the two images. Let N_r be the number of features extracted from I_r and N_t the number of features from I_t . The set of all hypothesized feature correspondences is $\mathcal{C} = \{c_{i,j}\}$ with $i \in [1, N_r]$, $j \in [1, N_t]$, $|\mathcal{C}| = N_r \cdot N_t$ and with $c_{i,j} = (f_i, f_j)$ being the pair of the features f_i in I_r and f_j in I_t .

The parameter set Θ defines the transform \mathcal{T} that aligns the two images and can be estimated based on the translation, scale and rotation invariance properties between f_i and f_j . The translational part between f_i and f_j can directly be estimated by: $\hat{\Theta}_{i,j}^T = \mathbf{p}_i - \mathbf{p}_j$, with \mathbf{p}_i and \mathbf{p}_j being the center positions of the i -th reference and j -th template features in physical space. The scale invariance in our case is not needed, as for 3D medical images, the voxel dimensions are provided by the DICOM (Digital Imaging and Communications in Medicine) header. To achieve rotation invariance, the rotation parameters are estimated by a local rigid body registration of the 3D salient feature regions based on their intensity values. The optimization is restricted to the rotational parameter subspace Θ^R and driven by an intensity similarity measure. Here we use the entropy correlation coefficient (ECC, see eq. 5), which is a specific form of normalized mutual information.^{16, 21, 22} Besides the improved robustness to the overlap domain, it provides some additional advantageous properties. Like mutual information, increasing values indicate an increasing dependency between the images and vice versa:

$$ECC(A, B) = 2 - \frac{2H(A, B)}{H(A) + H(B)} \quad (5)$$

The rotation invariance can therefore be formulated as an optimization problem for Θ^R :

$$\hat{\Theta}_{i,j}^R = \operatorname{argmax}_{\Theta^R} ECC(f_i, f_j^{\mathcal{T}_{\Theta^R}}) \quad (6)$$

The local similarity \mathcal{L}_{local} of two feature regions with respect to $\hat{\Theta}_{i,j} = \{\hat{\Theta}_{i,j}^T, \hat{\Theta}_{i,j}^R\}$ is used as a measure for the quality of a hypothesized correspondence $c_{i,j}$, where the samples for the evaluation of the ECC are drawn from \mathcal{R}_s with $s = \max\{s_i, s_j\}$:

$$\mathcal{L}_{local}(c_{i,j}) = ECC(f_i, f_j^{\mathcal{T}_{\hat{\Theta}_{i,j}}}) \quad (7)$$

Huang et al. suggest to order the set \mathcal{C} according to \mathcal{L}_{local} and to take M feature pairs from the top of the ordered set.¹⁶ A global image similarity measure \mathcal{L}_{global} is used to estimate the quality of each of the M pairs:

$$\mathcal{L}_{global}(c_{i,j}) = ECC(I_r, I_t^{T_{\hat{\Theta}_{i,j}}}) \quad (8)$$

where \mathcal{L}_{global} is evaluated on the entire overlap domain of the two images instead of just the local feature regions. Before continuing with the sub-pixel accurate registration on basis of this correspondence subset, the M pairs are reordered according to descending values of \mathcal{L}_{global} . One problem of this approach is the large number of false matches among the M pairs, which consequently leads to a small amount of good feature correspondences and more outliers in the resulting set after convergence. This is the case for the top row of figure 2. In addition, comparing all features with each other has a high computational complexity as $N_r \times N_t$ comparisons have to be performed, which is not suitable for a 3D approach due to the high computation time. In the following, we present a slightly different approach.

The correspondence search space can be drastically reduced from combinations of all pairs to combinations only between locally nearest feature neighbors. The set of neighbors is estimated by the ICP transform between the reference and template region feature sets (regarding the sets as point clouds of the region center positions). The result is used to transform all template features and store them into a new kD-tree. Now, for each salient feature in the reference image, the approximately nearest neighbors can be determined in a fast search on the tree. The number N_n of transformed template feature neighbors that are combined with each reference feature is set to a much smaller value than the entire cardinality of the set ($N_n \ll N_t$). This reduces the complexity to $N_r \times N_n$ based on the assumptions that the initial ICP transform is a good approximation of the actual alignment transform and that features are less likely corresponding if they are a greater distance apart. In our conducted experiments, a neighborhood size of $N_n = \frac{1}{10}N_t$ has been applied successfully to establish an initial search space for joint correspondences. Furthermore, we order the hypothesized correspondences not by the local measure \mathcal{L}_{local} (eq. 7) but by the global one \mathcal{L}_{global} (eq. 8) in the first place, which results in far less outliers in the estimated correspondence set. The second row of figure 2 shows the difference in the top four pairs found by the proposed method.

3.3. Sub-pixel Accurate Registration

The set of hypothesized correspondences $\mathcal{C} = \{c_{i_1,j_1}, \dots, c_{i_M,j_M}\}$ of size M that is computed in the previous step of the algorithm is used to estimate a transform \mathcal{T} between the two images. This transform is not accurate enough, because its parameters are computed on features that are bound to discrete image grid positions. As mentioned earlier, some feature pairs are additionally not located at the exact corresponding spatial positions. The resulting set may therefore contain outliers and inaccuracies that bias the transform in a negative way. In the following, Θ and \mathcal{C} are refined in a sub-pixel accurate iterative process in order to achieve a more accurate alignment.

The goal of this step consists of the optimization of a joint correspondence set $\mathcal{J} = \{c_{i_1,j_1}, c_{i_2,j_2}, \dots, c_{i_n,j_n}\}$, with $\mathcal{J} \subseteq \mathcal{C}$ and $n \leq M$ that contains sub-pixel accurately aligned feature pairs and ideally no outliers. The elements of the optimized joint correspondence set are used as input for an ICP algorithm in order to compute a transform that maximizes the global image similarity:

$$\hat{\mathcal{J}} = \underset{\mathcal{J}}{\operatorname{argmax}} \mathcal{L}_{global}(\mathcal{J}) = \underset{\mathcal{J}}{\operatorname{argmax}} ECC(I_r, I_t^{T_{\mathcal{J}}}) \quad (9)$$

In order to keep the number of feature pairs low and the registration efficient, an EM-type algorithm is used with a limited number of iteration steps. A transform $\mathcal{T}_{\mathcal{J}_k}$ is computed from a gradually refined joint correspondence set $\mathcal{J}_k \subset \mathcal{J}$ at each iteration. \mathcal{L}_{global} is used as a convergence criterion for the refinement process. The algorithm (illustrated in figure 3) is initialized with a joint correspondence set containing the two topmost pairs of \mathcal{C} : $\mathcal{J}_0 = \{c_{i_1,j_1}, c_{i_2,j_2}\}$. For these initial correspondences, usually the best two from the ordered set of feature pairs acquired in the previous step can be used.

Compared to the 2D algorithm of Huang et al.,¹⁶ all salient region feature centers in the current joint correspondence set have to be re-adjusted for each iteration step. A reason for this is the subsequent resampling

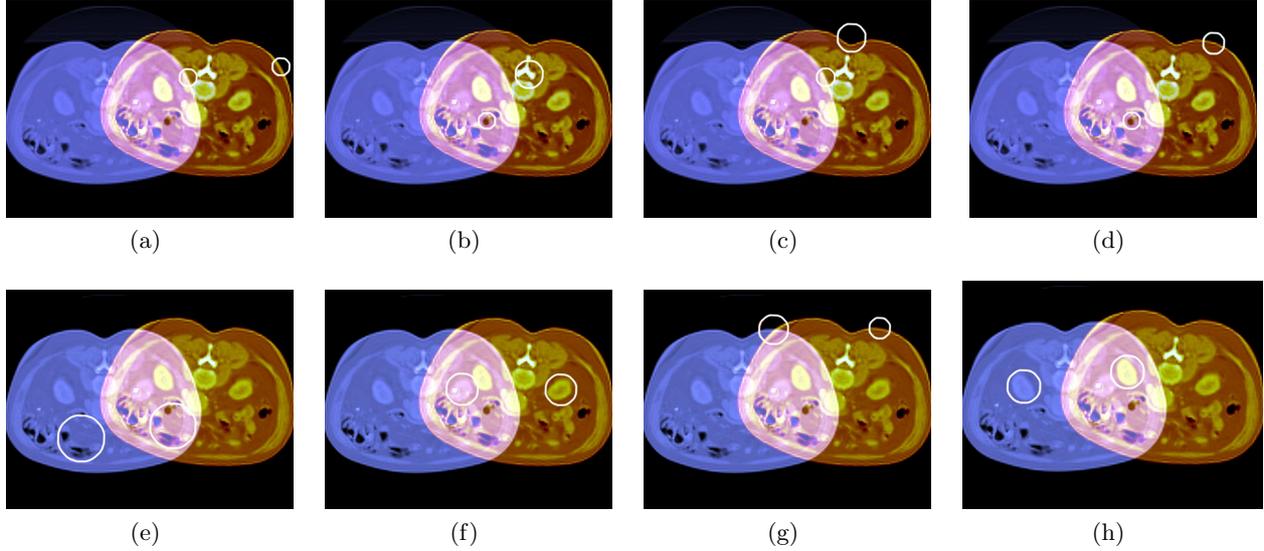


Figure 2. The figures show a slice of a CT volume (left part) that has been translated, rotated and overlaid onto the original slice. A circle represents a salient feature region with a specific scale. The top row shows the four best hypothesized correspondences after ordering according to \mathcal{L}_{local} . Only the first pair actually is a true correspondence. In the second row, the ordering has been performed according to \mathcal{L}_{global} and the entire set contained fewer outliers. For the sake of clarity, only the first four correspondences for each approach are shown.

register feature pairs in \mathcal{J}_0 with sub-pixel accuracy to refine centers	
let \mathcal{J}^* be the current set of sub-pixel refined feature correspondences	
E-STEP:	$\forall c_{i,j} c_{i,j} \in \mathcal{C} \wedge c_{i,j} \notin \mathcal{J}^*$
	compute $\mathcal{L}_{global}(\mathcal{J}^* \cup c_{i,j})$
M-STEP:	$c_{\hat{i},\hat{j}} = \max_{c_{i,j}} \mathcal{L}_{global}(\mathcal{J}^* \cup c_{i,j})$
IF	$\mathcal{L}_{global}(\mathcal{J}^* \cup c_{i,j}) \leq \mathcal{L}_{global}(\mathcal{J}^*)$
THEN	RETURN: transform $\mathcal{T}_{\mathcal{J}^*}$
ELSE	register feature pair $c_{\hat{i},\hat{j}}$ with sub-pixel accuracy to refine centers
	add the refined pair $c_{\hat{i},\hat{j}}^*$ to \mathcal{J}^* :
	$\mathcal{J}^* \leftarrow \mathcal{J}^* \cup c_{\hat{i},\hat{j}}^*$
	recompute $\mathcal{T}_{\mathcal{J}^*}$
UNTIL convergence criterion is met	

Figure 3. The EM-type algorithm for the optimization of both the joint correspondence set and the registration transform.

of the align image during the registration process, which results in subtle intensity changes within the feature regions. The re-adjustment is necessary for that method in order to maintain a high degree of accuracy. Unlike the 2D approach, the proposed 3D algorithm 3 does not need a subsequent resampling and is therefore more efficient. Once a region feature pair has been locally registered with sub-pixel accuracy, following registrations of this specific pair do not enhance the quality of this correspondence and are neglected. Computation time can therefore be saved by only refining the iteratively added feature pair locations during each iteration step.

4. RESULTS

The proposed algorithm has been tested on various intra-patient 3D medical images. The measurements have been performed on 11 PET-CT volume pairs that were acquired at different times, 3 CT volumes at different stages of the treatment, and 10 SPECT-CT volume pairs from a hybrid scanner (Siemens Symbia). The algorithm

had to compete with different modalities, noise, varying fields of view and image intensity artefacts in some of the PET-CT pairs, where some slices had different intensity scales that were not corrected during the import. The PET-CT and CT-CT registration quality was assessed by a medical expert by measuring the distances between several points of interest: lung right and left apex, cardiac apex, liver round end, left and right upper and left and right lower renal ends, see tables 1 and 2. As the 10 SPECT-CT images have been acquired by a state-of-the-art hybrid scanner, the physician manually de-registered the SPECT images rigidly with variations in x , y and z -direction from 10 to 50 mm and rotations around each axis ranging from 5 to 60 degrees. After registration, several distinguishable landmarks have been chosen by the medical expert on the CT and the SPECT images. The distances between the anatomical structure and the information in the SPECT are shown in table 3.

PET-CT Distances	x	y	z
Lung right apex	-0.54 ± 0.70	0.41 ± 0.92	-1.21 ± 1.84
Lung left apex	0.06 ± 0.38	-0.11 ± 0.68	-1.29 ± 2.18
Cardiac apex	0.88 ± 0.80	1.34 ± 0.60	0.48 ± 1.86
Liver round end	0.47 ± 0.77	0.84 ± 0.72	1.06 ± 1.83
Upper right renal end	-0.65 ± 0.78	-0.04 ± 0.80	0.50 ± 2.74
Upper left renal end	0.23 ± 0.74	0.01 ± 0.77	0.35 ± 1.75
Lower right renal end	0.12 ± 0.46	0.47 ± 1.30	1.92 ± 2.71
Lower left renal end	-0.47 ± 0.77	0.86 ± 1.50	1.61 ± 1.71

Table 1. Measured distances after registration for the PET-CT volume pairs in x , y and z -direction given in cm along with the standard deviation.

CT-CT Distances	x	y	z
Lung right apex	-0.28 ± 0.39	0.16 ± 0.13	-0.16 ± 0.18
Lung left apex	-0.25 ± 0.41	0.21 ± 0.24	-0.06 ± 0.14
Cardiac apex	-0.20 ± 0.51	0.52 ± 0.41	0.22 ± 0.31
Liver round end	0.11 ± 0.29	0.00 ± 0.73	0.37 ± 0.93
Upper right renal end	-0.30 ± 0.42	-0.12 ± 0.25	0.21 ± 1.24
Upper left renal end	0.01 ± 0.33	0.28 ± 0.44	0.15 ± 0.67
Lower right renal end	-0.5 ± 0.46	0.02 ± 0.87	0.89 ± 0.74
Lower left renal end	0.01 ± 0.23	-0.14 ± 0.40	0.46 ± 0.44

Table 2. The overall measured distances after registration for the CT-CT volume pairs in x , y and z -direction given in cm along with the standard deviation.

SPECT-CT Distances	x	y	z
Spine	0.07 ± 0.30	0.09 ± 0.23	-0.04 ± 0.22
Kidneys	0.01 ± 0.26	0.13 ± 0.50	0.03 ± 0.21
Others	0.04 ± 0.06	0.27 ± 0.37	0.13 ± 0.22

Table 3. The overall measured distances after registration for the SPECT-CT volume pairs in x , y and z -direction given in cm along with the standard deviation.

All experiments have been executed on real medical images that sometimes contained a high amount of noise or artefacts due to variations of the intensity scaling between the slices. We did not address these issues prior to the registration in order to test the algorithm with such data. Three slices taken from registration results are depicted in figure 4. The results were assessed by a medical expert using dedicated visualization software



Figure 4. Slices from fused registered result images. The images show the result of the proposed algorithm for a PET-CT 4(a), a CT-CT with intensity artefacts 4(b) and a SPECT-CT image pair 4(c). Although the latter CT image was acquired with a limited field of view and contains a high amount of noise, the proposed registration resulted in an acceptable accuracy. The remaining mis-registration may be addressed with a non-rigid transform model in future work.

(InSpace*) for the visualization and the measurements. For the evaluation, the medical expert had the choice between using centroids of 3D regions of interest and direct landmark to landmark positions. This task was supported by the integration of fusion visualization into the rendering software together with some additional measurement tools. In the PET-CT case, a higher standard deviation in the z -direction is evident. A reason for this may result from the differences between the acquisition models. The CT image shows one respiration snapshot, whereas the PET image is acquired over several respiration cycles and depicts more or less an average respiratory motion. Due to this motion of the diaphragm, some organs in the abdominal region are lifted and lowered, which causes the greater deviation seen in the data samples. Right now, the algorithm models only rigid transforms and does not cover such local deformations. For the CT-CT data, this effect is no longer dominating as the patient ideally inhales similarly in both acquisitions. The SPECT-CT data matches inherently well and a user-defined rigid transformation on the SPECT does not introduce local deformations. Therefore, a good registration result for these cases may well be expected.

In all the results, a specific measurement error is introduced because the medical expert has to specify the location manually by selecting the locations in the various slice views. In conducted experiments on this type of evaluation, however, the mean difference of specifying distances of points of interests in several measurement steps (both inter- and intra-observer) did not exceed 3 mm.²³

5. SUMMARY

The proposed algorithm demonstrates that salient region features are well suited not only for 2D, but also for 3D registration purposes. Valuable information about corresponding regions is obtained due to a robust optimization of a joint correspondence set. This optimization is based not only on the local intensity similarity between the feature regions based on the entropy cross correlation. We use the information gained from a global measure as well in order to remove outliers and order the correspondences more efficiently. Regarding the well-ordered and sub-pixel accurate salient feature correspondences as a set of points, the ICP algorithm has been used in order to compute the final registration transform. The accuracy of the registration has been evaluated by a medical expert and our experiments have shown that the registration is robust to the amount of overlap between the images. Varying coordinate systems between CT, SPECT and PET acquisition system may even lead to no initial overlap in some cases. The proposed algorithm does not rely on any form of pre-translation or pre-rotation and uses only the ICP transform between the salient feature regions as initialization for the optimization. The

*A commercial clinical 3D visualization package developed jointly by HipGraphics and Siemens

algorithm has been integrated into a clinical application prototype for robust initial registration between images of various modalities. Local mis-registrations may still occur due to the rigid nature of the transform model.

The extraction and the search for correspondences between salient features are the most time consuming aspect of the algorithm. Currently, the whole registration of a $512 \times 512 \times 512$ volume requires up to 15 minutes (including pre-processing) and solely intensity based approaches may solve the task of rigid registration in a few seconds. Nonetheless, there is a high potential for improvement concerning the runtime of the algorithm. The feature extraction that is currently the most computational demanding task is well suited for parallelization. In addition, the computation of the local correspondences and the ordering of the salient feature region pairs may be significantly improved by a faster similarity measure computation and optimization. Regarding the search space reduction for the joint correspondences, the nearest neighbor criterion for the pair combinations may not always be the best guess. It has to be investigated whether the scale as an additional hint for the correspondence may be exploited as well. Our future work will therefore concentrate on improving the performance and including the information about the feature correspondences into a non-rigid registration framework. Commonly, a rigid registration is performed before a non-rigid one as an initialization step. Using our approach, the resulting robust local correspondences between intensity regions of the two images can be directly used as additional information. This might for instance be utilized in form of an additional energy term within a common non-rigid registration functional.

6. ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG), grant number SFB, TP C10. The authors are also thankful to HipGrapics for providing the volume rendering software InSpace and Siemens Corporate Research (SCR) for both the FusionToolKit (FTK) and support.

REFERENCES

1. J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging* **22**(8), pp. 986–1004, 2003.
2. J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, *Medical Image Registration*, CRC Press, Boca Raton FL, USA, 2001. The Biomedical Engineering Series.
3. D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology* **46**(3), pp. R1–R45, 2001.
4. J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis* **2**(1), pp. 1–36, 1998.
5. S. Gilles, *Robust description and matching of images*. PhD thesis, University of Oxford, 1998.
6. T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision* **45**(2), pp. 83–105, 2001.
7. T. Kadir, D. Boukerroui, and M. Brady, "An analysis of the scale saliency algorithm," tech. rep., Technical Report OUEL No: 2264/03, University of Oxford, 2003.
8. T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *Proceedings of the 8th European Conference on Computer Vision, ECCV 1*, pp. 228–241, (Prague, Czech Republic), 2004.
9. F. Fraundorfer and H. Bischof, "Detecting distinguished regions by saliency," in *Image Analysis, 13th Scandinavian Conference, SCIA 2003*, pp. 208–215, Springer-Verlag, (Berlin, Germany), 2003.
10. F. Fraundorfer and H. Bischof, "Utilizing saliency operators for image matching," in *Proceedings International Workshop on Attention and Performance in Computer Vision*, pp. 17–24, (Graz, Austria), 2003.
11. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Incorporating connected region labelling into automated image registration using mutual information," in *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 23–31, (San Francisco CA, USA), 1996.
12. J. Modersitzki and B. Fischer, "Optimal image registration with a guaranteed one-to-one point match," in *Bildverarbeitung für die Medizin*, T. Wittenberg, P. Hastreiter, U. Hoppe, H. Handels, A. Horsch, and H.-P. Meinzer, eds., *Informatik aktuell* **80**, pp. 1–5, Springer Verlag, (Erlangen, Germany), 2003.

13. J. West, C. R. Maurer, and J. R. Dooley, "Hybrid point-and-intensity-based deformable registration for abdominal CT images," in *Proceedings of SPIE on Medical Imaging*, J. M. Fitzpatrick and J. M. Reinhardt, eds., **5747**, pp. 204–211, (San Diego CA, USA), April 2005.
14. R. Rohr, P. Cathier, and S. Woerz, "Elastic registration of gel electrophoresis images based on landmarks and intensities," in *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, pp. 1451–1454, (Washington DC, USA), 2004.
15. X. Wang and D. D. Feng, "Hybrid registration for two-dimensional gel protein images," in *Proceedings of the 3rd Asia-Pacific Bioinformatics Conference*, Y.-P. P. Chen and L. Wong, eds., **1**, pp. 201–210, (Singapore), January 2005.
16. X. Huang, Y. Sun, D. Metaxas, F. Sauer, and C. Xu, "Hybrid image registration based on configural matching of scale-invariant salient region features," in *IEEE Workshop on Image and Video Registration*, (Washington DC, USA), July 2004.
17. G. Wu, F. Qi, and D. Shen, "Learning best features for deformable registration of MR brains," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, J. S. Duncan and G. Gerig, eds., *Lecture Notes in Computer Science* **3750**, pp. 179–187, Springer, (Palm Springs CA, USA), October 2005.
18. J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM (CACM)* **18**(9), pp. 509–517, 1975.
19. J. L. Bentley and J. H. Friedman, "Data structures for range searching," *ACM Computing Surveys* **11**(4), pp. 397–409, 1979.
20. J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software* **3**(3), pp. 209–226, 1977.
21. F. Maes, *Segmentation and registration of multimodal medical images: From theory, implementation and validation to a useful tool in clinical practice*. PhD thesis, Catholic University of Leuven, 1998.
22. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition* **32**(1), pp. 71–86, 1999.
23. A. Nömayr, W. Römer, T. Hothorn, A. Pfahlberg, J. Hornegger, W. Bautz, and T. Kuwert, "Anatomical accuracy of lesion localization by retrospective interactive rigid image registration between F-18-FDG-PET and X-Ray CT," *Nuklearmedizin* **44**(4), pp. 149–155, 2005.