

# Utilizing Salient Region Features for 3D Multi-Modality Medical Image Registration

Dieter Hahn<sup>1</sup>, Gabriele Wolz<sup>2</sup>, Yiyong Sun<sup>3</sup>, Frank Sauer<sup>3</sup>, Joachim Hornegger<sup>1</sup>, Torsten Kuwert<sup>2</sup> and Chenyang Xu<sup>3</sup>

<sup>1</sup>Department of Pattern Recognition, Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Martensstr. 3, 91058 Erlangen, Germany

<sup>2</sup>Department of Nuclear Medicine, FAU, Krankenhausstr. 12, 91054 Erlangen, Germany

<sup>3</sup>Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA  
Email: dieter.hahn@informatik.uni-erlangen.de

**Abstract.** Besides common solely intensity- or feature-based image registration methods, hybrid approaches make use of two or more image properties. A 3D representation of *saliency* can be used to automatically locate distinct region features within two 3D images and establish a robust and accurate hybrid registration method. The extracted features contain information about the underlying saliency, the scale of the regions and enclosed voxel intensities. Similar anatomical or functional content results in similar salient region features that can be used to estimate an image transform based on corresponding feature pairs. The refinement of this estimate results in a robust and sub-pixel accurate set of joint correspondences. An evaluation by a medical expert on various clinical data using a medical application demonstrates that the approach is robust to image overlap, artefacts and different fields of view.

## 1 Introduction

Registration has grown to a major research topic in the field of medical image processing [1]. A major medical benefit can be withdrawn from the registration of images acquired with different modalities that can roughly be divided into two major categories: anatomical imaging (e.g. Computed Tomography (CT), Magnetic Resonance Imaging (MRI)) to retrieve the morphology in good spatial resolution and functional imaging (e.g. Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT)) to depict the underlying metabolism. Mono-modality registration approaches are for instance valuable for therapy control or tumor change diagnostics.

The concept of saliency can be used to define a measure for the local uniqueness of image elements based on the intensity complexity in a local neighborhood [2,3,4]. Applications of this descriptor range from content tracking across subsequent video frames, image matching for 3D reconstruction [5] to 2D image registration [6]. Salient image regions provide a hybrid source of information, as they describe the uniqueness of the underlying image content based on intensity

values and contain geometrical information of the region center and its scale. In the following, we describe a registration method that makes use of these properties for establishing correspondences between extracted features from a reference and a template image, which finally leads to a registration transform.

## 2 Methods

The proposed registration approach is based on finding corresponding salient region features. A feature in this context consists of a center position and a radius of a spherical neighborhood that describes a local saliency maximum. This radius is denoted as scale. According to [4], the saliency  $\mathcal{A}$  is defined as follows:

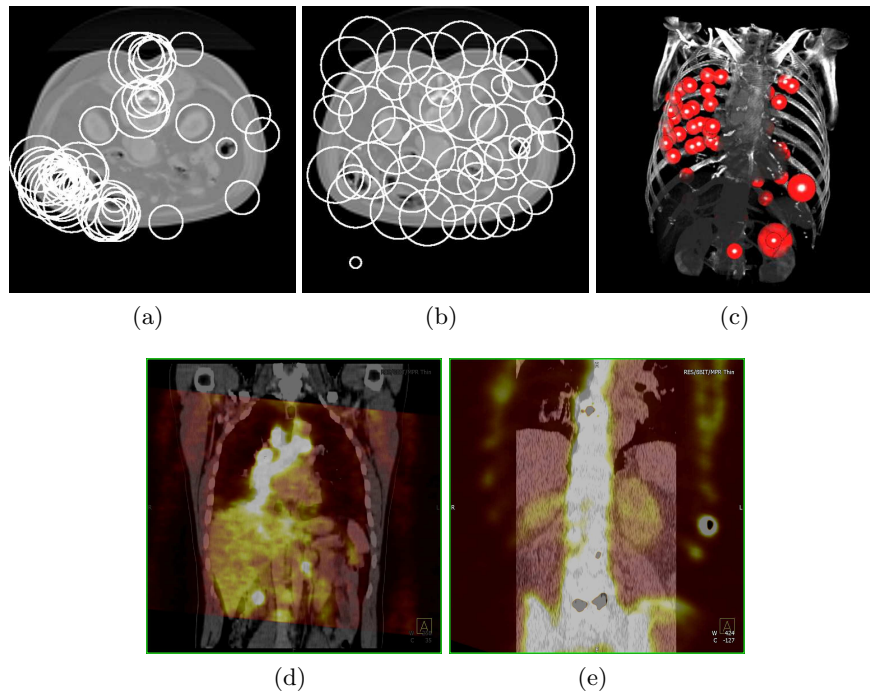
$$\mathcal{A}_D(s, \mathbf{x}) = - \int_{\mathcal{R}_s} p(i, s, \mathbf{x}) \log_2 p(i, s, \mathbf{x}) \, di \cdot s \int_{\mathcal{R}_s} \left| \frac{\partial}{\partial s} p(i, s, \mathbf{x}) \right| \, di \quad (1)$$

including the entropy with respect to the image intensity values  $i \in D$  within a spherical neighborhood region  $\mathcal{R}_s$  of scale  $s$  around a voxel  $\mathbf{x}$ . Here,  $p(i, s, \mathbf{x})$  is the probability density function (PDF) of  $i$  for  $\mathcal{R}_s$ .  $\mathcal{A}$  also contains a term for the similarity between PDFs with respect to the scale. Its values grow with increasing dissimilarity of the PDFs. The scale parameter  $s$  can be approximated by the scale that results in a local peak of the entropy for the region around  $\mathbf{x}$ .

Voxels containing high saliency values, i.e. high local unpredictability or local signal complexity, have to be extracted after computation of the eq. 1. Based on algorithms published in related literature [4,6], we added a region growing approach to find local maxima within the saliency values, however, in certain areas with high intensity variations, feature clustering is likely to occur. This may have a negative impact on the correspondence search. For instance if two images from different modalities with a different distribution of intensities are considered, one image may contain a dominant salient feature cluster in the upper right, the other one in the lower left part of the image but describing a totally different region of the body. A minimization of the mean square error (MSE) between the two feature sets then tends to align mainly the clusters. This is a problem that we address with a simple and efficient algorithm, because a set of features that is distributed more uniformly across the image domain provides a better initialization for the correspondence set optimization. A nearest point algorithm based on a kD-tree [7,8] is used for the storage of the feature centers of one set. The nearest neighbors of a specific element can therefore be retrieved efficiently. The scale parameter is used as a minimal distance requirement: all returned features within a distance to the query feature that is equal or less than its scale and with lower saliency are removed from the set. The list is padded with features of lower saliency in order to maintain a specific size of the set. After enforcing the restriction, the features are distributed more uniformly across the image. Examples of some of the results are given in figure 1.

Unlike a similar version of the algorithm[6], we reduce the search space prior to the correspondence computation. Initially, an iterative closest point (ICP) algorithm is used to estimate the transform that minimizes the MSE between the

**Fig. 1.** The images show the set of salient feature regions before 2(a) and after 2(b) clustering removal, the most significant 3D salient region features extracted from a CT 2(c) (windowed for better visualization), a registration result on a PET-CT 2(d) and a SPECT-CT volume pair 2(e).



region centers of the reference and the template feature sets. This transform is applied to the template features that are stored in a kD-tree afterwards. For each reference image feature, only the nearest transformed template feature neighbors are regarded as possible candidates for correspondences. Feature pairs with a large spatial distance apart are unlikely corresponding and can be removed from the search space. In order to quantify the correspondence between two features, the translation and rotation invariance properties are used. Translation invariance is given by the alignment of the region centers and the rotation parameters are achieved by a local rigid-body registration. The optimization is driven by the entropy correlation coefficient (ECC) [9] that is a specific form of normalized mutual information. Instead of ordering the correspondences according to local ECC values of the regions, we use the global ECC for each hypothesized correspondence between the reference and the template image ( $I_r, I_t$ ). This measure is computed for the parameter set  $\hat{\Theta}_{i,j}$  that is given by the translation and

**Table 1.** The overall measured distances in cm between the manually chosen landmarks for the PET-CT, CT-CT and SPECT-CT volume pairs in  $x$ ,  $y$  and  $z$ -direction along with the standard deviations.

PET-CT Distances	$0.013 \pm 0.68$	$0.47 \pm 0.91$	$0.43 \pm 2.08$
CT-CT Distances	$-0.17 \pm 0.38$	$0.12 \pm 0.43$	$0.26 \pm 0.58$
SPECT-CT Distances	$0, 04 \pm 0.21$	$0.16 \pm 0.37$	$0.04 \pm 0.22$

rotation invariance of each hypothesized feature pair  $c_{i,j}$ .

$$\mathcal{L}_{global}(c_{i,j}) = ECC(I_r, I_t^{\mathcal{T}_{\hat{\theta}_{i,j}}}) \quad (2)$$

The previous steps result in a correspondence estimate with less outliers. The final step of the algorithm consists in the sub-pixel accurate refinement of the feature pairs, which is achieved by an expectation maximization type algorithm that optimizes the correspondence set with respect to the global ECC and removes remaining outliers.

### 3 Results

The proposed registration approach has been evaluated on 11 PET-CT volume pairs acquired at different times, 3 CT volumes at different stages of the treatment and 10 SPECT-CT volume pairs from a hybrid scanning system (Siemens Symbia). The SPECT images have been artificially rotated and translated as they are initially very well registered. The algorithm had to compensate intensity artefacts, due to an uncorrected intensity scaling between several slices of some of the CTs, noise and different fields of view. The registration quality was assessed by a medical expert who measured the distances between several easy to locate points of interest: the apex of the left and right lung, cardiac apex, liver round end, left and right upper and lower renal ends and spine. The manual rigid transformation of the SPECT images varied between 10 to 50 mm translation in  $x$ ,  $y$  and  $z$ -direction together with a rotation around each axis of 5 to 60 degrees. The overall results of the evaluation are summarized in table 1.

For the evaluation, the medical expert could choose between using centroids of 3D regions of interest or direct landmark to landmark measurements in a medical application (InSpace). In the PET-CT results, a higher standard deviation in the  $z$ -direction is apparent. Reasons for this may be found in the longer acquisition time for the PET with arms-up, several respiratory cycles and the according movement of the diaphragm.

### 4 Discussion and Conclusion

We have demonstrated that salient region features are suitable for 3D registration. The operations for the extraction and correspondence search are most time

consuming and still require 10 to 15 minutes for high resolution images, however, there has been valuable additional information gained during the whole process: the set of joint feature correspondences that contain information about the extent, the parameters of the local deformation and the local signal complexity. As we used only a rigid-body transform model to evaluate the method, we did not address local deformations. Therefore, a certain amount of variation between the registered anatomical structures is inherent in the results. Like all manual measurements, the results are tainted with a measurement error. A recent study has shown that the overall upper limit for a similar measurement method is expected to be not more than 3 mm[10]. This has to be considered when drawing conclusions from the results. In future work, we will explore how the properties of the salient regions can be exploited for a non-rigid registration.

## 5 Acknowledgments

The authors acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG), grant number SFB, TP C10, HipGraphics for providing the medical application InSpace and Siemens Corporate Research for the Fusion-ToolKit (FTK).

## References

1. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* **22** (2003) 986–1004
2. Gilles, S.: Robust description and matching of images. PhD thesis, University of Oxford (1998)
3. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. *ECCV* **1** (2004) 228–241
4. Kadir, T., Brady, M.: Saliency, scale and image description. *International Journal of Computer Vision* **45** (2001) 83–105
5. Fraundorfer, F., Bischof, H.: Utilizing saliency operators for image matching. In: *Proceedings International Workshop on Attention and Performance in Computer Vision*. (2003) 17–24
6. Huang, X., Sun, Y., Metaxas, D., Sauer, F., Chenyang, X.: Hybrid image registration based on configural matching of scale-invariant salient region features. *Computer Vision and Pattern Recognition Workshop, Conference on* **1** (2004) 167–167
7. Bentley, J.L., Freidman, J.H.: Data structures for range searching. *ACM Computing Surveys* **11** (1979) 397–409
8. Freidman, J.H., Bentley, J.L., Finkel, R.A.: An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software* **3** (1977) 209–226
9. Maes, F.: 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)

10. Noemayr, A., Roemer, W., Hothorn, T., Pfahlberg, A., Hornegger, J., Bautz, W., Kuwert, T.: Anatomical accuracy of lesion localization by retrospective interactive rigid image registration between F-18-FDG-PET and X-Ray CT. *Nuklearmedizin* **44** (2005) 149–155