Anatomical accuracy of abdominal lesion localization

Retrospective automatic rigid image registration between FDG-PET and MRI

A. Kiefer¹; T. Kuwert¹; D. Hahn²; J. Hornegger²; M. Uder³; P. Ritt¹
¹Klinik für Nuklearmedizin, Universitätsklinikum Erlangen; ²Lehrstuhl für Mustererkennung, Friedrich-Alexander-Universität Erlangen-Nürnberg; ³Institut für diagnostische Radiologie, Universitätsklinikum Erlangen, Germany

Keywords
PET, FDG, MRI. image registration, image fusion

Summary
Software-based image registration can improve the diagnostic value of imaging procedures and is an alternative to hybrid scanners. The aim of this study was to evaluate the anatomical accuracy of automatic rigid image registration of independently acquired datasets of positron emission tomography with ¹⁸F-deoxyglucose and abdominal magnetic resonance imaging.

Patients, methods: Analyses were performed on 28 abdominal lesions from 20 patients. The PET data were obtained using a stand-alone PET camera in 14 cases and a hybrid PET/CT scanner in 9 cases. The abdominal T1- and T2-weighted MRI scans were acquired on 1.5 T MRI scanners. The mean time interval between MRI and PET was 7.3 days (0–28 days). Automatic rigid registration was carried out using a self-developed registration tool integrated in commercial available software (InSpace for Siemens Syngo). Distances between the centres of gravity of 28 manually delineated neoplastic lesions represented in PET and MRI were measured in X-, Y-, and Z-direction. The intra- (intraclass correlation 0.94) and inter- (intraclass correlation 0.86) observer repeatability were high.

Results: The average distance in all MRI sequences was 5.2 ± 7.6 mm in X-direction, 4.0 ± 3.7 mm in Y-direction and 6.1 ± 5.1 mm in Z-direction. There was a significantly higher misalignment in Z-direction (p < 0.05). The misalignment was not significantly different for the registration of T1- and T2-weighted sequences (p = 0.7).

Conclusion: The misalignment between FDG-PET and abdominal MRI registered using an automated rigid registration tool was comparable to data reported for software-based fusion between PET and CT. Although this imprecision may not affect diagnostic accuracy, it is not sufficient to allow for pixel-wise integration of MRI and PET information.

Schlüsselwörter
PET, FDG, MRI. Bildregistrierung, Bildfusion

Zusammenfassung

Ergebnisse: Die mittlere Schwerpunktstheorie für alle MRT-Sequenzen betrug 5,2 ± 7,6 mm in X-Richtung, 4,0 ± 3,7 mm in Y-Richtung und 6,1 ± 5,1 mm in Z-Richtung. Verschiebungen in Z-Richtung waren signifikant größer (p < 0.05). Es gab keine signifikant unterschiedlichen Abstände für die Registrierung der T1- und T2-gewichteten Sequenzen (p = 0,7). Schlussfolgerung: Die Ungenauigkeit der automatischen und starren Bildregistrierung zwischen FDG-PET und abdominalen MRT ist mit denjenigen für die retrospektive Registrierung zwischen PET und CT publizierten Daten vergleichbar. Obwohl diese Ungenauigkeit die diagnostische Zuverlässigkeit wohl nicht einschränkt, ist sie für eine pixelweise Verrechnung der Datensätze nicht ausreichend.

Anatomische Genauigkeit der retrospektiven, automatischen und starren Bildregistrierung zwischen FDG-PET und MRT bei abdominalen Läsionen
Nuklearmedizin 2011; 50: 147–154
doi:10.3413/nukmed-0364
received: October 29, 2010
accepted in revised form: April 28, 2011
prepublished online: May 19, 2011

Correspondence to:
Philipp Ritt
Clinic of Nuclear Medicine
University of Erlangen-Nürnberg
Krankenhausstr. 12, 91054 Erlangen, Germany
Tel. +49/(0)91 31/853 34 11
Fax +49/(0)91 31/853 92 62
E-mail: philipp.ritt@uk-erlangen.de

Downloaded from www.nuklearmedizin-online.de on 2011-08-11 | ID: 1000491814 | IP: 217.110.19.91
Note: Uncorrected proof, prepublished online
For personal or educational use only. No other uses without permission. All rights reserved.
Several groups report a significantly higher diagnostic accuracy of hybrid cameras combining X-ray computed tomography (CT) with single-photon emission computed tomography (SPECT) or positron emission tomography (PET) when compared to that of stand-alone SPECT or PET devices (9, 15, 19, 21, 27). The tremendous clinical and commercial success of these systems has stimulated efforts to develop also hybrid cameras between magnetic resonance imaging (MRI) and PET (2, 12, 33). First data in experimental animals and also in humans have recently proven the technical feasibility of an approach allowing the simultaneous acquisition of MRI and PET data (2, 13, 22, 23, 26). Nevertheless, only few human brain and even less whole-body PET-MRI systems have so far been installed and clinical data proving their superiority (except for some special applications) over PET/CT scanners are still lacking.

Software-based image registration of independently acquired image datasets is an alternative to hybrid scanners (28, 36). It may be more flexible in use and is probably cheaper. However, software-based image registration may lack the anatomical accuracy of its hardware-based counterpart since the position of the patient may vary considerably between two separately performed examinations (15). Furthermore, in the clinical context, registration is usually performed interactively and not in an automated fashion so that its reproducibility may be questioned.

Only few studies have so far investigated the anatomical precision of software-based registration between PET and MRI (8, 30). In this study, we evaluated the anatomical accuracy of automated rigid image registration of PET performed after injection of $^{18}$F-deoxyglucose (FDG) and abdominal MRI for datasets obtained independently from each other in the clinical routine.

Patients, material, methods

Patients

Analyses were performed on independently acquired FDG-PET and abdominal MRI data from 20 patients out of daily clinical routine who fulfilled the following criteria:

- at least one abdominal lesion visualized in the FDG-PET and the MRI,
- MRI not longer than 30 days before or after the FDG-PET,
- no surgical intervention between the two imaging procedures.

20 patients (12 men, 8 women) aged between 32 and 72 years (mean: 53 years) entered the study. Most of the 28 analyzed lesions were located in the liver (25), but also in the pancreas (1), spleen (1), and gallbladder (1). The mean time interval between the two imaging procedures was 7 days, ranging from 0 to 28 days.

Our retrospective study was in concordance with the guidelines of the ethical review board at our institution.

PET data acquisition

The PET image datasets were obtained using a stand-alone PET camera (ECAT Emerge, Siemens Medical Solutions, Knoxville, Tennessee, USA) in 11 cases and a hybrid camera combining a PET scanner with a 64-slice CT in 9 cases (Biograph 64, Siemens Medical Solutions, Knoxville, Tennessee, USA). The slice thickness and the in-plane matrix size/pixel spacing was 5.15 mm and 128×128 pixels (5.15×5.15 mm) for the stand-alone scanner and 3 mm and 168×168 pixels (4.07×4.07 mm) for the PET/CT. The technical performance of the stand-alone PET camera as well as the procedures used for data acquisition and image reconstruction have been described previously (16). The details of data acquisition were similar for the PET/CT system, although lower FDG activities could be used due to the higher sensitivity of its PET component compared to the stand-alone PET camera. All images were iteratively reconstructed (OSEM) and corrected for attenuation. The necessary transmission scans were acquired with Cs-137 rod sources in the case of the stand-alone PET or by use of the CT images in the case of the hybrid scanner.

MRI data acquisition

The patients were recruited out of the daily clinical routine. Therefore the parameters of the scans like slice thickness and the resolution varied considerably. Most T1 weighted scans were performed with a slice thickness of 3 or 5 mm, the T2 weighted with 7.2 mm. In plane resolution varied between 0.74×0.74 mm and 1.56×1.56 mm. The MRIs were either performed with the Magnetom Symphony 1.5 T (n = 12) or the Magnetom Avanto 1.5 T (n = 11) (both Siemens Medical Solutions, Erlangen, Germany).

All MRI scans were performed by using correction techniques for the breathing motion, either by using breath-hold or navigator-echo triggered sequences (Siemens PACE). For the T1 sequences a standard dose of intravenous contrast agent was used (0.1 ml Gadovist per kg body weight).

Image registration and fusion

The registration and fusion of the images were carried out using the InSpace application developed by HiPgraphics for the Siemens Syngo platform. PET and MRI volume images were simultaneously loaded in InSpace and displayed as a fused volume. Windowing and colour tables could be changed individually as well as the amount of opacity which the single volumes contribute to the fused image. The registration was embedded in a plug-in for InSpace. The plugin extends InSpace with the necessary user interface and algorithms in order to allow a fully automatic registration. The registration itself is realized as a rigid registration and therefore has 6 degrees of freedom (3 translations, 3 rotations). It is based on the pixel intensities of the image and the objective function (which is used as a measure for the quality of the registration) was the mutual information. The objective function is iteratively optimized by a hill climbing algorithm. The software incorporates a multi-resolution approach in order to speed up computational time. Further details can be found in Hahn et al. (11). All computations were carried out on machines with a Pentium IV 2.6 GHz CPU with 2 GB of RAM. The computational time on this platform was typically in the range of 5–10 seconds. The results of the automatic registration could be stored in a matrix and could be reloaded later for the fusion and re-evaluation of the same datasets.

Nuklearmedizin 4/2011 © Schattauer 2011
For the fusion display, a 10 step spectrum colour table was used for all PET images. The 10 step spectrum is intended to ensure a more standardized and reproducible delineation of the contours, which would be more difficult on a continuous look-up-table. Centre and width of the colour table where adjusted in such a way that the background and the healthy liver parenchyma where displayed in the same colours for all patients. Subsequently by using this method, we were able to delineate an iso-contour of every lesion at an activity value of 120% of the difference between background and healthy liver parenchyma for hot lesions and 80% for the delineation of cold lesions.

In case of the MRI scans a grey scale table was used: Centre and width were kept constant at 158 and 231 throughout the whole evaluation, for both, T1 and T2 sequences.

**Measurement**

**Distances**

The evaluation of PET/MRI misalignment was also performed using the InSpace platform by one physician and one physicist experienced in radiology and nuclear medicine. Our evaluation software allowed manual delineation of the contours of the lesions in three perpendicular (transversal, coronal, and sagittal) slices only, in both the MRI (Fig. 1a) and PET (Fig. 1b) images.

The images were evaluated in a 3-views MPR window, showing one transversal, one sagittal and one coronal cutting plane of the dataset. The planes were by default oriented perpendicular to each other (no oblique planes), thus they had one common point of intersection.

The manual steps in the measurement were as follows:

1. The MRI images were masked: Only the PET images are visible.
2. The intersection point was moved to the centre of the lesion in the PET image.
3. The lesion was delineated in one transversal, one coronal, and one sagittal plane of the PET image (Fig. 1b).
4. The PET image was masked, the MRI image was unmasked: Only the MRI image is visible.
5. The intersection point was moved to the centre of the lesion in the MRI image.
6. The lesion was delineated in one transversal, one coronal, and one sagittal plane of the MRI image (Fig. 1a).

Subsequently, the coordinates of the centre of gravity for every lesion were calculated automatically by the software on the basis of the delineation. In contrast to an approach where whole volumes of the lesions are segmented in every slice, our method delivered approximated coordinates for the centre of gravity of the lesions. Finally, the distances $d_x$, $d_y$, $d_z$ between the coordinates of the centres of gravity for MRI and PET images were automatically calculated and displayed by the program in.
The total distance (target registration error, TRE) between the lesions was determined by calculating the Euclidean distance out of the distances measured in the three dimensions (Fig. 1c) according to eq. (1).

\[
\text{TRE} = \sqrt{d_x^2 + d_y^2 + d_z^2}
\]  

(1)

The whole process was repeated with the saved registration result by the same reader for all patient images on separate days to gain insight on the repeatability of the manual delineation and evaluation. Another reader performed the evaluation separately, again using the saved registration result. The procedure itself was rather quick to perform (1–2 minutes per lesion).

**Lesion size**

Additionally we estimated the size of the lesions. We measured the extent \(s_x, s_y, s_z\) of each lesion with the same colour table and the same window adjustments as described in the previous two paragraphs in X- (R→L), Y- (A→P), and Z- (F→H) direction separately for the T1 MRI, T2 MRI, and the PET image. Subsequently we calculated the volume of each lesion in every modality by using eq. (2): We assumed that the lesions are of spheroidal shape.

\[
\text{volume} = \frac{4}{3} \times \pi s_x s_y s_z
\]  

(2)

**Statistical analysis**

The measurement results underwent the following statistical tests; in general, we considered p-values < 0.05 as significant.

We used a Friedman test for examining significant differences between the sets of absolute values of the X-, Y-, and Z- misalignment. If a significant difference was found, a following Nemenyi test was used to identify the differing measurement set.

In order to test for differences between the misalignment-measurements for T1 and T2 weighted sequences, we used Wilcoxon’s signed-rank test.

The intra-observer repeatability and inter-observer reproducibility were estimated by calculating intraclass correlation (two way mixed on single measures) between the separate measurement sets, each consisting of the directed differences in X-, Y-, and Z-direction. For reference, Pearsons’ correlation coefficients were calculated as well.

The approximated lesion volumes for the different modalities were tested for significant differences by using again a combination of Friedman and Nemenyi tests.

**Results**

The average distance for all MRI sequences was 5.2 ± 7.6 mm in X-direction, 4.0 ± 3.7 mm in Y-direction and 6.1 ± 5.1 mm in Z-direction (Fig. 2c). The average distance in the T1-weighted sequences was 5.0 ± 7.0 mm in X-direction, 3.7 ± 3.5 mm in Y-direction and 5.9 ± 5.0 mm in Z-direction (Fig. 2a). In the T2-weighted sequences the average distance was 5.4 ± 8.4 mm in X-direction, 4.4 ± 4.0 mm in Y-direction and 6.4 ± 5.4 mm in Z-direction (Fig. 2b). This leads to a TRE...
of 10.4 ± 8.4 mm for all sequences, a TRE of 9.8 ± 8.0 mm for T1-, and 11.0 ± 9.0 mm for T2 sequences.

The 95% confidence intervals for the misalignment were found to be
- 3.0 mm – 7.3 mm for the X-direction,
- 2.9 mm – 4.0 mm for the Y-direction,
- 4.7 mm – 7.6 mm for the Z-direction,
- 8.0 mm – 12.7 mm for the target registration error (TRE).

The relatively broad intervals indicate a high inter-patient variation of the misalignment. Figure 3 provides an overview for the 95% confidence intervals.

The Friedman test revealed a significant difference of the misalignment in the three spatial directions. The following Nemenyi test revealed a significant higher misalignment in Z-direction. The corresponding p-value was <0.001 for both, T1 and T2 weighted sequences.

With an associated value of p > 0.7, the Wilcoxon signed-rank test showed no evidence that the misalignment between PET/T1-MRI and PET/T2-MRI differ significantly.

The intraclass correlation was 0.94 (95% confidence interval: 0.92 – 0.96) for the intra-observer repeatability and 0.86 (95% confidence interval: 0.82 – 0.90) for the inter-observer reproducibility.

The Pearson correlation coefficient was r = 0.94 for the intra-observer repeatability and r = 0.87 for inter-observer reproducibility.

The volumes of the lesions (Fig. 4):
- 18.6 ± 31.0 cm³ for T1-weighted MRI,
- 24.2 ± 30.8 cm³ for T2-weighted MRI,
- 43.0 ± 51.3 cm³ for the PET images.
The volume in the PET scan was significantly higher than that of the T1- and T2-MRI.

**Discussion**

First of all, we found that T1 and T2 weighted images might be equally suited for retrospective rigid registration, at least for our setup. The negative result of the Wilcoxon signed rank test emphasizes this fact. It is in good concordance with the experience that registration results based on mutual information as a measure of distances are more sensitive to changes of the shape of objects than to changes of their pixel value (image intensity).

Most importantly, we found a significant misalignment for the automated retrospective registration between MRI and FDG-PET of abdominal lesions. Statistical tests revealed a higher misalignment in Z-direction than in X- and Y-direction. This is probably due to respiratory motion of abdominal organs, which was also indicated by other authors and for other combinations of modalities (10, 29, 30, 34, 35). When compared to other body regions like the skull, the abdominal region is more deformable.

In general, the misalignment was expected, if one keeps in mind that the MRI acquisitions were done in a single respiratory phase, either by using breath-hold protocols or triggered imaging, whereas the PET images were acquired under free breathing. An acquisition of the MRI images under free breathing is usually no option, as this would result in severe artefacts. The comparison of our results to those published in the literature showed that the magnitude of the TRE between PET and MRI is in the range of the results reported for retrospective registration between PET and CT (15, 34) but not as good as for hybrid machines (5, 15) (Fig. 5). Furthermore, the accuracy of retrospective registration between PET and MRI in sarcoma patients was not considerably different from ours (30) (Fig. 6). A more accurate registration might be achieved by using the CT component of a PET/CT scanner (30, 31). However, a recent study by Donati et al. using the CT component for the retrospective registration of PET to MRI reported TRE in the range of our study (8).

The relevance of the misalignment finally depends on the application of the co-registered PET and MR images.

Another study by Donati et al. (7) states that the diagnostic confidence is increased by the retrospective registration and fusion of PET and MRI. However, due to the already high sensitivity/specificity of MRI, this increase in confidence stays below statistic.
significance in this study. Still, it is commonly accepted that side-by-side or fused viewing of co-registered PET and MR images could improve the diagnostic accuracy, even despite some misalignment.

This changes when it comes to pixel-wise integration of the information from PET and MR images, like e.g. attenuation correction of the PET images by help of the MR acquisition.

The problem is in analogy to PET/CT: The PET images, due to their duration, are acquired under free breathing of the patient, whereas the CT images are acquired in a single respiratory phase for diagnostic reasons. Consequently, various literature references report errors of differing extent for PET attenuation correction by CT. These errors range from slight artefacts that do not affect the diagnostic statement (18) to artefacts that even led to wrong diagnoses (20, 25). The misalignments in these reports are in the same magnitude (10–20 mm) that we observed in our study. We therefore expect that automatic as well as manual retrospective rigid registration of abdominal PET and MRI is too inaccurate, not only for attenuation correction but also for other correction strategies that rely on an exact registration, e.g. variants of partial volume correction (24) and scatter correction (17).

An important reason for hybrid PET/MRI systems is to overcome such problems. We expect that side-by-side PET/MR systems with a common patient bed, but separated gantries will most likely provide better registration results than our approach. Without the need for a repositioning of the patient, the probability for differing patient positions between PET and MRI should be heavily reduced. Fully integrated PET/MRI scanners (common gantry and field-of-view) are expected to decrease those errors even further due to a shorter temporal delay between the acquisitions (simultaneous in optimal case).

However, both approaches are by nature also not able to correct for elastic deformations that occur when the acquisitions are done in different respiratory conditions. Here, additional corrections like dynamic attenuation correction (6), gated PET (14) or non-rigid registration could provide better solutions.

Unfortunately, studies about hybrid PET/MRI whole-body imaging are to our best knowledge still lacking in the literature as those systems have been available for a very short time only.

Limitations

Our study had several limitations: First of all, its poor case number has to be mentioned, it was for example necessary to pool the MRIs into T1 and T2 weighted groups without differentiating other imaging parameters namely fat saturation or imaging sequence. Secondly, we had a certain heterogeneity and variety of lesions and MRI sequences in our patient population. This mainly results out of the fact that all patients and images were re-
cruited out of daily clinical routine in our hospital. Furthermore, it may be asked whether the zones exhibiting high FDG uptake match those with MRI signal abnormalities since both modalities reflect different aspects of tumour pathology (2, 3).

After the registration procedure we noticed that the shape and the extension (Fig. 4) of the lesions in the two imaging procedures were different (the extension is significantly greater in PET) and the anatomical match was not perfect. The most probable explanation is the different windowing technique but this might also be due to the aforementioned effect (FDG-PET and MRI reflect different aspects of tumour pathology). Other explanations like tumour growth or especially deformation due to different reasons (breathing, filling of the stomach) seem reasonable. However, the diagnostic statement of the physicians who evaluated either the FDG-PET (nuclear medicine) or the MRI (radiologist) was in concordance for our cases.

At last, our manual measurement approach of the centres of gravity without any hard gold standard can be questioned. Nevertheless we think that our method is robust; the high intra-observer repeatability and the high inter-observer reproducibility underline this.

Conclusion

The misalignment between FDG-PET and abdominal MRI registered using an automated rigid registration tool was comparable to data reported for software-based fusion between PET and CT. Although this imprecision may not or only to a small degree affect diagnostic accuracy, it is not sufficient to allow for pixel-wise integration of MRI and PET information, like e.g. for attenuation correction.

Conflict of interest

The authors declare, that there is no conflict of interest.

References