Data-Driven Density Estimation applied to SPECT Subtraction Imaging for Epilepsy Diagnosis

Dieter A. Hahn[1], Volker Daum[1,2], Joachim Hornegger[1] and Torsten Kuwert[3]

[1] Chair of Pattern Recognition, Friedrich-Alexander University Erlangen-Nürnberg (FAU)
[2] Institute of Optics, Information and Photonics (Max-Planck Research Group), FAU
[3] Department of Nuclear Medicine, University Hospital, FAU

Abstract. In this article we present a subtraction imaging approach for the assessment of difference in the cerebral blood flow between intra- and inter-ictal SPECT images of epilepsy patients. The workflow consists of a rigid, automatic image registration of the SPECT images, an intensity normalization, and an alignment of the differences with an MRI. For the registration, the statistical measure of normalized mutual information is applied. The probability density estimation is sensitive to the input data, the sampling approach, and the kernels applied for Parzen-windowing. An adaptive scheme to estimate the required parameters is mandatory, because the system has to work reliably for a large number of images. We propose data-driven estimation techniques for a B-spline Parzen-window density estimation that is adapted to the variation within the random measures by an anisotropic binning approach. The optimal kernel widths are determined by a log-likelihood estimation.
The approach has been integrated into a commercially available software and applied to a collective of 26 epilepsy patients. Results are presented for a blind evaluation study with physicians from the Department of Nuclear Medicine of the University Hospital in Erlangen. The results show a good correlation of 81% between the certain outcomes of the proposed workflow and the standard procedure.

1 Introduction

In epilepsy surgery planning, inter- and intra-ictal SPECT images are acquired between epileptic seizures and closely afterwards. The goal is to find the location of the seizure onset by comparing the cerebral blood flow (CBF) inferred from the images. Standard side-by-side visual assessment techniques are not ideal, because the patient position and orientation may have changed between the acquisitions, and the images lack a normalization to standardized intensity values (e. g. as Hounsfield units in CT).

In [1], we introduced a workflow for SPECT epilepsy imaging. The basic idea is to use image subtraction similar to digital subtraction angiography (DSA) in order to visualize the differences between both images. The subtraction image then depicts changes in the CBF between both acquisitions. In order to be able to perform a difference operation on two SPECT images, we have to compensate

for patient motion and variations of the tracer uptakes between the acquisitions. A previously acquired MRI is integrated into the workflow and used to spatially localize the differences. In this article, we refine the concepts of the prior work and extend the registration, as well as the intensity normalization, by automatic, data-driven density estimation techniques.

The similarity measure used in the registration of the SPECT images has to be invariant to intensity variations, changes in blood flow activity, and structured noise. We apply normalized mutual information (NMI) [2] and use parameters that are adapted to the input data. This principally concerns the discrete probability density functions (PDF), which are computed by a novel, quasi-adaptive Parzen-window estimator based on cubic B-spline kernels. Structured noise in the background is handled by an automatic determination of background thresholds for both images, which can afterwards be reused within the intensity normalization. After the registration, we relate the intensities between both images by fitting an affine model into the joint PDF. The final result of the workflow is a subtraction image that can be fused with an MR scan.

We have applied the algorithm to a collective of 26 epilepsy patients and the results of the proposed workflow have been evaluated in a blind study.

2 Related Work

The workflow, presented in Hahn et al. [1], is based on the application of subtraction methods in the context of SPECT imaging. Other approaches make use of statistical parametric mapping, for example Chang et al. [3] and McNally et al. [4]. Their methods, however, require collectives of norm patients in order to identify those variations that are due to the epileptic disease. Koo et al. [5] proposed settings for the visualization of subtraction results in order to achieve a good correlation in the detection of the focal spots with other, well established techniques.

Image similarities can be modeled in a statistical framework. Here, the intensity values are regarded as random measures of an unknown distribution. Using nonparametric density estimation, efficient PDF estimators can be realized by a discretization of the Parzen-window technique [6, 7]. The density estimation is based on the work by Viola [8] and Hermosillo et al. [9]. Knops et al. [10] and Katkovnic and Shumulevich [11] investigated the effects of the kernel width parameter on the estimator and showed that state-of-the-art, isotropic binning is outperformed by adaptive techniques. Depending on the sampling pattern, numerical problems arise in the discretization of the estimator, as described by Maes [12] and Pluim et al. [13]. Thévenaz et al. [14] proposed quasi-random sampling based on Halton sequences in order to overcome these problems.

The structured noise within reconstructed medical images often poses problems to the registration, as the algorithm tends to align not only the interesting image content, but also the background noise. Some authors have tried to eliminate this problem by using intensity thresholds within the joint PDF [15, 16], or by masking the background region of the images [17]. Although this removes all influences of the background, mis-registrations with the background regions are partly or entirely disregarded.

3 Methods and Materials

This section is organized as follows. After a brief summary of the similarity measure for the registration in section 3.1, we present in part 3.2 an adaptive binning scheme that is used to re-quantize the image intensities for a given number of histogram bins. Based on a similar scheme, we describe in section 3.3 the automatic detection of the intensity thresholds for the background, which is assumed to contain structured noise. The density estimation is subject of part 3.4, where the efficient Parzen-window discretization using histograms is presented. The density estimation techniques are used for the registration, and afterwards, for the normalization of the SPECT intensities, as described in section 3.5.

3.1 Normalized Mutual Information

In an automatic, intensity-based image registration, a distance measure $\mathcal{D}$ is used as an objective function for the alignment between a reference image $\mathcal{R}$ and a template image $\mathcal{T}$. For a spatial transform $\Phi : \mathbb{R}^3 \mapsto \mathbb{R}^3$, which consists of rotation and translation parameters in our case, the term $\mathcal{T}_\Phi$ is used to refer to the transformed template image:

$$\mathcal{T}_\Phi(\boldsymbol{x}) = \mathcal{T}(\Phi(\boldsymbol{x})) \tag{1}$$

During the registration, we search for an optimal transform $\hat{\Phi}$ that minimizes the distance measure:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}}\, \mathcal{D}[\mathcal{R}, \mathcal{T}, \Phi] \tag{2}$$

Distance measures based on image intensity statistics are widely used for multimodal registration tasks and also for single modalities where the intensities are not normalized, for example the SPECT image pairs in our application. Based on Shannon's theory [18], the information content within the images is measured using the entropies of the marginal PDFs $p_\mathcal{R}$ and $p_{\mathcal{T}_\Phi}$, and the joint PDF $\boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}$:

$$\mathcal{H}(\mathcal{R}) = -\int_{\mathbb{R}} p_\mathcal{R}(r) \log p_\mathcal{R}(r)\ \mathrm{d}r \tag{3}$$

$$\mathcal{H}(\mathcal{T}_\Phi) = -\int_{\mathbb{R}} p_{\mathcal{T}_\Phi}(t) \log p_{\mathcal{T}_\Phi}(t)\ \mathrm{d}t \tag{4}$$

$$\mathcal{H}(\mathcal{R}, \mathcal{T}_\Phi) = -\int_{\mathbb{R}^2} \boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}(\boldsymbol{i}) \log \boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}(\boldsymbol{i})\ \mathrm{d}\boldsymbol{i} \tag{5}$$

where r, t, and $\boldsymbol{i} = (r, t)^T$ are intensity random measures of $\mathcal{R}$ and $\mathcal{T}$. In the following, we make use of the normalized mutual information (NMI) [2], which is less variant to overlap effects than the common mutual information, which was introduced by Wells et al. [19] and Maes et al. [20]:

$$\mathcal{D}^{\mathtt{NMI}}[\mathcal{R}, \mathcal{T}_\Phi] = -\frac{\mathcal{H}(\mathcal{R}) + \mathcal{H}(\mathcal{T}_\Phi)}{\mathcal{H}(\mathcal{R}, \mathcal{T}_\Phi)} \tag{6}$$

Here, $\mathcal{D}^{\mathtt{NMI}}$ is written as a distance measure, i. e. smaller values indicate a better result.

A common technique to estimate the intensity PDFs is Parzen-windowing. In a one-dimensional case with n random samples $x_1, x_2, \ldots, x_n$, the Parzen-window PDF estimator yields [7, 21]:

$$p_{\lambda,n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_\lambda(x - x_i) \ , \tag{7}$$

with K_λ being the kernel PDF of width λ. Unfortunately, this approach has a high computational complexity and the storage requirements needed for large numbers of samples are high. If the random samples are discretized into a histogram, a discretization error is introduced on the one hand, but on the other hand, a lot of computations can be saved. The non-parametric estimator then resembles the behavior of a mixture model with as many components as bins. The n samples are stored in a discrete histogram h_n with b bins ($b > 1$). Here, $h_n(x_i)$ provides the fraction of samples that fall into the bin corresponding to x_i. $\hat{p}_{\lambda,n}$ is the discretized PDF estimator that differs from $p_{\lambda,n}$ in (7) by the application of histogram binning:

$$\hat{p}_{\lambda,n}(c_j) = \sum_{i=1}^{b} h_n(c_i) K_\lambda(c_j - c_i) = (h_n \star K_\lambda)(c_j) \approx p_{\lambda,n}(c_j) \ , \tag{8}$$

where c_j is the intensity value corresponding to the center of the j-th bin, and '$\star$' the convolution operator. We assume in the following, that $\hat{p}_{\lambda,n}$ is an approximation of its continuous counterpart, as indicated in (8).

3.2 Adaptive Binning Scheme

In data-driven approaches that estimate the optimal kernel width, one can observe that the result is directly related to the uncertainty within the data. Due to the discrete nature of histograms, this uncertainty is reflected by a varying smoothness or degenerations. Estimators using constant kernel widths cannot distinguish between regions of high and low certainty within one histogram. Therefore, several authors suggest making this parameter spatially variant. A disadvantage of adaptive, anisotropic kernel widths is the increased computational complexity for both the estimator and the formulation of its derivatives. In medical imaging, this increase in complexity is prohibitive. In addition, the efficient evaluation scheme (8) cannot be applied to estimators with varying kernel sizes. Therefore, we propose a trade-off in favor of a higher computational efficiency.

Instead of determining different kernel widths for an equidistantly spaced histogram, the image intensities are initially sampled into a histogram of varying bin widths. The corresponding bin centroids define a quantization characteristic, which can be used to map the input intensities to re-quantized output values. These, in turn, can be represented with an equidistantly spaced histogram. A density estimation on this re-quantized intensity space then does not have to account for different bin widths of the histogram and the convolution-based estimator (8) can be applied. The nonlinear mapping is computed as a preprocessing

step in the beginning, which means that it has to be computed only once for each image, but requires a distance measure that is invariant to this type of intensity transform, a property that is fulfilled by $\mathcal{D}^{\mathtt{NMI}}$ (6).

In order to distribute the bin centers for the initial histogram with a minimal quantization error, we apply an approach introduced by Lloyd [22] and Max [23]. It minimizes the noise power $\mathcal{N}$ for a specific number of bins by an iterative refinement of the bin center locations. The spatial region of the i-th bin within the domain of the random variable is defined by the interval $[l_{i-1}; l_i]$ with the centroid c_i. The noise power of the re-quantization with respect to the signal PDF $p(x)$ is:

$$\mathcal{N} = \sum_{i=1}^{b} \int_{l_{i-1}}^{l_i} (c_i - x)^2 p(x) \, \mathrm{d}x \ . \tag{9}$$

Lloyd [22] proposed a fixed point iteration scheme to numerically minimize (9) with respect to the bin intervals and centroids. Again, $p(x)$ is unknown, but can simply be exchanged by the histogram of the entire image with full intensity resolution, or a suitable Parzen-window estimator.

3.3 Background Threshold Detection

Tomographic, medical images are the result of discrete, modality-specific reconstruction methods that are based on physical measurements. In practice, these measurements are affected by detector noise and many physical effects, which may impair the reconstruction result. Problems for image registration algorithms especially arise from structured noise in the reconstructed images. Thévenaz et al. [17] presented a robust technique to distinguish between the object and background region within an image. They used the aforementioned Max-Lloyd quantization algorithm on a low pass filtered version of the image. Combined with the filtering, the algorithm computes the bin widths for a discrete histogram of two bins. The boundary between these two bins is assumed to separate intensities in the background from object values. The authors used the algorithm to determine the background region within PET images. We apply the background values to down-weight the corresponding region within the joint PDF – instead of thresholding it – in order to reduce the influence of the background and the contained noise. The background thresholds are also used for the intensity normalization 3.5.

3.4 Density Estimation

Very common choices for the kernel PDF K_λ are the Gaussian or cubic B-spline [24, 25]. Using a cubic B-spline B yields the following Parzen-window kernel K_λ^{B}:

$$K_\lambda^{\mathrm{B}}(x) = \frac{1}{\lambda} \mathrm{B}\left(\frac{x}{\lambda}\right) \tag{10}$$

The B-spline function is commonly defined recursively by the Cox-de Boor recursion formula, however, in the case of degree three, the kernel may be written

as a non-recursive function:

$$K_\lambda^{\mathrm{B}}(x) = \begin{cases} \frac{1}{\lambda}\left(\frac{4}{3} - 2\frac{|x|}{\lambda} + \frac{x^2}{\lambda^2} - \frac{|x|^3}{6\lambda^3}\right) & , \text{ if } \frac{|x|}{\lambda} \in [1,2[\\ \frac{1}{\lambda}\left(\frac{2}{3} - \frac{x^2}{\lambda^2} + \frac{|x|^3}{2\lambda^3}\right) & , \text{ if } \frac{|x|}{\lambda} \in [0,1[\\ 0 & , \text{ otherwise} \end{cases} \tag{11}$$

K_λ^{B} can be discretized either by sampling of the kernel values or recursive filtering. This specific window function has some advantages over, for instance, a Gaussian: it has a local support and fulfills the partition of unity constraint [26].

A necessary requirement for the implementation of the estimator is the specification of λ. Unfortunately, this parameter is dependent on the data, i. e. the values and the number of the random samples. In order to get an optimal PDF estimator, it is necessary to apply data-driven estimation techniques. A leave-one-out estimator is usually plugged into a log-likelihood function with respect to the kernel width in order to measure how good it resembles the missing data. Let $p_{\lambda,n-1}^j$ be the estimator after deleting the j-th sample. The resulting log-likelihood objective function then yields [27]:

$$\mathcal{L}(\lambda) = \sum_{j=1}^{n} \log p_{\lambda,n-1}^j(x_j) \ . \tag{12}$$

An optimal value $\hat{\lambda}$ for the kernel width maximizes $\mathcal{L}$

$$\hat{\lambda} = \operatorname*{argmax}_\lambda \mathcal{L}(\lambda) \ . \tag{13}$$

To optimize (13), it is necessary to search for a zero crossing of the derivative of (12) with respect to λ. The problem can be solved using an iterative, nonlinear optimization scheme, e. g. Newton's method. The derivatives of $\mathcal{L}$ with respect to λ are:

$$\begin{aligned} \frac{\partial}{\partial\lambda}\mathcal{L}(\lambda) &= \sum_{j=1}^{n} \frac{1}{\hat{p}_{\lambda,n-1}^j(x_j)} \frac{\partial}{\partial\lambda}\hat{p}_{\lambda,n-1}^j(x_j) \\ \frac{\partial^2}{\partial^2\lambda}\mathcal{L}(\lambda) &= \sum_{j=1}^{n} -\frac{1}{\left(\hat{p}_{\lambda,n-1}^j(x_j)\right)^2}\left(\frac{\partial}{\partial\lambda}\hat{p}_{\lambda,n-1}^j(x_j)\right)^2 \\ &\quad + \frac{1}{\hat{p}_{\lambda,n-1}^j(x_j)} \frac{\partial^2}{\partial^2\lambda}\hat{p}_{\lambda,n-1}^j(x_j) \end{aligned} \tag{14}$$

For the B-spline kernel function K_λ^{B}, the derivatives of the estimator yield:

$$\begin{aligned} \frac{\partial}{\partial\lambda}\hat{p}_{\lambda,n-1}^j(x_j) &= \sum_{i=1}^{b} h_{n-1}^j(c_i) \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x_j - c_i) \\ \frac{\partial^2}{\partial^2\lambda}\hat{p}_{\lambda,n-1}^j(x_j) &= \sum_{i=1}^{b} h_{n-1}^j(c_i) \frac{\partial^2}{\partial^2\lambda} K_\lambda^{\mathrm{B}}(x_j - c_i) \end{aligned} \tag{15}$$

Similar to the kernel function (11), we formulate the derivatives of K_λ^{B} for three cases.

Case 1: $\frac{|x|}{\lambda} \in [1, 2[$

$$\begin{aligned}\frac{\partial}{\partial\lambda}K_\lambda^{\mathrm{B}}(x) &= \frac{1}{\lambda^2}\left[\frac{2|x|}{\lambda} - \frac{2x^2}{\lambda^2} + \frac{|x|^3}{2\lambda^3} - \lambda K_\lambda^{\mathrm{B}}(x)\right] \\ \frac{\partial^2}{\partial^2\lambda}K_\lambda^{\mathrm{B}}(x) &= \frac{2}{\lambda^3}\left[-\frac{2|x|}{\lambda} + \frac{3x^2}{\lambda^2} - \frac{|x|^3}{\lambda^3} - \lambda^2\frac{\partial}{\partial\lambda}K_\lambda^{\mathrm{B}}(x)\right]\end{aligned} \tag{16}$$

Case 2: $\frac{|x|}{\lambda} \in [0, 1[$

$$\begin{aligned}\frac{\partial}{\partial\lambda}K_\lambda^{\mathrm{B}}(x) &= \frac{1}{\lambda^2}\left[\frac{2x^2}{\lambda^2} - \frac{3|x|^3}{2\lambda^3} - \lambda K_\lambda^{\mathrm{B}}(x)\right] \\ \frac{\partial^2}{\partial^2\lambda}K_\lambda^{\mathrm{B}}(x) &= \frac{2}{\lambda^3}\left[\frac{3|x|^3}{\lambda^3} - \frac{3x^2}{\lambda^2} - \lambda^2\frac{\partial}{\partial\lambda}K_\lambda^{\mathrm{B}}(x)\right]\end{aligned} \tag{17}$$

Case 3: $\frac{|x|}{\lambda} \notin [0, 2[$

$$\frac{\partial}{\partial\lambda}K_\lambda^{\mathrm{B}}(x) = \frac{\partial^2}{\partial^2\lambda}K_\lambda^{\mathrm{B}}(x) = 0 \tag{18}$$

The equations (15) are again convolutions with the partial derivatives for the

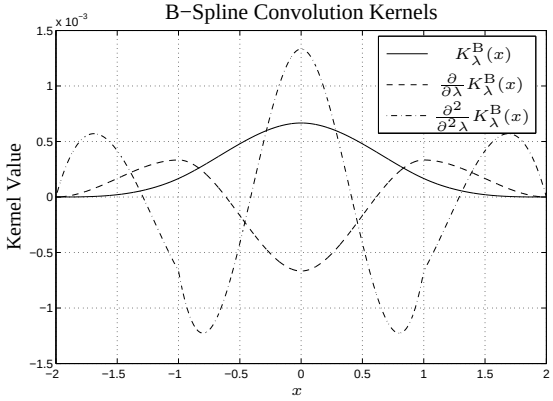
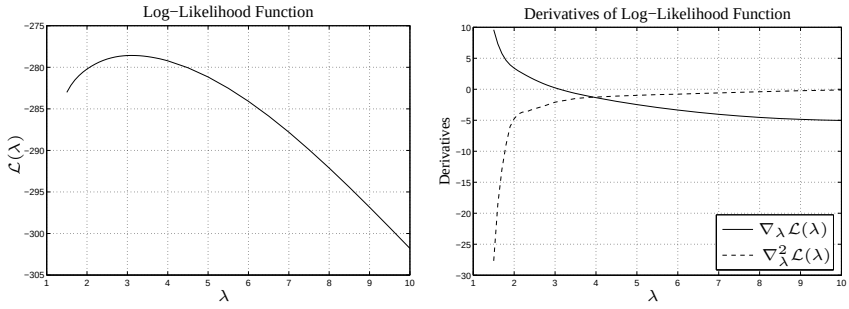


Figure 1. Convolution kernels for the kernel width estimation using a B-spline window function K_λ^{B} printed for a kernel width $\lambda = 1$. The kernel function, the first derivative, and the second derivative with respect to λ are plotted within the locally supported region.

kernel width of the Parzen-window kernel. Due to its local support, the spline kernel is very suitable for an implementation of the log-likelihood optimization using a discrete convolution operator and the kernel functions shown in Figure 1. The multivariate density kernel width estimation is realized analogue to the 1-D



(a) (b)

Figure 2. (a) Parzen-window log-likelihood function for a B-spline kernel K_λ^{B} cross-validation and an estimation of 100 samples drawn from a normal distribution with $\lambda = 4$. (b) Shows the corresponding first and second order derivative with respect to λ.

case and the 1-D kernels can be applied subsequently to each dimension of the histogram, because the multivariate B-spline kernel is separable. Figure 2 shows example curves that can typically be observed for the log-likelihood optimization of the kernel width. In this example, the width of a cubic B-spline kernel has been adapted to 100 sample values drawn from a Gaussian distribution with mean zero and a variance of 4.

3.5 Intensity Normalization

As mentioned above, for a correct interpretation of the differences between the SPECT images, the intensities have to be normalized to a common intensity range. This is necessary due to different acquisition times and changes in the overall tracer uptake within the human body. We model the mapping by an affine intensity transform similar to the proposals of Liao et al. [28]. In order to be invariant to the intensities of the background, we restrict the affine mapping to the region of the joint PDF above the background thresholds, i. e. the probabilities for joint intensities each belonging to intensities corresponding to brain tissue. The thresholds are determined using the approach described in section 3.3. The components of the intensity mapping are computed by linear regression within the joint PDF. This leads to an affine intensity transformation and, in the ideal case, a clustering at the diagonal entries of the joint PDF after the mapping.

4 Results

The proposed method has been applied to a collective of 26 epilepsy patients and assessed by physicians. Each patient underwent the standard diagnostic procedures, and, from the patient charts, the location as well as the number of the focal spots were known. The images have been anonymized within the clinics

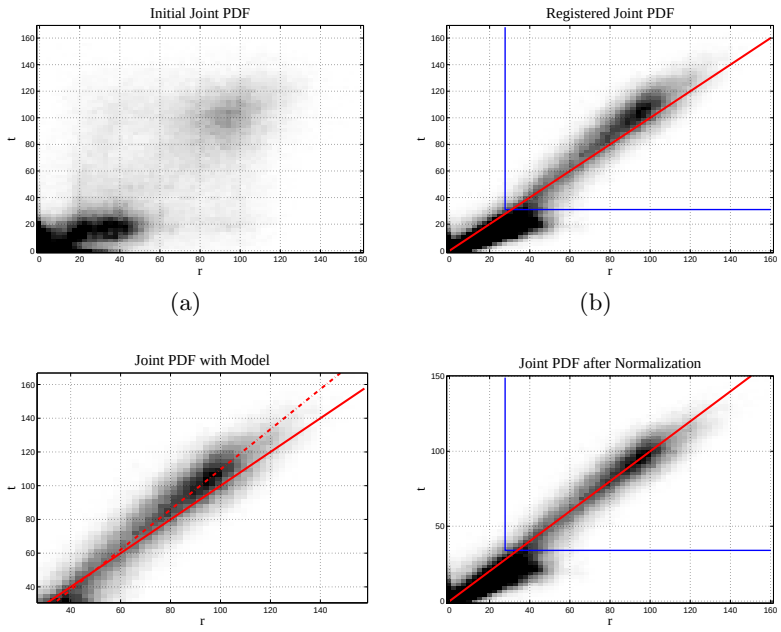
and handed blinded to us. The subtraction images were then generated with the proposed workflow and randomly ordered into a set of evaluation protocols. For the registrations, the number of bins has been computed automatically using a threshold for the quantization error of 0.05%. The number of samples has been 10% of the overlap domain, and at least 10.000 for lower resolutions in the multi-level optimization. No misregistrations were encountered during the experiments. Based on a fusion of the subtraction image with the corresponding MRI, the physicians had to specify the locations and the number of the focal spots, their certainty, and a rating of the image quality. No additional data was provided for the evaluation. The results of two physicians from the Department of Nuclear Medicine, University Hospital Erlangen, showed a good correlation of the proposed method with the conventional procedure. On an ordinal scale ranging from *very uncertain*, *uncertain*, *certain* to *very certain*, 45% of the results from the subtraction workflow were classified as *certain*. The observers reached a correlation value of 81% for correctly locating the focal spot in those cases. In the *uncertain* cases, still 53% of the focal spots have been located correctly. The overall image quality was rated *good* on an ordinal scale between *very bad*, *bad*, *good*, and *very good*. The intra-observer variability was 88% for the results computed from two evaluation passes.

An example for the density estimation and the normalization, using an affine model, is shown in Figure 3. Resulting subtraction images fused with the corresponding MR images are presented in Figure 4.

5 Discussion

In this article we have presented data-driven parameter estimation techniques for the Parzen-window estimation of intensity distributions, which are required for statistical image similarity measures. The proposed algorithms have been integrated into a normalized mutual information registration and applied to align intra- with inter-ictal SPECT images of epilepsy patients. After the registration, the estimated densities are used for the intensity normalization between the two SPECTs in order to allow for an image subtraction. The difference image is then fused with an MRI to spatially localize the focal spots.

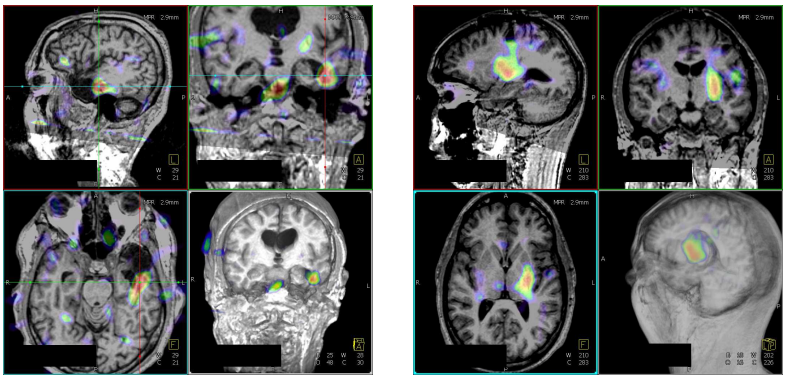
Regarding the evaluation, in the cases where the physicians indicated a high certainty, the results showed a high correlation with the standard evaluation method. In cases of low certainty, an inspection of the subtraction results revealed no failure of the algorithm, but instead often showed several weak focal spots that gave no clear information about the seizure onset. This might lead to the conclusion that the diagnosis of these patients with only the SPECT images is generally uncertain also in the standard approach. An evaluation based on real ground truth data (i. e. results from surgery) would be necessary to confirm this. Please note that only the *difficult* cases are undergoing a further analysis using multiple SPECT and MR images. A major advantage of the presented algorithm is the low amount of time that is required for the physician to assess the patient data in the beginning of the diagnosis. This suggests to combine the approach with the standard procedure, giving the physician a good indication of interesting regions within the data.



(a) (b)

(c) (d)

Figure 3. The figures show the joint PDF of two input SPECT images (a) before, (b) after the registration, (c) with the model fitted into the joint PDF without background content, and (d) the joint PDF after applying the intensity normalization. The figures also show the diagonal through the histogram space (solid red line), the fitted affine model (dashed red line), and the threshold region for the background (solid blue line).



(a) (b)

Figure 4. Two example patients taken from the collective. The images show the SPECT subtraction image fused with the corresponding MRI.

6 Ackowledgements

The authors would like to thank Prof. Dr. med. H. Stefan (Department of Neurology, University Hospital, Erlangen) for providing the patient data, PD Dr.

med. R. Linke (Department of Nuclear Medicine, University Hospital, Erlangen) for participating in the evaluation, and Dipl.-Ing. P. Ritt (Department of Nuclear Medicine, University Hospital, Erlangen) for his efforts in the patient data selection. We are also thankful to HipGraphics for providing the volume rendering software InSpace.

References

1. Hahn, D.A., Daum, V., Hornegger, J., Bautz, W., Kuwert, T.: Difference imaging of inter- and intra-ictal SPECT images for the localization of seizure onset in epilepsy. In Frey, E.C., ed.: IEEE Nuclear Science Symposium and Medical Imaging Conference, Honolulu, HI, USA, IEEE Nuclear & Plasma Sciences Society (October 2007) 4331–4335
2. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32**(1) (1999) 71–86
3. Chang, D.J., Zubal, I.G., Gottschalk, C., Necochea, A., Stokking, R., Studholme, C., Corsi, M., Slawski, J., Spencer, S.S., Blumenfeld, H.: Comparison of statistical parametric mapping and SPECT difference imaging in patients with temporal lobe epilepsy. Epilepsia **43** (2002) 68–74
4. McNally, K.A., Paige, A.L., Varghese, G., Zhang, H., Novotny, E.J., Spencer, S.S., Zubal, I.G., Blumenfeld, H.: Seizure localization by ictal and postictal SPECT. Epilepsia **46**(9) (2005) 1–15
5. Koo, C.W., Devinsky, O., Hari, K., Balasny, J., Noz, M.E., Kramer, E.L.: Stratifying differences on ictal/interictal subtraction SPECT images. Epilepsia **44**(3) (2003) 379–386
6. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1997)
7. Parzen, E.: On the estimation of probability density function and mode. Annals of Mathematical Statistics **33**(3) (1962) 1065–1076
8. Viola, P.: Alignment by Maximization of Mutual Information. PhD thesis, Massachusetts Institute of Technology (MIT) (1995)
9. Hermosillo, G., Chef d'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision **50**(3) (2002) 329–343
10. Knops, Z.F., Maintz, J.B.A., Viergever, M.A., Pluim, J.P.W.: Normalized mutual information based registration using k-means clustering and shading correction. Medical Image Analysis **10**(3) (2006) 432–439
11. Katkovnik, V., Shumulevich, I.: Kernel density estimation with adaptive varying window size. Pattern Recognition Letters **23**(14) (2002) 1641–1648
12. 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)
13. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Interpolation artefacts in mutual information-based image registration. Comput. Vis. Image Underst. **77**(9) (2000) 211–232
14. Thévenaz, P., Bierlaire, M., Unser, M.: Halton sampling for image registration based on mutual information. Sampling Theory in Signal and Image Processing **7**(2) (2008) 141–171
15. Rohlfing, T., Beier, J.: Improving reliability and performance of voxel-based registration by coincidence thresholding and volume clipping. In Hawkes, D.J., Hill, D.L.G., Gaston, R., eds.: Proceedings of Medical Image Analysis and Understanding 99, King's College (1999) 165–168

16. Rohlfing, T.: Multimodale Datenfusion für die bildgesteuerte Neurochirurgie und Strahlentherapie. PhD thesis, Technical University Berlin (2000)
17. Thévenaz, P., Ruttiman, U.E., Unser, M.: A pyramid approach to subpixel registration based on intensity. IEEE Transactions on Image Processing **7**(1) (1998) 27–41
18. Shannon, C.E.: A mathematical theory of communication (parts 1 and 2). The Bell System Technical Journal **27** (1948) 379–423,623–656
19. Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis **1**(1) (1996) 35–51
20. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16**(2) (April 1997) 187–198
21. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, New York (2001)
22. Lloyd, S.P.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2) (1982) 129–137
23. Max, J.: Quantizing for minimal distortion. IRE Transactions on Information Theory **6**(1) (1960) 7–12
24. Unser, M., Aldroubi, A., Eden, M.: B-Spline signal processing: Part I - theory. IEEE Transactions on Signal Processing **41**(2) (1993) 821–833
25. Unser, M., Aldroubi, A., Eden, M.: B-Spline signal processing: Part II - efficient design and applications. IEEE Transactions on Signal Processing **41**(2) (1993) 834–848
26. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. IEEE Transactions on Image Processing **9**(12) (2000) 2083–2099
27. Chow, Y.S., Geman, S., Wu, L.D.: Consistent cross-validated density estimation. The Annals of Statistics **11**(1) (1983) 25–38
28. Liao, Y.L., Chiu, N.T., Weng, C.M., Sun, Y.N.: Registration and normalization techniques for assessing brain functional images. Biomed Eng Appl Basis Comm **15** (June 2003) 87–94