

# Patch-based Label Fusion using Local Confidence-Measures and Weak Segmentations

André Mastmeyer<sup>a</sup>, Dirk Fortmeier<sup>a,c</sup>, Ehsan Maghsoudi<sup>a</sup>, Martin Simon<sup>b</sup>, and Heinz Handels<sup>a</sup>

<sup>a</sup>Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

<sup>b</sup>Clinic for Radiology and Nuclear Medicine, University Medical Center Schleswig-Holstein, Lübeck, Germany

<sup>c</sup>Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Germany

## ABSTRACT

A system for the fully automatic segmentation of the liver and spleen is presented. In a multi-atlas based segmentation framework, several existing segmentations are deformed in parallel to image intensity based registrations targeting the unseen patient. A new locally adaptive label fusion method is presented as the core of this paper. In a patch comparison approach, the transformed segmentations are compared to a weak segmentation of the target organ in the unseen patient. The weak segmentation roughly estimates the hidden truth. Traditional fusion approaches just rely on the deformed expert segmentations only. The result of patch comparison is a confidence weight for a neighboring voxel-label in the atlas label images to contribute to the voxel under study. Fusion is finally carried out in a weighted averaging scheme. The new contribution is the incorporation of locally determined confidence features of the unseen patient into the fusion process. For a small experimental set-up consisting of 12 patients, the proposed method performs favorable to standard classifier label fusion methods. In leave-one-out experiments, we obtain a mean Dice ratio of 0.92 for the liver and 0.82 for the spleen.

**Keywords:** multi-atlas based segmentation, label fusion, patch-based, local confidence, weak segmenter

## 1. INTRODUCTION

Multi-atlas based segmentation of abdominal organs has gained considerable attention in the last years.<sup>1-3</sup> Clinical applications comprise hepatic surgery and radiation treatment planning. The bottle-neck task for the planning phase is the manual expert delineation of the target organs in the patient data. In the abdominal area, the challenge lies in coping with the very large inter-patient shape and structure variations and dealing with pathological cases. Three steps usually define a multi-atlas segmentation: registration, label map warping and label fusion. While in the first step the accurate registration is very important, another research focus lies on the final classifier label fusion of the segmentation candidates.

Several new strategies have been published recently: On a global scale, Asman<sup>4</sup> proposed the combination of benefits from multi-atlas approaches and dedicated segmentation methods. Agarwal<sup>5</sup> proposed local confidence measures to extend the previously developed SIMPLE algorithm. Hao<sup>6</sup> uses local texture features to train a support vector machine for local fusion. Chen<sup>7</sup> learns several weak segmentations to build a strong one. Non-local STAPLE is a reformulation of the original STAPLE algorithm in a non-local means framework incorporating image intensities into the process.<sup>8</sup>

Preprocessing of our data consists of manually cropping the z-range of the data, thresholding and affine followed by non-linear registration. In this paper we use a diffeomorphic demons based registration approach.<sup>9</sup> After these preprocessing steps we obtain several candidate segmentations which fit the unseen patient more or less accurately. To unify the candidate segmentations, we present a new local label fusion technique using adaptive weights in a patch-based framework that takes gray value features of the targeted unseen patient data into account.

---

Further author information: (Send correspondence to André Mastmeyer; E-mail: mastmeyer@imi.uni-luebeck.de)

## 2. METHODS

### 2.1 Data and Preprocessing

The data used consists of 12 clinical routine scans with 512x512 pixels and 5 mm slice distance. The number of slices ranges from 126 to 157; pixel sizes are limited from 0.637 to 0.835 mm. Livers and spleens were manually segmented by our clinical partners.

To normalize data sets w.r.t. the contained structures a manual z-cropping procedure takes place. The region selected covers the cranial beginning of the iliac crest to the caudal slices of the heart. By this means, the number of slices used in the following steps is reduced to 50 on average and the abdominal organs are included in the remaining data. The next step consists of the removal of structures outside the body box, which could hinder the image registration process. Patient table, cables and spurious artifacts from the table are removed this way. Mainly, volume growing and morphological operations are used here. For the affine registration step, the data is thresholded by 0 HU to highlight important structures (bones, inner organs), which help the affine registration to converge fast and provide a coarse overlap of the structures. The thresholding step is motivated by the below zero Hounsfield values of fatty connective tissue embedding the inner organs. The distance measure chosen is the sum of squared differences (SSD). The result of the coarse registration step is an affine transformation matrix, which is co-applied to the label data.

On completion of these steps, the 11 target data sets are registered to the reference data using a non-linear registration approach with diffeomorphic demons.<sup>9,10</sup> In sum, 132 non-linear diffeomorphic registrations are carried out, 11 for each of the 12 leave-one-out experiments. The registrations are conducted on the cropped image data using a four resolution pyramiding scheme. The result of this phase are smooth deformation fields between the 11 target and the left-out reference image, which can be used to co-align the associated 11 segmentations to the unseen reference image.

The intersection and union (see fig. 1a, b) of the individual segmentations are calculated as assisting organ masks to estimate the organ gray value distribution in the reference image resulting in the mean  $\mu_{org}$  and standard deviation  $\sigma_{org}$  (see section 2.3). Another purpose of the masks is to reduce the search space for the proposed algorithm, see section 2.2.1.

On grounds of the rather anisotropic data all results from the previous steps are used in the resolution with:

- original 5 mm slices and
- 1 mm slices by resampling.

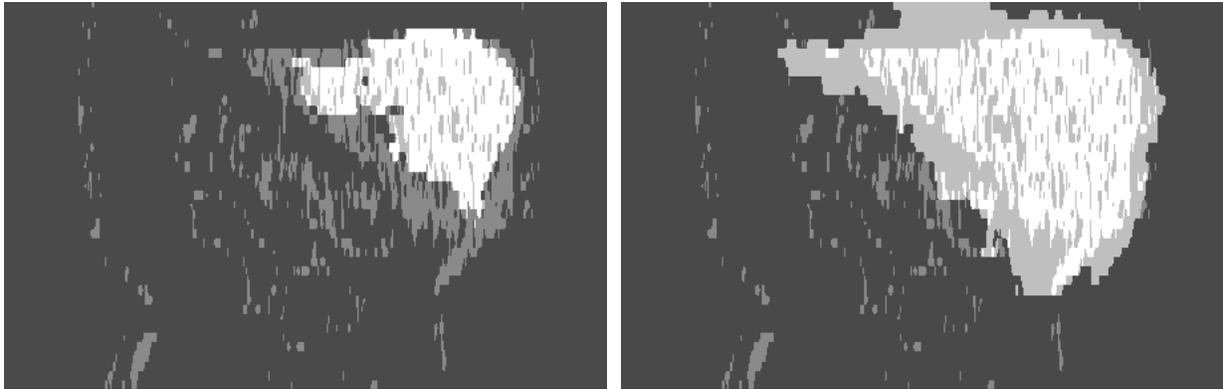
Resampling uses linear interpolation for gray value CT data and nearest neighbor interpolation for the segmentations.

### 2.2 Neighborhoods and Performance Cues

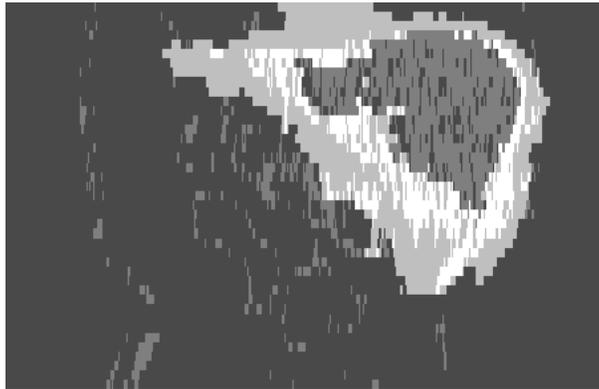
Majority Voting (MV) and Sum Rule (SR) fusion<sup>11,12</sup> are by far the fastest fusion methods and deliver very good results. The bottleneck of many new fusion algorithms is computation time. To speed up the computations, we define sparse neighborhoods. Moreover, computation takes place in a band near the supposed organ border only.

#### 2.2.1 Banded Calculation

For this means, the previously calculated organ masks are interpreted as (1) surely included voxels as seen in fig. 1a (intersection of the segmentations) and (2) candidate voxels for the voxels inside the union of all individual segmentations and outside the surely included voxels (see fig. 1c). Thus, subsequent calculations only take place in the band of candidate voxels, i.e. the result of an XOR between union and intersection of the segmentations.



(a) Intersection (white) vs. weak segmentation (gray)      (b) Union (light gray) vs. included weak segmentation (white)



(c) Band (light gray and white) vs. weak segmentation (gray)

Figure 1: Union, intersection and unsure candidate band: The union of all segmentations around the weak segmentation of one subject (a). The intersection of the segmentations (b). The resulting banded region of interest for the calculations (c).

### 2.2.2 Sparse Neighborhoods

The core of patch based label fusion is to compare a center patch in the unseen reference patient at a current voxel  $x_i$  to other patches in their neighborhood. These neighboring patches are extracted from the target subjects which are aligned via registration methods to the unseen patient.

Typically, the patch size is smaller than the neighborhood size, see fig. 2a. The definition of the neighborhood radius  $N_{rad}$  and patch radius  $P_{rad}$  strongly affects the performance of the algorithm. Smaller patch and neighborhood sizes boost the performance of the algorithm while the accuracy of the results may decrease.

Basically, the neighborhood of a voxel consists of adjacent voxels contained in a certain shape. In this work we restrict ourselves to cubical neighborhood and patch geometry.

The full neighborhood simply consists of all voxels around a center voxel contained in the neighborhood shape, see fig. 2a. In terms of offset vectors  $(m, n, o)$  given in image coordinates, the neighborhood set  $N(i)$  of a voxel  $x_i$  can be defined as follows:

$$N(i) = \bigcup_{m,n,o=-N_{rad}}^{N_{rad}} (m, n, o) \quad (1)$$

A sparse neighborhood only considers a smaller number of voxels compared to the full neighborhood. The first sparse neighborhood proposed here only takes into account every 2nd voxel (see fig. 2b):

$$N(i) = \bigcup_{m,n,o=-N_{rad}}^{N_{rad}} \begin{cases} (m, n, o), & m, n, o \text{ are all even numbers} \\ \emptyset, & \text{otherwise} \end{cases} \quad (2)$$

The second “very sparse” neighborhood only takes into account voxels at the outer edges and corners, see fig. 2c. In terms of offsets, the sparsest neighborhood can be defined as:

$$N(i) = \bigcup_{m,n,o=-1}^1 \begin{cases} (2 \cdot m \cdot P_{rad} + 1, 2 \cdot n \cdot P_{rad} + 1, 2 \cdot o \cdot P_{rad} + 1), & (m, n, o \text{ all } \neq 0) \vee (((i \oplus j) \oplus k) \\ \emptyset, & \text{otherwise} \end{cases} \quad (3)$$

where  $\oplus$  denotes the XOR operator. In this neighborhood the patches compared do not overlap but are still connected at a face or corner.

To account for anisotropic voxels as present in our data, calculations can be additionally restricted to the current slice that contains the voxel studied. The reason for this constraint is that with anisotropic voxels (5mm slices) the geometry of the neighborhood can span much more space in z-direction than in x- and y-directions. Therefore, a metrically more isotropic neighborhood can be achieved using this restriction. A significant computational performance boost is another benefit of it. The different neighborhood sizes are shown in tab. 1.

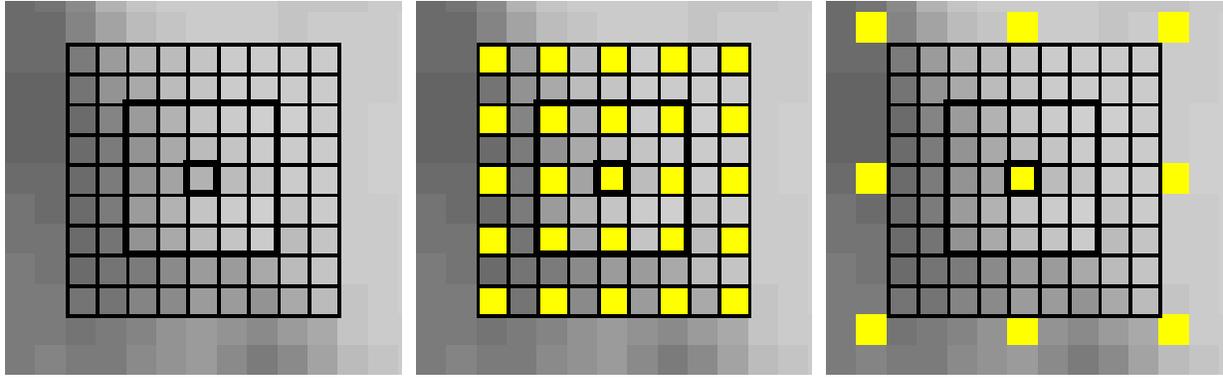
Neighborhood	$2 \cdot N_{rad} + 1 = 9$ Slices	Single Slice
Full	$9^3 = 729$	$9^2 = 81$
Sparse	$5^3 = 125$	$5^2 = 25$
Very Sparse	14	9

Table 1: Neighborhood sizes for  $N_{rad} = 4$  and  $P_{rad} = 2$ .

### 2.2.3 Selected algorithms

Based on the previous reasoning the following variations of the proposed method resp. input data are studied:

Our algorithm in tab. 2 is denoted by the short-cut PBLC for “Patch Based Local Confidence”. Furthermore we use the algorithms “Simultaneous Truth and Performance Estimation” (STAPLE),<sup>13</sup> “Selective and Iterative Method for Performance Level Estimation” (SIMPLE),<sup>14</sup> “Majority Voting” (MV) and “Sum Rule” (SR).<sup>11,12</sup>



(a) Full neighborhood

(b) Half sparse neighborhood

(c) Very sparse neighborhood

Figure 2: Center voxel (thick small square), neighborhood definition (grid with thin lines) and patch (thick square): Every voxel in the neighborhood region of a center voxel is visited (a). Only every second voxel (yellow small squares) is used for patch comparison (b). Only voxels at the corners and face centers are visited for comparison (c). The patch radius  $P_{rad}$  is set to a value of 2 and the neighborhood radius  $N_{rad}$  equals 4.

Alg.-Acronym	Neighborhood Density	In-Plane Restriction	Resampling to 1 mm Slices
PBLC 1	full	no	no
PBLC 2	full	yes	no
PBLC 3	sparse	yes	no
PBLC 4	very sparse	yes	no
PBLC 5	very sparse	no	yes
PBLC 6	sparse	no	yes
PBLC 7	full	no	yes

Table 2: Algorithms variants used and their acronyms.

### 2.3 Patch-based Label Fusion

Patch-based label fusion for hippocampus and ventricle segmentation was introduced by Coupé:<sup>15</sup> A voxel  $x_i$  is set to be the center of a cubical (1) neighborhood and (2) patch. Neighborhood size and patch size may differ to adapt the algorithm to structure sizes and spatial mismatches from the registrations. Image traversal visits all voxels, compares the patch under study with patches in the neighborhood and computes a weighted average label guess by:

$$L_{est}(x_i) = \frac{\sum_{s=1}^N \sum_{j \in N(i)} w(x_i, x_{s,j}) \cdot L(x_{s,j})}{\sum_{s=1}^N \sum_{j \in N(i)} w(x_i, x_{s,j})} \quad (4)$$

where  $L_{est}$  is the continuous label estimate for a voxel,  $N$  is the number of atlases,  $N(i)$  refers to the set of neighborhood voxels for  $i$ ,  $w$  are the weights dependent on the reference data voxel under study  $x_i$  and the voxel  $x_{s,j}$  in subject  $s$  at neighbor  $j$ . Finally,  $L(x_{s,j})$  represents the label in the atlas of subject  $s$  at position  $j$ . The label estimate  $L_{est}$  is a fuzzy estimate, i.e. with binary label data it ranges from 0 to 1, and the label guess is defined by thresholding with e.g. 0.5. In the weight  $w$  the comparison of the reference patient gray value patch  $P(x_i)$  centered at  $x_i$  and the target patient patch with center  $x_{s,j}$  takes place as  $w(x_i, x_{s,j}) = \exp\left(-\frac{\|P(x_i), P(x_{s,j})\|_{L_2}}{h}\right)$ . Here, the negative normalized  $L_2$ -norm divided by a decay parameter  $h$  is used as a patch distance measure, embedded in an exponential term. Consequently, the weight values lie in the continuum from 0 to 1 and low differences are preferred. The decay parameter  $h$  is recommended to be the minimal gray value patch distance in the  $L_2$ -norm.

In our proposed method, we use a weak segmentation as additional prior to the deformed atlas label data sets. Thus, we set the weights introducing local confidence measuring as follows:

$$w(x_i, x_{s,j}) = \exp\left(-\frac{1 - J(P_{wSeg}(x_i), P_{seg}(x_{s,j}))}{h}\right) \quad (5)$$

where  $P_{wSeg}(x_i)$  describes the label patch originating from the unseen reference image gray values that underwent some weak segmentation; and  $P_{seg}(x_{s,j})$  is the label patch from the target subject atlas data. Lastly,  $J$  calculates the Jaccard coefficient and  $h$  is the decay parameter. These weights correlate the deformed atlas segmentations to the hidden truth guessed by a thresholding criterion, our weak segmenter:

$$\forall x \in P_{wSeg}(x_i) : L(x) = \begin{cases} 1, & \text{if } \mu_{org} - l \cdot \sigma_{org} < G(x) < \mu_{org} + l \cdot \sigma_{org} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Here,  $L$  is the label at voxel  $x$  from the label patch,  $G$  describes the gray value in the reference image data,  $\mu_{org}$  and  $\sigma_{org}$  are the organ normal distribution parameters estimated in the intersection of the candidate segmentations (see fig. 1c). The standard deviation factor  $l$  is dependent upon the organ under study. The intersection voxels used to estimate  $\mu_{org}$  and  $\sigma_{org}$  are supposedly surely contained inside the liver resp. spleen of the reference patient.

On whole image scale, the results of the weak segmentation can look like point clouds where many positively labeled voxels concentrate in the area of the organ (see fig. 3a).

In summary, the described fusion method computes local similarity resp. confidence measures in the neighborhood of the studied voxel as weights for the neighboring labels to be averaged.

## 2.4 Study Set-Up

In a leave-one-out scheme we conduct 12 experiments: per experiment a multi-atlas segmentation is carried out (see section 2.1) for liver and spleen. It consists of an affine and variational registration of the image intensity data while co-warping the associated label maps. Afterwards the label maps are fused with the proposed seven method variants and compared to four standard algorithms.

The proposed method is sensitive to the decay parameter  $h$ , for which we found a reasonable setting is 0.1. This way, very similar patches in the weak segmentation and the warped atlas label data contribute over-proportionally to the averaging. The final segmentation is generated by thresholding the resulting fusion weight map with 0.618. The patch radius  $P_{rad}$  is set to a value of 2 and neighborhood radius  $N_{rad}$  is chosen as 4. The organ dependent standard deviation factor  $l$  for our weak segmenter is chosen as 1 resp. 3 for the liver resp. spleen.

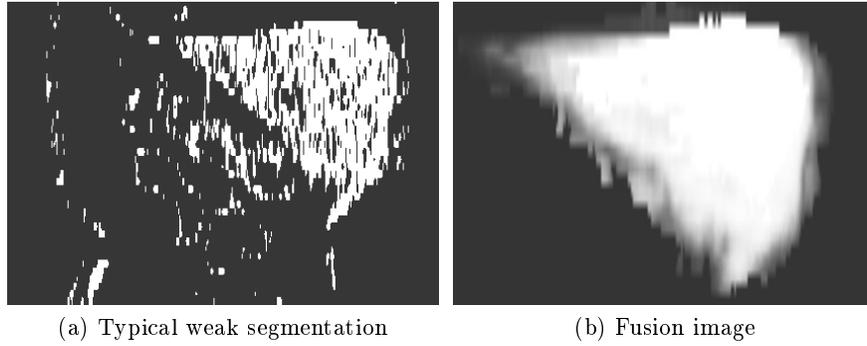
## 3. RESULTS

Figs. 3 and 5 show qualitative results for reference patient 1. Quantitative results for the liver and spleen are shown in tab. 3 and fig. 4.

Overall, consistent advantages in comparison to the standard methods can be observed. PBLC4-5 mainly are the overall winners of the contest, which holds true for all evaluation metrics used.

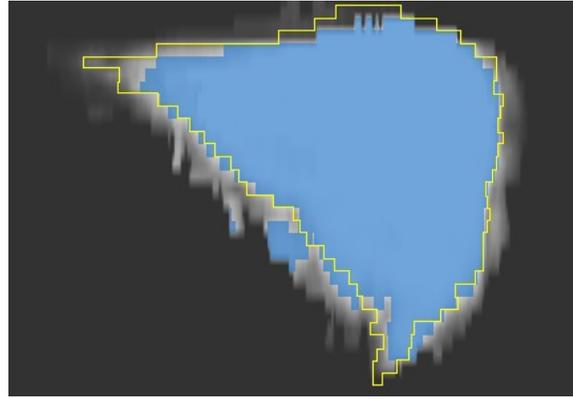
The Dice coefficients are generally higher for the liver. The Mean Surface Distance globally is lower than the slice distance of the original data and on average is lower for the larger organ, i.e. the liver, while the Hausdorff distances are smaller for the spleen.

We take a closer look at the surface distances for one patient to inspect the location of the main errors in this metric. In fig. 5a the medial area of the liver is shown, here we typically observe a higher distribution of errors than on the periphery. This is also indicated in fig. 3c where the segmentation in blue leaks out of the yellow contour which depicts the aimed reference segmentation. The same observation is true for the spleen, the medial area is depicted in fig. 5b. The organ ensemble is shown finally in fig. 5c, where we also note that surface errors are smaller for the spleen than for the liver. In fig. 5c some higher errors can also be noticed on some peripheral spots of the liver.



(a) Typical weak segmentation

(b) Fusion image



(c) Resulting (blue overlay) vs. reference (yellow contour) segmentation

Figure 3: Intermediate and final results: The weak segmentation roughly indicates the organ borders and its inside by positive label concentration (a). The fusion image shows the fuzzy labeling (b). The expert segmentation is compared to the guessed segmentation (c).

Organ Alg. \ Metric	Liver			Spleen		
	Dice	MSD [mm]	HD [mm]	Dice	MSD [mm]	HD [mm]
SIMPLE	0.88±0.06	3.30±1.47	31.48±9.30	0.78±0.12	3.49±1.93	27.90±13.04
VoteRule	0.90±0.04	2.99±1.17	29.81±12.99	0.79±0.15	3.33±2.23	24.85±12.03
SumRule	0.90±0.04	3.07±1.18	30.00±12.97	0.79±0.15	3.38±2.26	23.51±11.83
STAPLE	0.88±0.04	3.99±1.49	33.32±14.73	0.71±0.17	5.19±3.44	34.49±16.31
PBLC1	0.91±0.03	2.85±1.06	29.45±11.25	0.79±0.15	3.50±2.47	25.47±10.50
PBLC2	0.91±0.04	2.73±1.05	28.53±11.50	0.79±0.14	3.31±2.10	23.08±12.62
PBLC3	0.91±0.04	2.71±1.03	28.45±11.48	0.79±0.14	3.27±2.08	22.87±12.71
PBLC4	0.91±0.03	2.65±0.95	28.19±11.34	0.80±0.14	3.15±2.04	23.13±12.25
PBLC5	0.92±0.02	2.57±0.83	27.51±11.18	0.82±0.14	3.15±2.39	22.48±11.99
PBLC6	0.91±0.03	2.75±0.93	28.03±10.87	0.80±0.14	3.35±2.28	22.67±12.08
PBLC7	0.91±0.03	2.78±0.94	28.26±10.84	0.80±0.14	3.39±2.26	22.69±12.21
Winner	PBLC5	PBLC5	PBLC5	PBLC5	PBLC4	PBLC5
Winner w/o Resampling	PBLC4	PBLC4	PBLC4	PBLC4	PBLC4	PBLC3

Table 3: The table shows Dice coefficients, Mean Surface Distance (MSD) and Hausdorff Distance (HD) for four standard algorithms and seven variations of the proposed algorithm in our leave-one-out experiments. PBLC wins the competition with the fastest variant of the proposed algorithm (bottom rows).

Computation times for a single unseen patient on a Intel Xeon W3530 2.8 GHz workstation using 6 parallel threads are around one hour for PBL C1 and 2.5 minutes for PBL C4, which is the fastest variant of the algorithm w.r.t. the used data, see tab. 4.

Organ \ Alg. [min:sec]	PBL C1	PBL C2	PBL C3	PBL C4	PBL C5	PBL C6	PBL C7
Liver	50:16	6:38	2:25	1:28	12:52	44:58	242:19
Spleen	14:44	2:32	1:02	0:56	6:34	16:16	79:13

Table 4: Computation times for a label fusion targeting one patient on a Intel Xeon W3530 (6 threads) workstation in minutes:seconds.

Statistical testing of all results using one-way ANOVA showed no significant differences, but there is a clear tendency to favor the proposed method and thin slice CT data.

#### 4. CONCLUSION

The overall results for the new method underline to follow the proposed direction of research. We can also posit the benefit of thin slice CT data for label fusion algorithms in general, which we only have simulated in our experiment. Of course resampling can not provide the missing details present in fine resolution data sets.

The most accelerated variant PBL C4 (approximately more than 20 times faster compared to PBL C1) while supposedly more inaccurate due to using a sparse neighborhood is not inferior to the other competitors. In fact this algorithm only performs calculations in a slice-wise manner, which is appropriate for highly anisotropic data. While very sparse, the spatial extension of the neighborhood is highest for this algorithm variant.

Regarding geometry and volume, the size and shape of an organ are important factors for the outcome of the algorithms. Larger organs tend to have higher Dice coefficients and lower Mean Surface Distances. The Hausdorff Distance characterizes outliers which are more likely to occur in complex shaped organs such as the liver with its lobes, fissures and apices.

A closer inspection reveals that high surface distances occur in the fissures between the lobes of the liver which are not visible in the rendering and which are generally very difficult to segment resp. register. Thus organs with complicated geometric features such as fissures and lobes are prone to higher Hausdorff distances while compact organs such as the spleen are typically easier to tackle. The higher errors on the periphery of the liver might be due to some difficulties the registration algorithms face with aligning the rib cage correctly. We can see the mark of blurred rib lines in fig. 5c.

Summarizing, we present a new concept for a local patch-based confidence measure capable to combine dedicated strong or weak segmentation algorithms with multi-atlas label fusion.

The incorporation of reference gray value image features opens up a new perspective of connecting the worlds of dedicated segmentation algorithms and multi-atlas based label fusion.<sup>4</sup> From the perspective of a first study, here a very weak segmenter is used for reference image feature generation, but in sum with the local confidence based fusion approach shows quite promising results in this small scale study with 12 patients.

In future new weak segmentation algorithms could be used. The threshold based weak segmenter used here could also be easily replaced (1) by stronger segmentation methods or (2) can incorporate more sophisticated preprocessing algorithms, such as (1) Level-Sets<sup>16</sup> or Graph-Cuts<sup>17</sup> on one hand and (2) image region homogenizers (e.g. Anisotropic Diffusion<sup>18</sup>) on the other hand. Another line to follow would be the enlargement of the study to 20-30 patients to provide more shape variability to the system which should further improve the performance. Post-processing of the segmentations might also be of value, registration and fusion artifacts which occur as spurious voxel groups and spikes could be favorably smoothed out. This might be achieved by a dedicated segmentation algorithm such as Graph-Cuts<sup>17</sup> which could be fed with the obtained results to improve just the fringe of the segmentation. Because of the averaging nature of the fusion algorithms the major bottleneck is to capture the finer shape details as present in the medial and lower apex of the liver, while artifacts from the registration algorithms such as spikes need to be smoothed away.

For the proposed method, the trade-off of the influence between the weak segmenter and the warped expert segmentations is a delicate question. In this implementation the correlation of the experts takes place after the correlation of one expert with the weak segmentation is done. Finally, they are averaged with the same weight. In the averaging scheme the system could try to weight the individual correlation results using a performance measure to give better correlations a higher influence in the average (weighted averaging) or automatically sort out outliers as presented in the SIMPLE method.<sup>14</sup>

The main problem for the overall system remains the time complexity and quality of the non-linear registrations. For patient individual surgery planning, a system should be capable of delivering a segmentation result over night. With upcoming GPU-based massive parallel approaches<sup>19</sup> accurate pairwise non-linear registration will soon be a matter of less than ten minutes for a pair of data sets.

### ACKNOWLEDGEMENTS

This work is supported by the German Research Foundation (DFG-HA2355/10-1).

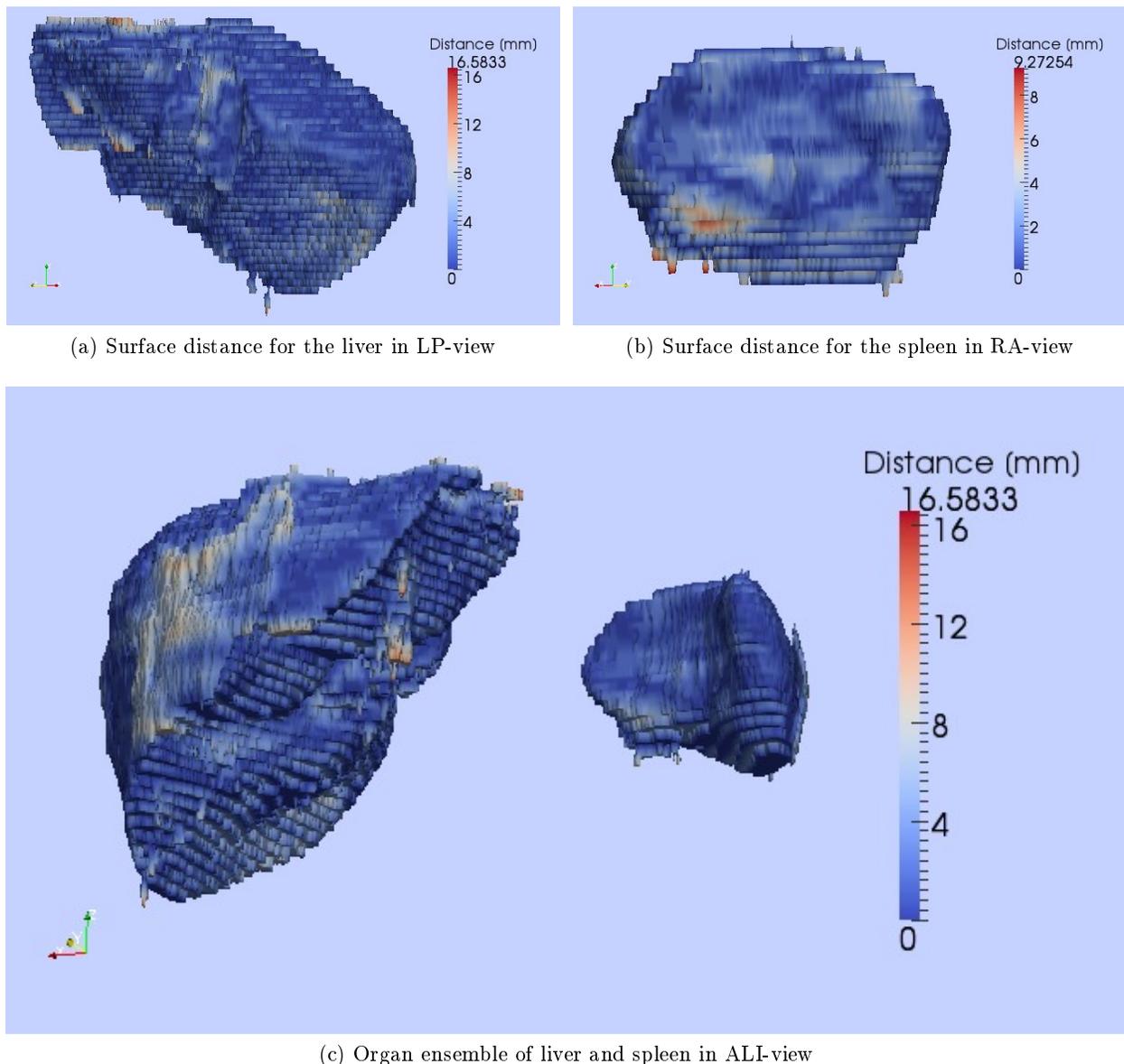
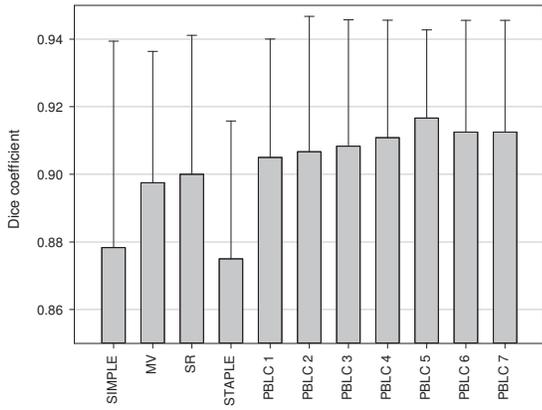
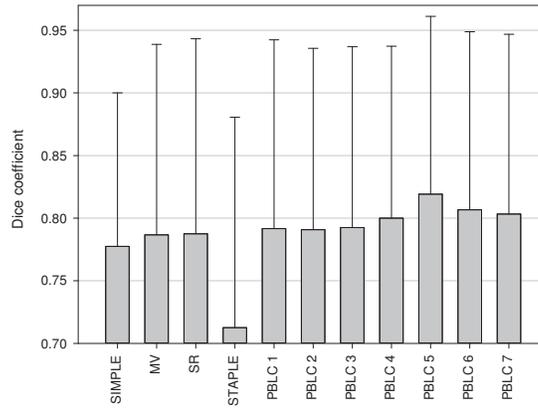


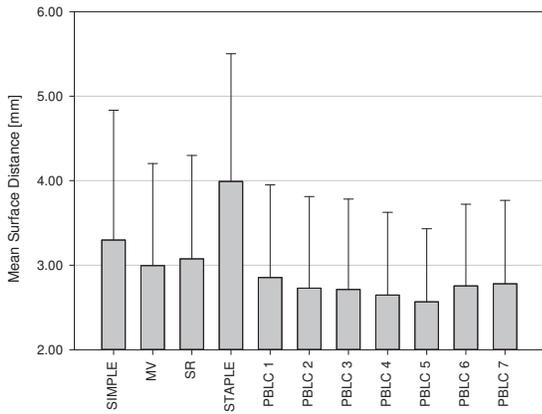
Figure 5: Surface distances for the different organs under study of a patient.



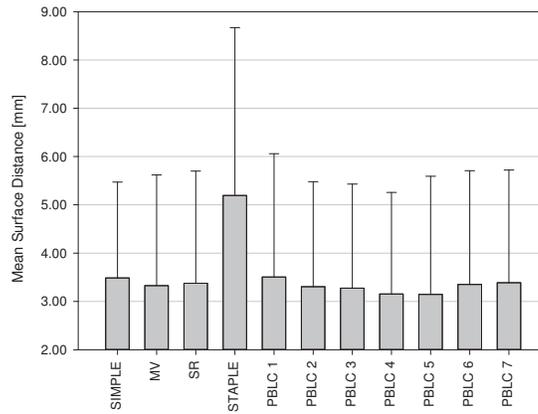
(a) Liver: Dice coefficients



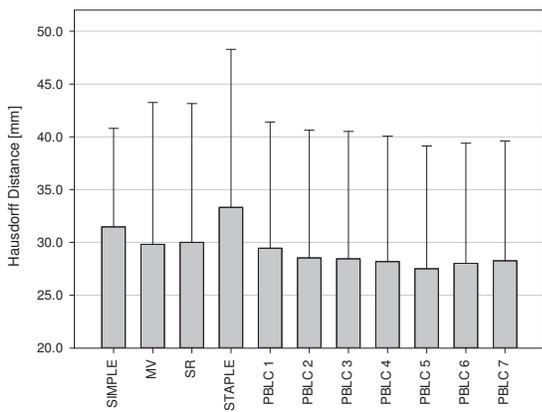
(b) Spleen: Dice coefficients



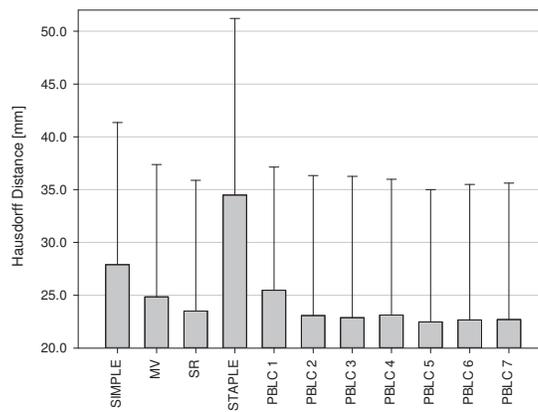
(c) Liver: Mean Surface Distances



(d) Spleen: Mean Surface Distances



(e) Liver: Hausdorff Distances



(f) Spleen: Hausdorff Distances

Figure 4: Bar charts with mean and standard deviation of the evaluation metrics.

## REFERENCES

1. Oda, M., Nakaoka, T., and Kitasaka, T., "Organ Segmentation from 3D Abdominal CT Images Based on Atlas Selection and Graph Cut," *Abdominal Imaging* **7029**, 181–188 (2011).
2. Wolz, R., Chu, C., Misawa, K., Mori, K., and Rueckert, D., "Multi-organ abdominal ct segmentation using hierarchically weighted subject-specific atlases," *Medical Image Computing and Computer-Assisted Intervention: MICCAI* **7510**, 10–17 (2012).
3. Suzuki, M., Linguraru, M., and Okada, K., "Multi-organ segmentation with missing organs in abdominal ct images," *Medical Image Computing and Computer-Assisted Intervention: MICCAI* **7512**, 418–425 (2012).
4. Asman, A. J. and Landman, B. a., "Simultaneous Segmentation and Statistical Label Fusion," *SPIE Medical Imaging 2012* **8314**, 83140Y–83140Y–8 (2012).
5. Agarwal, M., Hendriks, E. a., Stoel, B. C., Bakker, M. E., Reiber, J. H. C., and Staring, M., "Local SIMPLE Multi-Atlas-based Segmentation Applied to Lung Lobe Detection on Chest CT," *SPIE Medical Imaging 2012* **8314**, 831410–831410–7 (2012).
6. Hao, Y., Liu, J., Duan, Y., Zhang, X., Yu, C., Jiang, T., and Fan, Y., "Local Label Learning (L3) for Multi-Atlas based Segmentation," *SPIE Medical Imaging 2012* **8314**, 83142E–83142E–8 (2012).
7. Chen, T., Vemuri, B. C., Rangarajan, A., and Eisenschenk, S. J., "Mixture of Segmenters with Discriminative Spatial Regularization and Sparse Weight Selection.," *Medical Image Computing and Computer-Assisted Intervention: MICCAI* **14**, 595–602 (Jan. 2011).
8. Asman, A. J. and Landman, B. A., "Non-local STAPLE : An Intensity-Driven Multi-atlas Rater Model," *Medical Image Computing and Computer-Assisted Intervention: MICCAI* , 426–434 (2012).
9. Schmidt-Richberg, A., Ehrhardt, J., Werner, R., and Handels, H., "Diffeomorphic Diffusion Registration of Lung Images," *Workshop Proceedings, Medical Image Computing and Computer Assisted Intervention: MICCAI*, 55–62 (Sept. 2010).
10. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., and Handels, H., "Statistical Modeling of 4D Respiratory Lung Motion Using Diffeomorphic Image Registration.," *IEEE Transactions on Medical Imaging* **30**, 251–65 (Feb. 2011).
11. Kittler, J. and F.M. Alkoot, "Sum versus Vote Fusion in Multiple Classifier Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence on Pattern Analysis and Machine Intelligence* **25**(1), 110–115 (2003).
12. Rohlfing, T., Russakoff, D. B., and Maurer, C. R., "Performance-based Classifier Combination in Atlas-based Image Segmentation using Expectation-Maximization Parameter Estimation.," *IEEE Transactions on Medical Imaging* **23**, 983–94 (Aug. 2004).
13. Warfield, S. K., Zou, K. H., and Wells, W. M., "Simultaneous Truth and Performance Level Estimation (STAPLE): an Algorithm for the Validation of Image Segmentation.," *IEEE Transactions on Medical Imaging* **23**, 903–21 (July 2004).
14. Langerak, T. R., van der Heide, U. a., Kotte, A. N. T. J., Viergever, M. a., van Vulpen, M., and Pluim, J. P. W., "Label Fusion in Atlas-based Segmentation using a Selective and Iterative Method for Performance Level Estimation (SIMPLE).," *IEEE Transactions on Medical Imaging* **29**, 2000–8 (Dec. 2010).
15. Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., and Collins, D. L., "Patch-based Segmentation using Expert Priors: Application to Hippocampus and Ventricle Segmentation.," *NeuroImage* **54**, 940–54 (Jan. 2011).
16. Osher, S. and Sethian, J. A., "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics* **79**(1), 12–49 (1988).
17. Boykov, Y. and Kolmogorov, V., "An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision.," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**, 1124–37 (Sept. 2004).
18. Perona, P. and Malik, J., "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(7), 629–639 (1990).
19. Modat, M., Ridgway, G. R., Taylor, Z. a., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S., "Fast Free-form Deformation using Graphics Processing Units.," *Computer Methods and Programs in Biomedicine* **98**, 278–84 (June 2010).