Chapter 11
Automatic Multiorgan Segmentation Using Hierarchically Registered Probabilistic Atlases

Razmig Kéchichian, Sébastien Valette and Michel Desvignes

Abstract We propose a generic method for the automatic multiple-organ segmentation of 3D images based on a multilabel graph cut optimization approach which uses location likelihood of organs and prior information of spatial relationships between them. The latter is derived from shortest-path constraints defined on the adjacency graph of structures and the former is defined by probabilistic atlases learned from a training dataset. Organ atlases are mapped to the image by a fast (2+1)D hierarchical registration method based on SURF keypoints. Registered atlases are also used to derive organ intensity likelihoods. Prior and likelihood models are then introduced in a joint centroidal Voronoi image clustering and graph cut multiobject segmentation framework. Qualitative and quantitative evaluation has been performed on contrast-enhanced CT and MR images from the VISCERAL dataset.

11.1 Introduction and Related Work

Clinical practice today, especially whole-body CT and MR imaging, often generates large numbers of high-resolution images, which makes tasks of efficient data access, transfer, analysis and visualization challenging. This is especially true in distributed

---

R. Kéchichian (✉) · S. Valette
CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon,
Université de Lyon, Lyon, France
e-mail: razmig.kechichian@creatis.insa-lyon.fr

S. Valette
e-mail: sebastien.valette@creatis.insa-lyon.fr

R. Kéchichian · S. Valette
Université Claude Bernard Lyon 1, Lyon, France

M. Desvignes
GIPSA-Lab, CNRS UMR 5216, Grenoble-INP, Université Joseph Fourier,
Saint Martin d'Héres, France
e-mail: michel.desvignes@gipsa-lab.grenoble-inp.fr

M. Desvignes
Université Stendhal, Saint Martin d'Héres, France

A. Hanbury et al. (eds.), *Cloud-Based Benchmarking of Medical Image Analysis*, DOI 10.1007/978-3-319-49644-3_11

computing environments which have seen a growing use of hand-held terminals for interactive data access and visualization of anatomy. Therefore, there is great interest in efficient and robust medical image segmentation algorithms for the purposes of creating patient-specific anatomical models, clinical applications, medical research and education, and visualization and semantic navigation of full-body anatomy [3, 26].

Traditionally, single-object- or pathology-oriented, recent image processing methods [9, 12, 14, 15, 19, 23, 25, 27] have made the analysis and the segmentation of multiple anatomical structures increasingly possible. However, CT and MR images have intrinsic characteristics that render their automatic segmentation challenging. They are commonly degraded by various noise sources and artefacts due to limited acquisition time and resolution, and patient motion which all reduce the prominence of intensity edges in images. In addition, MR images suffer from spatial distortion of tissue intensity due to main magnetic field inhomogeneity. Regardless of the imaging modality and related artefacts, many anatomically and functionally distinct structures, especially those corresponding to soft tissues, have similar intensity levels in images and, furthermore, blend into surrounding tissues which have intensities close to their own. It is impossible to identify and segment such structures automatically on the basis of intensity information only. Hence, most advanced segmentation methods exploit some form of prior information on structure location [12, 19, 27] or interrelations [9, 14, 23, 25] to achieve greater robustness and precision. Hierarchical approaches to segmentation [23, 25, 32] rely on hierarchical organizations of prior information and algorithms that proceed in a coarse-to-fine manner according to anatomical level of detail.

Graph cut methods, which have been widely applied to single-object segmentation problems [4], rely on a maximum-flow binary optimization scheme of a discrete cost function on the image graph. For a particular class of cost functions which frequently arises in segmentation applications [16], these methods produce provably good approximate solutions in multiobject [5] and global optima in single-object segmentation. In addition, simultaneous multiobject segmentation approaches are superior to their sequential counterparts in that they raise questions neither on the best segmentation sequence to follow nor on how to avoid the propagation of errors of individual segmentations [9].

While widely used by the computer vision community, keypoint-based image description and matching methods, such as SIFT [20] and SURF [2], have found relatively few application proposals in medical image processing. These methods proceed by first detecting some points of interest (edges, ridges, blobs, etc.) within the image, then compute vectors describing local neighbourhoods around these points and use them as content descriptors. The approach has been successfully applied to image indexing, content-based image retrieval, object detection and recognition, and image registration [30]. In medical imaging, 3D versions of SIFT have been used in brain MR image matching [6], linear registration of radiation therapy data [1], and nonlinear (deformable) registration of thoracic CT [31] and brain MR [22] images. A SURF-based method [10] has also been successfully applied to the intermodality registration of 2D brain images. A review of keypoint-based medical image registration can be found in [28].

We propose a generic method for the automatic multiple-organ segmentation of 3D images based on multilabel graph cut optimization which uses location and intensity likelihoods of organs and prior information of their spatial configuration. The spatial prior is derived from shortest-path pairwise constraints defined on the adjacency graph of structures [14], and the organ location likelihood is defined by probabilistic atlases [24] learned from the VISCERAL training dataset [11]. We register organ atlases to the image prior to segmentation using a fast (2+1)D registration method based on SURF keypoints. Registered atlases are also used to derive organ intensity likelihoods. Prior and likelihood models are then introduced in a joint centroidal Voronoi image clustering and graph cut multiobject segmentation framework. We present the results of qualitative and quantitative evaluation of our method on contrast-enhanced CT and MR images from the VISCERAL dataset.

11.2 Methods

In the following, we present the different elements of our approach in detail, namely the keypoint-based image registration method and its use in organ atlas construction as well as its hierarchical application in segmentation. Image clustering and segmentation methods are detailed next, followed by a presentation of evaluation results in the subsequent section.

11.2.1 SURF Keypoint-Based Image Registration

We first outline our fast (2+1)D algorithm for the rigid registration of 3D medical images using content features. Our method has the following properties:

- Features are extracted in 2D volume slices. This has the advantage of being fast and easily parallelizable. Another advantage is that medical data are usually stored in a picture archiving and communication system (PACS) in the form of volume slices as opposed to full 3D volumes. Our method easily fits into such medical environments. Note that while feature extraction is done in 2D images, registration is still performed in 3D, hence the (2+1)D definition.
- Partial matching is well handled, thus making our algorithm suitable for general medical data.
- Total processing time is on the order of seconds.
- The (2+1)D paradigm currently restricts our method to image volumes with consistent orientations. A pair of images featuring patients with orthogonal orientations cannot be registered for now.

11.2.1.1 2D Feature Extraction and Matching

As previously mentioned, we extract features from 2D slices of the image volume. We currently use the SURF image descriptor [21]; however, our method is generic and would work with other descriptors as well. To reduce computation time, we first downsample the input volume to a user-specified size. As a rule of thumb, we isotropically resample each volume so that its second longest dimension is equal to the desired resolution R. For example, with $R = 100$, the VISCERAL training dataset image `10000108_1_CTce_ThAb` of dimensions $512 \times 512 \times 468$ and spacing $0.7\,\text{mm} \times 0.7\,\text{mm} \times 1.5\,\text{mm}$ is resampled to a $100 \times 100 \times 198$ volume with an isotropic spacing of 3.54 mm.

Next, we extract 2D SURF features from each slice. As these operations are completely independent, this step is carried out in a parallel manner. Figure 11.1 shows feature extraction results on a pair of axial slices from VISCERAL training dataset images `10000108_1_CTce_ThAb` (left) and `10000109_1_CTce_ThAb` (right). The top row shows all features extracted from these slices. The number of extracted features is 11500 and 9400, respectively.

Once all features are extracted, they are matched using the widely used second closest ratio criterion [20]. If the first volume V_1 contains n_1 slices and the second volume V_2 contains n_2 slices, we have to compute $n_1 \times n_2$ image matches, which again is easily carried out in a parallel fashion. The bottom row of Fig. 11.1 shows the nine matching couples (pairs of keypoints) found in both slices.

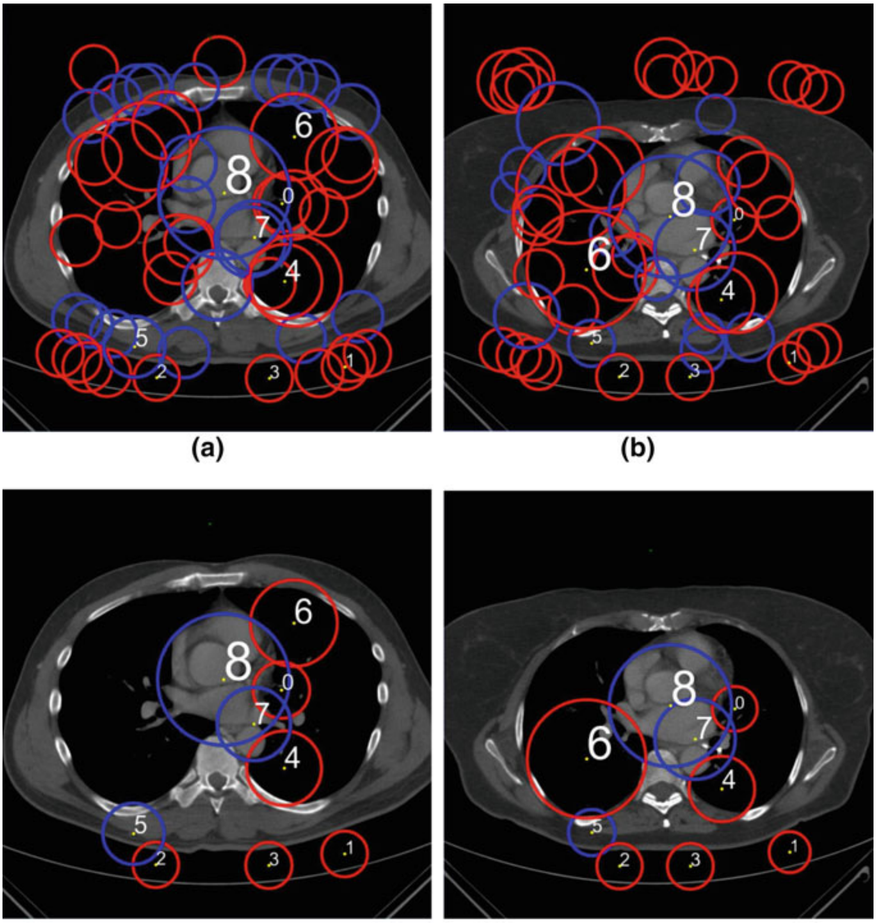
The output of the matching step is a similarity matrix S of $n_1 \times n_2$ of, possibly empty, matching couple sets $H_{i,j}$. Figure 11.2 illustrates the 2D matching procedure on test images `10000108_1_CTce_ThAb` and `10000109_1_CTce_ThAb`. Figure 11.3 shows similarity matrices reflecting the number of matching couples between any pair of slices for three settings of downsampling resolution R (grey level is inversely proportional to the number of matching couples). Matrix diagonals are clearly visible, confirming the fact that input volumes contain similar structures. In total, 2561 matching couples were found between this pair of volumes with $R = 100$.

11.2.1.2 (2+1)D Registration

Once 2D matches are found, we are able to proceed with volume registration. For robustness purposes, we use a simple "scale + translation" transformation model:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = s \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} . \tag{11.1}$$

We estimate the four parameters s, t_x, t_y and t_z in similar spirit to the RANSAC method [8], using the set of matching couples between the slices of the pair of volumes, computed as indicated in the previous subsection. RANSAC is an iterative



(a) (b)

(c) (d)

Fig. 11.1 Matching two slices of images. **a** `10000108_1_CTce_ThAb` and **b** `10000109_1_CTce_ThAb` from the VISCERAL training dataset. **a**, **b** Show all features found in both slices (a feature is represented by a *circle*). **c**, **d** Show the nine matching features between the two slices. *Blue* and *red circles* correspond to positive and negative Laplacian values [2]

parametric model estimation method known to be very efficient in the presence of outliers. One RANSAC iteration usually consists in randomly picking a small number of samples to estimate the model parameters, then counting the number of data samples consistent with the model, rejecting outliers. After performing all iterations, the model providing the highest number of consistent data samples is kept as the solution. In our case, we carry out parameter estimation in a two-stage fashion; first, we fix the parameters s and t_z which allow us to work only on a subset of S, then we estimate the remaining parameters t_x and t_y. More specifically, we carry out the n^{th} RANSAC iteration as follows:

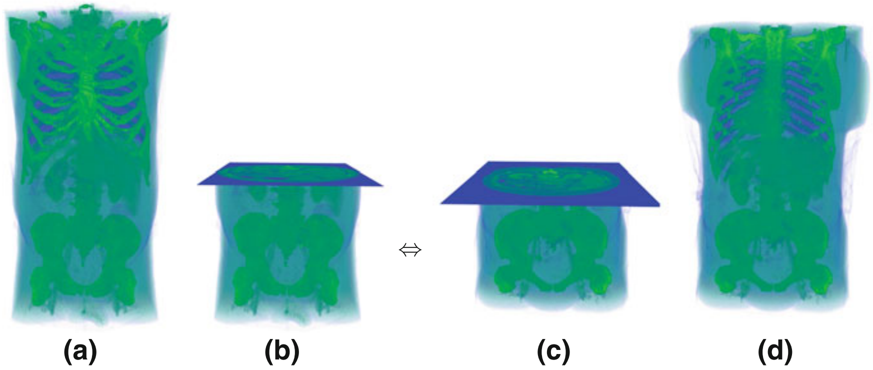


Fig. 11.2 Feature matching. The two input volumes (**a**, **b**) are sliced (**c**, **d**), and each slice from the first volume is compared against every slice from the second volume

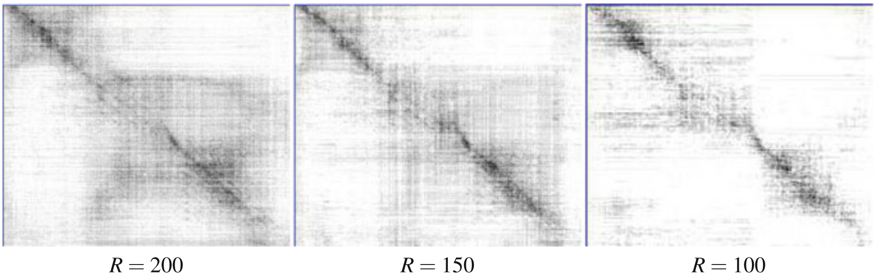
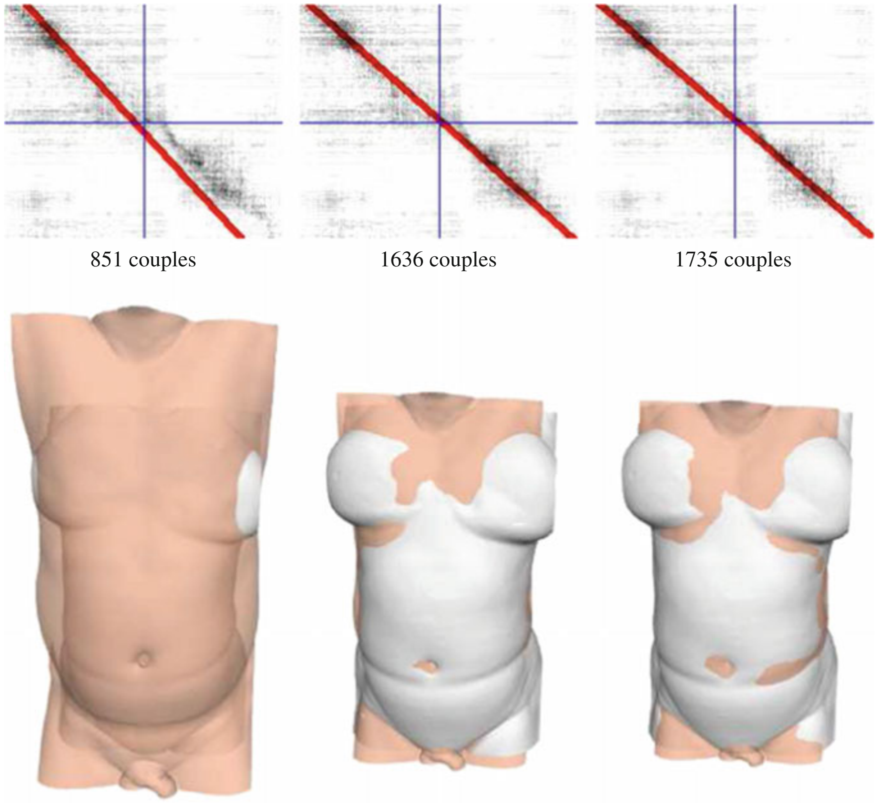


Fig. 11.3 Similarity matrices of test volumes `10000108_1_CTce_ThAb` and `10000109_1_CTce_ThAb` for different values of downsampling resolution R

- Randomly pick a line L_n crossing S. This fixes half of the transform parameters, that is, the parameters s and t_z. The top row of Fig. 11.4 shows three different randomly picked lines on S.
- Build the couple set M_n as the union of all couple sets $H_{i,j}$ in S within a distance d_L to L_n. In our experiments, we set d_L to 2.5. For the three cases illustrated in the top row of Fig. 11.4, couple sets M_n correspond to image pixels covered by the red lines.
- Randomly pick one couple from M_n, which allows to estimate the remaining transform parameters t_x and t_y.
- Count the number of couples N_n in M_n which are consistent with the transform, excluding outliers and forbidding any keypoints to appear in multiple matching couples. If f_1 and f_2 are the coordinates of a couple, consistency checking is done by transforming the coordinates of f_1 into f_1' using Eq. 11.1, and verifying that f_1' is within a fixed distance d_c from f_2. In practice, we set d_c to 20 mm.



851 couples 1636 couples 1735 couples

Fig. 11.4 (2+1)D registration of test volumes `10000108_1_CTce_ThAb` and `10000109_1_CTce_ThAb`. The *top row* shows three randomly picked lines L_n on the similarity matrix and their respective numbers of matching couples. The *bottom row* shows the corresponding registrations of test volumes, showing only patient body envelopes

In all experiments, we perform 2×10^5 iterations to register a pair of volumes. The bottom row of Fig. 11.4 illustrates the final registration of test images `10000108_1_CTce_ThAb` and `10000109_1_CTce_ThAb` showing only respective patient body envelopes.

11.2.2 Organ Atlas Construction

Using contrast-enhanced CT and MR images and available ground-truth annotations from the VISCERAL training dataset, we construct modality-specific probabilistic atlases for the following 20 structures: thyroid; trachea; sternum; liver; spleen; pancreas; gall bladder; first lumbar vertebra; aorta; urinary bladder; right and left lungs, kidneys, adrenal glands, psoas major and rectus abdominis muscle bodies. In addition, we create atlases for three additional image and body regions: background (BKG), thorax and abdomen (THAB) and a body envelope (ENV) from annotations generated automatically as follows. BKG is created by thresholding the image followed by morphological processing in order to isolate the background from the body region. THAB is created as the dilated union of the aforementioned 20 structures and their bounding 3D ellipse, from which the structures are subtracted after dilation. Finally, ENV is defined as the image minus BKG and THAB. Note that ENV is a crude body envelope that comprises skin, fat, muscle and bone structures. Figure 11.8c, f illustrate the additional annotations.

To create probabilistic atlases, we choose a representative image per modality from the dataset and use it as a reference onto which we register all remaining images in the modality via the method described in Sect. 11.2.1. We register each structure separately in a bounding box of a given margin in the intensity image, defined according to the corresponding annotation image, and apply the obtained transform subsequently to the annotation image. We accumulate annotations thus registered in a 3D histogram of reference image dimensions which is normalized to produce the corresponding probability map. Refer to Fig. 11.6a for an illustration of probabilistic atlases.

11.2.3 Image Clustering

The full-resolution voxel representation is often redundant because objects usually comprise many similar voxels that could be grouped. Therefore, we simplify the image prior to segmentation by an image-adaptive centroidal Voronoi tessellation (CVT), which strikes a good balance between cluster compactness and object boundary adherence and helps to place subsequent segmentation boundaries precisely. We have shown that the clustering step improves the overall run-time and memory footprint of the segmentation process up to an order of magnitude without compromising the quality of the result [14].

Let us define a greyscale image as a set of voxels $\mathscr{I} = \{v \mid v = (x, y, z)\}$ and associate with each voxel $v \in \mathscr{I}$ a grey level I_v from some range $I \subset \mathbb{R}$. Given a greyscale image $\mathscr{I}$ and n sites $c_i \in \mathscr{I}$, a CVT partitions $\mathscr{I}$ into n disjoint clusters C_i associated with each centroid c_i and minimizes the following energy:

$$F(v; c_i) = \sum_{i=1}^{n} \left(\sum_{v \in C_i} \rho(v) \left(\|v - c_i\|^2 + \alpha \|I_v - I_i\|^2 \right) \right) . \tag{11.2}$$

In Eq. 11.2, $\rho(v)$ is a density function defined according to the intensity gradient magnitude at voxel v, $\rho(v) = |\nabla I_v|$, α is a positive scalar and I_i is the grey level of the cluster C_i defined as the mean intensity of its voxels. Intuitively, minimizing Eq. 11.2 corresponds to maximizing cluster compactness in terms of both geometry

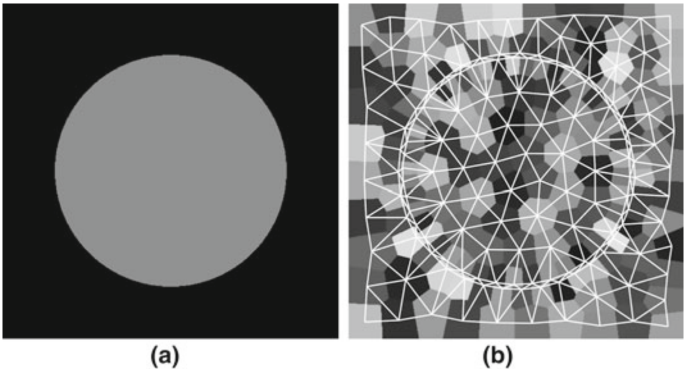


Fig. 11.5 An image-adaptive CVT clustering and its dual graph for a circle image

and grey level. Refer to Fig. 11.5 for an illustration in 2D. To minimize Eq. 11.2, we apply a variant of the clustering algorithm in [7], which approximates a CVT in a computationally efficient manner, involving only local queries on voxels located on boundaries of pairs of clusters.

For referral in later sections, we shall define the graph of a CVT, illustrated in Fig. 11.5b. Denote the surface of a cluster C_i by ∂C_i. Given a CVT clustering $\mathscr{C}$, let the set $\mathscr{S}$ index its clusters, and let $\mathscr{G} = \langle \mathscr{S}, \mathscr{E} \rangle$ be an undirected graph on cluster centroids where pairs of clusters having nonzero area common surface define the set of edges $\mathscr{E} = \big\{\{i, j\} \mid i, j \in \mathscr{S},\ |\partial C_i \cap \partial C_j| \neq 0\big\}$. Consequently, the neighbourhood of a node $i \in \mathscr{S}$ is defined as $\mathscr{N}_i = \big\{j \mid j \in \mathscr{S},\ \exists\, \{i, j\} \in \mathscr{E}\big\}$.

11.2.4 Multiorgan Image Segmentation

We formulate image segmentation as a labelling problem, defined as the assignment of a label from a set of labels L representing the structures to be segmented to each of the variables in a set of n variables, indexed by $\mathscr{S}$, corresponding to the clusters of a CVT-clustered image. Assume that each variable $i \in \mathscr{S}$ is associated with the corresponding node in the graph $\mathscr{G}$ of the CVT defined in Sect. 11.2.3. An assignment of labels to all variables is called a configuration and is denoted by $\ell \in \mathscr{L}$. An assignment of a label to a single variable is denoted by ℓ_i. We cast the labelling problem in a maximum a posteriori estimation framework and solve it by minimizing the following energy function of label configurations via the expansion moves multilabel graph cut algorithm [5], which has been shown to outperform popular multilabel optimization algorithms in terms of both speed and quality of obtained solutions [29]:

$$E(\ell) = t_1 \sum_{i \in \mathscr{S}} D_i(\ell_i) + t_2 \sum_{i \in \mathscr{S}} P_i(\ell_i) + \frac{1}{2} \sum_{i \in \mathscr{S}} \sum_{j \in \mathscr{N}_i} V_{i,j}(\ell_i, \ell_j) \ . \tag{11.3}$$

In Eq. 11.3, t_1 and t_2 are temperature hyperparameters, and $\mathscr{N}_i$ is the neighbourhood of the variable $i \in \mathscr{S}$. The first and second sums in Eq. 11.3 correspond, respectively, to organ intensity and location (atlas) likelihood energies, and the third is the energy of a prior distribution of label configurations expressed as a Markov random field [18] with respect to the graph $\mathscr{G}$. We shall define these terms in detail.

11.2.4.1 Spatial Configuration Prior

Pairwise terms of Eq. 11.3 encode prior information on interactions between labels assigned to pairs of neighbouring variables encouraging the spatial consistency of labelling with respect to a reference model. We define these terms according to the piecewise-constant vicinity prior model proposed in [14], which, unlike the standard Potts model, incurs multiple levels of penalization capturing the spatial configuration of structures in multiobject segmentation. It is defined as follows. Let $\mathscr{R}$ be the set of symmetric adjacency relations on pairs of distinct labels (corresponding to image structures), $\mathscr{R} = \{\mathrm{r} \mid a\,\mathrm{r}\,b,\ a, b \in L,\ a \neq b\}$. $\mathscr{R}$ can be represented by a weighted undirected graph on L, $\mathscr{A} = \langle L, W \rangle$, with the set of edges $W = \big\{\{a, b\} \mid \exists \mathrm{r} \in \mathscr{R},\ a\,\mathrm{r}\,b,\ a \neq b\big\}$, where edge weights are defined by $w(\{a, b\}) = 1$, such that $w(\{a, b\}) = \infty$ if $\nexists \mathrm{r} \in \mathscr{R},\ a\,\mathrm{r}\,b$.

Given the graph $\mathscr{A}$, we define the pairwise term in Eq. 11.3 as follows:

$$V_{i,j}\big(\ell_i, \ell_j\big) = |\partial C_i \cap \partial C_j|\, \omega\big(a, b\big), \quad \ell_i = a,\ \ell_j = b \ . \tag{11.4}$$

where $\omega\big(a, b\big)$ is the shortest-path weight from a to b in $\mathscr{A}$. The adjacency graph of structures according to which we define the spatial prior in our experiments is given in Fig. 11.6b. In Eq. 11.4, the area of the common surface of adjacent clusters $|\partial C_i \cap \partial C_j|$ is introduced, so that $\forall a, b \in L$ the sum of pairwise energies in (11.3) is equal to the area of the common surface between the corresponding pair of structures multiplied by the shortest-path weight. This definition ensures that the segmentation energy is independent of the CVT clustering resolution [13].

11.2.4.2 Intensity and Location Likelihoods

Unary terms of Eq. 11.3 measure the cost of assigning labels to variables. They are defined as negative log-likelihood functions derived from organ observed intensity and location probabilities:

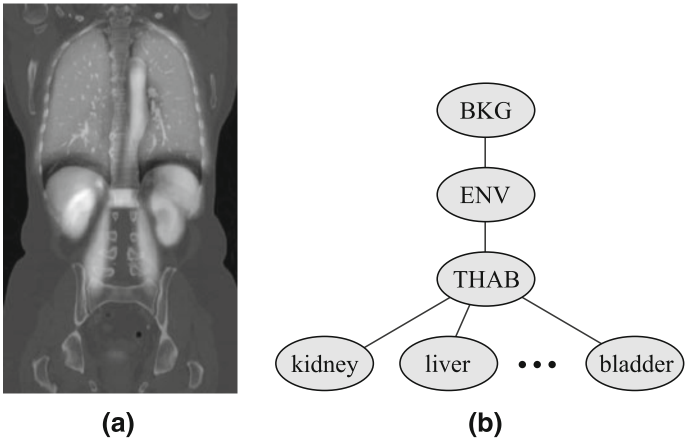


Fig. 11.6 **a** Registered organ atlases overlaid on a CT image and **b** the adjacency graph used to define the spatial prior

$$D_i(\ell_i) = -\ln \prod_{v \in C_i} \Pr(I_v \mid \ell_i)\ , \tag{11.5a}$$

$$P_i(\ell_i) = -\ln \prod_{v \in C_i} \Pr(X_v \mid \ell_i)\ . \tag{11.5b}$$

In Eq. 11.5b, X_v denotes the object-space coordinates of the voxel v. Conditional probabilities in Eq. 11.5a and 11.5b correspond, respectively, to those of voxel intensity and location given the structure ℓ_i. To estimate the conditional probability distribution $\Pr(I \mid l)$ for a given label $l \in L$, we first register the corresponding organ atlas to the image, then estimate the conditional probability as a Gauss-smoothed and normalized intensity histogram derived from voxels in high-probability regions of the registered atlas according to a threshold value. Conditional probability distributions $\Pr(X \mid L)$ are defined directly from registered atlases. The next section outlines our hierarchical registration method which maps organ atlases to an image prior to its segmentation.

11.2.4.3 Hierarchical Registration of Organ Atlases

We register probabilistic organ atlases, constructed as described in Sect. 11.2.2, to an image in a three-step hierarchical fashion starting at the full image scale, then on an intermediate level corresponding to the THAB region and finally on individual organs. After performing registration on each scale, we apply the obtained transform to the corresponding atlas as well as to those of organs contained in the registered region. As in Sect. 11.2.2, we register each structure separately in a bounding box of a given margin in the intensity image, defined according to the corresponding atlas.

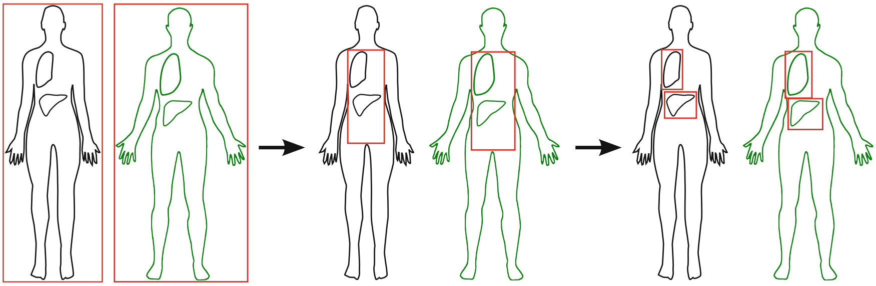


Fig. 11.7 An illustration of the proposed hierarchical registration procedure

Figure 11.7 illustrates the hierarchical registration procedure, and Fig. 11.6a gives an example of registered organ atlases overlaid on the CT image to which they have been registered.

11.3 Evaluation Results and Discussion

We have carried out qualitative evaluation on several contrast-enhanced CT and MR images from the VISCERAL training dataset. Figure 11.8 shows a pair of segmentations on images `10000109_1_CTce_ThAb` and `10000324_4_MRT1cefs_Ab`. Their dimensions, respectively, are 512 × 512 × 428 and 312 × 72 × 384. For this pair of images, the number of CVT clusters is set, respectively, to 3 and 20% of image voxel count. In all experiments, we set temperature parameters t_1 and t_2 so that intensity and location likelihood-based unary terms have the same magnitude in Eq. 11.3. Likewise, for intensity likelihood estimation in all experiments, we fix the probability threshold mentioned in Sect. 11.2.4.2 to 0.9 times that of the maximum probability of the registered probabilistic atlas. The spatial configuration prior is defined according to the adjacency graph given in Fig. 11.6b. We note that, due to a smaller field of view, VISCERAL dataset contrast-enhanced MR images exclude thoracic organs, namely the pair of lungs, the trachea, the sternum and the thyroid. Naturally, we do not construct probabilistic atlases for these structures nor do we take them into account for the segmentation of MR images.

Table 11.1 presents the results of quantitative evaluation of our segmentation method on contrast-enhanced CT images during the VISCERAL Anatomy 2 Benchmark and those of its more recent evaluation on contrast-enhanced MR images. We report results corresponding to the best setting of temperature parameters out of the allowed five. For CT images, the settings for t_1 are as follows: 0.15, 0.20, 0.25, 0.30 and 0.40. For MR images, tested settings of this parameter are as follows: 0.6, 0.8, 1.0, 1.2 and 1.4. The parameter t_2 was set to $0.2\,t_1$ in both cases. These ranges of parameter values were experimentally found to give the best results in offline evaluations on the VISCERAL training dataset. For each structure, the Table 11.1 gives the

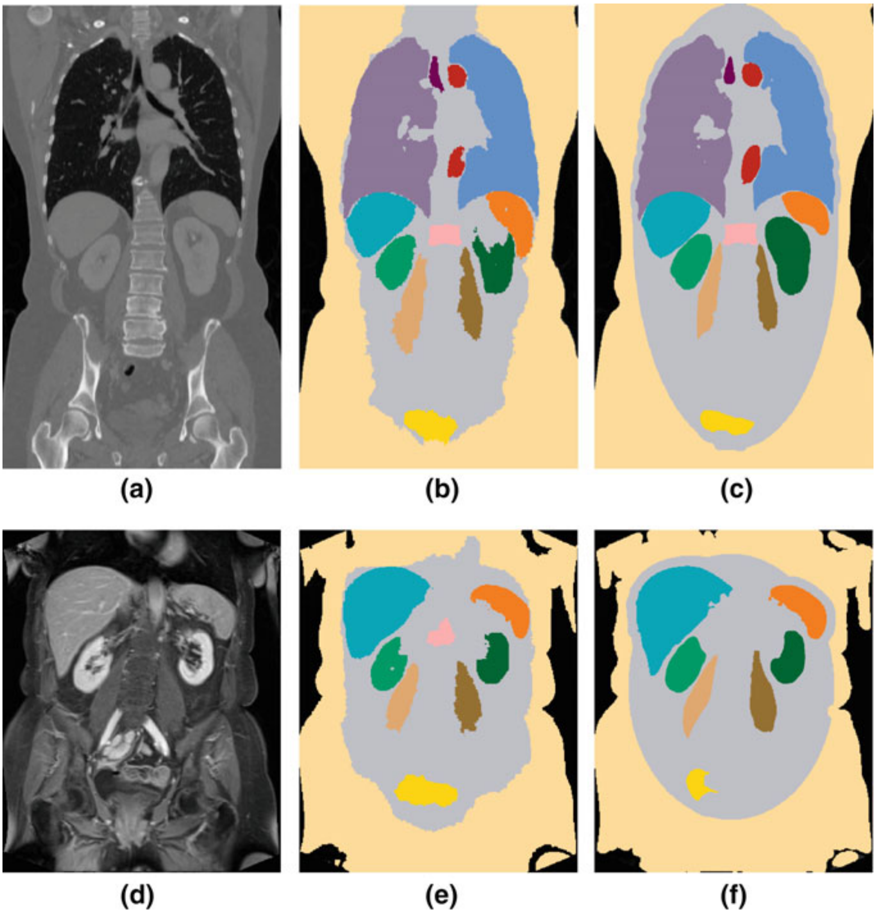


Fig. 11.8 Segmentation of VISCERAL training dataset images `10000109_1_CTce_ThAb` (top row) and `10000324_4_MRT1cefs_Ab` (*bottom row*). Coronal sections correspond to **a**, **d** the image, **b**, **e** its segmentation and **c**, **f** the associated ground truth with additional labels for BKG, ENV and THAB regions

number of produced segmentations out of an attempted 10, mean Dice and average distance (in millimeters) measurements. “N/A” indicates an absent structure, while a dash “–” indicates one for which the segmentation was missed or was not attempted.

Mean run-time and memory footprint figures of our algorithm are given in Table 11.2. These measurements are taken on the 20 contrast-enhanced CT images in the VISCERAL training dataset, the average dimension of which is 512 × 512 × 426. The number of CVT clusters is set to 5% of image voxel count. The algorithm is run on a cluster computer of heterogeneous nodes with an average CPU speed of 2.1 GHz, an average number of cores of 20 and an average memory size of 87 GB.

Table 11.1 Quantitative evaluation results of the proposed method on contrast-enhanced CT and MR images

	CT			MR		
Structures	#	Dice	Avg. Dist.	#	Dice	Avg. Dist.
Trachea	9	0.62	18.56	N/A	N/A	N/A
Lung (R)	10	0.95	0.30	N/A	N/A	N/A
Lung (L)	10	0.96	0.20	N/A	N/A	N/A
Pancreas	7	0.35	11.45	6	0.37	11.99
Gall bladder	2	0.14	21.82	1	0.30	1.90
Urinary bladder	10	0.77	1.08	10	0.40	3.67
Sternum	10	0.63	6.59	N/A	N/A	N/A
Lumbar vertebra	10	0.49	9.74	7	0.26	6.65
Kidney (R)	10	0.81	1.81	10	0.80	3.90
Kidney (L)	10	0.86	0.89	8	0.74	1.69
Adrenal gland (R)	–	–	–	–	–	–
Adrenal gland (L)	–	–	–	–	–	–
Psoas major muscle (R)	10	0.71	2.70	10	0.69	1.73
Psoas major muscle (L)	10	0.79	1.22	10	0.66	2.28
Rectus abdominis muscle (R)	9	0.26	30.25	–	–	–
Rectus abdominis muscle (L)	10	0.13	24.43	–	–	–
Aorta	10	0.58	5.43	3	0.27	17.40
Liver	10	0.93	0.34	10	0.77	1.91
Thyroid	3	0.04	13.77	N/A	N/A	N/A
Spleen	10	0.84	1.29	9	0.53	3.31

Table 11.2 Mean memory footprint and run-time figures of proposed algorithms measured on 20 contrast-enhanced CT images from the VISCERAL training dataset

Memory (MB)	Registration (s)	Clustering (s)	Segmentation (s)	Total run-time (m)
10520.87	4294.60	8995.20	2598.48	264.80

From these results, we can readily see that our method performs better on CT than on MR images. This is due to the fact that tissues in CT images have consistent appearances, whereas in MR images, they suffer intensity inhomogeneity. In addition, MR images in the VISCERAL dataset have lower resolution compared to CT images. We can observe furthermore that our method performs better on larger, well-contrasted structures than on smaller, low-contrasted ones such as the gall bladder, the thyroid and the adrenal glands. This is mainly due to the inaccurate localization of these structures by our registration method and the subsequent flawed estimation of the structure intensity likelihood. For most structures however, even though our hierarchical approach of mapping atlases to the image relies on a rigid registration method, unlike many hierarchical methods which use non-rigid deformable registration [17], it helps localizing structure boundaries in segmentation quite well. This is because location information roughly registered atlases provide is complemented by intensity similarity and spatial consistency criteria. Furthermore, full-body modelling by the introduction of BKG, ENV and THAB annotations not only complements location information and allows for hierarchical registration, but also increases the discriminative power of the spatial prior by a higher penalization of inconsistent configurations.

11.4 Concluding Remarks and Future Work

It should not go without notice that without the VISCERAL platform and the dataset, we would not have been able to test and understand the limits and the properties of our algorithms, to improve them and to develop new ones. We hope that our active participation in benchmarks and our regular feedback on software and data have been valuable for the VISCERAL project.

We are currently scrutinizing our hierarchical registration method in view of multiresolution extensions, possibly bypassing anatomical hierarchy, which would help better localize structures, especially small, low-contrasted ones. We are also investigating the introduction of a better, more robust intensity likelihood estimation method. If an inaccurate registration could be detected and quantified, then it may be possible to "correct" it. Otherwise, with a large training dataset to draw upon, techniques from machine learning could easily be used. Another interesting venue for future research is the extension of the spatial prior model to express other types of relations, such as spatial directionality, and the possibility of taking into account the uncertainty of relations.

References

1. Allaire S, Kim JJ, Breen SL, Jaffray DA, Pekar V (2008) Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. In: IEEE CVPRW, pp 1–8
2. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). Comput Vis Image Underst 110(3):346–359
3. Blume A, Chun W, Kogan D, Kokkevis V, Weber N, Petterson RW, Zeiger R (2011) Google body: 3D human anatomy in the browser. In: ACM SIGGRAPH 2011 Talks, p 19:1
4. Boykov Y, Funka-Lea G (2006) Graph cuts and efficient N-D image segmentation. Int J Comput Vis 70(2):109–131
5. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell 23(11):1222–1239
6. Cheung W, Hamarneh G (2009) n-SIFT: n-dimensional scale invariant feature transform. IEEE Trans Image Process 18(9):2012–2021
7. Dardenne J, Valette S, Siauve N, Burais N, Prost R (2009) Variational tetraedral mesh generation from discrete volume data. Vis Comput 25(5):401–410

8. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
9. Fouquier G, Atif J, Bloch I (2012) Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. Comput Vis Image Underst 116(1):146–165
10. Gu Z, Cai L, Yin Y, Ding Y, Kan H (2014) Registration of brain medical images based on SURF algorithm and RANSAC algorithm. TELKOMNIKA Indones J Electr Eng 12(3):2290–2297
11. Hanbury A, Müller H, Langs G, Weber MA, Menze BH, Fernandez TS (2012) Bringing the algorithms to the data: cloud–based benchmarking for medical image analysis. In: Catarci T, Forner P, Hiemstra D, Peñas A, Santucci G (eds) CLEF 2012. LNCS, vol 7488. Springer, Heidelberg, pp 24–29. doi:10.1007/978-3-642-33247-0_3
12. Iglesias JE, Konukoglu E, Montillo A, Tu Z, Criminisi A (2011) Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In: Székely G, Hahn HK (eds) IPMI 2011. LNCS, vol 6801. Springer, Heidelberg, pp 25–36. doi:10.1007/978-3-642-22092-0_3
13. Kéchichian R (2013) Structural priors for multiobject semiautomatic segmentation of three-dimensional medical images via clustering and graph cut algorithms. PhD thesis, Université de Lyon
14. Kéchichian R, Valette S, Desvignes M, Prost R (2013) Shortest-path constraints for 3D multi-object semi-automatic segmentation via clustering and graph cut. IEEE Trans Image Process 22(11):4224–4236
15. Kohlberger T, Sofka M, Zhang J, Birkbeck N, Wetzl J, Kaftan J, Declerck J, Zhou SK (2011) Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In: Fichtinger G, Martel A, Peters T (eds) MICCAI 2011. LNCS, vol 6893. Springer, Heidelberg, pp 338–345. doi:10.1007/978-3-642-23626-6_42
16. Kolmogorov V, Zabih R (2004) What energy functions can be minimized via graph cuts? IEEE Trans Pattern Anal Mach Intell 26(2):147–159
17. Lester H, Arridge SR (1999) A survey of hierarchical non-linear medical image registration. Pattern Recognit 32(1):129–149
18. Li SZ (2009) Markov random field modeling in image analysis. Springer, Berlin
19. Linguraru MG, Pura JA, Pamulapati V, Summers RM (2012) Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. Med Image Anal 16(4):904–914
20. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
21. Lukashevich P, Zalesky B, Ablameyko S (2011) Medical image registration based on SURF detector. Pattern Recognit Image Anal 21(3):519–521
22. Moradi M, Abolmaesoumi P, Mousavi P (2006) Deformable registration using scale space keypoints. In: Medical imaging, international society for optics and photonics, p 61442G
23. Okada T, Linguraru MG, Yoshida Y, Hori M, Summers RM, Chen Y-W, Tomiyama N, Sato Y (2012) Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations. In: Yoshida H, Sakas G, Linguraru MG (eds) ABD-MICCAI 2011. LNCS, vol 7029. Springer, Heidelberg, pp 173–180. doi:10.1007/978-3-642-28557-8_22
24. Park H, Bland PH, Meyer CR (2003) Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Trans Med Imag 22(4):483–492
25. Seifert S, Barbu A, Zhou SK, Liu D, Feulner J, Huber M, Suehling M, Cavallaro A, Comaniciu D (2009) Hierarchical parsing and semantic navigation of full body CT data. In: SPIE Medical Imaging, Lake Buena Vista, FL, USA
26. Seifert S, Kelm M, Moeller M, Mukherjee S, Cavallaro A, Huber M, Comaniciu D (2010) Semantic annotation of medical images. In: SPIE medical imaging, international society for optics and photonics, p 762, 808
27. Song Z, Tustison N, Avants B, Gee J (2006) Adaptive graph cuts with tissue priors for brain MRI segmentation. In: IEEE ISBI, pp 762–765

28. Sotiras A, Davatzikos C, Paragios N (2013) Deformable medical image registration: a survey. IEEE Trans Med Imag 32(7):1153–1190
29. Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C (2008) A comparative study of energy minimization methods for markov random fields with smoothness-based priors. IEEE Trans Pattern Anal Mach Intell 30(6):1068–1080
30. Tuytelaars T, Mikolajczyk K (2008) Local invariant feature detectors: a survey. Found Trends Comput Gr Vis 3(3):177–280
31. Urschler M, Bauer J, Ditt H, Bischof H (2006) SIFT and shape context for feature-based nonlinear registration of thoracic CT images. In: Beichel RR, Sonka M (eds) CVAMIA 2006. LNCS, vol 4241. Springer, Heidelberg, pp 73–84. doi:10.1007/11889762_7
32. Wolz R, Chu C, Misawa K, Mori K, Rueckert D (2012) Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In: Ayache N, Delingette H, Golland P, Mori K (eds) MICCAI 2012. LNCS, vol 7510. Springer, Heidelberg, pp 10–17. doi:10.1007/978-3-642-33415-3_2