Electronic Letters on Computer Vision and Image Analysis 8(1):1-14, 2009

Automatic Abdominal Organ Segmentation from CT images

Paola Campadelli*, Elena Casiraghi*, Stella Pratissoli*, Gabriele Lombardi*

* *Department of Computer Science, Universitá degli studi di Milano, Via Comelico 39/41, Milan, Italy*

Received 26 May 2008; accepted 13 January 2009

Abstract

In the recent years a great deal of research work has been devoted to the development of semi-automatic and automatic techniques for the analysis of abdominal CT images. Some of the current interests are the automatic diagnosis of liver, spleen, and kidney pathologies and the 3D volume rendering of the abdominal organs. The first and fundamental step in all these studies is the automatic organs segmentation, that is still an open problem. In this paper we propose our fully automatic system that employs a hierarchical gray level based framework to segment heart, bones (i.e. ribs and spine), liver and its blood vessels, kidneys, and spleen. The overall system has been evaluated on the data of 100 patients, obtaining a good assessment both by visual inspection by three experts, and by comparing the computed results to the boundaries manually traced by experts.

Key Words: Abdominal CT images, 3D organs segmentation, histogram analysis, graph cut, α-expansion.

1 Introduction

Imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) are nowadays a standard instrument for diagnosis of abdominal organs pathologies. Among these techniques, CT images are often preferred by diagnosticians since they have high Signal-to-Noise ratio and good spatial resolution, thus providing accurate anatomical information about the visualized structures. These good image qualities, and the advances in the digital image processing techniques, motivate the great deal of research work aimed at the development of computerized methods for the automatic abdominal organ analysis and 3D volume rendering. More precisely, some of the current interests are the automatic detection of liver cancer or other diseases [1], the measurement of kidneys and liver volumes [2], and the 3D volume rendering of abdominal organs [3].

In this paper we describe an automatic system that processes abdominal CT images, in order to satisfy the request of radiologists working in two hospitals in Milan; they need a completely automatic procedure to perform the 3D visualization of liver, spleen and kidneys (which has been shown to be helpful both for surgical planning [4], and for radiation treatment programs [5]), and to measure their volumes (liver volume is a necessary information in case of living donor liver transplantation, to choose the best candidate among different donors; spleen and kidney volumes are helpful indexes to follow the organ conditions after they have been transplanted into the receiver). The system should be robust with respect to database variations, it should have limited computational costs, so as to work on a normal PC, and it should deal with images in JPG format, since this is the format they use to export CT volumes from their Picture Archiving and Communication System (PACS).

Correspondence to: <casiraghi@dsi.unimi.it>

To realize such a system the first and fundamental step is the organs segmentation, which is currently performed by expert radiologists who trace the organs contour on each slice of the CT data with the aid of semi-automatic tools [6, 7]. These segmentation techniques let the user have a full control over the segmentation process, while having the computer do most of the detail work. For this reason, the quality of the segmentation result is heavily dependent on the operator's skill, suffering from its error and biases; as a result, the inter and intra observer repeatability is low [8] (see figure 1). This problem, together with the boring and time consuming user interaction (on average, one hour is needed to segment the liver), motivates the great deal of research work recently devoted to automatic abdominal organ segmentation. Notwithstanding, the problem is still open [9, 10], due several factors such as low contrast and blurred edges that characterize CT images; they are caused by partial volume effects resulting from spatial averaging, patient movement, beam hardening, and reconstruction artifacts. Moreover, neighboring organs (e.g. liver, spleen and stomach) might have similar gray levels, since the gray tones in CT images are related to the tissue density, often similar in different organs. In the meantime, the same organ may exhibit different gray level values both in the same patient, due to the administration of contrast media, and in different ones, for different machine setup conditions.

In [10] we reviewed the most relevant works on automatic liver segmentation from CT scans. Although several segmentation techniques have been experimented (e.g. gray level based techniques, learning techniques, model fitting techniques, probabilistic atlases, and level set) this problem is still open.
Indeed, although the gray level based techniques proposed so far [11, 12, 5, 13, 14, 15, 16, 17] obtain the most promising results, they are not robust to database variations; this is because their basic step of organ gray level estimation does not take into account the high variability observed both in the same and in different CT volumes. For this reason, when tested on larger and complex data sets, these methods' performance could decrease significantly. Besides, almost all of them still need some manual intervention and require critical parameters to be experimentally set; this obviously affects their robustness.
Learning techniques [18, 19, 20] are strongly dependent on the training set, which must be chosen carefully.
Model based techniques [21, 22, 23, 24] and probabilistic atlases [25, 26, 27, 28] suffer from the same difficulty, since they need a lot of data to be properly collected, and manually segmented, to produce the model; this is because the obtained model is strongly affected both by the training set and by user errors and biases. About their performance, it must be noted that the initial placement of the model might affect the segmentation result; moreover, these algorithms might fail when processing not standard shapes, or they might require too much computing time [24].
Regarding level set approaches [29, 30, 27, 9], their main drawback is the difficulty in defining a proper speed function, and its parameters.
Furthermore, our survey revealed that a comparison among different systems would not be meaningful due to the lack of a common dataset with its gold standard, i.e. a commonly accepted manual segmentation, and a unique measure of discrepancy between the automatic and the manual segmentation of each abdominal organ (for clarity of presentation all the adopted measures are described in appendix A). Besides, the private datasets employed by most authors are very small (less than 10 patients).
Given these considerations, and based on the fact that radiologists require a robust, simple system (that is with a limited computational cost), requiring no manual intervention, we chose to develop a gray level based technique which is easy to be defined and might produce the most robust, promising, and stable results.

Our first results were presented in [31], where we proposed two gray level based liver segmentation methods, that overcome the weaknesses of the gray level based algorithms presented in the literature. We evaluated our systems performance over a set of 40 CT images, and the best performing system achieved promising results.
In this paper we propose both an improvement and an extension of the method, that is an automatic system to segment abdominal organs (i.e. heart, bones, liver, kidneys, and spleen*). Each organ is segmented by em-

*Note that we also segment heart and bones, although these organs' analysis was not required by radiologists; this is because these structures allow to provide a more accurate spatial information about the relative location of organs of interests.

ploying a general framework composed of two steps: at first, the organ gray level is automatically estimated to produce an initial coarse segmentation by thresholding; secondly, the extracted volume is refined by morphological operators and by exploiting both gray level and edge information. The system works in a hierarchical fashion, for it uses anatomical information provided by already detected volumes to locate and extract a significant gray level sample of the organ to be segmented; as a result, the estimation of the organ gray level range, which is important for it affects the number of mis-segmented voxels, is robust and reliable.
Robustness and reliability are also ensured by the fact that we estimate each organ gray level separately on each patient, to account of gray level variabilities among different CT volumes.
The paper is organized as follows: section 2 describes the dataset used; in section 3 we report the heart and bones (i.e. spine and ribs) segmentation algorithms; section 4 describes our liver segmentation method, that accomplishes liver vessels extraction too; section 5 and 6 describe the kidney and spleen segmentation algorithms, while section 7 reports results and future work.

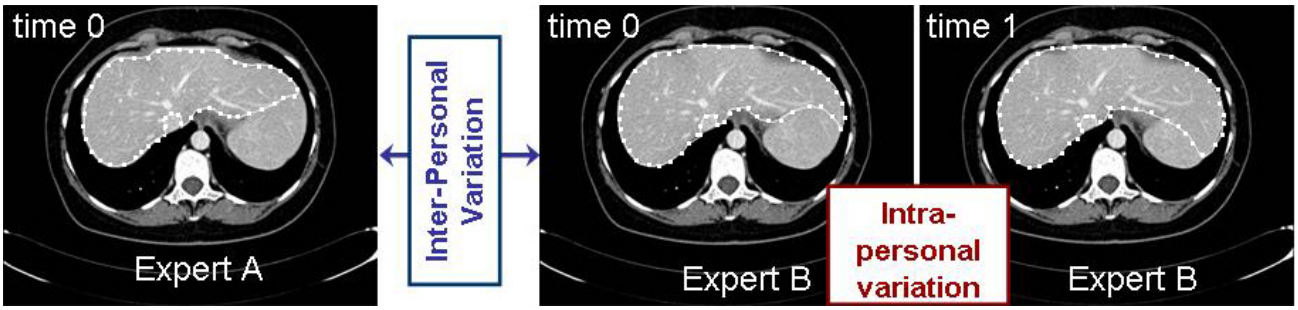

Figure 1: High variability among liver contours traced by Expert A and Expert B (inter-observer variability), and by Expert B in two different times (intra-observer variability).

2 Material and methods

Our dataset is composed of 100 abdominal contrast enhanced CT images of the portal venous phase. They have been acquired at two different hospitals in Milan, with different multi-detector spiral CT machines, after the injection of contrast material. The images are stored into a PACS system in DICOM format, with 4096 gray levels in the range $[-1024, +3071]$, directly related to the Hounsfield units (HU). For each patient a set of about 80 axial slices with a 3 mm interval is acquired, starting by the mid of the heart; each slice has a 1024×1024 pixel size, and a 0.625×0.625 mm pixel resolution.

Before processing, all the images have been scaled to the range $[0, 255]$. To this aim, the histogram of voxels whose HU (gray level) is in the range $[-500, +500]$[†] has been computed; it shows a unique peak correspnding to soft tissue voxels. Therefore, we select the HU values, H_{min} and H_{max}, that limit this peak by selecting the first two values, at its left and right side, where the first derivative is less than 25 degrees. Finally, the image to be processed $g(\mathbf{v})$ ($\mathbf{v} = (x, y, z)$) is computed from the original image $DICOM(\mathbf{v})$, as:

$$g(\mathbf{v}) = 0, \text{ if } DICOM(\mathbf{v}) \leq H_{min},$$
$$g(\mathbf{v}) = 255, \text{ if } DICOM(\mathbf{v}) \geq H_{max},$$
$$g(\mathbf{v}) = \frac{DICOM(\mathbf{v}) - H_{min}}{H_{max} - H_{min}} * 255, \text{ otherwise.}$$

Next, the images have been reduced to 256×256 pixels, and a 3×3 median filter has been subsequently applied to remove impulsive noise; notice that, as reported in section 7, the down sampling factor (that is $f_4 = \frac{1}{4}$) has been chosen experimentally, in order to reduce the computational costs without lowering the performance of

[†]The range $[-500, +500]$ HU surely includes soft tissue voxels, that is voxels belonging to abdominal organs [32].

our system, aimed at liver volume extraction.

The 3D coordinate system used in this work has the Z axis parallel to the body axis and oriented from the topmost to the bottommost slice, while the X and Y axis are oriented respectively along the width (from left to right) and the height (from top to bottom) of the 2D axial slices. While 'axial' slices are those obtained as cross sections on the Z axis, 'sagittal' and 'coronal' slices are those obtained as cross sections along the X and Y axis, respectively.
Our dataset contains patients with normal, fatty, cirrhotic, overextended livers, and livers with cancer, so that we must deal with a big anatomical and gray level variability in the data. As an example, figure 2 shows two axial slices of two patients, taken at the same vertical position on the Z axis; on their right, the gray level histograms, H_{Vol}, of the whole patients' volumes are shown, where the liver peak is pointed by an arrow.

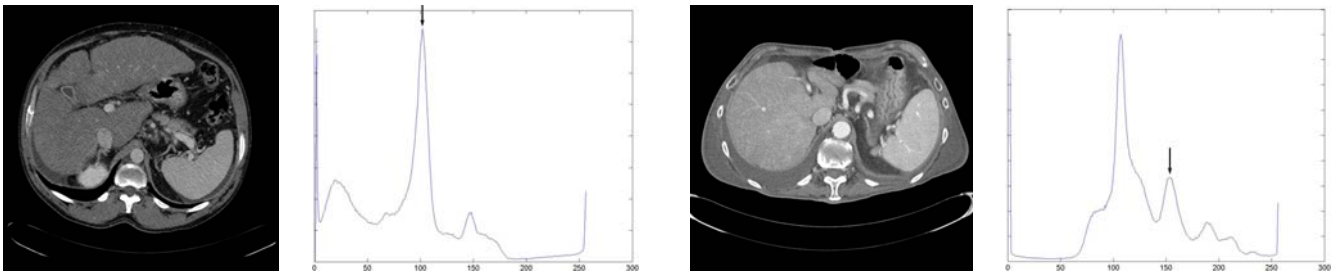

Figure 2: Axial slices of two patients, and the gray level histograms of the whole patients' volumes.

3 Heart and bones segmentation

At first we segment the 3D volume strictly including the patient's body. To extract it from the dark background, we threshold the CT data with the gray level value corresponding to the leftmost local minimum in the histogram H_{Vol}. The patient's body is contained in the 3D 'body box' that includes the biggest 3D connected component in the thresholded result (see figure 3). All the following computation steps will be applied only to this 'body box' (simply referred as CT image, or CT volume).

3.1 Heart segmentation and edge map construction

To find the heart we initially define a 2D bounding box, B_H (see figure 3, Left), based on anatomical knowledge about the heart position in the patient's body; this is used to localize, in the first axial slice, a coarse heart region, H_1, as follows:
1) convolve the image with a 2D gaussian filter (with $\sigma = 2$),
2) select the 10% of the pixels with the highest gray levels,
3) select H_1 by finding, in the thresholded image, the biggest connected region that intersects B_H.
A similar procedure is applied to each following axial slice, i, where the heart region H_i is identified by selecting, in the thresholded image, the biggest region that intersects H_{i-1}, detected in the previous slice. This process is repeated until the selected region H_i is less than $0.3 \times area(H_1)$.
The heart regions detected in successive slices form an initial 3D heart volume, V_H, that is further refined by a 3D region growing algorithm. It takes as seed points the voxels on the surface of V_H, and considers the 6-connected 3D neighborhood of each seed. Each analyzed voxel, $\mathbf{v}$, is included into the heart volume, and it is used as a new seed, if:
(i) it has not been considered yet;
(ii) its gray level $g(v)$ is such that $||g(v) - \nu|| < c\,\sigma$, where c is a constant set to 2.0, ν and σ are the mean and the standard deviation of the gray levels in V_H;

(iii) it is not an edge point in the binary edge map, $Edge_{3D}$, that is computed by applying to each axial, coronal, and sagittal slice the first order derivative of a gaussian function (with $\sigma = 0.5$) evaluated in eight directions. To keep only the significant edge pixels in each direction, the result is thresholded with hysteresis, by using 0.15 and 0.05 of the maximum gradient value as the high and the low threshold, respectively.
The region growing stops either when it finds no more voxels that can be added to the heart volume, or when a maximum number of 100 iterations has been reached.

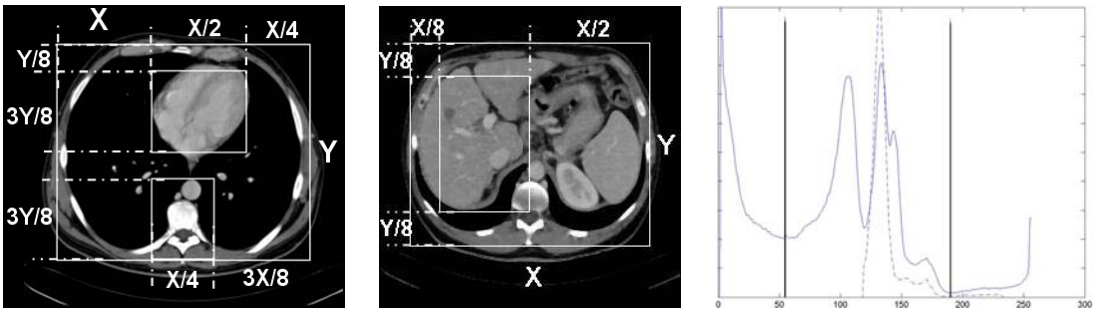

Figure 3: Left: the axial section of a patient's 'body box', the two dimensional heart, B_H, and spine, B_S, bounding boxes; in the image we show the relationships among them and the two dimensions of the 'body box' in the X-Y plane. Center: an axial section of the 3D liver bounding box used to extract the liver sample, and its relationships to the axial section of the 'body box'. Right: the gray level histogram of the whole volume H_{Vol} (solid line) and the thresholds used to identify the body box and the bones, and the gray level histogram of the liver sample (dashed line).

3.2 Bones segmentation

Bones (i.e spine and ribs) segmentation is carried out by exploiting the information provided both by gray levels and by anatomical knowledge regarding the position of ribs and spine in the body. We start by selecting the gray level value $minG_B$ corresponding to the rightmost local minimum the histogram H_{Vol} (see figure 3, Right). Next, we define a 2D spine bounding box, B_S, based on anatomical knowledge about the spine position in the patient's body (see figure 3, Left). Then, for each axial slice we compute a binary image, $Ibin$, by selecting all the pixels whose gray level is bigger than $minG_B$; after discarding from $Ibin$ all the connected regions smaller than 9 pixels, we select the spine region (it is the biggest connected region that intersects B_S), and the rib regions (their centroid is located more than $DimY/4$ pixels away from the body box centroid, where $DimY$ is the smallest dimension of the body box).
When all the axial slices have been processed, the obtained 3D binary image still contains some undesired organs, that may be attached to either the ribs or the spine. To detach and discard them, we perform a 3D binary erosion with a ball structuring element of 2 voxels radius, we select the most external volumes in the 3D result, and we finally perform a 3D dilation with the same structuring element used to erode.
Although very simple, this procedure allows to obtain a very good representation of the bones in the abdominal part, as shown in figure 7.

4 Liver segmentation

To cope with inter-patient gray level variability we process each patient separately; thus, we obtain a reliable liver gray level estimate, that is crucial as it affects the number of voxels that are wrongly segmented as liver by the following algorithm. Furthermore, we overcome problems due to intra-patient gray level variability by automatically extracting, from the patient's volume, a significant liver sample set; to this aim we exploit anatomical knowledge about liver size and position (that is below the heart, at its right side - left in the image).

Thus, we define a 3D box located below the segmented heart volume, that surely contains the liver tissue; the height of this box along the Z axis is 20 voxels, while its position and dimensions in the horizontal X-Y plane are related to the body axial slice dimensions, as shown in the center of figure 3. In the same figure the gray level histogram of the defined sample is plotted with a dashed line; it always shows a unique peak, corresponding to a narrow range of liver gray levels, that is used to correctly identify the liver peak in the histogram of the whole volume, H_{Vol} (solid line in the right of figure 3). A proper liver gray level range $[MinG, MaxG]$ is defined by finding the nearest local minima at the left and at the right side of the liver peak.
To robustly identify the liver we interpret the segmentation as a 3D labeling problem, and we search for 5 classes; this is because we noted that the volume histogram generally shows 5 peaks corresponding to liver, bones and kidneys, spleen and heart, stomach and noise, background. Labels are assigned according to both gray levels and spatial relationships between neighboring voxels, by minimizing the following energy function:

$$E(L) = \sum_{i=1..V} E_1(L(i)) + \sum_{i,j \in Neigh} E_2(L(i), L(j)) \quad (1)$$

where L is a labeling, V is the number of voxels, $E_1(L(i))$ sets the cost of assigning the label $L(i)$ to the voxel i depending on its gray level, $g(i)$, and $E_2(L(i), L(j))$ imposes spatial smoothness as it defines the cost of assigning the labels $L(i)$ and $L(j)$ to the voxels i, j in the 3D 6-connected neighborhood $Neigh$. To minimize $E(L)$ we use the α-expansion algorithm described in [33], whose formulation allows to transform a n-classes partitioning problem into a set of n binary ones, each solved via a graph-cut algorithm [34]. More precisely, given a labeling L and a particular label $\alpha \in [1, .., 5]$, another labeling L' is defined to be an α-expansion move from L if, for each voxel i, $L'(i) \neq \alpha \Rightarrow L'(i) = L(i)$. The α-expansion algorithm cycles through the 5 labels in random order, to find, for each α, a binary assignment, B, that identifies the voxels that must be labeled as α, given the current labeling, L, and the input data, $G(i)$.
In our implementation $G(i)$ is set to $G(i) = |g(i) - W|$, and scaled to the range [1,5], where W is the weighted mean of the liver gray levels: $W = \frac{\sum_{k=MinG}^{MaxG} k\ H_{Vol}(k)}{\sum_{k=MinG}^{MaxG} H_{Vol}(k)}$. The initial labeling, L_0, is produced by rounding the values $G(i)$ (so that the liver class has the lowest label). B is computed via graph-cut algorithm, to minimize the energy function

$$E(B) = \sum_{i=1..V} E_1(B(i)) + \sum_{i,j \in Neigh} E_2(B(i), B(j))$$

whose terms E_1, E_2 are defined as:

$$E_1(B(i)) = |G(i) - L(i)(1 - B(i)) - \alpha B(i)|$$
$$E_2(B(i), B(j)) = |L(i)(1 - B(i)) + \alpha B(i) - L(j)(1 - B(j)) - \alpha B(j)|$$

After each binary assignment, the new labeling, L', substitutes the current L if $E(L') <= E(L)$[‡]. The algorithm terminates when $\forall\, \alpha\ E(L') > E(L)$; the biggest volume corresponding to the lowest label is the searched liver. At this stage, the computed liver volume, $Liv(x, y, z)$, might contain parts of neighboring organs, such as heart, stomach, portal vein, and spleen. The next steps of the algorithm have been developed to remove these unwanted parts, and to smooth the boundaries.
At first, we remove those voxels contained also in the segmented heart and bones (see section 3.1). Then, we create three 2D images by projecting $Liv(x, y, z)$ onto the Y-Z, X-Z, and X-Y plane as follows:

$$SUM_X(y, z) = \sum_{i=1}^{N} Liv(i, y, z)$$
$$SUM_Y(x, z) = \sum_{i=1}^{M} Liv(x, i, z)$$
$$SUM_Z(x, y) = \sum_{i=1}^{P} Liv(x, y, i)$$

where N, M, P are the sizes of the CT data on the X,Y, and Z axis, respectively. As shown in figure 4, voxels belonging to not-liver organs can be identified because their projections have the lowest values in either one of

[‡]The terms in $E(L)$ are simply defined as: $E_1(L(i)) = |G(i) - L(i)|$, and $E_2(L(i), L(j)) = |L(i) - L(j)|$.

the three SUM images. The wrongly segmented voxels are then removed by applying the following steps to the three images separately; without loss of generality we will refer to SUM_X only:
1) in the SUM_X image, find the 2D coordinates, (y_{del}, z_{del}), of the pixels whose value is less than $\frac{max(SUM_X)}{10}$;
2) delete from the liver volume, $Liv(x, y, z)$, all the voxels v with coordinates (x, y_{del}, z_{del}), $\forall\ x = [1, .., N]$;
3) select from the resulting liver volume the 3D biggest connected component.
After smoothing the boundaries by a 3D morphological opening operation with a digital sphere with radius of 1.5 voxels, Liv is processed to consider the relationships among neighboring axial slices in the CT data. The

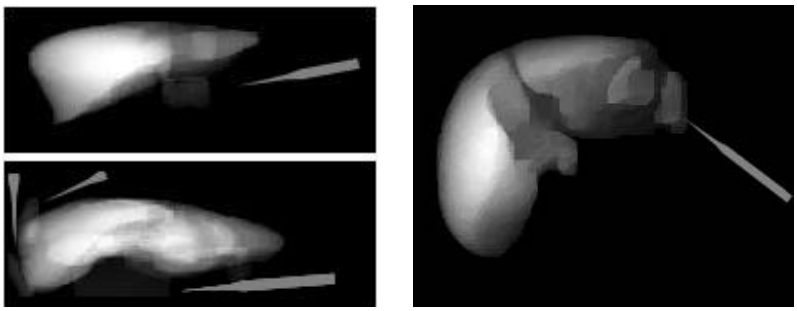

Figure 4: First column: SUM_Y and SUM_X images; Second column: SUM_Z image.

procedure that checks consecutive slices in Liv, is initialized with a starting slice, where the liver is correctly segmented. The methods presented in the literature [14] need the user to select it, so that they are affected from his errors and biases and the repeatability of the results is low; for this reason we automatically find a starting slice, by applying a compactness criterium. To this aim, for each axial slice $i = 1, .., P$, we compute the number of pixels segmented as liver in i, Num^i_{Liv}, and a coefficient, C^i_{Liv}, as:

$$C^i_{Liv} = Num^i_{Liv}/(DimX^i * DimY^i)$$

where $DimX^i$ and $DimY^i$ are the sizes of the bounding box including all the liver pixels in the i-th slice. When all the Num_{Liv} and C_{Liv} values are computed, they are normalized to the range [0, 1], and a 'compactness' factor, $Q^i = Num^i_{Liv} \cdot C^i_{Liv}$ is computed for each slice. The axial slice with the highest Q is chosen as the one to start an iterative process. It runs both toward the top and the bottom, and uses the information contained in the previous, and correctly segmented slice, S_{Prev}, to delete the regions in the current slice, S_{Curr}, that do not intersect any of the liver regions in S_{Prev}.
The same iterative procedure is repeated also along the X and the Y axes, to also consider the spatial relationships among neighboring sagittal and coronal slices.
Although these correction steps are always successful in removing the unwanted parts, they might also remove some liver voxels; to recover these regions we apply a refinement process. In the literature [9, 17], this is usually done by complex techniques (eg.: snakes, level set methods), that might take much computational time and need a cost function to be defined; besides they are applied separately to each slice, neglecting the 3D relationships among neighboring slices. To overcome all these limitations we apply the 3D region growing algorithm described in section 3.1, where ν and σ are computed based on the liver volume, c is set to 3, and the maximum number of iterations is set to 100. The 3D region growing method is simple, very fast, and considers inter-slice and intra-slice relationships.

Once the liver has been segmented an outline of the liver vessels is at first extracted by thresholding the liver voxels with hysteresis; more precisely we use, as the high and the low threshold, those values corresponding respectively to the 1% and the 5% of the liver voxels with the highest gray level values. After removing 3D connected regions smaller than 150 voxels, we produce a final vessel segmentation by applying 3D region growing, with $c = 3.0$ and a maximum number of iterations equal to 100.

5 Kidney Segmentation

The information provided by the organs segmented so far, is helpful to accomplish kidney segmentation. In particular, we exploit anatomical knowledge to firstly extract, from the CT image, a 3D 'right kidney box', B_{Rk}; its axial section is shown in the left of figure 5, it starts 20 axial slices below the first liver slice and goes until the bottommost patient's slice. To have an estimate of the kidney gray levels we remove from B_{Rk} the voxels in the segmented liver and bone volumes (their 2D contour in the starting axial slice of the kidney box is shown in the left of figure 5), and we build the gray level histogram, H_{Rk}, of the remaining voxels in B_{Rk}. Notice that, although kidneys are characterized by a wide range of gray levels, their external tissue is much brighter than the rest of the abdominal organs partially included into B_{Rk}; for this reason, the kidney gray level peak is chosen as the brightest peak in H_{Rk} (that is the rightmost, as shown in figure 5, Right), and it is limited by the nearest local minima, $minG_K, maxG_K$, at its left and its right side. Since the two kidneys are characterized by similar gray levels, $minG_K$ and $maxG_K$ are employed to detect both of them. To this aim, we at first obtain a 3D image, $Bin3D$, by binarizing the CT data with $minG_K$ and $maxG_K$ as the lower and upper thresholds, and removing the bone, heart, and liver voxels from the result. The kidney volumes in $Bin3D$ are quite fragmented, for the gray level variability of their texture; hence, we fill them by applying 3D region growing, with $c = 1.0$ and maximum number of iterations set to 100, followed by a 3D erosion operations with a $3 \times 3 \times 3$ cube.
Next, to identify the kidneys, we extract the four biggest 3D volumes in the binary image $Bin3D$: two are selected in its right half, and two in its left half. Each volume, $i = 1, .., 4$, is characterized by the normalized vector, V_i, composed of the first four moments (i.e. mean, standard deviation, skewness, kurtosis) of its gray level histogram; the left and the right kidneys are then chosen as the two most similar volumes[§], located respectively in the left and in the right half of $Bin3D$.
Finally, we refine them by applying 3D region growing, with $c = 2$ and maximum number of iterations set to 100.

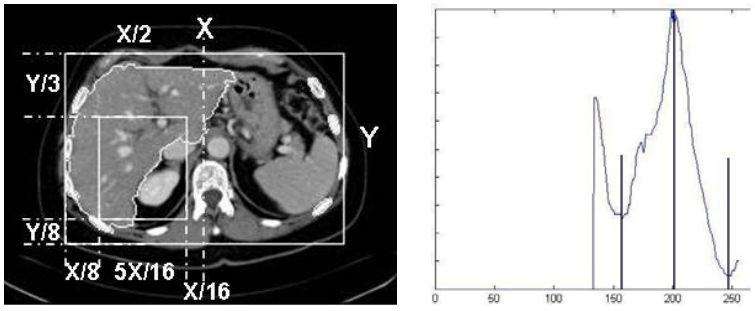

Figure 5: Left: the 'right kidney box' and its relationships to the axial slices of the body box. Right: the gray level histogram of the right kidney box where the liver and bone voxels have been removed; the rightmost peak corresponds to the kidney gray levels.

6 Spleen segmentation

Having segmented almost all the abdominal organs we have investigated whether a simple technique could be effective in finding the spleen volume.
At first, we extract a 'spleen box' whose position in the X-Y plane is shown in the left of figure 6; on the Z axis the box occupies 20 slices, starting at the 10th liver slice. The deletion of the voxels belonging to the already segmented abdominal organs allows to collect a meaningful spleen sample, whose histogram is used to estimate

[§]the similarity among two volumes, i and j, is computed as the euclidean distance among their normalized vectors, V_i and V_j.

the range, $[minG, maxG]_{spleen}$, of the spleen gray levels (see figure 6, Right). A rough segmentation is then obtained by selecting the biggest volume, inside the spleen box, composed by voxels whose gray level is in the range $[minG, maxG]_{spleen}$. Finally, to obtain a well defined spleen volume, we apply the 3D region growing algorithm, described in section 3.1 where c is set to 2, and the maximum number of iterations is set to 100.

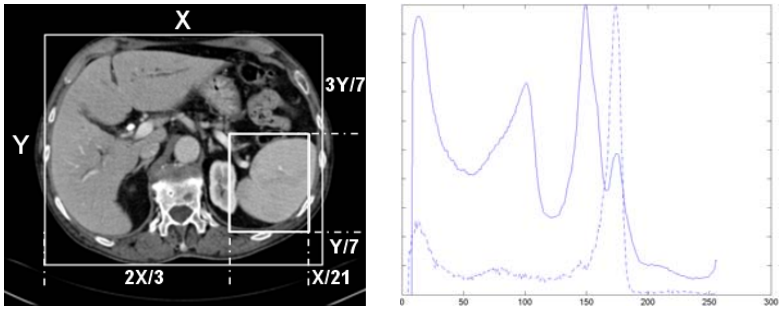

Figure 6: Left: the 'spleen box' and its relationships to the axial slices of the body box. Right: the histogram of the whole volume (solid line), and the histogram of the spleen box (dashed line); it shows a unique peak, easily identifiable.

7 Results and future work

The liver and spleen segmentation methods have been objectively evaluated by comparing the automatically detected volumes, V_{Aut}, to the ground truths, V_{Man}, manually traced by three experts, to compensate human errors and biases. We note that the ground truth has been delineated on the original images, whose size is not down-sampled. Therefore, the automatically detected volumes must be resized, to have the original size (that is 1024×1024 pixels per slice), before the performing any comparison.

To account of both over segmentation and under segmentation errors we measured the average 'symmetric volume overlap' $SVO = 1 - SVD$ (see appendix A, equation 2); moreover, to get an exhaustive evaluation of the system performance in terms of over segmentation and under segmentation errors, we measured the average sensitivity ($Sens_{Ratio}$) and overlap ($Over_{Ratio}$) ratios.

Regarding the liver segmentation algorithm, we achieved average $SVO = 95.01\%$ ($\pm 2.6\%$), $Sens_{Ratio} = 95.42\%$ ($\pm 2.7\%$), and $Over_{Ratio} = 94.87\%$ ($\pm 3.4\%$). Visual inspection of the obtained results (see figure 7) shows that over segmentation errors are due to voxels belonging either to the stomach, or to the body muscles; on the other hand, some parts of the liver might be lost near the liver edges in case of fatty liver. Nevertheless, it must be noted that the liver is always properly segmented in its first and last slices, that are generally the most critical.

The good quality of these results is proved by the fact that they are comparable to both the mean intra-personal (96%) and inter-personal variation (95%). These two measures were evaluated on the available manual segmentations, by computing the SVO between two liver volumes of the same patient, produced respectively by the same expert in two different times, and by two different experts. Besides, our results are comparable to those obtained in [24], where the author achieves a mean SVO of about 95%. Indeed, the author himself specifies that his dataset contains only livers with standard shapes, and without any pathology; hence this comparison might not be fair.

Future work will be aimed at improving the robustness of the liver segmentation algorithm by integrating the edge information in the energy function used by the α-expansion algorithm.

Regarding the spleen segmentation the achieved performance is also promising; indeed, we obtain on average $SVO = 94.17\%$ (± 3.6), $Sens_{Ratio} = 94.97\%$ (± 2.5), and $Over_{Ratio} = 93.68\%$ (± 3.9), meaning that the segmentation errors are mainly due to over segmentation.

Unfortunately, the liver vessels, heart, bones, and kidneys segmentation results could be evaluated only by visual inspection, since no ground truth was provided; radiologists found no noticeable errors.

In the following we report the experiments performed to choose the down-sampling factor. At first we must note that image sub-sampling clearly deletes image details, thus reducing the accuracy of the organs' surface location. Nevertheless, our application has two main purposes: to compute a good estimate of the abdominal organs' volume, and to extract their segmentation to allow their volume rendering. Therefore, details such as the precise vessel location, can be lost without affecting the quality of the required result. To choose the best sub-sampling factor, we run experiments on images reduced with the following factors[¶], and compared the SVO values achieved[‖] for liver and spleen segmentation:
$f_1 = 1$ (original size),
$f_2 = \frac{1}{2}$ (each slice has 512×512 pixels),
$f_4 = \frac{1}{4}$ (each slice has 256×256 pixels).
We noted that the system achieves comparable performance when the down-sampling factor is either f_1 ($SVO_{Liver} = 95.24\%$, $SVO_{Spleen} = 94.38\%$), or f_2 ($SVO_{Liver} = 94.99\%$, $SVO_{Spleen} = 94.63\%$), or f_4 ($SVO_{Liver} = 95.01\%$, $SVO_{Spleen} = 94.17\%$). Furthermore, if f_4 is used as the down-sampling factor the system takes about 90 seconds when running on a Pentium IV, 3,2GHZ/775[**], and this time is obviously increased when f_1 or f_2 are used. Given these considerations, radiologists are currently experimenting the system, described in this paper, that employs slices down-sampled to 256×256 pixels.

7.1 Conclusion

In this paper we have presented a fully automatic system which extracts the volume of the biggest organs from abdominal CT scans. The system is currently being experimented, in the clinical practice, to perform organ visualization and, above all, to compute an estimate of the organs' volumes, which is helpful in cases of living donor transplantation.
Note that, since the system works on sub-sampled images to reduce the computational costs, it cannot achieve high precision results. Therefore, it might still be improved, e.g. by applying a multi-scale approach, to segment also fine organs' details, such as the hepatic vascular system; an increased segmentation precision would improve the quality of visualization, and it is necessary to perform automated diagnosis of different pathologies.

Finally, we note that a necessary condition for an objective comparison among different volume extraction methods is the choice of a performance measure to be adopted as a reference; of course a common database with its gold standard is equally important.

A Appendix: Performance evaluation measures

The evaluation measures used by different authors can be divided into three classes.
1) The volume error based measures compare the computer output to those drawn by radiologists either by comparing the automatic to the ground truth volume ($VolErr_{3D}$), or by measuring the percentage of error among the automatic and the manual liver areas in each slice, and then averaging over all the patient's slices ($VolErr_{2D}$). These evaluation measures completely neglects the shape similarity; the ground truth and the automatically segmented liver are considered equal if they have the same volume.
2) The *overlap* ($Over_{Ratio}$) and *sensitivity* ($Sens_{Ratio}$) *ratios* measure the percentage of mismatching voxels between the automatic and manual segmentation. They are computed by normalizing the number of cor-

[¶] The size of the morphological operators employed by the algorithm have been changed according to the size of the down-sampled images being processed.

[‖] Note that the comparison has always been performed with the high resolution ground truth volume for each case.

[**] α-expansion and graph cut algorithms are implemented in C++, while the other steps are implemented in Matlab.

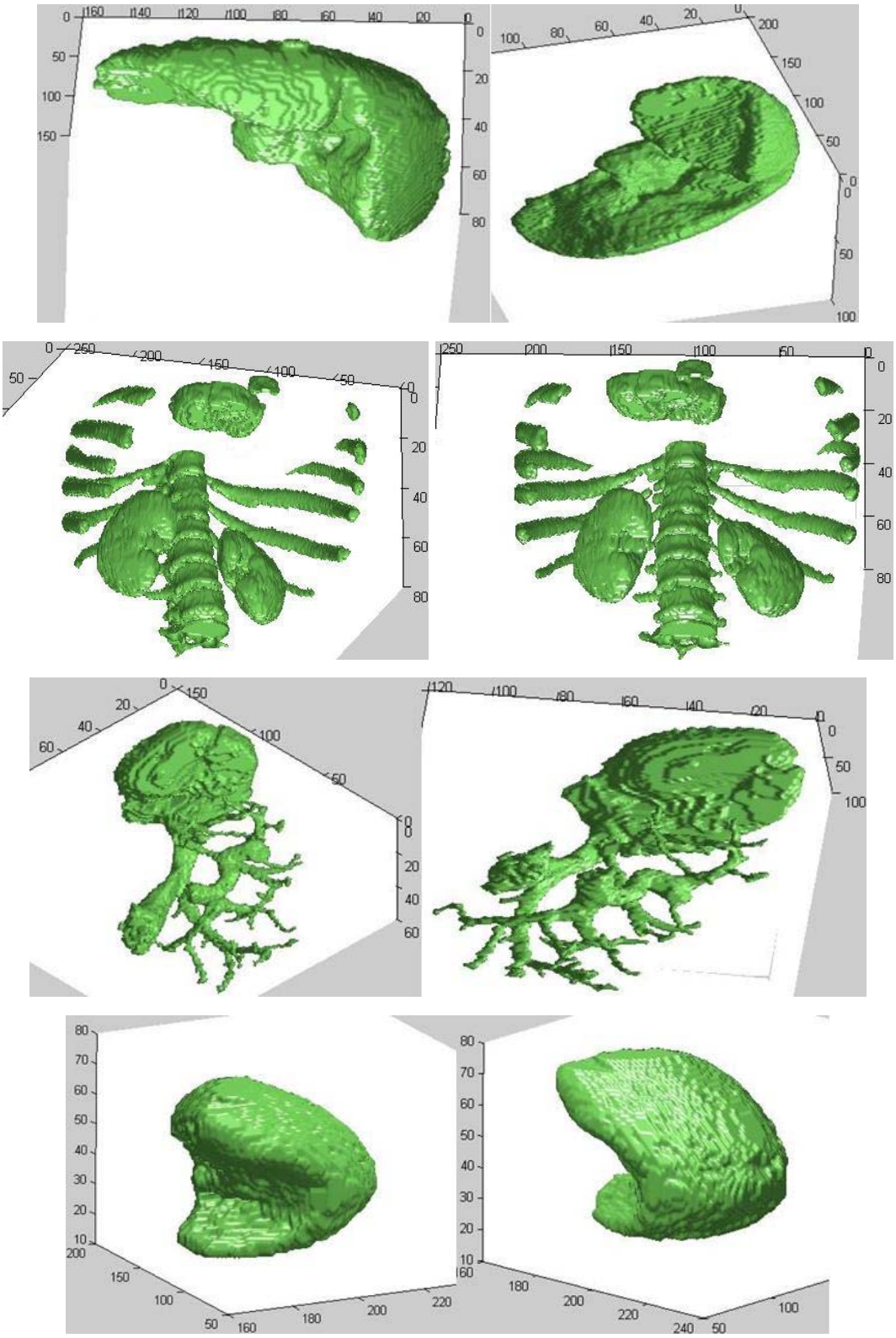

Figure 7: First row: 3D rendered liver volumes; second row: the 3D rendering of a patient's heart, spine, and kidney volumes; third row: the heart and its vessels, and the liver vessels; fourth row: the 3D rendered spleen volume.

rectly segmented liver voxels with respect to either the manual ($Over_{Ratio}$) or the automatic segmentation ($Sens_{Ratio}$). In [24] Lamecker noted that, being normalized with respect to either the ground truth or the computed volume, the first measure accounts of under segmentation errors, while the second detects the over segmentations only. Therefore, he introduced a symmetric 3D volume difference (SVD) as:

$$SVD = 1 - \frac{|V_{Aut} \cap V_{Man}|}{\frac{1}{2}(|V_{Aut}| + |V_{Man}|)} \tag{2}$$

where V_{Aut} and V_{Man} are respectively the computed liver binary volume and the binary volume of the ground truth.
3) The distance based measures evaluate distances from the points on the automatically detected liver contour (liver surface) to the points on the contour (surface) of the ground truth. The 2D (3D) distance from a point x to a contour (surface) S is usually computed as:

$$d_S(x, S) = min_{p \in S}(d(x, p))$$

where $d(\cdot, \cdot)$ may be any 2D (3D) point-to-point distance. In the 2D case some authors employ the 3/4 Chamfer distance [35], or the city block distance [36], for they are faster to compute than the euclidean one, which is generally used in the 3D case. The distances are then summed over all the points of the automatically obtained contour (surface), and the result is normalized with respect to either the ground truth or the automatic result (these details are often not specified by the authors). In [24] the author noted that also these measures are not symmetric, and he proposed an average symmetric surface distance as:

$$d_{Mean}(S, S') = \frac{1}{|S| + |S'|} \left(\int_{x \in S} d_S(x, S) + \int_{x \in S'} d_S(x, S') \right)$$

where $|S|$ denotes the area of the surface.

References

[1] C.-C. Lee, P.-C. Chung, Y.-J. Chen, Classification of liver diseases from ct images using bp-cmac neural network, Proceedings of 9th International Workshop on Cellular Neural Networks and Their Applications (2005) 118–121.

[2] Y. Nakayama, Q. Li, S. Katsuragawa, R. Ikeda, Y. Hiai, K. Awai, S. Kusunoki, Y. Yamashita, H. Okajima, Y. Inomata, K. Doi, Automated hepatic volumetry for living related liver transplantation at multisection ct., Radiology 240 (3) (2006, September) 743–748.

[3] J. Harms, M. Bartels, H. Bourquain, H. Peitgen, T. Schulz, T. Kahn, J. Hauss, J. Fangmann, Computerized ct-based 3d visualization technique in living related liver transplantation., Transplantation Proceedings 37 (2005) 1059–1062.

[4] S.-T. Fan, C.-M. Lo, C.-L. Liu, B.-H. Yong, J. K.-F. Chan, I.-L. Ng, Safety of donors in live donor liver transplantation using right lobe grafts., Archives of surgery 135 (2000) 336–340.

[5] M. Kobashi, L. G. Shapiro, knowledge-based organ identification from ct images., Pattern Recognition 28 (4) (1995) 475–491.

[6] A. Schenk, G. Prause, H. Peitgen, Efficient semiautomatic segmentation of 3d objects in medical images., Proceedings of Medical image computing and computer-assisted intervention (MICCAI 2000) 1 (2000) 186–195.

[7] A. Schenk, G. Prause, H. Peitgen, Local cost computation for efficient segmentation of 3d objects with live wire., Proceedings of SPIE Medical Imaging 4322 (2001) 1357–1364.

[8] L. Hermoye, I. Laamari-Azjal, Z. Cao, L. Annet, J. Lerut, B. Dawant, B. Van Beers, Liver segmentation in living liver transplant donors: Comparison of semiautomatic and manual methods., Radiology 234 (2005) 171178.

[9] A. Foruzan, R. Zoroofi, Y. Sato, M. Hori, T. Murakami, H. Nakamura, S. Tamura, Automated segmentation of liver from 3d ct images., International Journal of Computer Assisted Radiology and Surgery (CARS 2006) 1 (7) (2006) 71–73.

[10] P. Campadelli, E. Casiraghi, E. Esposito, Liver segmentation from ct scans: a survey and a new algorithm., To appear in Journal of Artificial Intelligence in Medicine.

[11] K. T. Bae, M. L. Giger, C. T. Chen, C. E. Kahn, Automatic segmentation of liver structure in ct images., Medical Physics 20 (1993) 71–78.

[12] L. Gao, D. Heath, B. Kuszyk, E. Fishman, Automatic liver segmentation technique for three-dimensional visualization of ct data., Radiology 201 (1996) 359–364.

[13] T. Kaneko, L. Gu, H. Fujimoto, Recognition of abdominal organs using 3d mathematical morphology., Systems and Computers in Japan 33 (8) (2002) 75–83.

[14] F. Liu, B. Zhao, P. Kijewskic, L. Wang, L. Schwartz, Liver segmentation for ct images using gvf snake, Medical Physics 32 (12) (2005, December) 3699–3706.

[15] S.-J. Lim, Y.-Y. Jeong, C. Lee, Y.-S. Ho, Automatic segmentation of the liver in ct images using the watershed algorithm based on morphological filtering., Proceeding of SPIE 5370 (2004) 16581666.

[16] S.-J. Lim, Y.-Y. Jeong, Y.-S. Ho, Segmentation of the liver using the deformable contour method on ct images., Proocedings of SPIE, Lecture Notes in Computer Science 3767 (2005) 570–581.

[17] S.-J. Lim, Y.-Y. Jeong, Y.-S. Ho, Automatic liver segmentation for volume measurement in ct images., Journal of Visual Communication and Image Representation 17 (4) (2006, August) 860–875.

[18] J. E. Koss, F. D. Newman, T. K. Johnson, D. L. Kirch, Abdominal organ segmentation using texture transforms and a hopfield neural network., Correspondences of IEEE Transaction on Medical Imaging 18 (7) (1999, July) 640–648.

[19] C.-C. Lee, P.-C. Chung, Recognizing abdominal organs in ct images using contextual neural network and fuzzy rules, Proceedings of the 22th Annual EMBS International Conference 0-7803-6465-1 (2000) 1745–1748.

[20] C.-C. Lee, P.-C. Chung, H.-M. Tsai, Identifying multiple abdominal organs from ct image series usinmg a multimodule contextual neural network and spatial fuzzy rules, IEEE Transaction on Information Technology in Biomedicine 7 (2003, September) 208–217.

[21] L. Gao, D. Heath, E. Fishman, Abdominal image segmentation using three-dimensional deformable models., Investigative Radiology 33 (6) (1998) 348–355.

[22] J. Montagnat, H. Delingette, Volumetric medical images segmentation using shape constrained deformable models., CVRMed-MRCAS, Springer Verlag Publisher LNCS 1205 (1996) 13–22.

[23] L. Soler, H. Delingette, G. Malandain, J. Montagnat, N. Ayache, C. Koehl, O. Dourthe, B. Malassagne, M. Smith, D. Mutter, J. Marescaux, Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery., Computed Aided Surgery 6 (3) (2001) 131–142.

[24] H. Lamecker, T. Lange, M. Seebass, Segmentation of the liver using a 3d statistical shape model., ZIB-Report 04-09 04-09 (2004, April) 1–25.

[25] J. Boes, T. Weymouth, C. Meyer, Multiple organ definition in ct using a bayesian approach for 3d model fitting., Proceedings of SPIE, Vision Geometry IV. 2573 (1995) 244–251.

[26] H. Park, P. Bland, C. Meyer, Construction of an abdominal probabilistic atlas and its application in segmentation, IEEE Transactions on Medical Imaging 22 (4) (2003, April) 483–492.

[27] A. Shimizu, R. Ohno, T. Ikegami, H. Kobatake, S. Nawano, D. Smutek, Multi-organ segmentation in three dimensional abdominal ct images, International Journal of Computer Assisted Radiology and Surgery (CARS 2006) 1 (7) (2006) 76–78.

[28] X. Zhou, T. Kitagawaa, K. Okuo, T. Hara, H. Fujita, R. Yokoyama, M. Kanematsu, H. Hoshi, Construction of a probabilistic atlas for automated liver segmentation in non-contrast torso ct images., Proceedings of Computed Assisted Radiology and Surgery (CARS 2005) (2005) 1281 (2005) 1169 1174.

[29] S. Pan, B. Dawant, Automatic 3d segmentation of the liver from abdominal ct images: a level-set approach., Proceedings of SPIE Medical Imaging 2001 4322 (2001) 128–138.

[30] A. Shimizu, T. Kawamura, Y. Mekada, Y. Hayashi, D. Deguchi, S. Nawano, Preliminary report of cad system competition for liver cancer extraction from 3d ct imaging and fusion of the cads., International Journal of Computer Assisted Radiology and Surgery 1 (2005) 525–526.

[31] P. Campadelli, E. Casiraghi, G. Lombardi, S. Pratissoli, Automatic liver segmentation from abdominal ct scans., Proceedings of International Conference on Image Analysis and Processing (ICIAP2007).

[32] S. Jackson, R. Thomas, CROSS-SECTIONAL IMAGING MADE EASY, Elsevier Health Sciences, 2004.

[33] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 26 (9) (2004) 1124–1137.

[34] V. Kolmogorov, R. Zabih, What energy functions can be minimized via graph cuts?, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2) (2004) 147–159.

[35] G. Borgefors, Hierarchical chamfer matching: a parametric edge matching algorithm., IEEE Transaction on Pattern Analysis and Machine Intelligence 10 (6) (1988) 849–865.

[36] A. Rosenfeld, A. Kak, Digital Picture Processing 2 (1982) 205–219.