Improvement of Liver Segmentation by Combining High Order Statistical Texture Features with Anatomical Structural Features

Suhuai Luo¹, Xuechen Li¹, Jiaming Li²
¹The University of Newcastle, Australia
²The CSIRO ICT Centre, Australia
Email: suhuai.luo@newcastle.edu.au

1. Introduction

Automatic and accurate liver segmentation in medical images such as computed tomography (CT) and magnetic resonance imaging (MRI) is one of the most important concentrations in medical image processing. Segmentation of liver from its surrounding organs and tissues are a crucial yet very difficult task in building a surgical planning system for liver transplantation and resection. This is because the boundary between the liver and its neighbouring structures such as the heart is sometimes barely noticeable in CT images, and the liver is nonrigid in shape and variant in position.

Various algorithms have been proposed to deal with liver segmentation, including live wire-based, gray level-based, neural networks-based, model fitting-based, probabilistic atlas-based, graph cut, deformable model-based, level set-based, and machine learning-based [1-9]. Although much progress has been achieved in recent years, challenges remain on the aspects of segmentation accuracy, robustness and automation.

This paper presents an automatic liver segmentation by combining high order statistical texture features with anatomical structural features. Section two describes the algorithm in detail, including texture analysis, liver distribution image calculation with support vector machines, and liver organ localization with a group of morphological operations. Section three gives the details of the liver segmentation experiment on CT images, including the experiment setting, performance validation and discussion, and future work in the area.
Abdominal CT images

Texture feature analysis

Liver distribution image calculation

Liver region localization

Segmented liver organ

Figure 1. The diagram of the proposed automatic liver segmentation.

2.1. Texture Feature Analysis

In abdominal CT images, liver shows similar grey scales and textures to its neighbouring structures such as heart and stomach. Therefore, three important considerations are taken in developing the proposed algorithm. First, liver segmentation based on greyscale parameters alone is not sufficient; second, high order texture parameters can better deal with liver segmentation; lastly, optimal liver segmentation can be achieved when global anatomical structural features are used.

Since homogeneity and consistency characterize liver segmentation where multiple slices and different patients are dealt with, texture features are considered. Textures are complex visual patterns that are composed of entities or subpatterns that have characteristic brightness, colour, slope, size, etc. [10]. It can be regarded as a similarity grouping in an image. Methods of texture analysis can be broadly classified into four categories, including: structural approach, which represents texture by defined primitives; statistical approach, which represents texture by non-deterministic properties that govern the distributions and relationships between grey levels of an image; model-based approach, which uses fractal and stochastic models to interpret image texture; and transform approach, which represents image in a space where texture is well characterised. The statistical approach and the transform approach are investigated and adopted in the proposed segmentation method, mainly based on the facts that the statistical approach has the advantage of representing texture inexplicitly and the transform approach has the advantage of representing texture at various scales.

2.1.1. Grey Level Co-occurrence Matrix and Haralick Texture Descriptors

In characterising the distribution and relationship of pixels, i.e., texture, in a grey scale image, the joint probability distribution of pairs of pixels is used. It is defined as the co-occurrence matrix. Its normalised form is noted as $C_{ij}(d, \theta)$ [11].

In an image $I_m$ of size $H$ by $W$ pixels and with $G$ intensity levels, for every pixel centered on a neighborhood $I(x,y)$ of size $N$ by $N$, $C_{ij}(d, \theta)$ is defined as the total numbers of times that, within the $N$ by $N$ neighborhood:

$$I(x_i,y_j) = i \text{ and } I(x_1+d\cos\theta, y_1+d\sin\theta) = j \quad (1)$$

where $x_i = 0, 1, ..., N-I$, is the row number in the neighborhood; $y_j = 0, 1, ..., N-I$, is the column number in the neighborhood; $i = 0, 1, ..., G-I$, is the row number in the co-occurrence matrix; $j = 0, 1, ..., G-I$, is the column number in the co-occurrence matrix; $d = 1, ..., N-I$, is the displacement distance along $\theta$; $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, is the angle between the pair.

Figure 2 illustrates the calculation of $C_{ij}(d, \theta)$ at a pixel in an abdominal CT image. The co-occurrence matrix is square and has a size $G$, which is the intensity level in original image. Note that in Figure 2 (b), for the sake of easy description, the maximal intensity value is supposed to be 5, much less than that of original liver CT image. For a normalised CT image, the maximal intensity value is usually 255. In the figure, the curved line connecting (b) and (c) shows how $C_{00}(1, 0)$ is calculated: for $d = 1$ and $\theta = 0$, within the 5 by 5 neighborhood, the total number of times that a pair of 0-value pixels appears is one, so $C_{00}(1, 0) = 1$.

![Figure 2](image-url)
The co-occurrence matrix basically keeps track of all the pixel-pair counts. It is also called spatial dependence matrix. Since representing an image with its co-occurrence matrix will result in much more data (e.g., for an image of size 1024 by 1024 pixels and with 256 intensity levels, the co-occurrence matrix will be of size 256 by 256 by 1024 by 1024), a set of features with much less size yet reflecting the co-occurrence characters was proposed, known as Haralick texture descriptors [11]. The nine Haralick texture descriptors can be defined and calculated as below.

Entropy: measures the randomness of gray-level distribution:

\[- \sum_i \sum_j C_{ij} \log C_{ij}\] (2)

Energy: measures the occurrence of repeated pairs:

\[\sum_i \sum_j C_{ij}^2\] (3)

Contrast: measures the local contrast:

\[\sum_i \sum_j (i-j)^2 C_{ij}\] (4)

Sum Average: measures the average of the gray-level:

\[\frac{1}{2} \sum_i \sum_j (iC_{ij} + jC_{ij})\] (5)

Variance: measures the variation of gray-level distribution:

\[\frac{1}{2} \sum_i \sum_j ((i - \mu_i)^2 C_{ij} + (j - \mu_j)^2 C_{ij})\] (6)

Correlation: measures a correlation of pixel pairs:

\[\frac{\sum_i \sum_j (i - \mu_i)(j - \mu_j) C_{ij}}{\sqrt{\sigma_i^2 \times \sigma_j^2}}\] (7)

Maximum Probability (MP): gives the most predominant pixel pair:

\[\frac{\max_{i,j} C_{ij}}{2}\] (8)

Inverse Difference Moment (IDM): measures the smoothness:

\[\sum_i \sum_j C_{ij} \frac{1}{1+|i-j|}\] (9)

Cluster Tendency: measures the grouping of pixels that have similar gray-level values:

\[\sum_i \sum_j ((i - \mu_i + j - \mu_j)^2 C_{ij}\] (10)

where \(C_{ij}\) is the normalised co-occurrence matrix with displacement distance \(d\) and angle \(\theta\); \(\mu_i, \mu_j, \sigma_i^2, \text{ and } \sigma_j^2\) are the means and variance of row and column in \(C_{ij}(d, \theta)\).

2.1.2. Wavelet Coefficients

Wavelet coefficients are the output of wavelet transform (WT) [12] which is the decomposition of a signal into a set of basis functions consisting of contractions, expansions and translations of a mother wavelet \(\psi\).

The wavelet transform of a signal \(f(s)\) is defined as

\[Wf(u,s) = \int_{-\infty}^{\infty} f(t) \overline{\psi'}(\frac{t-u}{s}) dt\] (11)

where the mother wavelet \(\psi\) is a zero average function, centered around zero with a finite energy. The family of vectors is obtained by translations and dilatations of the mother wavelet:

\[\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi(\frac{t-u}{s})\] (12)

In image processing applications, the wavelet transform is usually computed with dyadic wavelet transform which is implemented by filter banks. The filtering is done along both row and column with pairs of lowpass filter and highpass filter [12]. Figure 3 illustrates the process of deriving wavelet coefficients for an image using the dyadic wavelet transform. Figure 3 (a) gives a one-scale wavelet decomposition result which has four blocks of components: LL is the downsampling of the lowpass filtering along both row and column, LH is the downsampling of the lowpass filtering along row and highpass filtering along column, HL is the downsampling of the highpass filtering along row and lowpass filtering along column, and HH is the downsampling of the highpass filtering along both row and column. Such filtering or decomposition can be done further on LL, resulting a two-scale wavelet decomposition of an image as shown in Figure 3 (b). Note that the number of total wavelet coefficients equals to the number of the pixels in the image, no matter being a one-scale decomposition or two-scale decomposition. In general, there will be \(4 + 3^S\) \((S-1)\) blocks for an \(S\)-scale decomposition.
Comparing to other transforms such as Fourier [13] and Gabor [14], Wavelet transform has two advantages in segmentation application. One is that it can represent textures at the most suitable scale by varying the spatial resolution. The other is that wavelets best suit for texture analysis in a specific application can be chosen because of a wide range of choices for the wavelet function.

2.1.3. Combining High Order Statistical Texture Features with Anatomical Structural Features

As discussed before, the grey level co-occurrence matrix and related Haralick texture descriptors are second-order statistical texture features. They have the advantages of describing the statistic relationships among neighboring pixels. However, in practical segmentation applications, they are confined by two factors – small local range coverage and huge computation load. The small local range coverage is the fact that the co-occurrence matrix is calculated within a neighboring of $N$ by $N$ pixels. $N$ is usually a single digit value, e.g., 5, considering the computation load. To consider the statistical texture representations for an image, the computation load is huge. For example, for an image of size 1024 by 1024 pixels and with 256 intensity levels, if three kinds of pixel-pairs are considered (i.e., the displacement distances $d = 1, 2, 3$), and only one direction is considered (i.e., angle between the pair $\theta = 0$), $3 \times 9 \times 1024 \times 1024$ Haralick texture descriptors will be calculated, with each calculation being proportional to the task of deriving the co-occurrence matrix of size 256 by 256.

Wavelet coefficients can compensate Haralick descriptors in specifying texture, by providing features to describe anatomical structure at a large scope with various resolutions. For example, for a WT of 3 scales and filter length 9, a coefficient in the lowest resolution block can represent the texture of $8 \times 9$ pixels, which will well cover the important anatomical structure around liver.

Therefore, to fully take the advantages of high order statistical texture features and anatomical structural features, both Haralick texture descriptors and WT coefficients are used as the inputs to liver distribution image calculation stage.

2.2. Liver Distribution Image Calculation

The liver distribution image of an abdominal CT image is a binary image. In the distribution image, the values of the pixels are one if the pixels have the most possibility of being liver, whereas the values of the other pixels are zero. Support vector machines (SVMs) [15] are implemented as classifier to pixelwisely derive the distribution image. SVMs are a set of discriminative classifiers which are defined by an optimal separating hyperplane. Viewing input data as two sets of vectors in an n-dimensional space, the hyperplane will maximize the margin between the two data sets.

The SVMs classifiers are built in a training process. In the process, assume the training set is $\{(x_i, y_i), i = 1, 2, \ldots, l\}$, where $x_i$ is the input with $x_i \in \mathbb{R}^d$, $y_i$ is the output with $y_i \in R = \{-1, +1\}$, and $l$ is the number of input samples. Then an optimal hyperplane in canonical form must satisfy the following constraints:

$$\alpha \phi(x) + b = 0$$

where $b \in \mathbb{R}$, $\alpha$ is a normal vector, and $\phi(x)$ is an inner product induced feature map that maps the input space into a high dimension linear space. SVMs convert the task of finding the optimal hyperplane into a task of quadratic programming problem as:

$$\min(\frac{1}{2} ||\alpha||^2 + C \sum_{i=1}^{l} \xi_i)$$

subject to

$$y_i (\alpha \phi(x_i) + b) \geq 1 - \xi_i, \ y_i \in \{-1, 1\}$$

Applying Lagrange multipliers, the optimal quadratic programming problem can be solved as the following dual optimal problem:

$$\max \left\{ \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\}$$

subject to

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{l} \alpha_i y_i = 0$$

where $\alpha_i$ is support value, the $x_i$ corresponding to $0 \leq \alpha_i \leq C$ is support vector (SV), and the $x_i$ corresponding to $0 < \alpha_i < C$ is normal support vector (NSV).

$$b = \frac{1}{N_{NSV}} \sum_{x_i \in SV} \left( y_i - \sum_{x_j \in SV} \alpha_i y_j K(x_i, x_j) \right)$$

where $N_{NSV}$ is the number of NSV, $K(x_i, x_j)$ is kernel function. Typical kernel functions are linear, polynomial, radial basis function, and sigmoid.

The training process will derive $\alpha_i, b$, and $K(x_i, x_j)$. Then the SVM as a classifier can classify any input data $x$ with the following classify function:

$$f(x) = \text{sign} \left( \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b \right)$$

2.3. Liver Region Localization

The liver distribution image derived with SVMs is a binary image. It can indicate most of the liver correctly. Figure 4 illustrates one such example, where the left is an original abdominal CT image and the right is the output of SVMs on the image. In the image, liver is at the top-left corner, indicated with the white curve.
false classified as the liver by the proposed method, as a fraction of the total amount of pixels that are identified as the liver in the benchmark. It can be expressed as:

\[
FPVF = \frac{|L_{\text{LSVM}} - L_{\text{man}}|}{|L_{\text{man}}|}
\]

where \(L_{\text{man}}\) denotes the total amount of pixels that are identified as the liver by benchmark. \(L_{\text{LSVM}}\) denotes the total amount of the pixels that are classified by the proposed method as the liver. \(|L_{\text{LSVM}} - L_{\text{man}}|\) is the set difference between \(L_{\text{LSVM}}\) and \(L_{\text{man}}\).

- False negative volume fraction (FNVF)

\[
FNVF = \frac{|N_{\text{man}} - N_{\text{LSVM}}|}{|N_{\text{man}}|}
\]

where \(N_{\text{man}}\) denotes the total amount of pixels that are identified as non-liver in the benchmark. \(N_{\text{LSVM}}\) denotes the total amount of the pixels that are classified by the proposed method as non-liver. \(|N_{\text{man}} - N_{\text{LSVM}}|\) is the set difference between \(N_{\text{man}}\) and \(N_{\text{LSVM}}\).

- True positive volume fraction (TPVF)

\[
TPVF = \frac{|L_{\text{proposed}} \cap L_{\text{man}}|}{|L_{\text{man}}|}
\]

where \(L_{\text{proposed}}\) denotes the total amount of the pixels that are classified as the liver by both the proposed method and in the benchmark, as a fraction of the total amount of pixels that are identified as the liver in the benchmark. It can be expressed as:

\[
TPVF = \frac{L_{\text{proposed}} \cap L_{\text{man}}}{|L_{\text{man}}|}
\]

3. Experiments and Discussions

The proposed automatic liver segmentation algorithm was applied to human abdominal CT images obtained from [17]. All the images were enhanced with contrast agent and scanned in the central venous phase on a variety of scanners ranging from 4 to 16 and 64 detector rows. All the data were acquired in transversal direction. The pixel spacing varied between 0.55 and 0.80 mm, the inter-slice distance varied from 1 to 3 mm. In the experiments, eight images from one subject were chosen as training set to train the SVM classifier, and testing set were the images from another subject.

Segmentation performance validation was done by comparing the automatic segmentation results with the benchmark provided by the data supplier. Three metrics are designed to evaluate the algorithm as below.

- False positive volume fraction (FPVF)

FPVF is defined as the amount of the pixels that are falsely classified as the liver by the proposed method, as a fraction of the total amount of pixels that are identified as the liver in the benchmark. It can be expressed as:

\[
FPVF = \frac{|L_{\text{LSVM}} - L_{\text{man}}|}{|L_{\text{man}}|}
\]

where \(L_{\text{man}}\) denotes the total amount of pixels that are identified as the liver by benchmark. \(L_{\text{LSVM}}\) denotes the total amount of the pixels that are classified by the proposed method as the liver. \(|L_{\text{LSVM}} - L_{\text{man}}|\) is the set difference between \(L_{\text{LSVM}}\) and \(L_{\text{man}}\).

- False negative volume fraction (FNVF)

\[
FNVF = \frac{|N_{\text{man}} - N_{\text{LSVM}}|}{|N_{\text{man}}|}
\]

where \(N_{\text{man}}\) denotes the total amount of pixels that are identified as non-liver in the benchmark. \(N_{\text{LSVM}}\) denotes the total amount of the pixels that are classified by the proposed method as non-liver. \(|N_{\text{man}} - N_{\text{LSVM}}|\) is the set difference between \(N_{\text{man}}\) and \(N_{\text{LSVM}}\).

- True positive volume fraction (TPVF)

TPVF is defined as the amount of the pixels that are classified as liver by both the proposed method and in the benchmark, as a fraction of the total amount of pixels that are identified as the liver in the benchmark. It can be expressed as:

\[
TPVF = \frac{L_{\text{proposed}} \cap L_{\text{man}}}{|L_{\text{man}}|}
\]

where \(L_{\text{proposed}}\) denotes the total amount of the pixels that are classified as the liver by both the proposed method.

The procedure of the experiments are so designed that the performance comparison is done between the method using high order statistical texture features only and the method using both high order statistical texture features and anatomical structural features. Two experiments had been done. In experiment 1, nine Haralick texture descriptors (as defined in equations 2 to 10) were used to derive the liver distribution image. Where \(d = 2\), \(\theta = 0\), and the intensity was normalized to 256 levels. In experiment 2, in addition to the nine Haralick texture descriptors, Wavelet coefficients were used. Where scale number \(S = 3\). In both the experiments, the parameters for SVMs are the same, including using a polynomial kernel function.

Table 1 shows the performance comparison of the two experiments. From the table, it can be seen that when both the high order statistical texture features and anatomical structural features are used, the total segmentation performance is apparently improved than high order
Table 1. Performance metrics (%) of the two experiments.

<table>
<thead>
<tr>
<th></th>
<th>FPVF</th>
<th>FNVF</th>
<th>TPVF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>14.7</td>
<td>6.3</td>
<td>93.8</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>11.1</td>
<td>5.1</td>
<td>97.3</td>
</tr>
</tbody>
</table>

4. Conclusions

This paper presents an accurate liver segmentation algorithm. The main focus of the discussion is how to improve segmentation performance by selecting most suitable image features. There are three major steps in the proposed method, including texture analysis which results in a suitable set of texture features, calculation of liver distribution image using support vector machines, and accurate liver organ localization using a group of morphological operations to locate the liver organ. The novelty of the approach is resided in the fact that the features are so selected that both local and global texture distributions are considered. Out of detailed methodology description and segmentation experiments, it has shown that the proposed method can accurately segment liver in CT image, achieving as high as 97.3% on true positive volume fraction.

REFERENCES