Medical Volume Image Summarization [*]

Feng Ding Hao Li Yuan Cheng Wee Kheng Leow
Dept. of Computer Science, National University of Singapore
{dingfeng,lihao,cyuan,leowwk}@comp.nus.edu.sg

Abstract

Medical volume images are large in size. They cannot be efficiently transmitted and visualized as candidates for medical image retrieval and relevance feedback. On the other hand, 2D images that are small in size and rich in 3D details can be efficiently transmitted and visualized as candidates. This paper presents an algorithm that summarizes the 3D details in a volume image into a single 2D image. It applies soft segmentation to highlight the anatomy of interest in the volume, and automatically selects a salient view that contains the most amount of semantic information as the summarization of the volume image. Experimental results show that the proposed method can well summarize medical volume images of different anatomical structures. Compared to representation of volume images using 2D slices and conventional volume rendering, our summarized images are rich in 3D details, and they can be transmitted and visualized very efficiently.

1. Introduction

With the advancement of technologies for acquiring and storing digital medical images, there is now an explosion of medical images in any moderate-sized hospital. Access to medical images provided by standard medical databases is very limited. Therefore, there is an increasing interest in the research of medical image retrieval systems.

A typical image retrieval system allows the user to query the system using keywords or image features [2, 10]. The system retrieves a list of candidate images and ranks them according to their relevance to the user's query. Some systems allow the user to indicate the relevance of the retrieved candidate images, and use relevance feedback technique to refine the retrieval results.

There are two technical difficulties in the retrieval and relevance feedback of a list of candidate 3D volume images: (1) transmit and (2) display of volume images. Transmitting a typical volume image of 100MB can take more than one minute depending on the network bandwidth and traffic. Transmitting a single 2D slice in the volume is much faster but it does not contain sufficient information for the user to interpret the 3D structure of the objects in the volume. Moreover, it is impossible to display all the slices of all the candidate volume images on the screen at the same time. On the other hand, rendering the volume image in 3D reveals the 3D structure, but volume rendering takes a lot of memory space and computation power. For instance, the volume rendering result shown in Figure 1(g) of a heart CT with 355 slices (Fig. 1(a)) takes about 10 seconds for loading and displaying on a 2.33 GHz Core 2 Duo PC with 4 GB memory. Consequently, a typical PC cannot render multiple volume images simultaneously. Therefore, volume images are not appropriate for medical image retrieval and relevance feedback.

A naive approach to reducing transmission and display time is to reduce the resolution of the volume images. This approach is not ideal because important details about 3D objects are lost in low-resolution images. In addition, the anatomical objects of interest may be occluded by other objects in the volume rendering result, even with a carefully adjusted opacity transfer function which maps voxel intensity to opacity. In this case, essential information of the volume cannot be revealed adequately.

Instead of reducing volume image resolution, this paper proposes an approach to *summarize* the details in a medical volume image into a single 2D image which is small in size and rich in 3D details. Our approach consists of three key ideas:

(1) soft segmentation of the volume image to produce an opacity value for each voxel based on simple user markups (e.g., Fig. 1(b)) indicating the foreground and background objects (Sec. 3.1, 3.2),

(2) volume rendering the volume image according to the voxel-wise opacity values (Sec. 3.3) in order to hide the unimportant anatomical structures that may occlude the objects of interest, and

(3) selecting a 2D rendered view called the *salient view* that reveals the most amount of information (Sec. 3.3).

[*] This research is supported by A*STAR SERC 0521010103 (NUS R-252-000-319-305).

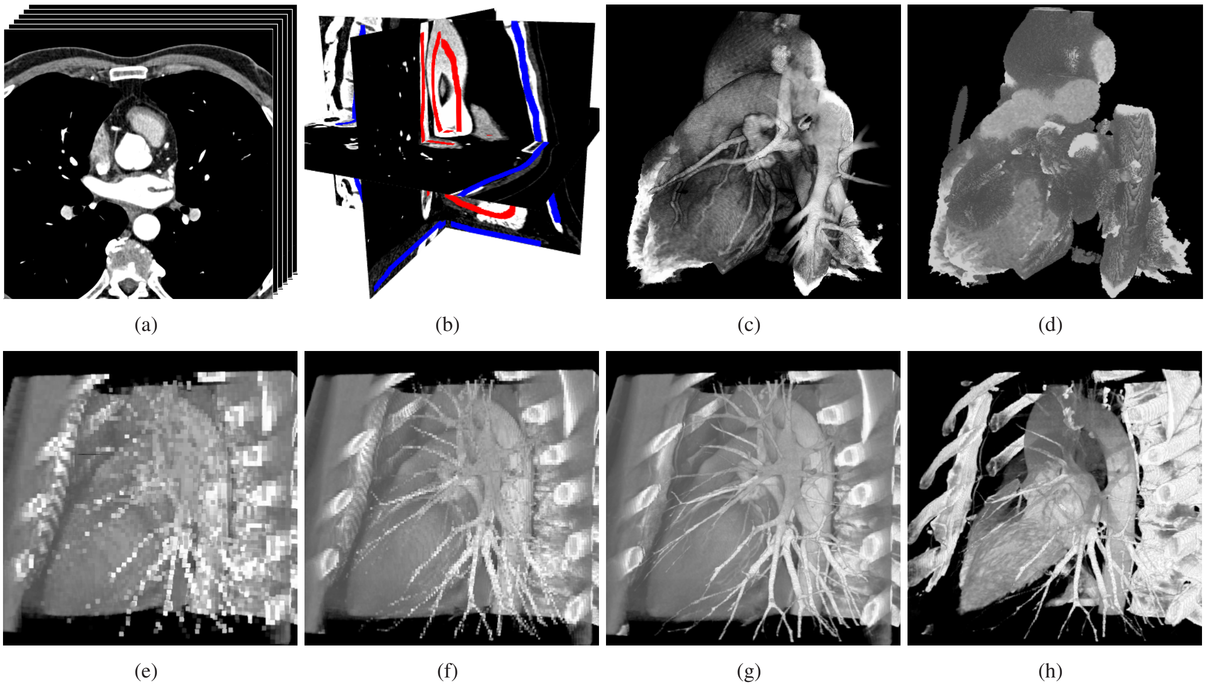


Figure 1. Summarization of heart CT volume image. (a) A sequence of heart CT slices. (b) User markups in the axial, sagittal, and coronal views to indicate foreground (red) and background (blue) objects. (c) Our summarization result: details such as the coronary arteries on the surface of the heart are clearly visible. (d) Hard segmentation. (e–h) Volume rendering. (e) 1/256 down-sampled volume. (f) 1/32 down-sampled volume. (g) Original volume image. (h) Volume rendering with adjusted opacity transfer function.

The selected salient view (e.g., Fig. 1(c)) can thus serve as the summarization of the volume image.

2. Related Work

The idea of volume image summarization is analogous to video summarization [3, 8, 12, 14], whose objective is to allow the user to gain maximum information from the video in a limited time, thus facilitating the browsing and navigation of video collections. The summary of a video is usually a sequence of keyframes or video skims. These keyframes or video skims are usually selected based on the temporal and semantic information of the video [3, 8]. In addition, user behavior while viewing the videos may be taken into consideration for generating the keyframes or video skims [14].

In contrast to video summarization, it is very difficult to define key slices in a volume image for summarization. In volume rendering, an opacity transfer function is usually used to render unwanted portions of the volume transparent. The transfer function can be set manually or derived using soft segmentation algorithms such as [1, 4, 9].

For determining the salient view of 3D mesh models, Lee et al. [6] proposed to measure saliency in a multi-scale manner such that the saliency in each scale is the difference of Gaussian mean curvatures in the current scale and a coarser scale. Váquez et al. [13] proposed viewpoint entropy to measure the goodness of a viewpoint. For 2D images, Itti et al. [5] proposed to measure saliency based on the intensity, color, and textual orientation in the corresponding pyramidical images. These techniques cannot be directly applied to determine the salient view of volume rendered medical images because there is neither pre-constructed mesh models nor color/texture information in the volume image.

3. Summarization Method

Our summarization algorithm performs soft segmentation of the objects of interest in the input CT/MR volume images. The segmented images are then volume rendered. The salient viewing angle that reveals the most amount of information is determined and the rendered 2D image at the salient viewing angle is the *summarization*.

3.1. Foreground / Background Characteristics

To perform soft segmentation, the characteristics of foreground and background in the volume image need to be determined. They can be obtained from user's markups (Fig. 1(b)) on the axial, sagittal, and coronal views of the input images, or derived from prior markups on training data. The user markups also function as the seed regions for soft segmentation (Sec. 3.2). A different set of foreground and background characteristics is required for summarizing a different type of volume images, e.g., CT of head, chest and abdomen.

In general, our soft segmentation algorithm can work with any image features including, but not restricted to, intensity, gradient, edge, texture, etc. In this paper, distribution of voxel intensities is used because the anatomical objects of interest have inhomogeneous voxel intensities and some object boundaries are indistinct. More sophisticated features such as textures are also possible but they are computationally more expensive.

Given a volume image I, the intensity probability density functions $P(I(\mathbf{x})|F)$ and $P(I(\mathbf{x})|B)$ of the foreground F and the background B regions are estimated by constructing the Gaussian Mixture Models (GMM):

$$g(x) = \sum_i a_i f_i(x), \tag{1}$$

where x is the voxel intensity, a_i are coefficients such that $\sum_i a_i = 1$, and $f_i(x)$ are Gaussian distributions with parameters space (μ_i, σ_i). Parameters a_i, μ_i and σ_i are estimated by Expectation Maximization (EM) algorithm. Preprocessing of the input image such as noise removal and contrast enhancement may be applied.

To facilitate the convergence of the EM algorithm, clustering is first carried out. The clustering results are then used to initialize EM. To avoid manual specification of the exact number of clusters, i.e., number of Gaussians in our GMM, the proposed algorithm performs clustering using the adaptive binning algorithm [7]. Given the estimated class radius and class separation, the adaptive binning algorithm automatically determines the optimal number of clusters. This property is highly desirable since the number of clusters are usually hard to be determined based on the user markups.

3.2. Soft Segmentation

The fast marching algorithm [11] is used to perform soft segmentation given the foreground and background characteristics. It performs segmentation by propagating fronts from both foreground and background seed regions simultaneously. The propagation of the fronts is described by

$$\begin{array}{ll} |\nabla T_F| S_F = 1 & \text{for foreground,} \\ |\nabla T_B| S_B = 1 & \text{for background,} \end{array} \tag{2}$$

where T_F and T_B are the arrival times of the foreground and background fronts, and S_F and S_B are corresponding speed functions. If S is constant (e.g., $S = 1$), T corresponds to the Euclidean distance transform of the user markups. In the case of soft segmentation, S_F should be large in the foreground region and small or zero in the background region, similarly for S_B. Therefore, the speed functions are proportional to the posterior probabilities (Sec. 3.1),

$$S_F(\mathbf{x}) \propto P(F|I(\mathbf{x})), \tag{3}$$

$$S_B(\mathbf{x}) \propto P(B|I(\mathbf{x})), \tag{4}$$

where

$$P(F|I(\mathbf{x})) = \frac{P(I(\mathbf{x})|F)P(F)}{P(I(\mathbf{x}))}, \tag{5}$$

$$P(B|I(\mathbf{x})) = \frac{P(I(\mathbf{x})|B)P(B)}{P(I(\mathbf{x}))}. \tag{6}$$

$P(I(\mathbf{x}))$ is the normalization constant. $P(F)$ and $P(B)$ can be determined *a priori* according to the objects of interest in the volume image and the type of the image.

If the foreground front arrives at voxel $\mathbf{x}$ earlier, i.e., $T_F(\mathbf{x}) < T_B(\mathbf{x})$, $\mathbf{x}$ is more likely a foreground voxel, otherwise a background voxel. In soft segmentation, an opacity value α is assigned to each voxel based on $T(x)$,

$$\alpha(\mathbf{x}) = \begin{cases} \dfrac{T_F^m - T_F(\mathbf{x})}{T_F^m} & if \quad T_F(\mathbf{x}) < T_B(\mathbf{x}) \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where T_F^m is the largest arrival time of the foreground front across the volume image I.

3.3. Salient View Selection

After soft segmentation, the whole image is volume rendered and the viewing angle that reveals the most amount of information, called the *salient view*, is automatically determined. Unlike conventional volume rendering which applies a global opacity transfer function to map each intensity value to an opacity value, the proposed algorithm renders the volume image using the voxel-wise opacity value $\alpha(\mathbf{x})$.

Finding the salient viewing angle by rotating the camera with respect to the volume image I is equivalent to fixing the camera while rotating I. The centroid of I is aligned with the camera's optical axis, and the scale is normalized so that the projected 2D image of I has a fixed size. Rotation of I with respect to its centroid is represented as $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \theta_3\}$. In-plane (image) rotation θ_3 is set to 0 since it has no effect on the saliency measure of the projected image. Therefore, the problem is to find the out-of-plane rotation θ_1 and θ_2 that maximize saliency.

Given object rotation $\boldsymbol{\Theta}$, we denote the projected 2D image of I as $J(\Theta)$. The saliency M of J consists of two terms,

$$M(J) = E(J) + wG(J) \tag{8}$$

where w is the weighting coefficient that balances the entropy $E(J)$ and the average gradient magnitude $G(J)$, which are defined as

$$E(J) = \sum_i -p_i \log_2 p_i, \tag{9}$$

$$G(J) = \frac{1}{N} \sum_x |\nabla J(x)|, \tag{10}$$

with p_i the probability of intensity i, and N the number of pixels in the projected image J. The entropy term $E(J)$

measures the local variation of intensity values in a view. It is large when the view shows the part of the volume that is rich in both large-scaled surface features such as edges and ridges, and small-scaled surface features such as corners and saddle points. However, noise can be confused for small-scaled surface features. The average gradient magnitude $G(J)$ measures the global variation of intensity values in a view. It is large when the view shows the part of the volume that is rich in large-scaled surface features, and it is more resilient to noise than $E(J)$. By combining $E(J)$ and $G(J)$, their individual weakness can be mitigated. Therefore, the combined measure $M(J)$ facilitates the detection of salient view that is rich in 3D surface features while remaining resilient to noise. Analogous entropy and gradient measures have been used in existing work for the determination of salient views for 3D mesh models (Sec. 2).

The orientation that maximizes saliency M is determined by using an iterative maximization algorithm that is analogous to gradient ascent:

```
Repeat for a fixed number of iterations:
1. Randomly initialize $\mathbf{\Theta}$, and set the size of neighborhood
   $\delta$ to an initial fixed constant.
2. Repeat until $\delta$ is small enough:
   a. Find the neighbor $\mathbf{\Theta}'$ of $\mathbf{\Theta}$ with the largest saliency.
   b. If $M(J(\mathbf{\Theta}')) > M(J(\mathbf{\Theta}))$, set $\mathbf{\Theta} \leftarrow \mathbf{\Theta}'$; otherwise,
   reduce $\delta$.
```

The neighbors of $\mathbf{\Theta} = \{\theta_1, \theta_2, 0\}$ are $\mathbf{\Theta}' = \{\theta_1 \pm \delta, \theta_2 \pm \delta, 0\}$. The algorithm iteratively determines the angle $\mathbf{\Theta}$ that maximizes the saliency $M(J)$.

4. Experiments and Results

Summarization tests were conducted on 50 medical volume images to evaluate the effectiveness of the proposed algorithm and its applicability to different types of images. The input volume images include 20 sets of head CT, 20 sets of abdominal CT, 5 sets of heart CT, and 5 sets of brain MR images, each containing 120–360 slices of resolution 512×512. The objects of interest range from hard objects like bones to soft tissues such as heart, liver, blood vessels etc. Both qualitative evaluation of visual clarity (Sec. 4.1) and quantitative evaluation of transmission and display time (Sec. 4.2) were performed.

4.1. Visual Clarity

To evaluate the visual clarity of our summarized images, we compare them with single 2D slice, conventional volume rendering, and hard segmentation. As described in Section 1 that one way to reduce transmission and display time is to reduce the resolution of the volume image, we also compare the visual clarity of our summarized images with conventional volume rendering of down-sampled volume images. Due to space limitation, only a selected set of comparison results are presented in this paper.

Figure 1 illustrates our summarized image of a heart CT with 355 slices, compared with hard segmentation result generated by a GMM-based classifier (Fig. 1(d)), conventional volume rendering (Fig. 1(g, h)), and volume rendering of down-sampled volume images (Fig. 1(e, f)).

Hard segmentation is inadequate for summarization purposes because it is very difficult to develop a general hard segmentation algorithm that can accurately segment various anatomical objects based on simple user inputs, especially for soft tissues whose boundaries are indistinct. Consequently, as hard segmentation maps the opacity of segmented portions to 1 and the rest to 0, it does not produce visually pleasant summarization image for inaccurate segmentation result as shown in Figure 1(d). In contrast, even though soft segmentation does not produce ideal segmentation results, it retains the visual clarity of the objects of interest and allows the viewer to obtain a perception of the input image with sufficient details (Fig. 1(c)).

Conventional volume rendering requires the user to interactively adjust a global opacity transfer function. Since the function is global, objects with the same intensity are mapped to the same opacity. Consequently, the objects of interest may be occluded by other objects with similar intensity values. For instance, the cardiac tissues have similar intensity values as the skin, fat, and bronchi. Therefore, the details on the heart surface may be occluded by the skin and the bronchi as shown in Figure 1(g). On the other hand, over-adjusting the opacity transfer function resulted in the lost of cardiac tissues (Fig. 1(h)). Our algorithm computes the opacity of each voxel according to Eq. 7, making it possible to remove the occlusions which have similar intensity values as the objects of interest. Fig. 1(c) shows that our summarized image is visually pleasant and clearly illustrates details of the heart surface such as the coronary arteries. Compared to the down-sampled volume images (Fig. 1(e, f)), our summarized image reveals significantly higher amount of detailed information.

Figure 2 shows more examples of the summarization tests. The objects of interest include bones (row 1), liver (row 2), multiple organs (row 3), hepatic vein (row 4) and bronchi (row 5). These objects have significantly different anatomical structures. Our summarized images shown in Fig. 2(c) present very informative views about the objects, compared to visualization by a single 2D slice (Fig. 2(a)) and by conventional volume rendering (Fig. 2(b)) using 3D-DOCTOR (http://www.ablesw.com/3d-doctor/).

Summarized images from sub-optimal viewing angles are demonstrated in Fig. 2(d). They either reveal less shape information about the objects of interest compared to those with salient viewing angles (skull and liver), or cause self-occlusions (multiple organs, hepatic vein and bronchi).

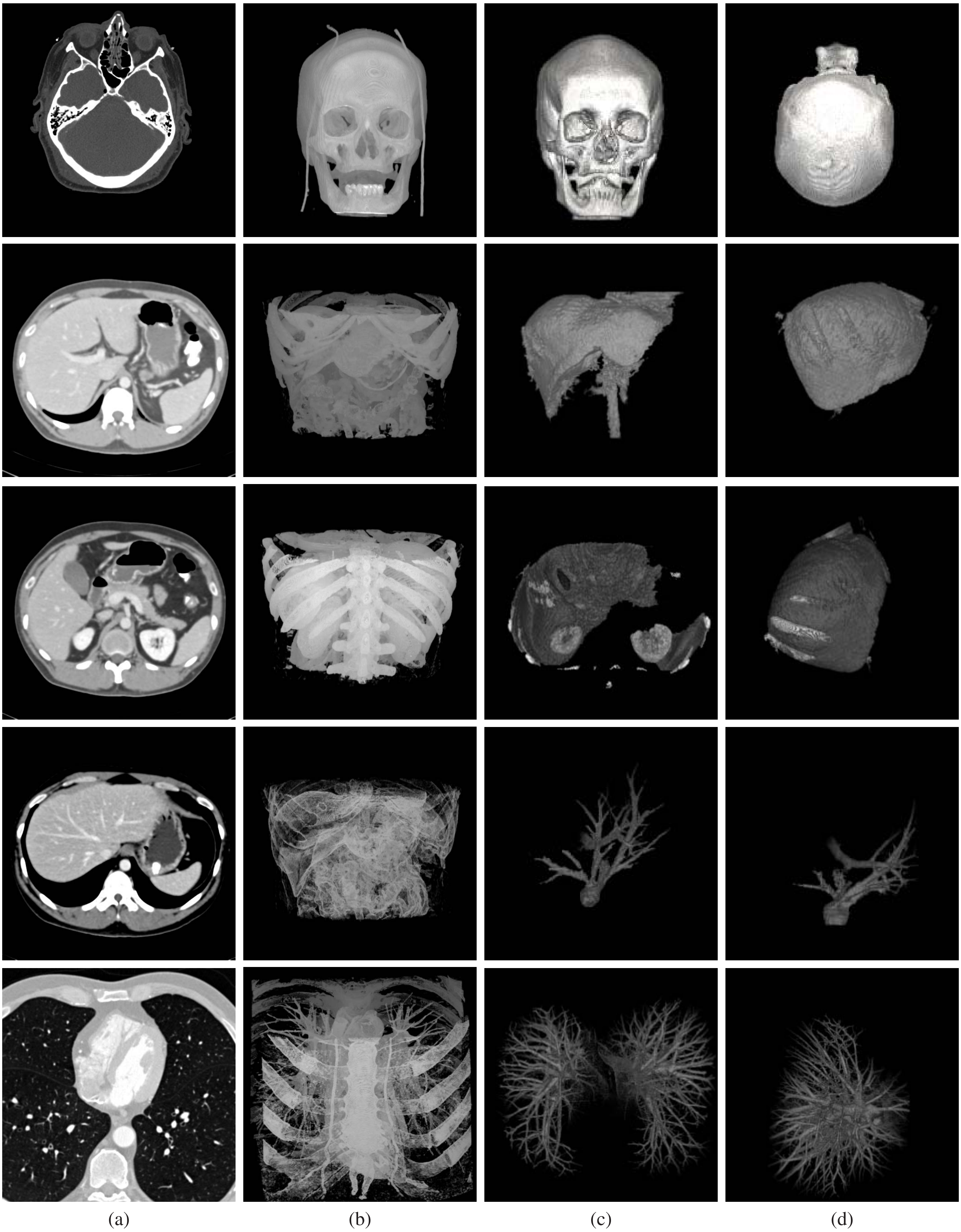
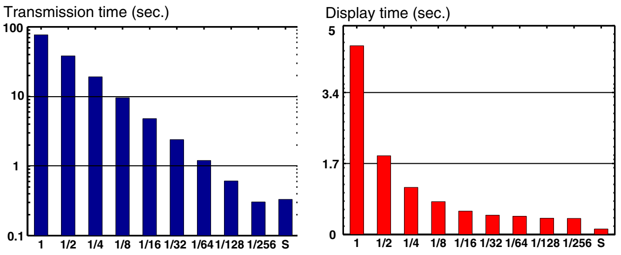


Figure 2. Summarization results of volume images. First row to last row: skull, liver, multiple organs, hepatic vein and bronchi. (a) Single 2D slice. (b) Conventional volume rendering. (c) Proposed algorithm with salient viewing angles and (d) other sub-optimal viewing angles.



(a) (b)

Figure 3. Transmission and display time. S: Summarized image. Transmission time is log-scaled for visual clarity.

These results show that our algorithm can indeed determine salient viewing angles.

4.2. Transmission and Display Time

To evaluate the efficiency of our summarization method, we compare the transmission and display speed of the summarization images with those of the down-sampled volume images. Each test data was down-sampled to $1/2$, $1/4$, $1/8$, $1/16$, $1/32$, $1/64$, $1/128$, and $1/256$. They were transmitted over a 100 Mbps LAN and volume rendered by VTK (http://www.vtk.org) on a 2.33 GHz Core 2 Duo PC with 4 GB memory.

The transmission and display times were measured and averaged over all the test data as shown in Figure 3. Transmitting the original volume (average size 98.3 MB) took 76.86 seconds, while loading and displaying it took 4.52 seconds. Down-sampling the image reduced the size of the image, thus reducing the transmission and display times. But important details were also lost in the down-sampled images as discussed in Section 4.1.

The size of our summarization result is 0.43 MB, which is close to that of the $1/256$ down-sampled image (0.38 MB). Therefore, their transmission times (0.33 and 0.31 seconds) are also similar. However, since volume rendering consumes more computation power, displaying our 2D summarized image takes only 0.13 seconds, which is only $1/3$ of the time taken by volume rendering $1/256$ down-sampled image (0.39 seconds).

5. Conclusion and Future Work

This paper presented a volume image summarization algorithm for medical image retrieval. It applies soft segmentation to derive a voxel-wise opacity function that highlights the objects of interest in volume rendering, and determines a salient view based on intensity gradient and entropy of the volume rendered image. The rendered image at salient view serves as the summarization of the volume image. Test results show that the summarized images generated by our algorithm are small in size for transmission and display, while rich in 3D details compared to 2D slices and conventional 3D volume rendering techniques. They are also visually clearer and more pleasing than hard segmentation results.

The algorithm presented in this paper deals with static volume images of various modalities and anatomies with different structures. Future work will be conducted on the summarization of dynamic volume images such as fMRI, which involves temporal changes of the volume. In that case, video summarization techniques may be incorporated into the current algorithm.

References

[1] X. Bai and G. Sapiro. Geodesic matting: A framework for fast interactive image and video segmentation and matting. *Int. J. Comput. Vision*, 82(2):113–132, 2009.

[2] I. Cox, M. Miller, T. Minka, T. Papathomas, and P. Yianilos. The bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE Trans. Image Proc.*, 9(1):20–37, 2000.

[3] D. DeMenthon, V. Kobla, and D. Doermann. Video summarization by curve simplification. In *Proc. ACM Multimedia*, pages 211–218, 1998.

[4] L. Grady. Random walks for image segmentation. *IEEE Trans. PAMI*, 28(11):1768–1783, 2006.

[5] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. PAMI*, pages 1254–1259, 1998.

[6] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh saliency. In *Proc. ACM SIGGRAPH*, pages 659–666, 2005.

[7] W. K. Leow and R. Li. The analysis and applications of adaptive-binning color histograms. *Computer Vision and Image Understanding*, 94(1–3):67–91, 2003.

[8] J. Nam and A. H. Tewfik. Dynamic video summarization and visualization. In *Proc. ACM Multimedia*, pages 53–56, 1999.

[9] C. Rother, V. Kolmogorov, and A. Blake. “GrabCut”: Interactive foreground extraction using iterated graph cuts. In *Proc. ACM SIGGRAPH*, pages 309–314, 2004.

[10] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, 1999.

[11] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1999.

[12] B. T. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1):3, 2007.

[13] P.-P. Váquez, M. Feixas, M. Sbert, and W. Heidrich. Viewpoint selection using viewpoint entropy. In *Proc. Vision, Modeling, and Visualization*, pages 273–280, 2001.

[14] B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang. Video summarization based on user log enhanced link analysis. In *Proc. ACM Multimedia*, pages 382–391, 2003.