US 20160104293A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2016/0104293 A1**

(43) **Pub. Date: Apr. 14, 2016**

Gering

(54) **SYSTEM AND METHOD OF VOICE ACTIVATED IMAGE SEGMENTATION**

(71) Applicant: **David Thomas Gering**, Waunakee, WI (US)

(72) Inventor: **David Thomas Gering**, Waunakee, WI (US)

(21) Appl. No.: **14/874,425**

(22) Filed: **Oct. 3, 2015**

Related U.S. Application Data

(60) Provisional application No. 62/071,897, filed on Oct. 3, 2014.

Publication Classification

(51) **Int. Cl.**

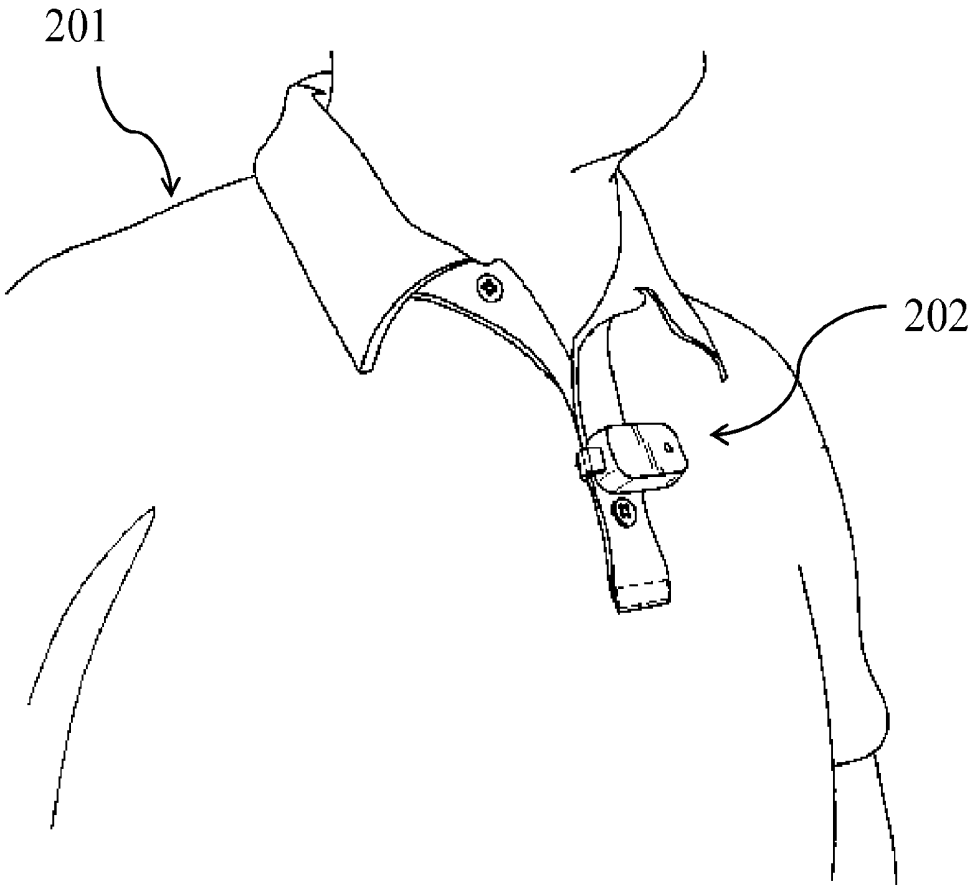
G06T 7/00	(2006.01)
A61B 8/08	(2006.01)
A61B 6/03	(2006.01)
G06T 11/00	(2006.01)
G06F 3/16	(2006.01)

(52) **U.S. Cl.**

CPC ***G06T 7/0081*** (2013.01); ***G06T 7/0012*** (2013.01); ***G06T 7/0097*** (2013.01); ***G06T 11/001*** (2013.01); ***G06F 3/167*** (2013.01); ***A61B 6/032*** (2013.01); ***A61B 6/037*** (2013.01); ***A61B 8/085*** (2013.01); *G06T 2200/04* (2013.01); *G06T 2207/20104* (2013.01); *G06T 2207/30096* (2013.01); *G06T 2207/30196* (2013.01); *G06T 2207/10072* (2013.01); *G06T 2207/20141* (2013.01); *G06T 2207/20036* (2013.01)

(57) **ABSTRACT**

A method and system for incorporating voice commands into the interactive process of image segmentation. Interactive image segmentation involves a user pointing at an image; voice commands quicken this interaction by indicating the purpose and function of the pointing. Voice commands control the governing parameters of the segmentation algorithm. Voice commands guide the system to learn from the user's actions, and from the user's manual edits of the results from automatic segmentation.



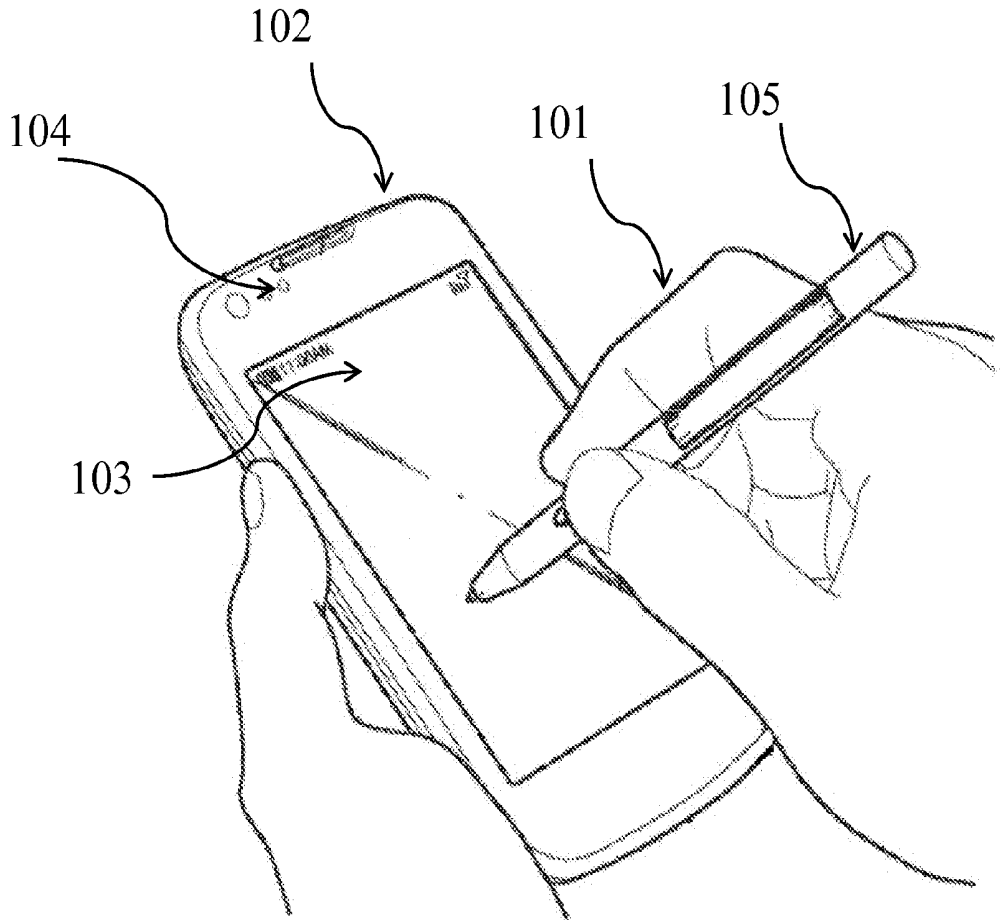


Fig. 1

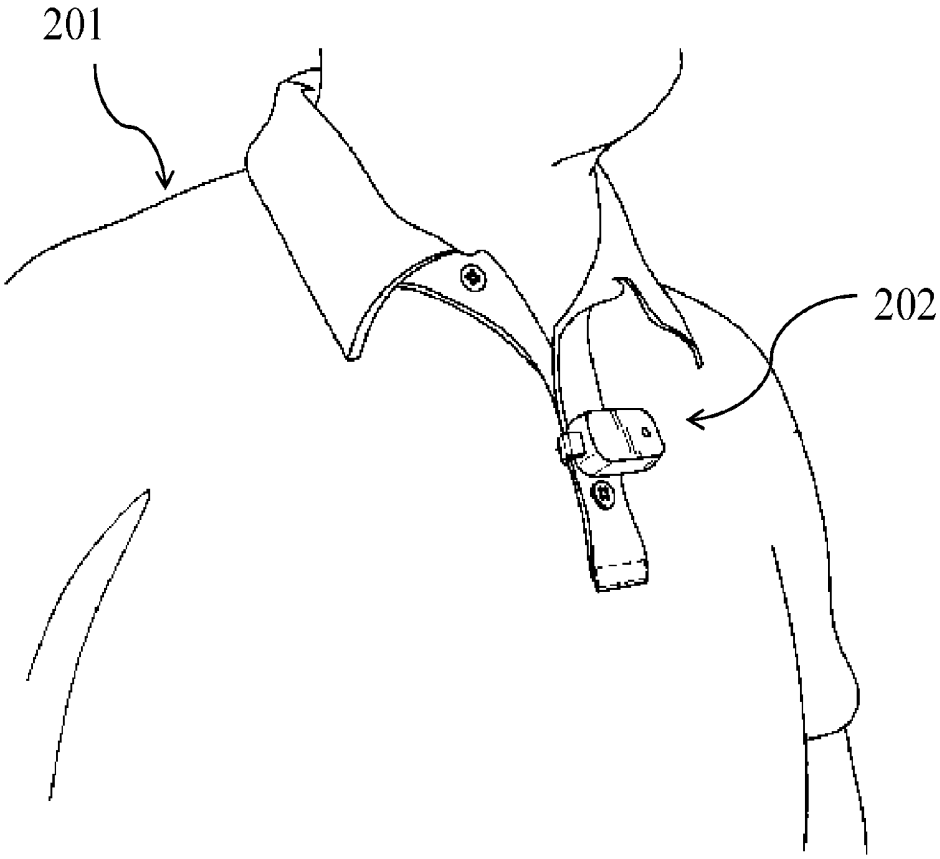


Fig. 2

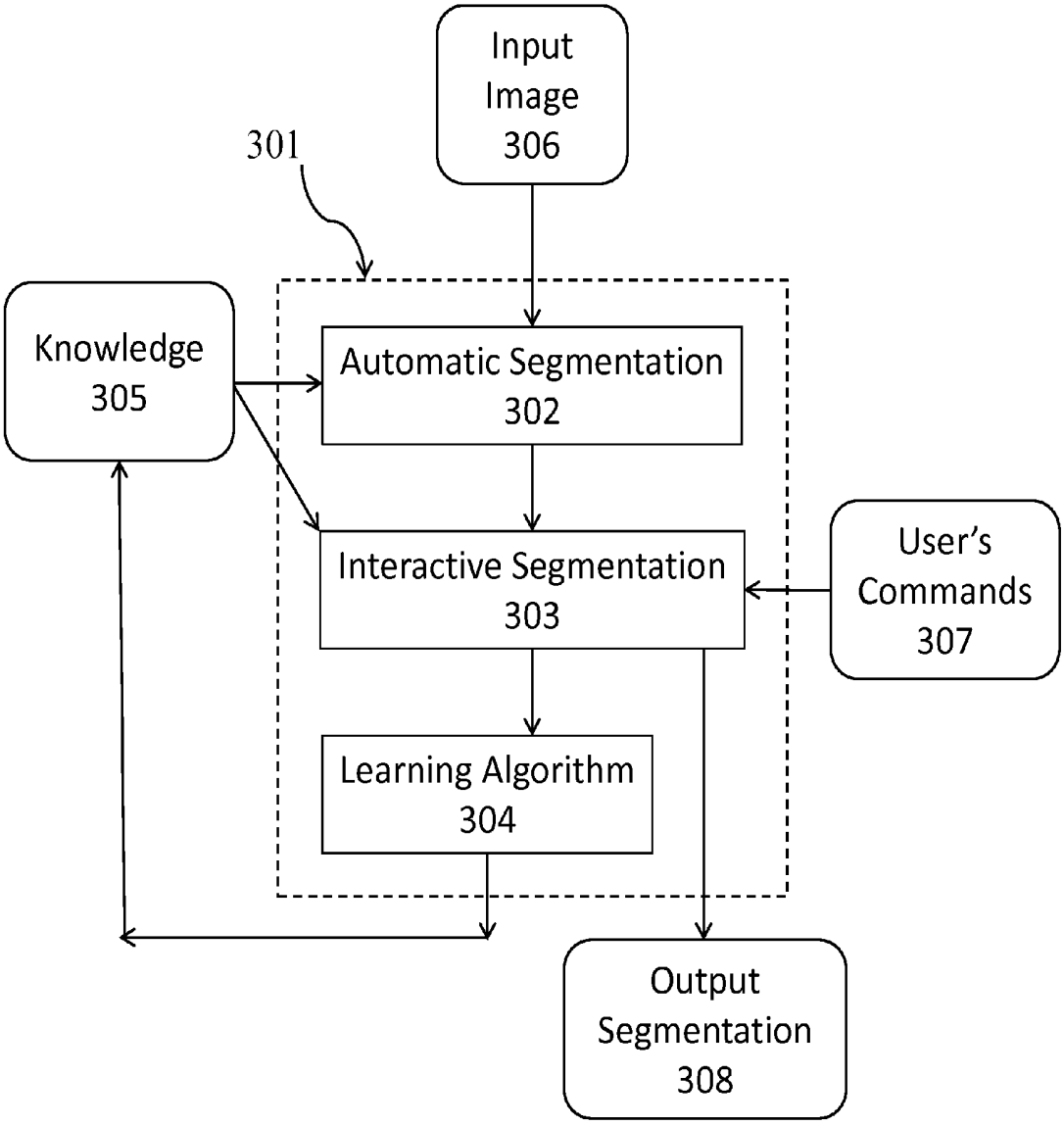

Input
Image
306
301
Knowledge
305
Automatic Segmentation
302
Interactive Segmentation
303
User's
Commands
307
Learning Algorithm
304
Output
Segmentation
308

Fig. 3

SYSTEM AND METHOD OF VOICE ACTIVATED IMAGE SEGMENTATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present invention relates to the US provisional patent application with an identical title, application No. 62/071,897 and filing date of Oct. 3, 2014.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] Non-applicable.

REFERENCE TO SEQUENCE LISTING, A TABLE, OR A COMPUTER PROGRAM LISTING COMPACT DISC APPENDIX

[0003] Non-applicable.

BACKGROUND OF THE INVENTION

[0004] The present invention relates generally to image segmentation, or the process of assigning labels to the elements in an image. More specifically, it relates to semi-automatic 3D medical image segmentation, where there is an interactive process for labeling each voxel according to tissue type represented.

[0005] Image segmentation has been an active area of research in the computer vision community, and the medical image analysis community. The ultimate goal is often a fully automatic algorithm that need not require input from a human user. In practical applications, however, stringent goals for accuracy may compel assistance from an expert. Semi-automatic algorithms may query the user for seed points to initiate region growing, or training points to initialize probability distributions (used by Bayesian or kNN classification), or threshold levels to govern expansion terms of level set methods, or strokes of a virtual paint brush to indicate foreground and background objects (used by level set or GrowCut algorithms), or bounding boxes to encapsulate regions of interest to provide spatial constraints. Furthermore, they may allow interactive edits of automatically-computed results by relocating control points of active contours, or by manipulating tools for repositioning contours. It may behoove the user to manually redraw an incorrect border astride the perimeter of a structure. Fully manual segmentation involves a person drawing all the boundaries of all structures, but such tedious monotony is prone to error and inter-observer variability. Any method less than nearly fully automatic could be prohibitively expensive to deploy in clinical settings due to how much time is consumed by healthcare personnel. The user interface device is usually a computer mouse, stylus, or touch screen, but could be a trackball, haptic interface, or eye-gaze detector in academic settings.

[0006] Segmented images are essential for various clinical applications that stand to benefit from the presence of images where each relevant anatomic structure has been delineated. Segmentation can be a valuable ally in treating cancer, whether by radiotherapy, chemotherapy or surgical resection. Image guided radiation therapy (IGRT) uses cross-sectional images of the patient's internal anatomy to better target the radiation dose to the tumor while sparing exposure of healthy organs. The radiation dose delivered is controlled with intensity modulated radiation therapy (IMRT), which involves changing the size, shape, and intensity of the radiation beam to conform to the size, shape, and location of the patient's tumor. IGRT and IMRT simultaneously improve control of the tumor while reducing the potential for acute side effects due to irradiation of healthy tissue surrounding the tumor. Segmentation is widely employed for IGRT and IMRT because the process of planning the delivery is a quantitative and numerical exercise best suited for a computer. Chemotherapy, in contrast to radiotherapy, tends to follow a more qualitative planning process whereby the tumor's response to the treatment regimen is visually monitored, such as by a CT scan every couple months. Precise quantification of tumor extent would be useful for decision making, but oncologists are too short on time to be guiding semi-automatic segmentation methods, and they're unlikely to be trained in using expensive analysis workstations. Surgical resections and biopsies benefit from image segmentation by rendering 3D views of the spatial relationships between organs for surgical planning and guidance. Beyond treating cancer, image segmentation is utilized in longitudinal studies that track quantitative measurements such as anatomic dimensions, cross-sectional areas, or volumes.

[0007] Recent improvements in speed, accuracy, and automation of segmentation algorithms have nearly obviated human intervention in certain research applications. These applications tend to focus on tissue that appears normal, such as quantitative measurements of neuroanatomy. Disease can vary in unexpected ways that are complicated to model, and disease often presents special cases and outliers that extend beyond the understanding of computer software. What the software needs is interaction with a keen physician, quick and clever, to astutely manipulate facts.

[0008] Even if fully automatic algorithms could become sufficiently accurate for routine clinical use, certain physicians vary in personal approach and requirements, so algorithms would still benefit from some manner of catering to individual preferences. When the full knowledge and artful discernment of the physician(s) is reflected in the output of the segmentation, then the downstream processes to which segmentation is an input can become effectual instruments.

[0009] The foregoing discussion highlights the need for new semi-automatic strategies that can incorporate the expertise of the physician(s) into the segmentation process. The key enabler is to employ their penetrating intellect with a minimum of time and expense. The present invention proposes voice-activation as this key enabler. Voice recognition has a history of employment by the medical profession for dictation and medical transcription. Healthcare researchers have also experimented with voice-activated image retrieval, operating an imaging scanner by voice commands, and hands-free manipulation of a display of 3-D angiography by a surgeon in the operating theater.

BRIEF SUMMARY OF THE INVENTION

[0010] A method for voice activated image segmentation is introduced, which allows the physician to quickly and easily interact with the computerized segmentation process, thereby imparting his/her skilled expertise to the segmented result.

[0011] In some embodiments, the system is equipped to automatically respond to voice commands, such as "Grow more anteriorly toward ventricle," because it segments not only the "target" structures, but also the "situational" structures, or surrounding anatomy, to which the physician may refer. Both target and situational structures are identified in the initial automatic segmentation, which is subsequently

presented to the physician for feedback. If the accuracy is deemed sufficient, then no interaction occurs, aside from the physician pronouncing, "It's good." Otherwise, the physician indicates which changes to make, and ideally, these changes are made in real-time, but if processing power is an issue, then the updates could be coarse during interaction, but refined to full resolution after conversation concludes.

[0012] In some embodiments, the physician's feedback need not be given by voice alone, but also by "pointing", such as via touch screen or directional eye gaze. For example, the physician may dictate, "Remove this", while pointing to an island of segmentation labels to be erased. The physician could speak, "Add", or, "Grow", if the pointer were being used to add more image elements to the segmented structure. In this manner, the pointer's function is changed from a virtual brush to a virtual eraser without the need for the user to click on a menu, or press a button. Such actions are a great distraction to the user during the interaction, and prolong the segmentation process. The interaction time can be greatly reduced by using voice commands to indicate the pointer's purpose and function, thereby avoiding interruptions.

[0013] In some embodiments, the system is a cloud-based solution where intense processing occurs in a high-performance computing environment while the physician interacts with it elsewhere on a mobile device.

[0014] In some embodiments, the physician's sequence of interactions is recorded in the form of an annotated video that a human segmenter can watch in case the physician indicates that the computer misunderstood his/her instruction. Thus, the "cloud" could be a system comprising computer servers and a team of humans working in tandem to satisfy physicians positioned at many hospitals around the globe. When the cloud is unsure of something, it can present questions to the physician in the form of annotations on the original automatic segmentation, such as "Is this lesion or cyst?" The physician would answer by "pointing" to the labeled region and answering, "Lesion." For truly complicated cases, the interaction between physician and cloud can take the form of real-time video chat, where the device screen shows the image, as well as annotations marking any pointing/drawing activity, as well as the faces of the physician and segmenter overlaid in small "bubbles" on the perimeter so they can speak with one another clearly.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] FIG. **1** illustrates a system diagram of an embodiment of the present invention where the pointing device is a stylus;

[0016] FIG. **2** illustrates a system diagram of a preferred embodiment of the present invention where the user wears a clip-on microphone;

[0017] FIG. **3** is a block diagram of a method according to a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0018] Before any embodiments of the invention are explained in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the following drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Also, it is to be understood that the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," or "having" and variations thereof herein is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

[0019] FIG. **1** is a system diagram of an embodiment where the user **101** interacts with the segmentation process using a voice-activated segmentation system, the system comprising a processing computer **102**, an image display **103**, a microphone **104**, and a pointing device **105**. The computer **102** could be a smart phone, tablet, laptop, or desktop computer. The act of pointing at image elements can take many forms; it can be the click of a mouse or trackball; it can be the touch of a finger or stylus on a touchscreen; it can be a gesture toward a very large display; it can be the point toward which the eyes are gazing when eye-tracking is available on the device (such as a camera with real-time computer vision algorithms). In the preferred embodiment, the pointing device is a stylus because interacting with an image feels most natural when the user feels like he/she is drawing directly on the screen that displays the image.

[0020] FIG. **2** extends the system diagram of FIG. **1** to add another component, which is for the user **201** to wear a clip-on microphone **202**. By placing a microphone close to the mouth, the user is able to talk in quieter tones, so as to not be a distraction to others nearby. In a preferred embodiment, the wearable microphone **202** is in addition to another microphone, referred to as the "background mic". The background mic could be located on the computer **102** (as shown as **104**) or the stylus **105**, or worn on the back of the user, such as on a headset. While the microphone nearest the mouth records the users spoken words, the background mic collects ambient noise and uses that information for noise cancellation. In this manner, the user could be segmenting in a crowded noisy room without the segmentation system becoming confused by background noise.

[0021] FIG. **3** depicts a block diagram of the method for the preferred embodiment of the invention. The processing system **301** computes an initial segmentation **302** automatically, and shows it to the physician. The physician comments on the results by speaking and pointing to offer commands **307**. The device shows its understanding by highlighting the object to which the point was directed, and displaying its interpretation of what was spoken as text. The physician and computer interact until the segmentation is complete **308**. The learning algorithm **304** analyzes the editing operations that the physician made during the interaction, and uses this information to update the a priori knowledge **305** used to perform the automatic segmentation **302**. Therefore, the system learns from earlier interactions to become smarter for future interactions.

[0022] The precise form of the learning algorithm **304** depends on the type of segmentation algorithm **302**. For example, if the segmentation algorithm is based on Bayesian classification, then the knowledge **305** consists of prior probability distributions and spatially varying priors. These were initially computed from training data, and each new image that is processed can be added to this training set in order to update the probability distributions. As another example, if the segmentation algorithm **302** is based on statistical shape models, then the probability distributions governing those models (such as curvature, smoothness, angular measures, and radii) may be updated with each successfully completed segmentation. As another example, the distance of the target object from surrounding anatomical landmarks can be

extremely helpful to a segmentation algorithm. The difference between the initial distances, and the distances following the user's edits can be noted for the future.

[0023] In some embodiments, the ability of the system to learn from the user's past edits is fully automatic, and tailored to the chosen segmentation method. In some embodiments, the learning responds to voice command. For example, a patient could be an outlier, in that there is some exceptional anatomy that the physician wishes to segment without impacting the learning process. The physician would indicate "Don't learn this" or "Exclude this patient."

[0024] In some embodiments, vocal commands can be used not only to direct the segmentation, but also to view the results. Segmentations of medical images are often presented as 2D cross-sectional slices and 3D surface renderings side-by-side. Navigating the 2D display involves selecting the orientation of the slices (e.g.: axial, coronal, sagittal), scrolling through the slices along this direction, and zooming and panning within a slice. Navigating a 3D display involves rotating, zooming, panning, and changing the opacity and visibility of occluding structures. It also involves enlarging the 2-D or 3-D views, meaning altering the layout of where things are displayed on the screen. These navigational commands can be given by spoken word in a manner more intuitive than using a mouse. For example, the user can change the slice by saying, "Next slice" or "Previous slice". The user can quickly advance through slices by chaining commands, such as "Next . . . next . . . next . . . next . . . go back . . . back again, stop." Likewise, the user could rotate the viewpoint of a 3-D rendering by saying "Rotate left, more, more, a little more." In situations such as this, the word "more" can be interpreted to mean a typical increment, such as 10 degrees. Then "a little more" would be half the usual increment, or 5 degrees. The user can program the system by directly defining the meaning of commands, "When I say 'Rotate', move 10 degrees."

[0025] In those embodiments that include a pointing device, vocal commands serve to alter the pointing mode. This means that the same pointing motion, such as touching an object, will have a different effect depending on what the user says as the user points. For example, to add move image elements (2D pixels or 3D voxels) to a segmented tumor object, the user would say "Add" as while clicking on objects, and to erase them, the user would say "Remove" or "Erase" or "Delete". Short one-word commands chosen from a limited vocabulary will be easier for a voice-recognition system to understand correctly. For example, a type of region-growing for liver lesions can be initialized or edited simply by the user pointing at each lesion while saying either "Lesion." or "Not lesion." As another example of simplified vocabulary, the GrowCut algorithm takes input from the user in the form of brush strokes on the foreground and background objects. The user can provide these inputs with seamless hand motion by drawing with the pointer while speaking the name of the object being touched, which is either "Background", or "Foreground."

[0026] In addition to altering the mode of the pointer, vocal commands can alter the form of the pointer. Suppose the pointer is being used as a digital paintbrush, then the user can change the radius of the brush by saying "enlarge brush" or "shrink brush". Some edits of segmentation are precise manual drawing, in which the user would say, "Precisely this", while other edits are rough guidelines that the user wants the computer to use as a starting point for finding the boundary from there (based on image intensities and anatomical models), so the user might say, "Roughly this" while drawing that edit.

[0027] In some embodiments, voice commands control the governing parameters of the segmentation process. For example, level set methods use curvature for regularization, and the user can dictate "Smaller curvature" or "Larger curvature . . . larger . . . larger . . . good."

[0028] The automatic segmentation **302** of anatomic landmarks can be leveraged to make it possible for the user to reference anatomy in the spoken commands. For example, while interacting with a level set or region-growing algorithm, the user may notice that the segmentation "leaked" out of the desired organ into a nearby organ (imagine the liver leaking out between the ribs). The user would say "Avoid ribs", and the computer would then construct an avoidance mask, or region into which the segmentation is not allowed to leak, and then re-apply the region-growing algorithm with this constraint in place. By "mask", we refer to an image that represents a binary segmentation, 1's for foreground and 0's for background. A preferred embodiment allows the user to vocally construct these anatomical masks by saying the names of the organs to include in the mask, and also saying how to employ the mask. For example, the command "Stay below the hyoid", would result in the computer constructing an avoidance mask by first copying the binary segmentation of the hyoid bone onto a blank image, and then filling in all voxels above (superior to) the hyoid. The user could continue to add other organs and directions, such as "And stay left of sternum."

[0029] Some embodiments are cloud-based. Some "clouds" could actually be human technicians ready to respond to physicians. As the physician interacts with the segmentation algorithm, a video can be generated automatically that shows all the pointing and drawing strokes that the physician is making on the image. The physician's voice is superimposed over these actions. Given such a video, a human technician or medical resident could perform some meticulous and time-consuming manual image segmentation tasks in response to just a few seconds of a physician's instructions via video. This can be a significant time-saver for the practicing clinician.

[0030] The video communication can also go the opposite direction, from cloud to physician. In this case, the cloud, whether a human technician, or automatic algorithm, or some combination thereof, would record a video including voice that requests clarification from the physician. For example, while pointing at a certain object, the video could ask "Is this lesion?", or "I'm unsure about this." The physician can then respond very quickly with a video message that combines voice recording with annotated images to say, for example, "Lesion" or "Edema". Note that this is also a form of system learning, even when the system comprises human technicians, because the technicians are learning to become better segmenters from their interactions with the physicians.

1. A computerized method for image segmentation, the computerized method comprising:

accessing a set of at least one image to be segmented;

initiating an interactive segmentation process on the acquired set of images;

receiving voice commands and pointing inputs from the user; and

incorporating the received voice commands and pointing inputs into the interactive segmentation process.

2. The computerized method of claim 2, wherein the acquired set of images further comprises:

one or more MRI image, CT image, PET image, X-ray image, ultrasound image.

3. The computerized method of claim 2, wherein the interactive segmentation process delineates the boundary of one or more tumor, lesion, nodule, organs at risk.

4. The computerized method claim of 1, wherein the voice commands control the interpretation of the pointing.

5. The computerized method of claim 1, wherein the voice commands control the governing parameters of the segmentation process.

6. The computerized method of claim 1, wherein the voice commands dictate algorithmic steps for the interactive segmentation process to perform.

7. The computerized method of claim 1, wherein the voice commands direct how to grow or shrink a segmented structure.

8. The computerized method of claim 4, wherein the voice-controlled interpretation of the pointing can be one or more of specifying the type of structure being pointed to, placing a seed point for a region-growing algorithm, locating an anatomic landmark, drawing a limiting boundary to a region of interest, drawing like a brush, indicating the size of the radius of the brush, changing the color of the brush, indicating whether the brush drawing adds or subtracts from the structure, indicating whether to draw precisely or smoothly, indicating whether providing positive or negative training examples to an adaptive algorithm.

9. The computerized method of claim 5, wherein the governing parameters can be one or more of threshold levels, smoothness of structure boundaries, radius of morphological erosion, radius of morphological dilation, sizes of holes to fill, sizes of islands to erase, level-set curvature parameter, level set threshold levels.

10. The computerized method of claim 6, wherein the algorithmic steps can be one or more initiate the segmentation process, conclude the segmentation process, undo previous action, repeat previous action, define a region of interest by listing a set of bounding anatomic landmarks, threshold within a region of interest, run connected component analysis, perform morphological erosion, perform morphological dilation, run level-set evolution, train a grow-cut algorithm, close a drawn contour.

11. The computerized method of claim 7, wherein the vocal commands to grow or shrink a segmented structure can be one or more of indicating the amount to grow or shrink as a percentage, indicating the amount to grow or shrink as a distance, indicating the amount to grow or shrink as an area, indicating the amount to grow or shrink as a volume, indicating the direction in which to grow or shrink as being toward or away from certain anatomic landmarks, indicating regions to avoid growing into as anatomic structures.

12. The computerized method of claim 11, wherein the vocal interaction further comprises repeating actions in a series of increments with commands that can be one or more of "more", "less", "again", "repeat", "closer", "further", "bigger", "smaller", "smoother", "finer", "darker", "brighter".

13. The computerized method of claim 12, wherein the voice commands further comprise directing visualization of the current results of the interactive image segmentation process through one or more of show next slice, show previous slice, show slice of different orientation, zoom 2D view, pan 2D view, adjust window/level, rotate 3D view, zoom 3D view, pan 3D view, show structures in 3D view, hide structures in 3D view, hide structures in 3D view, adjust opacity of structures in 3D view, alter window layout.

14. A system for voice-activated interactive image segmentation, the system comprising:

a graphical user interface configured to display images;

a pointing device;

a voice recognition module;

an interactive image segmentation process that responds to user input;

wherein the voice recognition module converts voice commands into inputs for the interactive segmentation process;

wherein the voice recognition module interprets voice commands to change the behavior of the pointing device as the pointing device provides inputs to the interactive segmentation process;

wherein the segmentation process continues to perform interactive segmentation and to incorporate user input until the user indicates satisfaction.

15. The system of claim 14, wherein the pointing device can be one of computer mouse, stylus, trackball, touch screen, haptic interface, eye gaze detector, gesture recognition system.

16. The system of claim 15, wherein the voice commands perform one or more of controlling the interpretation of the pointing, controlling the governing parameters of the segmentation process, dictating algorithmic steps for the segmentation process to perform, directing how to grow or shrink a segmented structure, recording image annotation as a communication video.

17. A computerized method for image segmentation, the computerized method comprising:

accessing a set of at least one image to be segmented;

applying an automatic segmentation process on the acquired set of images;

receiving user input;

incorporating the received user input into an interactive segmentation process on the set of images;

repeating the action of receiving and incorporating until the segmentation of the acquired set of images satisfies user requirements; and

storing knowledge gained from the interactive segmentation process for further use by the automatic segmentation process.

18. The computerized method of claim 17, wherein the automatic segmentation process further comprises applying knowledge in the form of one or more of probability distributions, statistical models, spatially varying priors for Bayesian classification, distance transforms, warp fields, shape parameters, spatial relationships to other structures, curvature of boundaries, physiological angles, physiological distances, polynomial coefficients, profiles along rays emanating from the boundaries of segmented structures.

19. The computerized method of claim 18, wherein the user input further comprises:

speech recognition of commands that direct the interactive segmentation process.

20. The computerized method of claim 19, wherein the speech recognition further comprises:

one or more controlling the interpretation of the user input, controlling the governing parameters of the segmentation process, dictating algorithmic steps for the segmen-

tation process to perform, directing how to grow or shrink a segmented structure, indicating which actions to learn from and which actions to not learn from.

* * * * *

专利名称(译)	语音激活图像分割的系统和方法		
公开(公告)号	US20160104293A1	公开(公告)日	2016-04-14
申请号	US14/874425	申请日	2015-10-03
[标]申请(专利权)人(译)	GERING DAVID THOMAS		
申请(专利权)人(译)	GERING，DAVID THOMAS		
当前申请(专利权)人(译)	GERING，DAVID THOMAS		
[标]发明人	GERING DAVID THOMAS		
发明人	GERING, DAVID THOMAS		
IPC分类号	G06T7/00 A61B8/08 A61B6/03 G06T11/00 G06F3/16		
CPC分类号	G06T7/0081 G06T2207/20036 G06T7/0097 G06T11/001 G06F3/167 A61B6/032 A61B6/037 A61B8/085 G06T2200/04 G06T2207/20104 G06T2207/30096 G06T2207/30196 G06T2207/10072 G06T2207/20141 G06T7/0012 A61B6/461 A61B6/468 A61B8/00 G01R33/546		
优先权	62/071897 2014-10-03 US		
其他公开文献	US9730671		
外部链接	Espacenet USPTO		

摘要(译)

一种用于将语音命令结合到图像分割的交互过程中的方法和系统。交互式图像分割涉及用户指向图像;语音命令通过指示指向的目的和功能来加速这种交互。语音命令控制分段算法的控制参数。语音命令引导系统从用户的动作中学习，并从用户手动编辑自动分段的结果。