Interactive tube finding
on a Virtual Workbench

Tim Poston, Luis Serra, Wayne Lawton, Beng Choon Chua

Centre for Information-Enhanced Medicine

Institute of Systems Science

National University of Singapore

```
{tim, luis, wlawton,
 bcchua}@iss.nus.sg
```

Abstract

Medical images of such branching structures as blood vessels are clear to the human visual system, but easily confuse computer vision programs. We present an intuitive, hand-eye coordinated, reach-in interface that allows the user to sketch central curves for arteries, nerves, etc., detectable in volume data and in 3D space, and have this position/shape estimate refined by active contour methods.

CR Categories and Subject Descriptors: I.3.2 [**Computer Graphics**]: Graphics Systems; I.3.6 [**Computer Graphics**]: Methodology and Techniques - Interaction techniques.

General terms: Human Factors

Additional Key Words and Phrases: Dexterity, hand-eye coordination, calibration, medical.

1. Introduction

We describe a system for machine-assisted user mapping of tubular structures in 3D data, whose advantages to image interpretation are threefold. First, in this environment one can localize, zoom, change viewing angle, etc., as easily as one manipulates 2D images with standard packages. Second, the results of local study (for instance, of a complex branching network) are intuitively input to the system and displayed by it as curves, so that much of the burden of integrating local analysis to global understanding moves from the user to the machine. ("Is point X downstream of point Y?", for instance, becomes a more tractable question on the extracted network.) Finally,

since user curve estimates provide a good starting point for computation, the path is open for users to put questions such as "Is there numerical evidence for stenosis in this section of artery?" to the system; a path we intend to develop.

Medical images are full of branching tubular structures; blood vessels, lungs, nerves, microtubules within cells, and so on. The human visual system can perceive these clearly, and usually sees their topology correctly—more so than existing computer vision programs, which are easily confused by noise or apparent gaps. (A human knows that contrast agent in an angiography image must have somehow crossed an apparently missing segment of artery, and will join the broken ends. To make a computer program 'know' this is surprisingly difficult.)

Except in cases where they are laid out on or near a surface, as are the vessels that serve the heart, the image is usually full of apparent crossings (Figure 1). This makes the task of computer vision design much harder, but does not greatly confuse a human with good stereo perception if $6°$-separated views are available. (Computed stereo fusion remains a partly solved problem, successful mainly where there are sharp 'features' like machine corners for the correspondence matcher to focus on.) This is a case for human-machine synergy, with the human sketching the topology and the machine adding precision and integrating the results with other data (perfusion, brain activity revealed by PET, volumetric change in the heart or lung cycle . . .).

In a 2D layout it is a simple matter to sketch the network with a standard mouse for further, automated analysis, though mouse-sketching is made harder by the dissonance between hand and cursor positions, sensed via muscles and eyes. (Few people can convincingly forge their own signatures in a Paint program.) In 3D one needs to pass depth information to the computer also, in a comfortable, hand-eye coordinated input scheme.

We describe a robust environment for general 3D manipulation tasks in Section 2, with the specifics for tube tracing in Section 4, after Section 3 outlines our approach to bringing volume displays to interaction speed. We have explored its use with clinical volume data, but it is not yet embedded in a more general system for the use of clinicians.

Most 3D interfaces are based either (a) on the head mounted display 'immersive' paradigm, (e.g., [3] [13]), with its low resolution, long drawing delays for complex objects (and consequent simulator sickness), and calibration problems, that combine to make accurate work extremely hard; or (b) on a stereo display viewed directly (not through a mirror) [1] [4], so that the hand and the sensor it holds cannot go behind any part of the display without obscuring the screen and destroying stereo; furthermore, when such systems use negative parallax to bring the display 'in reach' in front of the screen, they create severe conflict between the stereo and accommodation depth cues. The calibration of the Virtual Workbench is straightforward [8], and the depth cues from the single-eye focus and two-eye stereo fusion are in better accord.

Section 5 briefly describes the numerical variational approach by which the user's sketch can be turned into a system of branched 3D curves that more precisely matches the data, enabling quantitative analysis of (for instance) blood flow along the tubes centred on these curves; a more detailed mathematical account will appear elsewhere. Conclusions are given in Section 7.

2. The Virtual Workbench

This section briefly describes a general environment for close, careful work, intended first for 3D image manipulation and in due course for interaction with the finite element models of flexible objects (such as tissues and minimally invasive surgery equipment) whose creation is part of the CI$_E$MED[1] programme. We have set up tools for 3D image rotation, volume of interest selection, etc., more comfortably than with a 2D mouse interface. (For a longer discussion of the overall environment, including issues of calibration, etc., see [8]). Section 3 describes its application to the problem of locating branched one-dimensional structures in volume data: like the heart wall slice contour editor described in [11], this illustrates its potential for dextrous interaction with the computer to produce medically useful output (beyond display) from the user's input.

We have prototyped a Virtual Workbench, a form of VR workstation modelled more on the binocular microscope

familiar to medical workers than on the immersion paradigm, but 'hands on' in a way that a microscope is not. We currently use a fixed viewpoint, which is not suitable for immersive VR (where typical view distances are beyond one metre, and parallax is the most important depth cue), but excellent for objects within the skilled worker's natural work volume, where stereo fusion is the dominant cue. We intend shortly to add head tracking, but for work within this volume we have given priority to stereo. Currently the Virtual Workbench is based on an SGI Reality Engine, but we expect soon to have a version based on a high-end PC. We have constructed an API in C++ that makes building such Virtual Workbench applications reasonably straightforward.

Our display configuration (Figure 2), using stereo achieved by time-splitting the display with Crystal Eyes™ glasses, allows the real hands to move in the positions at which the eyes seem to look. (Placing a hand directly in front of the screen would make it unmaskable by graphical objects that should be between it and the eye.) Mammals have multiple 3D systems—visual, neuromuscular, aural—and human dexterity is based on unusually strong integration of the first two. Fine control depends on exploiting this.

Schmandt's system [10], the closest (and earliest) analogous work we know of, used a half silvered instead of a full mirror, mixing the view of the hand with that of the graphics. This approach suffered from conflicts between graphical and physical objects which he tried to resolve by controlling the brightness of the graphical objects as well as by controlling the illumination of the hand. His system did not lead to substantial follow up, amongst other things, because the workspace corresponding to ours was set too low, accounting for their "lack of arm support and limited hand stability", and created accuracy problems when drawing was attempted. (An elbow on the table improves control and reduces fatigue: for the most delicate work, we are also studying the design of an adjustable elbow support.)

We have experimented with several types of sensor to be the 'tool' held by the hand. A mechanically linked sensor (the Immersion Probe™) is awkward in certain positions, but is more stable, accurate and reproducible in its output than electromagnetic and acoustic sensors. Unfortunately it does not include a button as a standard feature, and the custom added button with which we obtained it is far inferior to the engineering standard of the basic design, being uncomfortably hard and narrow to the finger, and subject to failure. Slightly less accurate, and subject to interference from the monitor, is the electromagnetic Polhemus FasTrak™ stylus. This is more costly, but its freedom of movement and more comfortable grip make it our instrument of choice when its precision is sufficient. The stereo display shows an image of the tool (as pointer, scalpel, loop, . . .) but not

---

1. That is, the Centre for Information-enhanced Medicine established jointly by Johns Hopkins University and the Institute of Systems Science of the National University of Singapore.

the hand itself; it is vital to see and feel where the tool is, but a visual display of the hand merely obscures the workpiece at an unnecessary cost in coding and CPU time. The hand is supported at the elbow, or for special tasks at the forearm near the region where delicate work is to be done, avoiding problems of fatigue. Good, high-resolution stereo within the natural work volume requires only an appropriate monitor and glasses, bought in increasingly competitive markets. (In an immersive volume, vision remains costly, high resolution is not yet available, and adequate hand-eye calibration is problematic.)

This is our environment of choice for dextrous work.

3. Working in a displayed volume

The bare minimum speed for interaction that can allow dexterity is about ten images per second; anything less makes the problem of overshooting hand movements intolerable for serious work, and more is highly desirable. The stereo requirement makes this twenty.

Very few medical images are displayed at these speeds. In most systems the fastest displays are of extracted surfaces; but extraction by (for instance) the Marching Cubes algorithm[2] typically produces hundreds of thousands of triangles, when applied to a medical volume image. Even the SGI Reality Engine's million triangles/sec.—achievable under somewhat atypical conditions—cannot achieve ten stereo views/sec. in this situation. In any case, extracting a surface that separates grid points above a somewhat arbitrary threshold τ from those below τ begs important questions in the analysis of uncertain data. Humans appear to see tubes better than isosurface algorithms do.

It is thus appropriate to work with volume rendered images. Until recently these have been displayed significantly more slowly than triangulated surfaces, but the new hardware support for '3D textures' [2] [7] [14] has made a great advance in rendering speed. In a cardiac context, for example, we have a Virtual Workbench tool [12] that allows the user to move a cutting plane through a beating heart volume cine-loop at interactive speeds, using a 3D texture display on a Reality Engine at 15 stereo frames/second, when restricted to the volume containing the heart. This, however, is a volume that (in the dataset we are using) includes only 20 slices of 256x256 data points each. An angiography record like that in Figure 1 contains 64 slices of 256x256 data points, and cannot be redrawn at such comfortable speeds. The solution is to avoid re-rendering, whenever possible.

---

2. We have found an alternative algorithm [9] that yields 25% fewer triangles, but this does not make most medical images interactively rotatable.

Figure 3 shows that an interactive tool, and simple changing objects such as curves, can be visually embedded in a 3D volume display without continuously repeating the volume rendering computations. While working from a particular viewpoint, it is sufficient to store the left- and right-eye images of the volume as bitmaps, which can be moved rapidly from a buffer to the screen, and the 'active' elements of the display with a small number of lines and translucent polygons. A steady image is in any case appropriate when one is trying to reach in and select a point for change or deletion, or sketch a curve. When a change of viewpoint is wanted, the user selects a 'rotator' tool and turns the image by pulling it around . . . at a far lower spatial resolution. By invoking 3D texture slices at twice the spacing while the turning occurs, the system displays 10 stereo frames/sec., sufficient for comfortable interaction and choice of a new angle of view. Once this has been chosen, the system takes 0.195 sec. to produce and store a fully detailed volume rendering, after which the user can continue to interact at 15 frames/sec. with the display.

The ideal system would allow head tracking, so that parallax would give an added depth cue; but the close neural coupling between the inner ear's motion sensors and the eyes—providing an involuntary eye movement to compensate for head motion—makes the latency problem particularly acute in this case. Ten views per second is far from enough to overcome the jitter from head movement if one is trying to concentrate on a delicate task, and failure to keep pace with the eyes' corrective motion often results in 'simulator sickness'.

We also reduce the perceived latency by using a predictive filter on user input [6], so that the stylus is displayed closer to where the tool is than to where it was last read.

4. The curve editor

The Curve Editor enables the intuitive creation, modification and deletion of 3D curves by means of the interactive tool (or 'stylus'). The user creates curves as 3D splines by sketching them with 3D stylus motions, matched to the different tubes visible in the stereo maximum intensity images. One can later edit these curves, while looking from the same or a different direction, to achieve better matches. (Errors in depth perception can thus be corrected from a side view.)

The virtual 'tool rack' (Figure 4) holds the buttons that enable tools and activate different modes of operation.

The Curve Editor provides the following tools:

Sketch   as the tip moves, a spline curve grows to follow it, adding nodes either equidistantly, while the stylus switch is held down, or (selected by the button curve+) whenever the stylus switch is clicked. In the first case, a new curve begins

whenever the switch is released and re-pressed; in the second, curve+ must be clicked again for a new curve.

Curve– delete a selected curve.

Edit   grab and drag an element of a curve. In Figure 3, the line from the stylus tip shows the ends of the segment that would currently be grabbed by holding the stylus switch; a new node would appear, held by the stylus tip. When near enough to a node to highlight it, holding the switch grasps that node instead.

Cut   break a curve into two at a node.

Link   either end to end, or branching, merge two curves into a single connected structure.

Clone   copy a connected curve structure.

Move   select a connected curve structure and move it around rigidly, like a carrot on a fork.

Mirror   produce a mirror image of a connected curve structure. (As in Figure 3, approximate symmetry is common; it is faster to edit a reflected curve—which cannot be produced by rotating a clone— than to create the opposite curve from scratch.).

Rotate   when the stylus switch is held down, the displayed complex of volume display and curves rotates to follow a 'rubber' string joining its centre C to the stylus tip, each step being a rotation in the plane formed by C and two successive stylus positions. Intuitively, the feel is 'reach in and pull it around its centre'. (Volume resolution is lowered during this operation, in the interests of interactive speed.).

Each button's icon shows whether it chooses a manipulator like edit or rotate, an action such as clone, or a change of toolrack. When the stylus is near the toolrack, the button nearest the tip is highlighted and can be activated by pressing the switch. Alternatively, a constrained mouse controls this 1D row of buttons when the stylus is not close to it.

This two-handed approach provides an effective interaction scheme, in which 3D-intensive operations are performed with the 3D stylus in the dominant hand, while coarse 1D selections are left to the other.

5. Variational improvement

It is tedious and difficult to mark the centre of a tube (in a 2D or a 3D image) with great precision. Fortunately, the human topological sense—hard to recreate in a computer—can be complemented by the computer's power of precise estimation. Given a good 'initial guess' at the position of a curve, so that the algorithm is not distracted by aspects of the image that are created by noise or real tubes at a different depth from the tube in question, numerical methods can greatly improve the description of its position. (It is an interesting power of the human visual system to ignore noise in a 3D data set when it is perceived in a stereo display to be at an irrelevant depth. By passing approximate depth information to the computer in an initial guess, much of this power is shared with the system.)

The initial estimate for the centre of a tube is a parametrized space curve

$$u_o : [0,L] \to R^3 \qquad (1)$$

that interpolates a sequence of user-chosen points, as above. This initial guess can be improved by the *active geometry* methods developed in [5], as follows.

First, construct an energy function $E(u)$ (for a class of space curves $u(s)$), such that $E(u)$ is small when $u(s)$ is smooth and its parallel projections to two different planes match the centres of tubes in each stereo image. (The latter involves constructing an 'unlikeness-to-tube-centre' function on each image, as in Figure 5, and integrating it along each curve.) An image potential defined on an entire 3D volume can be set up, but for reasons of both RAM and CPU its construction and use are very slow. Using 2D potentials on paired projections is an innovation we have introduced, that makes the timing practical. If only a single pair of stereo views is available[3], the active geometry can use these (and all of the editing tools in Section 3 remain usable, except that the volume cannot be rotated), but when volume data are present it is faster and more robust to use projections in orthogonal directions.

Second, use an iterative minimization algorithm to compute a sequence of parametrized curves u^n that converges to a curve u_{min} that minimizes E. The curve u_{min} is a refined centre-of-tube estimate.

The external force consists of two terms, one from each of the image potentials. Thus the image derived forces attempt to force the two planar projections of the spatial contour to simultaneously fit the two projected images. Figure 6 illustrates the resulting change.

6. Further work

The variable blur in CT images of arteries due, for instance, to pulse motion during the scan) makes threshold isosurface extraction a less reliable tool for diagnostic detection stenosis (narrowing) than visual inspection of MIP images [newref]. With a variationally-enhanced estimate of the centre curve, we are designing numerical tools that study cross-sections orthogonal to it, to estimate tube capacity more robustly than is possible with an arbitrary-threshold based approach.

---

3. A stereo pair of X-rays, for instance, can be collected faster and with less radiation load than a full CT scan.

7. Conclusions

Proper management of what is rendered when, and suitable choice of opacities, makes it possible to do delicate work involving medical volume data, displayed in full detail, at interactive speeds.

We have described a hand-eye coordinated 'reach-in' environment for such work; in particular, for the frequently important task of tracing the central curves of tubular structures in the data set, applicable also to other near-one-dimensional features such as the sutures joining parts of the cranium: vital landmarks for medical work. We have implemented a variational method for improving these user-supplied tracings, and we are working on such methods for the location of higher-dimensional landmarks.

8. Bibliography

[1] Deering, M., High Resolution Virtual Reality, *Computer Graphics* **26**, 1992, pp. 195-201.

[2] Cullip, U., and Neumann, U., Accelerating Volume Reconstruction with 3D Texture Hardware. *UNC Tech. Report TR93-0027* (1993)

[3] Halliday, S., and Green, M., A Geometric Modeling and Animation System for Virtual Reality. In *Proceedings ACM VRST'94*, (1994), World Scientific, pp. 71-84.

[4] Iwata, H., Artificial Reality with Force-Feedback: Development of Desktop Virtual Space with Compact Master Manipulator, *Computer Graphics* **24**, 1990, pp. 165-170.

[5] Lawton, W., Mathematical Methods for Active Geometry, in *Proceedings of International Conference on Computer Aided Geometric Design*, Penang, Malaysia, July 4-8, 1994, Baltzer, 1995.

[6] Lawton, W., Poston, T. and Serra, L., Time-lag reduction in a medical Virtual Workbench, in *Proceedings of British Computer Society conference "Applications of Virtual Reality"*, Leeds, June 1994.

[7] Perlin, K., and Hoffert, E.,M., Hypertexture, *Computer Graphics (SIGGRAPH '89 Proceedings)* **23**, (1989), pp. 253-262.

[8] Poston, T. and Serra, L., "The Virtual Workbench: Dextrous VR", In *Proceedings ACM VRST'94*, (1994), World Scientific, pp. 111-122.

[9] Poston, P., Nguyen, H.T., and Heng, P.A., 'Skeleton Climbing': hole-free isosurfaces with fewer triangles. submitted to *Visualisation'95*.

[10] Schmandt, C., Spatial Input/Display Correspondence in a Stereoscopic Computer Graphic Work Station, *Computer Graphics* **17**, 1983, pp. 253-259.

[11] Serra, L., Poston, T., Ng, H., Heng P.A., and Chua B.C., "Virtual Space Editing of Tagged MRI Heart Data", in *Proceedings CVRMed'95 - First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine* April 3-5, 1995 Nice, France, pp. 70-76.

[12] Solaiyappan, M., Poston, T., Heng, P.A., Zerhouni, E.A., McVeigh, E.R., and Guttman, M.A., Interactive Visualization for Speedy Non-invasive Cardiac Assessment, submitted to *IEEE COMPUTER and IEEE Computer Graphics and Applications,* Special January 1996 Issues on Computer Applications in Surgery and Therapeutic Procedures.

[13] Slater, M., Usoh, M., and Steed, A., Steps and Ladders in Virtual Reality. In *Proceedings ACM VRST'94*, (1994), World Scientific, pp. 45-54.

[14] Wilson, O., van Gelder, A., and Williams, J., Direct Volume Rendering via 3D Textures, *UC Santa Cruz technical report*, UCSC-CRL-94-19, June 1994.

Figure 1. A typical angiography image, obtained from volume data by maximum intensity projection. Note that the apparent lumps and branching of the arteries displayed become less deceptive upon fusion of this stereo pair. (The combination of human depth perception and a good stereo display is capable of resolving the structure of a far more complex network than this, but given the available hard copy technology we have chosen a relatively sparse image for clarity.)
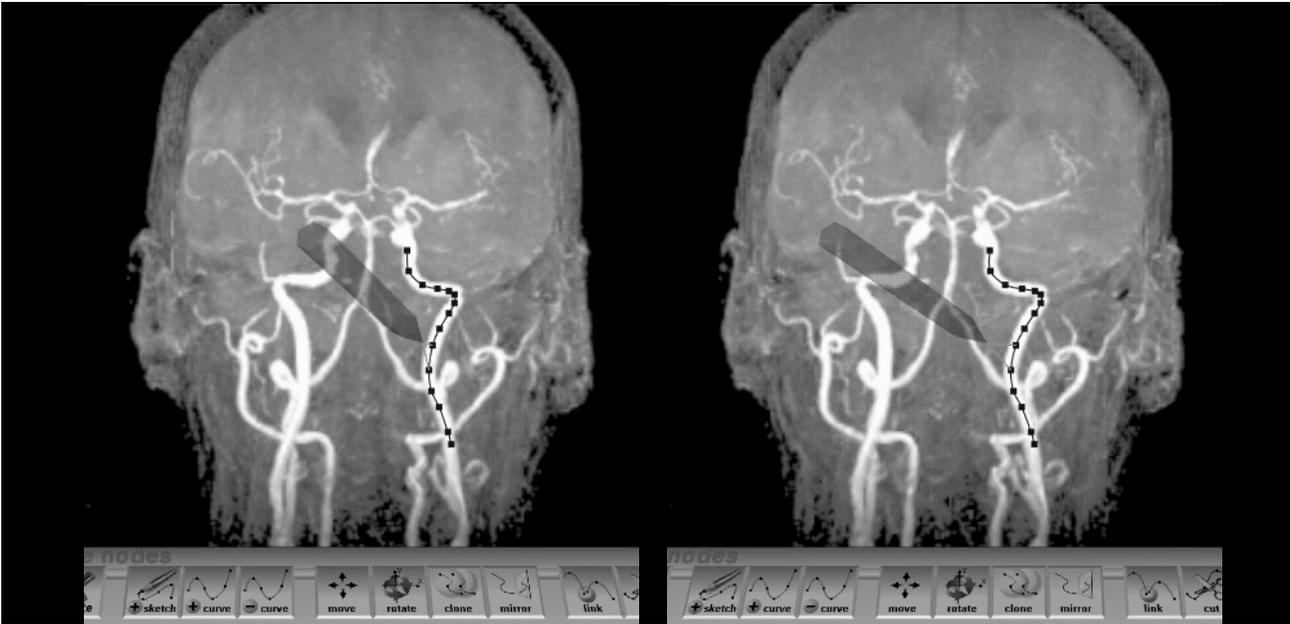

Figure 3. The Virtual Workbench 3D curve editing tool. Upon stereo fusion one sees the arteries as translucent, with the 'hand-held' stylus showing through them, though graphically the reverse is true: the left and right eye views are captured as pixel background images, and the stylus rendered with $\alpha = 0.6$ on top of them.
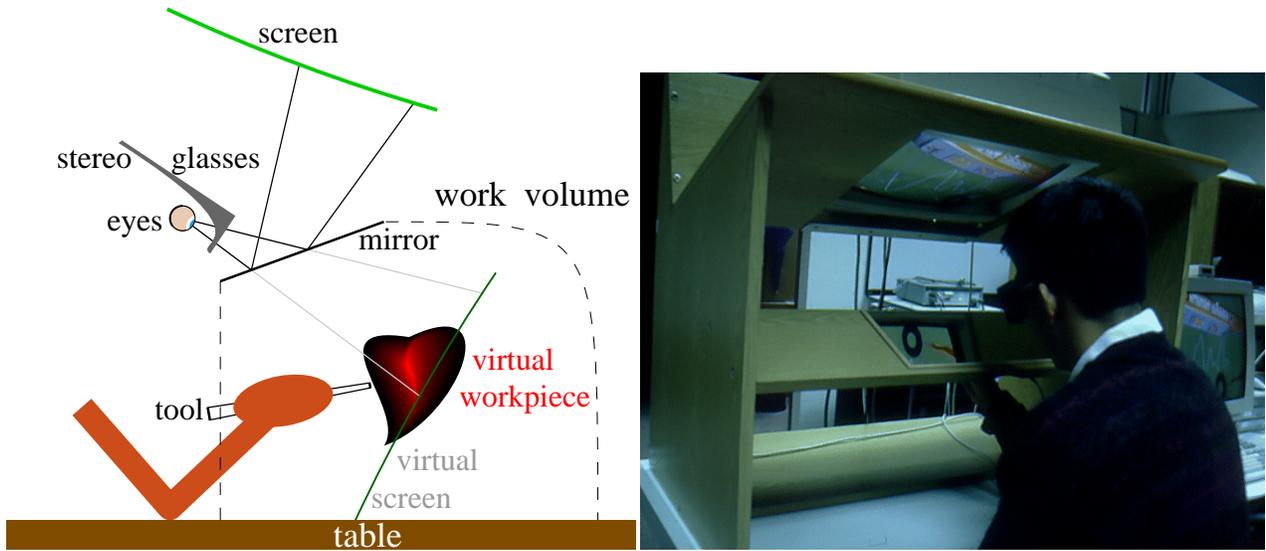
Figure 2. The Virtual Workbench. The left view shows a schematic, the right a photograph of the hardware implementation. The user sees a virtual tool, in the stereo display in the mirror, exactly where the hand holds the real sensor-equipped handle: the hand is invisible.
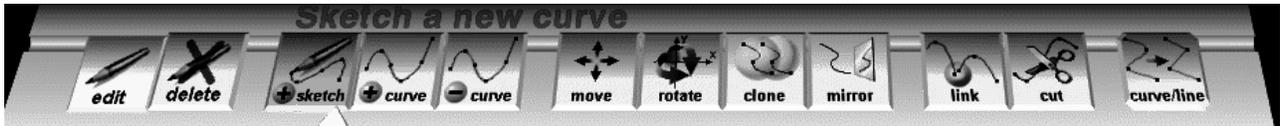

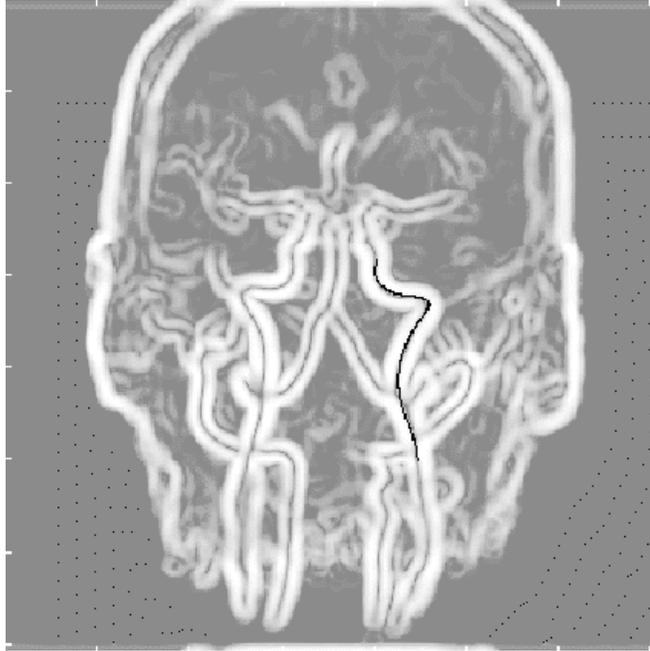Figure 4. Selection buttons controlling the editor shown in Figure 3.

Figure 5. A 2D image potential derived from the right image in Figure 3, with lowest values (shown dark) at points most plausibly lying on the central curve of some tube 12 pixel units thick, in some direction. The black curve is a dynamic contour, drawn toward the tube-central dark points: clearly, a good initial guess is vital.

(Not stereo)

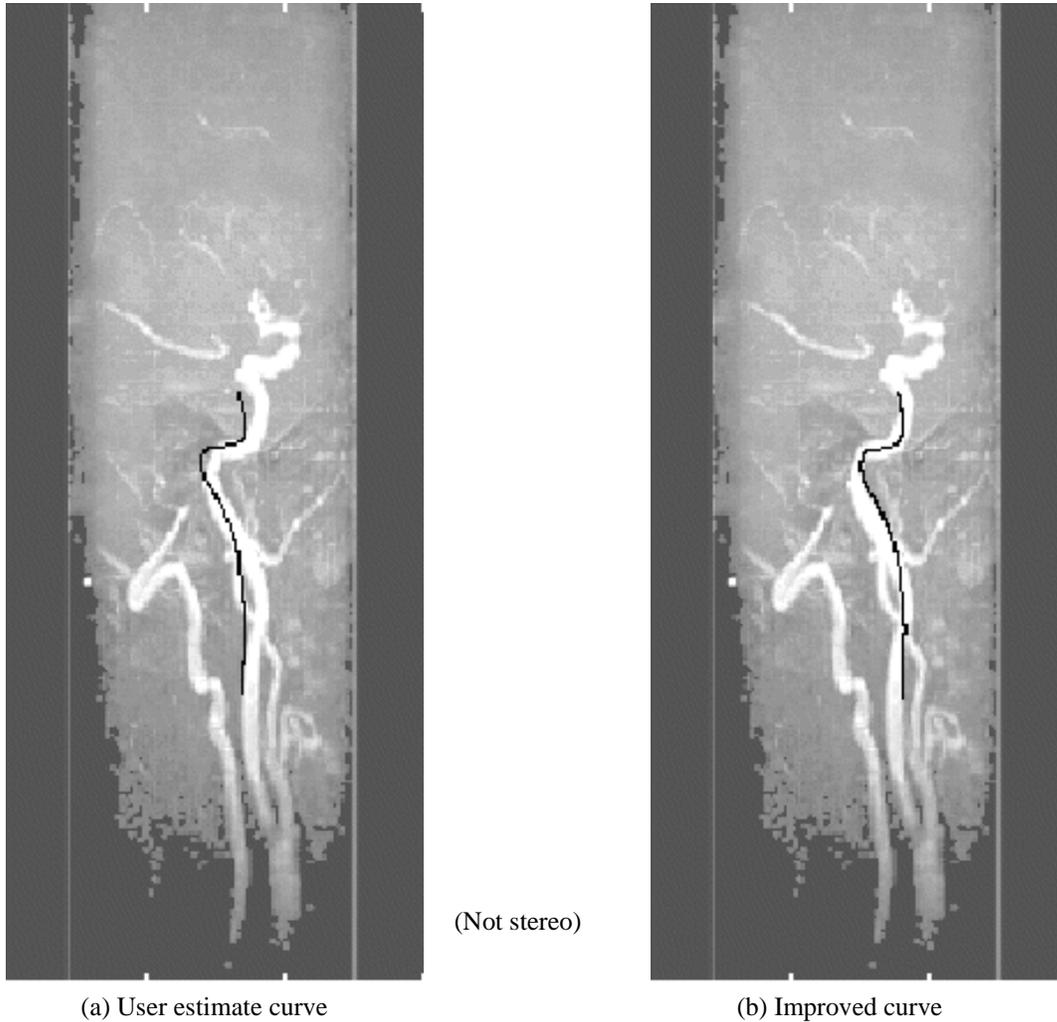(a) User estimate curve   (b) Improved curve

Figure 6. A side view of a sub-block of the volume in Figure 3, containing the selected artery and unconfused by the arteries in the other side of the head. The initial user estimate (with 'depth' now showing as left *vs.* right) is displayed at left, the algorithmically improved one at right. Note that the apparent crossing near the bottom of the curve would confuse a 2D active contour, leaving it on the loop at left; the stereo variational approach converges correctly to the main artery.