Medical Applications of Image Understanding

W.E.L. Grimson, Artificial Intelligence Laboratory, Massachusetts Institute of Technology

MANY SURGICAL PROCEDURES REQUIRE HIGHLY PRECISE 3D LOCALIZATION TO EXTRACT DEEPLY BURIED TARGETED TISSUE WHILE MINIMIZING COLLATERAL DAMAGE. THE AUTHOR DESCRIBES A SYSTEM THAT BUILDS ON A WIDE RANGE OF IU METHODS TO ACHIEVE THIS END.

IMAGE UNDERSTANDING RESEARCH traditionally has focused on problems in areas such as aerial photo-interpretation, industrial automation, and autonomous navigation. These applications range from automatic target recognition and robotic hand-eye coordination, to landmark recognition and free-space detection. Researchers designed these IU methods to work with electro-optical sensors—such as standard visible-light cameras—and geometric objects that are rigid and easily articulated, such as polyhedra and simple quadric surfaces.

Medical imaging, on the other hand, presents a different set of challenges. Medical imaging problems typically use sensing methods with very different underlying physics: *magnetic resonance imaging* (MRI) or *computed tomography* (CT). These problems often deal with geometrically intricate objects, such as the surface of the cortex or the bronchial structure of the lungs. Medical problems also deal with flexible, deformable objects, as for example the nonrigid motion of the beating heart. Despite the apparent differences in underlying bases, developments over the past few decades in computer vision, especially through ARPA's Image Understanding Program, are beginning to revolutionize the use of medical imagery: in surgery, in diagnosis, and in therapy evaluation.

This applicability of IU techniques to medical problems holds for many of the central problems in traditional IU: extracting key features from the imagery, registering data sets, predicting images of object models from arbitrary viewpoints, and fitting parameterized surface models to data. By tailoring IU techniques designed for traditional vision to the special circumstances of medical imagery, systems are emerging that support effective use of medical image data.[1] To demonstrate the range of roles that IU methods play in medical image utilization, we describe an end-to-end system for image-guided surgery. This system directly builds on a wide range of IU methods to provide surgeons with visualization and guidance during surgical procedures.

A demonstration: image guided surgery

Many surgical procedures require highly precise 3D localization to extract deeply buried targeted tissue while minimizing collateral damage to adjacent structures. Image-guided surgery emerged to meet this need. While MRIs and CTs image and display the body's 3D structure, the surgeon must still relate what he or she sees in the 3D display with the patient's anatomy.

Surgeons usually examine 2D slices of MRI or CT imagery and then mentally transfer that information to the patient. Thus, there is a clear need for registered visualization techniques in which 3D reconstructions of internal anatomy are exactly overlaid with the surgeon's view of the patient. This allows

0885-9000/95/$4.00 © 1995 IEEE

An example scenario

We describe a scenario that demonstrates the use of our image guided methods here.

First, the staff scan a patient requiring surgical therapy with a three-imensional, high-resolution, internal anatomy scanner, such as MRI or CT. Our system automatically segments the scan into different tissue types, and generates graphical models of desired structures, such as tumors, ventricles, skin surface, bone, and white matter.

The patient relocates to the operating room, which is equipped with the following:

- a system for obtaining 3D positional information from the patient's skin surface at the locus of surgery;
- enhanced reality visualization equipment, such as a video or digital camera, mixer, and display monitor; or a head-mounted display with trackable landmarks;
- medical instrument holders containing trackable landmarks;
- an operating table that may contain fixed raised landmarks that will remain viewable and in the same position relative to the patient during surgery;
- landmark tracking equipment.

Prior to draping, the staff scans the patient with the 3D sensor. This sensor could be, for example, a laser range scanner or a passive stereo system. The system calculates the 3D locations of any table landmarks to identify their location relative to the patient.

The system then automatically registers the MRI or CT scan to the patient skin surface depth data that the range scanner obtained. This provides a transformation from MRI/CT to patient. Matching the video images of the range points on an object to the actual 3D data determines the position and orientation of a video camera relative to the patient. This provides a transformation from patient to video camera.

The enhanced reality visualization displays the registered internal anatomy to "see" inside the patient. In particular, the system uses the two previously computed transformations to transform the 3D model into the same view as the video image of the patient, so that video mixing allows the surgeon to see both images simultaneously. The staff drapes the patient and performs the surgery. The enhanced reality visualization neither interferes with the surgeon, nor requires him or her to do anything out of the ordinary. Rather, the system provides additional visualization information that greatly expands her limited field of view.

The system can continually track the location of the table landmarks to identify changes in the position of the patient's attitude, relative to the visualization camera. Visualization updates occur after updating the MRI/CT to patient transformation. The system can also continually track the viewer location to identify any changes in the position of the viewer. In the case of a stationary video camera, this is straightforward. In the case of head-mounted displays, such tracking is both more relevant and more challenging. Visualization updates take place when the system updates the patient-to-viewer transformation. The system may track the medical instruments as well to align them with predetermined locations as displayed in the enhanced reality visualization.

In general, the surgical procedure is executed with an accurately registered enhanced visualization of the entire relevant anatomy of the patient. By providing this information, the surgeon will be able to execute the procedures more efficiently and effectively, with reduced side effects to the patient.

the surgeon to directly visualize important structures and plan accordingly, with the images guiding the surgeon's execution of the procedures. This method requires that we automatically extract relevant image information, convert it to a form most useful to the surgeon, and present that information seamlessly to the surgeon by registering it to the patient. See Figure 1 for an example of what our system can deliver.

An enhanced reality surgical visualization system. Using a range of IU methods from visual object recognition, we have created a system that registers the clinical image data with the patient's position on the operating table at the time of surgery. We have combined the method with an enhanced reality visualization technique, in which we display a composite image of the 3D anatomical structures with a view of the patient (see Figure 1). After the surgeon analyzes the segmented 3D preoperative data, this registration enables the transfer of preoperative surgical plans to the operating room. The surgeon then graphically overlays the plans onto video images of the patient. This method allows the surgeon to apply carefully considered surgical plans and to make internal landmarks that guide the surgery's progression.

The following input are the conditions of the specific problem that the system solves: Given a video view of the patient, together with an MRI or CT model of the anatomy of the patient, each defined in its own coordinate system. We want the system to produce the following output: Segment the MRI/CT model into distinct, relevant anatomical structures; find a transformation aligning the model with the patient, and a transformation describing the position and orientation of the video camera relative to

Figure 1. Example of combining the registration of MRI to depth data and the calibration of a video camera relative to the depth data to provide an enhanced reality visualization of a patient. The tumor and the ventricles are displayed in exact registration with a live view of the patient. Using this registered data, the surgeon can mark the key structure positions, plan procedures, and check progress at intermediate stages of the procedure.

Best Copy Available

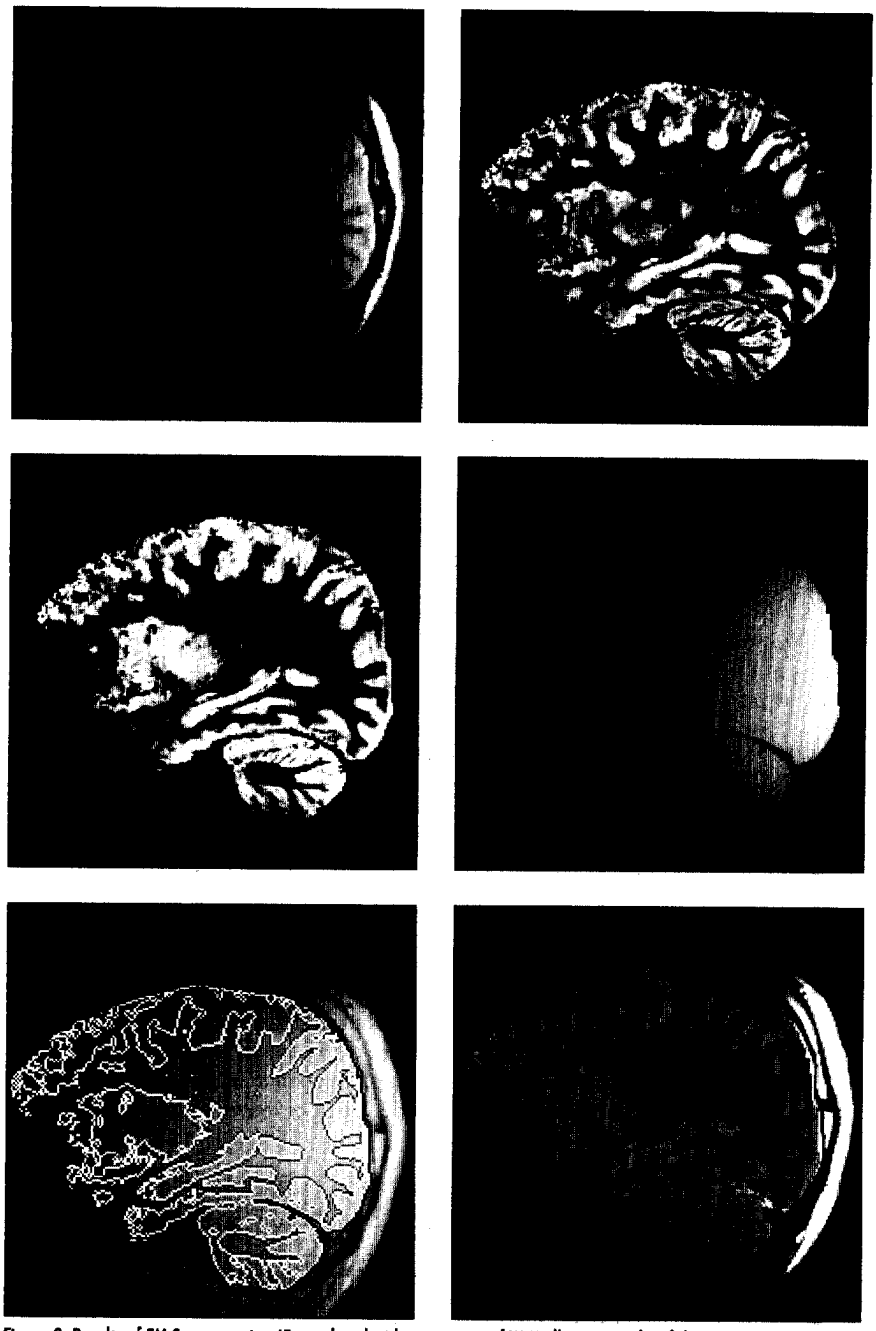

Figure 2. Results of EM Segmentation (Reproduced with permission of W. Wells III). Results of the EM segmenter on a sagittal section obtained using a surface coil. In the top row, on the left is the input image and on the right is the initial gray matter probability. In the second row, on the left is the final probability for gray matter obtained using EM segmentation, and on the right is the bias estimate. In the third row, on the left is the resultant classification of the input image, and on the right is the final intensity corrected image obtained from EM segmentation.

the patient; use these components to register and overlay the segmented model with the video image.

Once we have these transformations, we can present visualizations of internal structures to the surgeon. Extensions of the method also allow us to adaptively re-register the patient's video image to the 3D anatomical data as the patient moves or as the video source moves.

How IU is essential to this problem. In the following sections I discuss how well-developed IU techniques critically support most stages of this end-to-end system.

Segmenting the volume into tissue types. The first stage of our scenario is standard practice. The results of the scan are a set of 3D voxels, each with some associated intensity information, that reflects the interaction of the sensor with the tissue type. MRI and CT scans are often displayed as images. However, the underlying physics by which tissue type generates a signal that can be converted to a brightness value for display is quite different than in conventional visible spectrum sensors. Nonetheless, it is easy to adjust the IU techniques developed for the latter case to handle such sensing modalities.

The next step is the first stage in dealing with the volumetric medical image. Here it is necessary to convert signal values at voxels into tissue labels, a problem known in the vision community as *segmentation*. Under ideal conditions, this is a straightforward problem. We simply need to train a classifier, by identifying a small number of points in the imagery whose tissue type is known from anatomy. We can use the extracted set of intensity values associated with each tissue type as a basis for classification; all other elements in the volume are assigned a tissue label based on the cluster of known labels to which they are closest.

Unfortunately, things are not quite so nice in practice. First, the sensors can often have large gain effects, in which the underlying signal is corrupted and thus can take on a value inconsistent with its actual tissue cluster. Second, many tissue types can have very similar intensity values, thereby defeating the purpose of the classifier.

To overcome these problems, we can use a suite of methods from the IU community. First, consider the gain artifact. If we know the gain artifact, we could correct the signal and then use our classifiers to identify tissue type. Conversely, if we know the tissue type, we could predict the ideal signal and use that to solve for the gain artifact. But we don't know either. The solution is to use a technique that researchers have successfully used in IU problems, based on the Expectation/Maximization algorithm.[3] In short, we assume some initial value for the gain field and fix that value while solving for the best estimate of the tissue types. We then use those estimates to re-estimate the gain field, and iterate this process to convergence. The result is a highly reliable tissue classifier that has proven to be at least as good as highly trained expert segmenters. See Figures 2 and 3 for examples of EM segmentation.

This technique gives us a tissue labeling on a voxel by voxel basis. To turn this into structures of more direct utility, we need to extract connected structures with distinctive anatomical significance. Employing IU methods once again,[4] we can use morpho-

logical operators to isolate the common tissue type's connected volumetric components.

Image morphology provides a way to incorporate neighborhood and distance information into algorithms. The basic idea in morphology is to convolve an image with a given mask (known as the structuring element) and to binarize the result of the convolution using a given function. The choice of convolution mask and binarization function depends on the particular morphological operator being used. Standard morphological operations include:

- **Erosion:** An erosion operation on an image I containing labels 0 and 1, with a structuring element S, changes the value of pixel i in I from 1 to 0. This occurs if the result of convolving S with I, centered at i, is less than some predetermined value. The structuring element (also known as the *erosion kernel*) determines the details of how a particular erosion thins boundaries.
- **Dilation:** Similar to erosion, a dilation operation on an image I containing labels 0 and 1, with a structuring element S, changes the value of pixel i in I from 0 to 1. This occurs in dilation if the result of convolving S with I, centered at i, is more than some predetermined value. The structuring element (also known as the *dilation kernel*) determines the details of how a particular dilation increases boundaries in an image.
- **Conditional dilation:** This is a dilation operation with an added condition. Only pixels that are 1 in a second binary image, I_c, (the image on which the dilation is conditioned), will be changed to 1 by the dilation process. This operation is equivalent to masking the results of the dilation by the image I_c.
- **Opening:** An opening operation consists of an erosion followed by a dilation with the same structuring element.
- **Closing:** A closing operation consists of a dilation followed by an erosion with the same structuring element.

For example, to extract the intercranial cavity of the brain, we can use the following combination of morphological operations:

- Perform an erosion operation on the output of the EM segmenter—in other words, find and mark all the voxels with a particular tissue label. Perform this operation with a spherical structuring element with radius corresponding to the thickness of the connectors between brain and the cranium. The use of this radius eliminates connections from the brain to any misclassified nonbrain structure.
- Find the largest 3D connected component with tissue labels corresponding to the brain.
- Dilate the brain component obtained in the previous step by a structuring element comparable in size to the one used in the erosion, conditioned on the brain labels in the input image. This corresponds approximately to restoring the boundaries of the brain component that were distorted in the erosion step.

Figure 3. Results of EM Segmentation (Reproduced with permission of W. Wells III). Results of the EM segmenter on a double-echo axial MRI. Top left is a proton-density image, top right is a T2-weighted image. Segmentation on the bottom left is using a conventional statistical classifier. The EM segmentation on the bottom right is significantly better. Here, white is white matter, blue is cerebro-spinal fluid, and orange is other head tissue.

The result of this process is a set of segmented structures. The process, however, often blurs out fine geometric detail on the surface of the extracted structures. To refine this extracted set of structures, we use another very common IU tool: deformable models (also known as *snakes* or *balloons*).[5] Snakes, for example, are a common IU technique; they help to fit contours or surfaces to image data. Snakes operate by emulating a controllable elastic material, much like a thin, flexible sheet. We can initially position the model by using information from anatomical atlases; the model is then allowed to relax to a stationary position. This minimum energy position seeks to find the best position in which to trade off internal and external forces. The internal forces are due to the elastic nature of the material and the external forces stem from sharp boundaries in image intensities.

A deformable contour is a planar curve which has an initial position and an associated objective function. Witkin, Kass, and Terzopoulos[5] introduced the special class of deformable contours called *snakes* in which the user interactively specifies the initial position. This deformable contour's objective function is referred to as the *energy* of the snake. The energy of the snake (E_{snake}) is expressed as:

$$E_{snake} = E_{internal} + E_{external}. \quad (1)$$

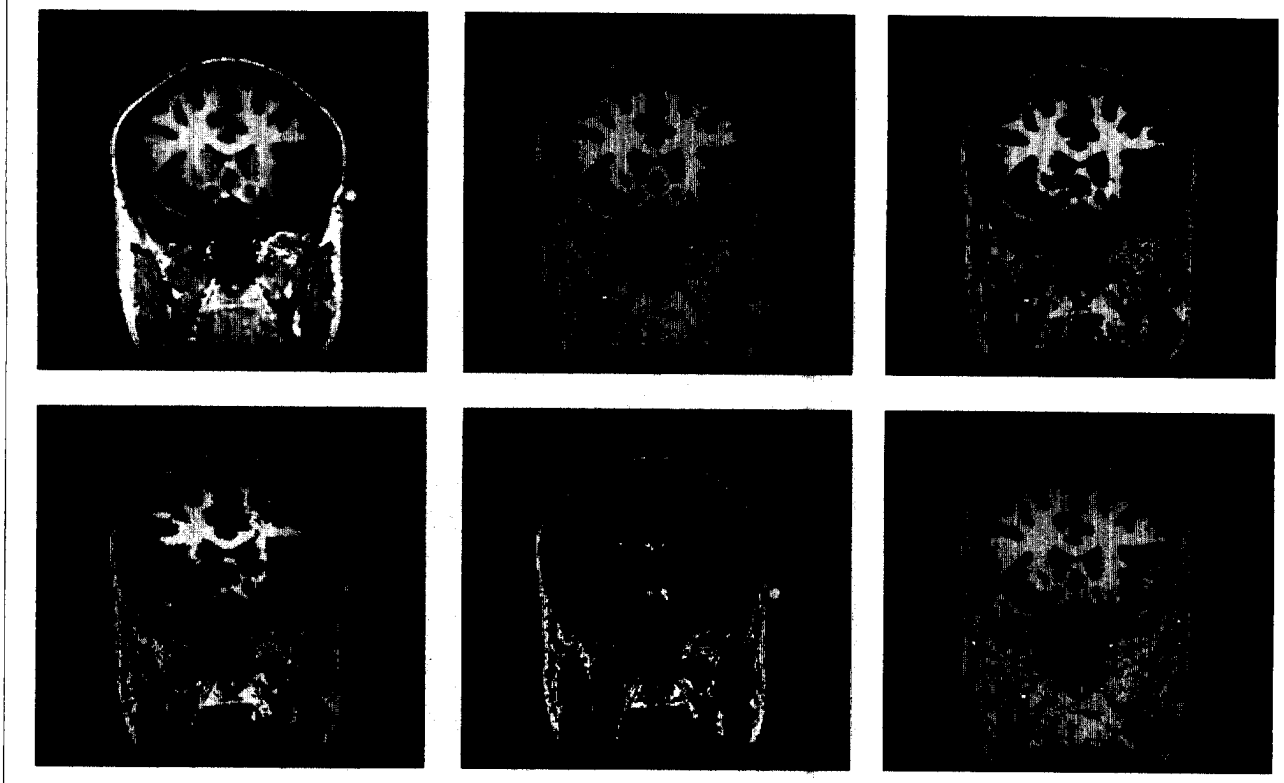

Figure 4. Top to bottom, left to right: Input image and tissue classification generated by successive iterations of the EM segmenter (white matter is brightest, gray matter is medium gray, and csf and air are black). (Courtesy of Tina Kapur.)

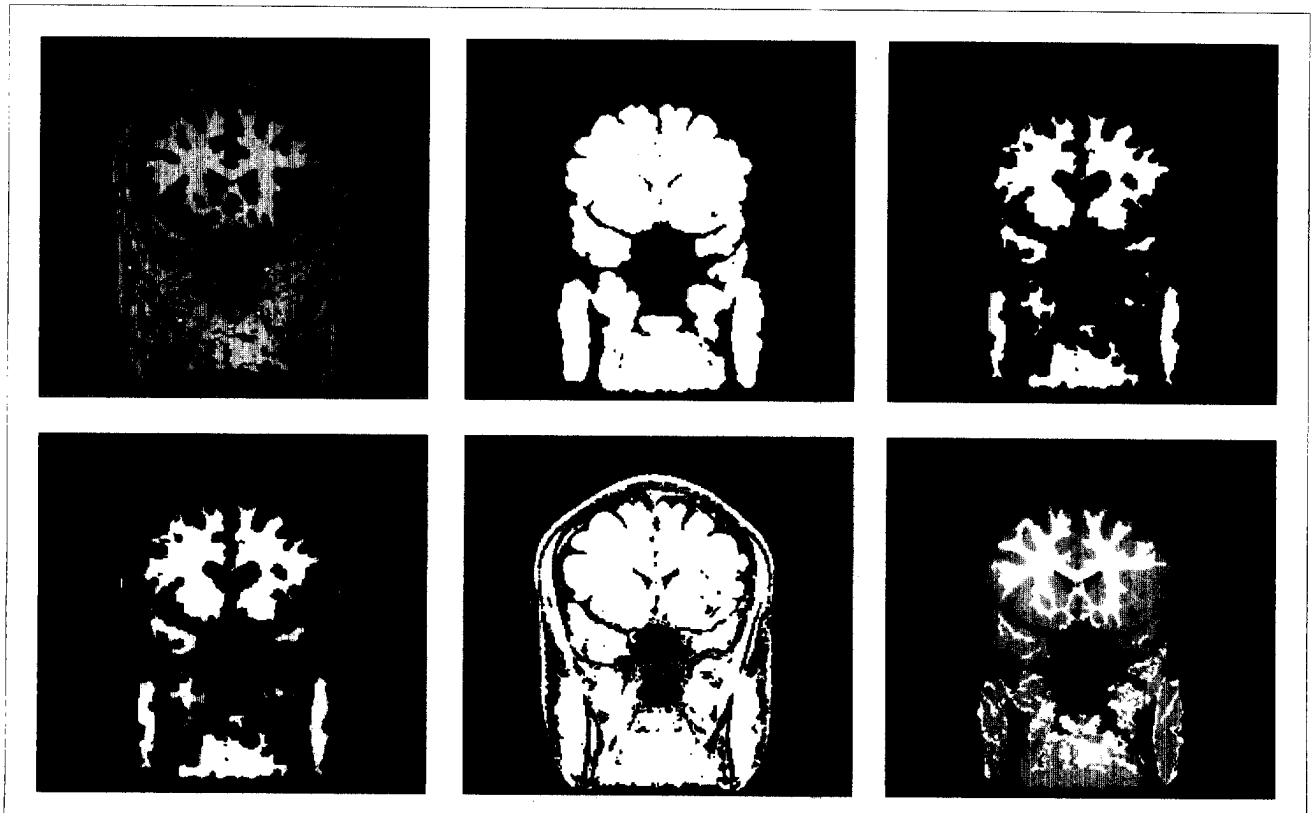

Figure 5. Top to bottom, left to right: EM segmentation from Figure 4, binarized image, eroded image, largest connected component in eroded image, dilated connected component, conditionally dilated labeled connected component. (Courtesy of Tina Kapur.)

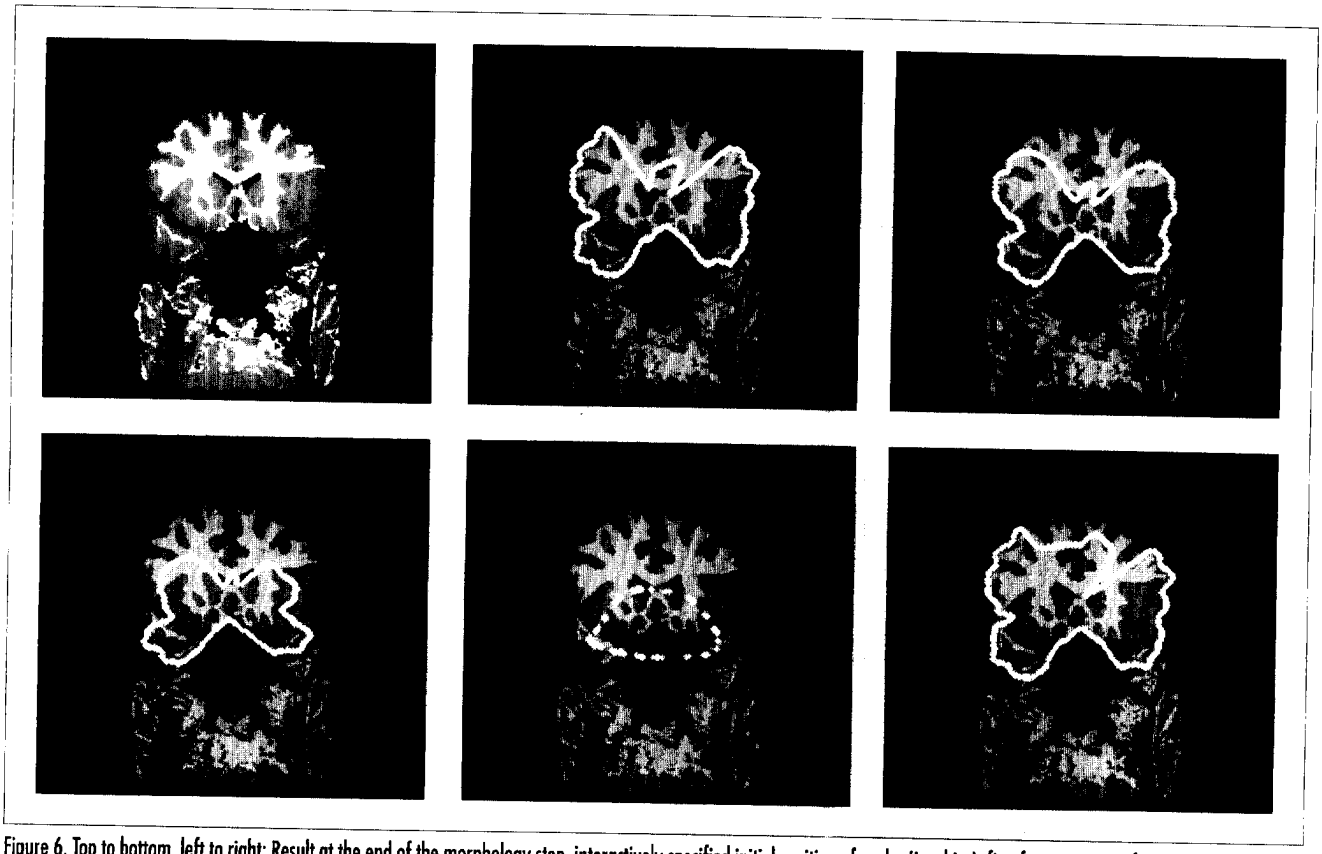

Figure 6. Top to bottom, left to right: Result at the end of the morphology step, interactively specified initial position of snake (in white), first few iterations of a customized balloon model. (Courtesy of Tina Kapur.)

The internal energy term imposes a regularization constraint on the contour as follows

$$E_{internal} = \int_s \left(w_1(s)\|v'(s)\|^2 + w_2(s)\|v''(s)\|^2 \right) ds, \quad (2)$$

where s is arclength, derivatives are with respect to s, and $v(s)$ stands for the ordered pair $(x(s),y(s))$, which denotes a point along the contour. The choice of w_1 and w_2 reflects the penalty associated with first and second derivatives along the contour respectively.

The external energy term in Equation 1 is responsible for attracting the snake to interesting features in the image. The exact expression for $E_{external}$ depends on the characteristics of the features of interest.

Finding a local minima for E_{snake} from Equation 1 corresponds to solving the following Euler-Lagrange equation for v:

$$-(w_1 v')' + (w_2 v'')'' + , E_{external}(v) = 0, \quad (3)$$

In this equation we need to use boundary conditions specifying whether the snake is a closed contour or the derivatives are discontinuous at the end points. This equation is then written in matrix form as $Av = F$, where $F(v) = -, E_{external}$. Here A is a pentadiagonal banded matrix, v is the position vector of the snake, and F is gradient of the external energy of the snake, or the external force acting on it. We solve the evolution equation $\frac{dv}{dt} - Av = F$ to obtain the v that is closest to the initial position. As $\frac{dv}{dt}$ tends to zero, we get a solution to the system $Av = F$.

Formulating this evolution problem using finite differences with time step τ, we obtain a system of the form:[6]

$$(I + \tau A)v^t = v^{t-1} + \tau F(v^{t-1}), \quad (4)$$

where v^t denotes the position vector of the snake at time t, and I is the identity matrix. The system reaches equilibrium when the difference between v^t and v^{t-1} is below some threshold.

The balloon model for deformable contours introduces improvements on the snake model.[6] It modifies the snake energy to include a "balloon" force, which can either be an inflation force, or a deflation force. The external force F is changed to

$$F = k_1 n(s) + k \frac{\nabla E_{external}}{\|\nabla E_{external}\|}, \quad (5)$$

where $n(s)$ is a unit vector normal to the contour at point $v(s)$, and $|k_1|$ is the amplitude of this normal force. See Figures 4–7 for examples of this segmentation.

Extracting depth information. The next stage in our process is to extract information about the position of the patient in the operating room. This will allow us to obtain an accurate reconstruction of points from the skin surface of the patient. Researchers have used two standard IU methods to get this data. In one method we apply stereo matching techniques to a pair of views of the patient. In the second, we use a laser striping device.

Stereo vision is a well-developed IU technique that has been used in a range of applications, especially the automated construction of terrain from aerial photography and the navigation of autonomous vehicles. Stereo, in brief, consists of taking two views of a scene from a pair of cameras, then determining the correspondence between these views. This means for each pixel in one image, the viewer finds the pixel (if any) in the other image that represents the projection of the same scene point in the first image. Finally, drawing on the relative orientation of the two cameras, the difference in projection between corresponding pairs of points

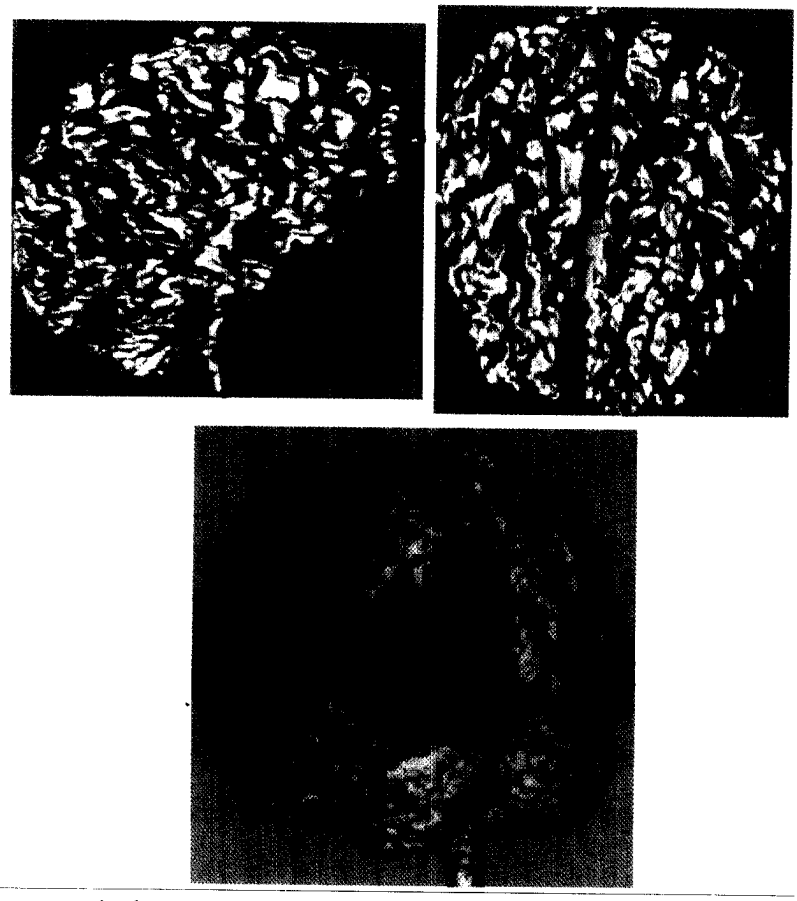

Figure 7. Examples of segmentation. By combining the E/M stochastic segmenter, followed by morphological extraction and deformable models, the system can automatically extract intricate structures, such as the surface of the brain, from the volumetric data. Top case is normal brain. Bottom case highlights a tumor.

functions to determine the actual distance of the scene point from the cameras.

Researchers have developed and successfully applied a wide range of stereo algorithms to problems such as accurate terrain reconstruction. In the domain of registration in medical imagery, Figure 8 shows the reconstructed depth map of a patient's face from a pair of stereo views[7] that is then registered to an MR model.[8]

A laser striper is an alternative to stereo for extracting depth information. This system projects a tightly collimated beam of laser light through a cylindrical lens and onto an object. By controlling the orientation of this plane of light, and by recording the position of the reflected laser light in a carefully calibrated camera, one can again deduce the 3D position of points in a scene, relative to the laser system.

In either case, the result of this stage is a set of 3D data points from the skin surface of the patient, measured in some coordinate system relative to the sensor.

Registration. This stage of our process requires that we match or register our depth data from the live view of the patient to the skin surface of the previously segmented MRI scan. To solve this problem, we can use a series of IU algorithms originally designed for problems such as target recognition problems.

Matching data sets. When finding the transformation between the two data sets, we first separate the depth data of the patient's head from background data. This is a traditional IU problem known as the *figure/ground problem*, and requires separating data corresponding to one object from data corresponding to all other objects. In this case, relatively straightforward techniques will suffice to separate out the patient data from background data.

To get an initial alignment of the two data sets, we execute the following technique from the IU field. As a preprocessing step, we hash all pairs of MRI points based on the distance between them. Furthermore, at each MRI skin point, we estimate the surface normal to the skin by a local fit of the neighboring data. To find an initial transformation, we select any two depth data points; we ensure stability by selecting two widely separated points. Using the distance between these two depth points, we access the hash table to find possible matching MRI points. For each such pair in the hash table, we consider the hypothesis that the depth points match the MRI points (there are two such matches).

This then determines 5 of the 6 degrees of freedom associated with the coordinate frame transformation. The missing parameter is the rotation about the axis connecting the two points. To solve for this parameter, we can estimate the normal to the skin surface at the depth point by fitting a plane to the neighboring data. We then need the rotation about the axis between the points that will rotate the depth normal to align with the MRI normal. Such a rotation may not exist, in which case we can discard this pair. Similarly, after solving for the rotation, we can confirm that the application of this rotation to the normal at the other point also causes it to agree with its matching normal. If not, the pair is discarded.

By cycling over all the possible pairings of MRI points to depth points, as defined by the entries in the hash table, we can collect the set of feasible initial transformations. We rank these transformations on the basis of the root mean square (RMS) fit of the transformed depth data to the MRI data. We can further process the resulting rank-ordered list of hypotheses using the methods I describe in following sections, stopping when we find a sufficiently accurate fit.

We can also use a related approach, interpretation tree search, to match triples of visible sampled model points to the three selected depth points. This method basically searches over all possible ways of matching three depth points that are intentionally spread out to three points selected from the sample MRI model. For each pairing of model and data points triples, the method tests whether the pairwise distances between model points and depth points are roughly the same. If all such tests are valid, the match is kept. We then compute the coordinate frame transformation that maps the three depth points into their corresponding model points. These transformations form a set of hypotheses. Note that due to the sampling of the model data, the actual object points corresponding to the selected depth points may

not exist. Therefore these hypothesized transformations are at best approximations to the actual transformation. For efficiency, we hash pairs of points on distance, and only retrieve likely candidates for testing.

In either case, these IU methods give us a set of hypotheses as to the transformation needed to align the MRI model with the actual patient.

We can use another IU registration algorithm, called the Alignment Method, to filter these hypotheses. For each hypothesis, we transform all the depth points by the hypothesized transformation. We then verify that the fraction of the transformed depth points that do not have a corresponding model point within some predefined distance are less than some predefined bound. We discard those hypotheses that do not satisfy this verification.

For each verified hypothesis, we refine the alignment of the two data sets by minimizing an evaluation function that measures the amount of mismatch between the two data sets.

For all transformed depth points, the first stage sums a term that is itself a sum of the distances from the transformed point to all nearby model points, where the distance is weighted by a Gaussian distribution. This Gaussian weighted distribution is a method for roughly interpolating between the sampled model points to estimate the nearest point on the underlying surface to the transformed depth point. More precisely, if l_i is a vector representing a depth point, m_j is a vector representing a model point. Furthermore, if T is a coordinate frame transformation, then the evaluation function for a particular pose (or transformation) is

$$E_1(T) = -\sum_i \sum_j e^{-\frac{|Tl_i - m_j|^2}{2\sigma^2}} \qquad (6)$$

This objective function is similar to a posterior marginal pose estimation (PMPE) method, and to the use of elastic net constraints. One can visualize this objective function as if we placed a Gaussian distribution of some spread σ at each model point, then summed the contributions from each such distribution at each point in the volume. Then the contribution of each transformed depth point towards the evaluation function is simply the summed value at that point. Because of its formulation, the objective function is generally quite smooth, and thus facilitates "pulling in" solutions from moderately removed locations in parameter space. Moreover, it bears some similarity to the radial basis approximation schemes used for learning and recognition in other parts of computer vision.

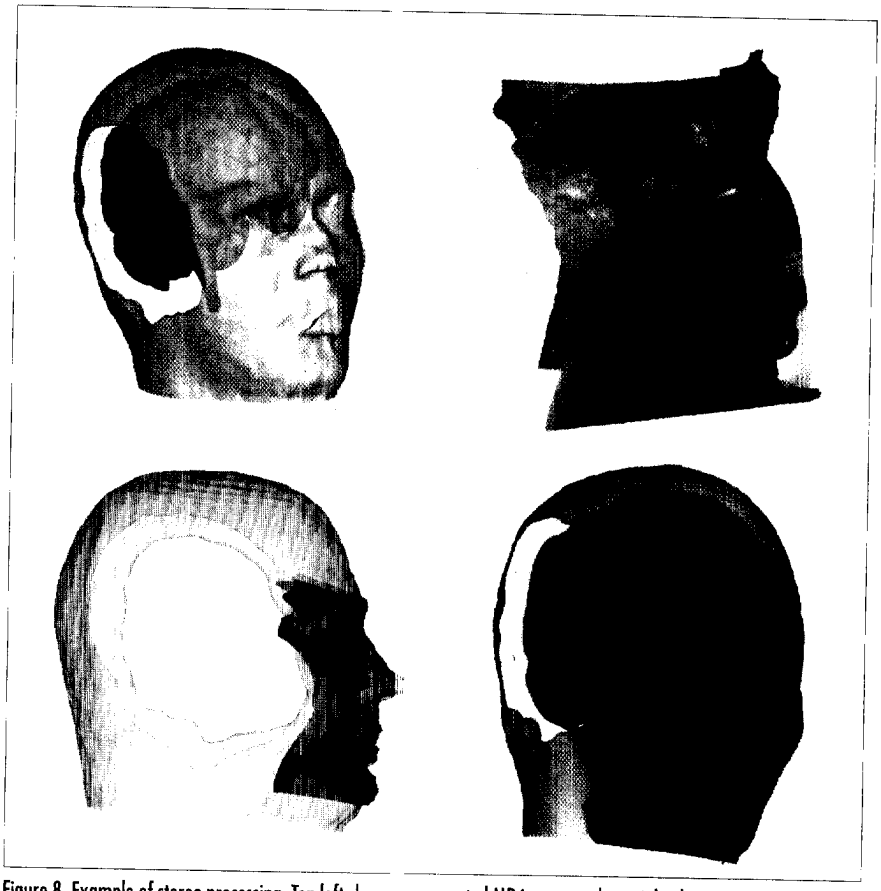

Figure 8. Example of stereo processing. Top left shows a segmented MR image, and top right shows a stereo reconstruction of the patient. The registration problem is to match these two data sets. Bottom left shows the result of such a registration, with the registered stereo surface overlaid on the MRI model. Bottom right shows a fusion of the underlying image texture onto the reconstructed depth, together with the segmented MRI. (Figure courtesy of Nicholas Ayache.)

To minimize this evaluation function, we use standard numerical methods. For example, we can use the Davidon-Fletcher-Powell quasi-Newton method to find the best transformation. The DFP method iteratively builds up a good approximation to the inverse Hessian matrix. We then apply this approximation to the data to solve for the parameters yielding a minimum of evaluation function. Solving this minimization problem yields an estimate for the pose of the depth points in model coordinates.

We execute this refinement and evaluation process using a multiresolution set of Gaussians. We initially use a broad-based Gaussian to allow influence over large areas. This results in a coarse initial alignment, but one which can be reached from a wide range of starting positions. Subsequently, we can use more narrowly tuned Gaussian distributions to refine the pose, while focusing on only nearby model points to derive the pose.

Using the resulting pose of this refinement, we repeat the pose evaluation process, now using a rectified least-squares distance measure. In particular, we evaluate each pose by measuring the distance from each transformed depth point to the nearest model point, (with a cutoff at some predefined maximum distance). We evaluate the pose by summing the squared distances of each point. We minimize using the DFP method to find the least-squares pose solution. Here the evaluation function is

$$E_2(T) = \left[\frac{1}{n} \sum \min \left\{ d_{\max}^2, \min_j |Tl_i - m_j|^2 \right\} \right]^{\frac{1}{2}} \qquad (7)$$

where $d_{\max}$ is some preset maximum distance, and where n is the number of depth

Figure 9. Example of tracking a registered instrument. The tip has been placed inside a plastic skull. Here we see three cross-sectional views of the position of the tip relative to a full CT model of the skull, as well as a 3D graphical illustration of the position of the instrument tip relative to the full model.

points l_i. This objective function is essentially the same as a maximum a posteriori model-matching scheme. It acts much like a robust chamfer-matching scheme, or an iterative closest point-matching scheme. We expect this second objective function to be more accurate locally, since it is composed of saturated quadratic forms. However, this function also tends to get stuck in local minima. Hence, we add one more stage.

While the preceding method always gets very close to the best solution, it can get trapped into local minima in the minimization of E_2. To improve upon this, we take the pose returned by the preceding step, perturb it randomly, and then repeat the minimization. We continue to do this, keeping the new pose if its associated RMS error is better than our current best. We terminate this process when the number of trials that have passed since last improving the RMS value becomes larger than some threshold.

The best found solution is a pose and a measure of the residual deviation of the fit to the model surface. Once we find the final solution, we can measure the residual error at each point and remove the depth points with large residual errors from the data set. This automatically deletes possible outliers. We can then rerun the final stages of the registration process using the remaining data, to obtain a tighter fit to the surface.

We collect such solutions for each verified hypothesis, over all legal-view samples, and rank-order them by smallest RMS measure. The result is a highly accurate transformation of the MRI data into the depth sensor's coordinate frame.

Camera calibration. Once we have a registration, we need to relate it to a view of the patient. A video camera can approximate the surgeon's viewpoint, as though it were looking over her shoulder. By calibrating the position and orientation of this camera relative to the depth coordinate system, we can render the aligned MRI or CT data relative to the camera's view. Mixing this rendering with the live video signal gives the surgeon an enhanced reality view of the patient's anatomy. The surgeon may use this method to plan a craniotomy or a biopsy, or to define the margins of an exposed tumor for minimal excision.

Camera calibration, especially solving for the camera's orientation and position relative to a set of known world points, is a well known IU problem. Several techniques will suffice to handle this part of the problem. A straightforward method is to solve for the camera model by minimizing the error between a set of known fiducial marks' image coordinates and the predicted positions of those known fiducials under the camera model's current estimate.

Visualization. By combining the camera calibration and the registration of the data sets, we can visualize the data. In particular, applying the data set transformations brings the MRI or CT model into alignment with the patient in the coordinate frame of the depth system. We can then project that model into the video camera's coordinate frame by applying the computed camera model. This gives us a virtual view of the MRI model, as seen by that camera. We then mix the virtual view with an actual video view of the camera, thus allowing the surgeon to use it as a visualization tool.

Tracking. The foregoing method registers an MRI or CT data set against a static view of a patient, providing a visualization of the patient's internal anatomy. In many surgeries, a static view serves the surgeon's needs. More generally, however, surgery calls for movement between the viewpoint and the patient, either because the patient moves, or because the viewpoint of the surgeon changes. The latter situation is critical for visualization displays using goggles or other display devices that are worn by the surgeon. These devices change viewpoint as the surgeon does.

Thus we need a method for tracking the patient and visualization system position changes. Fortunately, there are a wide range of IU methods specifically devoted to measuring motion. Some methods track known fiducial marks, while others use imagery variation to measure optical flow and, therefore, changes in viewpoint. In our system, we place a small number of fiducial markers on the frame supporting the patient and then track changes in the position of those fiducials relative to the visualization camera. By tracking these changes, we can rapidly update the camera model and reregister the visualization of the MRI model to the current view.

Once we have registered a segmented model to a patient, we can track other objects relative to that model. For example, suppose that we consider a rigid surgical instrument,

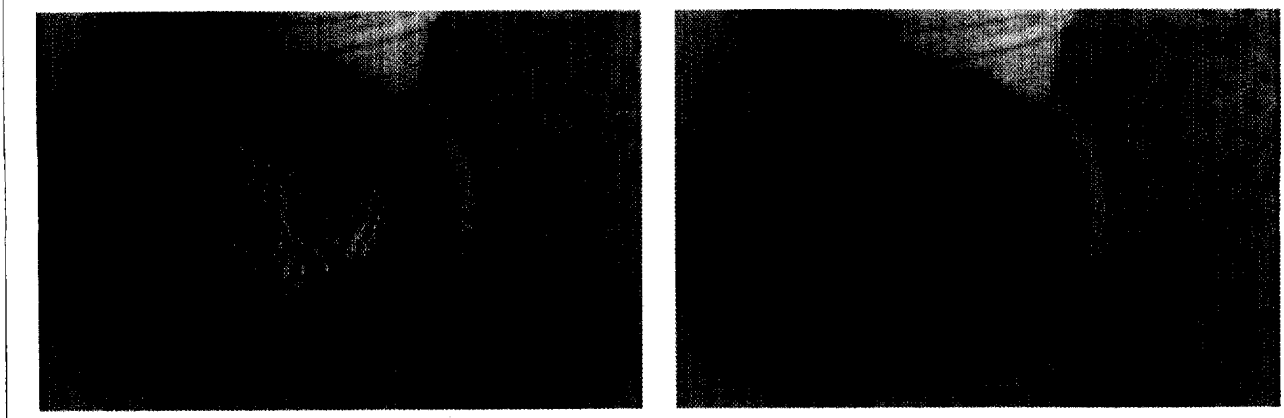

Figure 10. Examples of combining the registration of MRI to depth data and the calibration of a video camera relative to the depth to provide an enhanced reality visualization of patients. In the left case, the tumor and brain are displayed, in the right case, the tumor and the ventricles are displayed in registration with the patient.

one end of which we mark with a set of visible marks, such as infrared light-emitting diodes. There are standard systems that visually track markers and determine the instrument's position and orientation relative to some coordinate frame. If we register the instrument to the 3D scanner used in the operating room, then we can track the position of that instrument, both relative to the patient, and relative to the full 3D scanned model of the patient. This allows the surgeon to see imagery of the position of instrument's tip, even if it is buried inside the patient. Using this visualization, the surgeon can track the instrument's tip relative to the desired target areas deep inside the body. See Figure 9 for an example of tracking a registered instrument.

Application of the visualization method

We have run a series of trials of this registration and visualization system with actual neurosurgery patients. Figure 10 displays a registration of the depth data against a patient's MRI model. We have highlighted the tumor and ventricles.

We have been using this registration and visualization method to transfer presurgical plans to the patient. In our current approach, we use our registration method to provide a visual overlay of the view of the patient with internal structures that the surgeon has selected. By viewing this overlay on a live video monitor, the surgeon can trace with a marker the outline of key structures on the patient's scalp. This enables the surgeon to mark locations for planned surgical steps, prior to placing the patient in the operating room. To date, we have used this procedure on a small number of neurosurgical patients at Brigham and Women's Hospital in Boston. Perturbation studies of the method show that it has a repeatability on the order of 200–300 μm.

During surgery, with operator guidance, we perform a second registration that provides the surgeon with guidance and feedback. The surgeon uses the full 3D segmented model for locating targets, as well as for guidance to critical structure proximity.

Other IU applications in medicine

Visualization in image guided surgery is only one way in which IU tools can solve medical applications. IU techniques are providing important leverage in other medical areas as well. In the following sections I briefly describe some of these applications.

Deformable models. I have already mentioned the use of deformable models, such as snakes, for structure extraction as part of our surgical visualization system. Many other medical applications also use deformable models. These models are especially useful for capturing shape representations of structures that are themselves flexible, such as the lungs or heart. For example, by fitting a deformable model to a motion sequence of a beating heart, one can capture a physical model of the deformations of the heart over its full cycle. Such models facilitate pathology diagnosis, as well as the impact of injuries to the heart wall.

Change detection. Registration methods also have application in clinical settings. For example, we took two MRI scans of the same patient, each several months apart. These scans are part of an ongoing National Institute of Health study of multiple sclerosis (MS) at Brigham and Women's Hospital aimed at determining the optimal frequency for performing MR imaging of MS patients. Under this study researchers image patients with varying disease stages at different frequencies to identify changes in MS lesion activity. To support this analysis, it is necessary to register the MRI scans from different points in time and compare them to detect relevant changes. We have applied our technique to this task, using the surface of the intracranial cavity in different MRI scans as the basis for the registration.

Given this alignment, we can transform the second data set into the coordinate frame of the first data set and then resection the data to obtain 2D slices equivalent to those of the first data set. With these new sections, we can then compare individual slices of the first data set to the resectioned second data set, and do image differencing to find noticeable changes. An example of this, highlighting the growth of a lesion in the patient, is shown in Figure 11 (next page).

This registration makes it easy to measure changes in structures, especially when combined with automatic segmentation techniques. Researchers could use these tools to track the progression of a disease, or to track the effectiveness of a therapy in controlling a tumor or disease.

Feature detection. IU techniques also aid disease diagnosis. For example, Cerneaz and Brady use edge detection and contour ex-

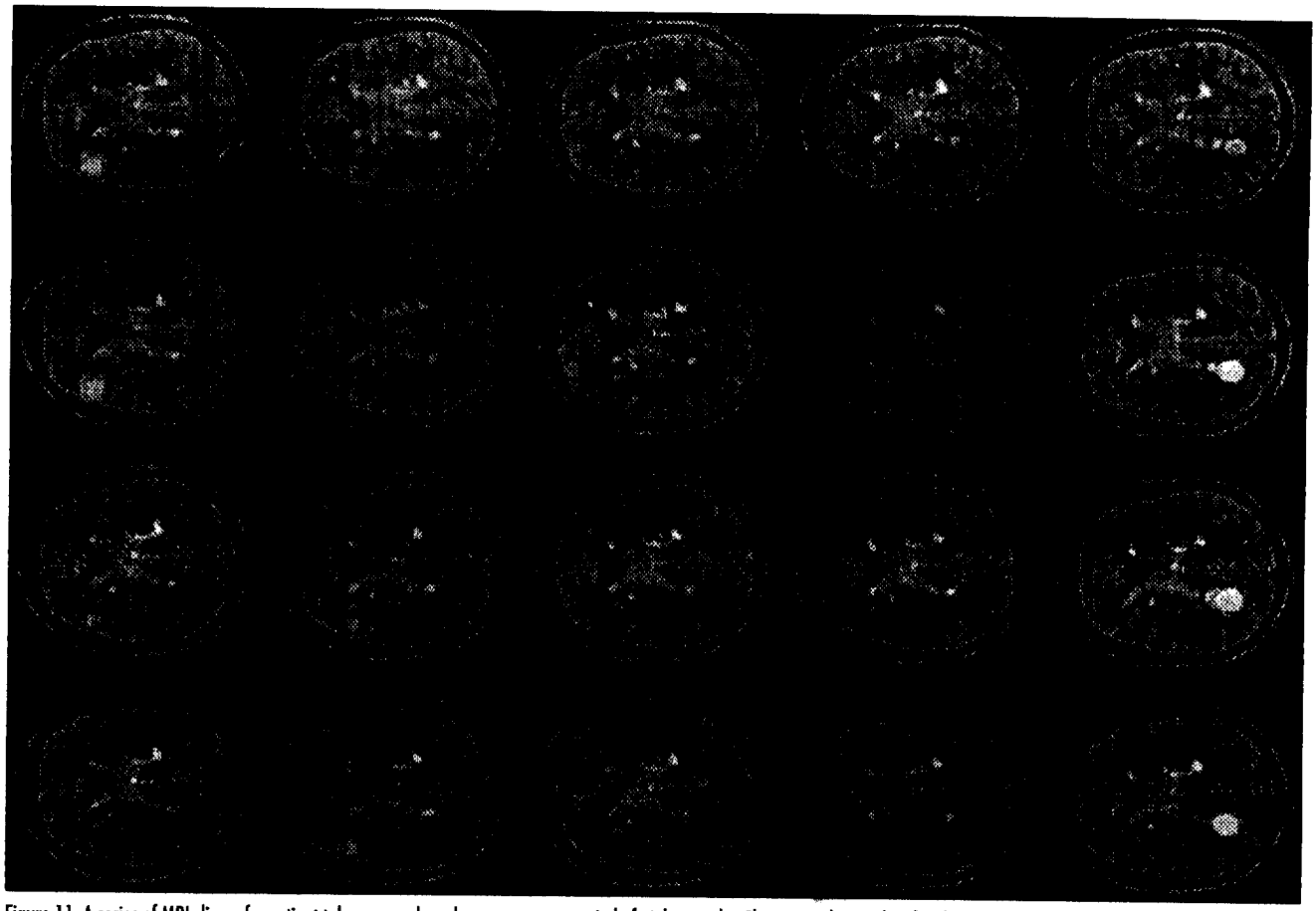

Figure 11. A series of MRI slices of a patient taken several weeks apart, over a period of eight months. Shown are the results of reslicing and normalizing each subsequent MRI scan, relative to the first scan, so that, in principle, the same slice of the anatomy is shown. The reader can easily see the lesion changes, in the lower left and lower right, over time. The difference image (right) shows positive (green) and negative (red) change (center top) indicates the apparent growth of a lesion.

traction tools developed for traditional IU work as a basis for constructing diagnostic screening methods for mammograms.[9]

TRADITIONAL IU TECHNIQUES, originally developed to address problems such as photo-interpretation, industrial automation, and autonomous-vehicle navigation, have proven effective in several medical applications. Although the underlying bases for medical imaging problems are very different, researchers have successfully retailored IU techniques to such medical applications as image guided surgery, therapy evaluation, and diagnostic screening.

References

1. R.H. Taylor et al., eds., *Computer-Integrated Surgery*, MIT Press, Cambridge, Mass., 1995.
2. W.E.L. Grimson et al., "Evaluating and Validating an Automated Registration System for Enhanced Reality Surgery," *Proc. Conf. Computer Vision, Virtual Reality, and Robotics for Medicine*, Springer-Verlag, New York, 1995, pp. 3–12.
3. W.M. Wells III, et al., " Adaptive Segmentation of MRI Data," *Proc. Conf. Computer Vision, Virtual Reality, and Robotics for Medicine*, Springer-Verlag, New York, 1995, pp. 59–69.
4. T. Kapur, W.E.L. Grimson, and R. Kikinis, "Segmentation of Brain Tissue from MR Images," *Proc. Conf. Computer Vision, Virtual Reality, and Robotics for Medicine*, Springer-Verlag, New York, 1995, pp. 429–433.
5. A. Witkin, M. Kass, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision*, Vol. 1, No. 4, June 1988, pp. 321–331.
6. L. Cohen, "On Active Contour Models and Balloons," *Computer Vision, Graphics and Image Processing: Image Understanding*, Vol. 53, No. 2, Mar. 1991, pp. 211–218.
7. F. Devernay and O.D. Faugeras, "Computing Differential Properties of 3D Shapes from Stereoscopic Images Without 3D Models," *Proc. Conf. Computer Vision and Pattern Recognition*, CS Press, Los Alamitos, 1994, pp. 208-226.
8. F. Betting et al., "A New Framework for Fusing Stereo Images with Volumetric Medical Images," *Proc. Conf. Computer Vision, Virtual Reality, and Robotics for Medicine*, Springer-Verlag, New York, 1995, pp. 30–39.
9. N. Cerneaz and M. Brady, "Finding Curvilinear Structures in Mammograms," *Proc. Conf. Computer Vision, Virtual Reality, and Robotics for Medicine*, Springer-Verlag, New York, 1995, pp. 372–382.

W.E.L. Grimson is a professor of computer science and engineering at the Massachusetts Institute of Technology. He received a BSc in mathematics and physics from the University of Regina, Saskatchewan, in 1975, and a PhD in mathematics from MIT, Cambridge, MA, in 1980. His research interests include machine vision, human vision, robotics, artificial intelligence, and finite mathematics. He has been an associate editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence*, a General Chair for the 1995 International Conference on Computer Vision, and is a member of the IEEE and AAAI. He can be reached at The Massachusetts Institute of Technology's Artificial Intelligence Lab, Cambridge, MA 02139; welg@ai.mit.edu.