Manuscript submitted to *International Journal of Neural Systems*, Version 2, March 5, 2002.

A Semi-Localized Elastic Net for Surface Reconstruction of Objects from Multislice Images

Robert I. Damper*, Stuart J. Gilson† and Ian Middleton‡

*Image, Speech and Intelligent Systems (ISIS) Research Group,
Department of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ, UK.
Email: rid@ecs.soton.ac.uk

† Laboratory of Physiology,
Parks Road,
University of Oxford,
Oxford OX1 3PT, UK.
Email: stuart.gilson@physiol.ox.ac.uk

‡Microsoft Corporation,
One Microsoft Way,
Redmond,
WA 98052
USA.
Email: ianmid@microsoft.com

Address for correspondence:

Dr. R. I. Damper,
Image, Speech and Intelligent Systems (ISIS) Research Group,
Department of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ,
UK.

Tel: +44 1703 594577
FAX: +44 1703 594498

Email: `rid@ecs.soton.ac.uk`

Abstract

The traveling salesman problem (TSP) is an prototypical problem of combinatorial optimization and, as such, it has received considerable attention from neural-network researchers seeking quick, heuristic solutions. An early stage in many computer vision tasks is the extraction of object shape from an image consisting of noisy candidate edge points. Since the desired shape will often be a closed contour, this problem can be viewed as a version of the TSP in which we wish to link only a subset of the points/cities (i.e., the 'noise-free' ones). None of the extant neural techniques for solving the TSP can deal directly with this case. In this paper, we present a simple but effective modification to the (analog) elastic net of Durbin and Willshaw which shifts emphasis from global to local behavior during convergence, so allowing the net to ignore some image points. Unlike the original elastic net, this *semi-localized* version is shown to tolerate considerable amounts of noise. As an example practical application, we describe the extraction of 'pseudo-3D' human lung outlines from multiple preprocessed magnetic resonance images of the torso. An effectiveness measure (ideally zero) quantifies the difference between the extracted shape and some idealized shape exemplar. Our method produces average effectiveness scores of 0.06 for lung shapes extracted from initial semi-automatic segmentations which define the noise-free case. This deteriorates to 0.1 when extraction is from a noisy edge-point image obtained fully-automatically using a feedforward neural network.

1 Introduction

The traveling-salesman problem (TSP) is a member of the large and important class of NP-complete problems [1, 2]. The TSP owes its importance to the facts that (1) it is easy to state and (2) any NP-complete problem can be transformed into any other. Hence, the TSP can act as an easily-understood prototypical form of the whole class. To state the TSP: Given N cities in (usually two-dimensional) space, with distance d_{ij} between them, find the minimum-length tour that visits each exactly once and returns to its start point.

Since there are no known exact solutions for NP-complete problems other than selecting the optimum from an exhaustive enumeration of all possible solutions, and the computational costs of such enumeration are subject to combinatorial explosion with problem size, heuristic solutions are mandatory in most practical cases. Neural solutions have proved very popular, dating back to the seminal work of Hopfield and Tank [3], who imposed soft constraints on a recurrent network (fully-connected except for the absence of self-loops). These constraints were derived from a cost function formally equivalent to the network's Liapunov ('energy') function, and including tour length as a term. In principle, the activation of the (i, j)th neuron is either 1 or 0, according to whether city j is visited at step i of the tour, or not. As the energy is guaranteed to reduce with each neural update until equilibrium is achieved, so the cost function is minimized. However, their original formulation was unduly sensitive to the choice of parameter values for the cost function and to problem size [4], thereby rendering Hopfield-Tank optimization impractical in general. It was also discovered that: "Using binary threshold units does *not* work well ... So we must use either stochastic units with simulated annealing, or continuous-valued units" [5, p. 78]. Although many authors have suggested ways of improving the robustness of Hopfield-Tank neural optimization (e.g., [6, 7, 8]), it may be better to enforce the constraints differently. One such attempt to do this was the elastic net of Durbin and Willshaw [9].

Durbin and Willshaw describe their approach as "more geometrical" in that: "The algorithm is a procedure for the successive recalculation of the positions of a number of points in the plane in which the cities lie" (p. 690). However, as pointed out by Simic [10], it shares the same basis in statistical mechanics as Hopfield-Tank optimization (an insight which allows Simic to correct a small error in the original elastic net formulation). It can be viewed as 'neural' in its use of units corresponding to points on the tour whose coordinates in the plane act like a pair of 'activations'. Hence, updating neurons during iteration corresponds to movement of points—ideally in the direction of defining a low-cost tour. Each unit is subjected to two competing forces: one pulling it towards cities and the other pulling it towards its neighbors. By its use of a topological, neighborhood-preserving mapping between the spaces of points and cities, it is apparent that there is a close relation to Kohonen nets [11, 12], as discussed by Hertz, Krogh and Palmer [5, pp. 244–6]. So it is no surprise that Kohonen nets have also been used to find heuristic solutions to the TSP (e.g., [13, 14, 15]). Further, by the use of two competing forces, there is also seen to be a close relation to snakes [16, 17] with their 'external' and 'internal' energies. (See Abrantes and Marques [18] for additional remarks on relations between snakes, elastic nets, and Kohonen maps).

In this paper, we consider the problem of extracting the shape of a closed object from a digital image. This is a difficult computer vision task [19] for two main reasons. First, there is an infinite number of viewpoints, each generating a possibly unique image. Second, signal corruption due to the imaging process results in noise which is usually difficult to characterize and cannot be ignored. Any generic shape-extraction method must cope with shifts and rotations of the object with respect to the viewpoint. Alternatively, a specialized method could be devised to extract shapes in constrained situations, such as identifying positions of components on a machine production line. A further problem is trying to extract three-dimensional (3D) shape, either from 3D range data, from multiple viewpoints, or from transverse slices through the object, which introduces additional complications in the form of the correspondence problem, the tiling problem, and the branching problem [20]. Noise corruption can be compensated for either through image processing (such as blurring and smoothing) or by using shape-extraction techniques which treat their input data as inherently uncertain (e.g., Bayesian methods, fuzzy systems and neural networks).

Here, we adopt a neural-optimization approach based on a view of edge-linking as solving a 'noisy' version of the TSP. That is, we wish to link only a subset of the points—those that are true edge points—omitting those arising from noise in imaging or subsequent processing. We use a modification to the Durbin-Willshaw elastic net for this purpose, which alters the global behavior of the net to make it *semi-localized*. That is, the update rule allows global extraction to be traded for more local behavior during convergence. The choice of the elastic net was based on earlier empirical study [21] which showed it (and the Kohonen net also) to be a competent edge linker when applied to a range of synthetic images. The advantage over the Kohonen net is that an explicit cost function is associated with the approach [5, p. 244]. The elastic net also offers the ability to generalize in sparse areas of the data space, and is data-driven in the sense that there is no requirement for any explicit model of the object. The latter is beneficial for generic shape extraction. Given 2D slices through the imaged object, the technique can be applied to many slices simultaneously to extract 'pseudo-3D' shape information. By 'pseudo', we mean that the technique operates on all slices at once but points are constrained to remain in the plane of their respective slice. The emphasis in this paper is on identifying and extracting the shape, leaving any desired recognition of the object to a later stage.

As a demonstration, we consider the extraction of lung outlines from $T1$-weighted magnetic resonance (MR) images [22] of the human thorax. This is an important and difficult problem, since medical images are typically hard to process: The imaged object is highly complex, access to internal body structures is poor and image quality is often compromised in favor of reduced capture times. It is also a real-world application, and successful automatic extraction could eliminate the time and cost currently required for an expert to process the MR scans manually. The main novelty is the semi-localization, which bridges the gap between the geometrically-flexible, global nature of the elastic net and the noise-tolerant, local behavior of techniques such as snakes.

The rest of this paper is organized as follows. Section 2 reviews previous work on shape extraction. Section 3 introduces the semi-localized elastic net technique for 2D shape extraction while Section 4 describes the extensions for pseudo-3D extraction. Section 5 details the example MR imaging application, which we have chosen to illustrate the method, and Section 6 concludes.

2 Background to Shape Extraction

As stated above, the approach adopted here is arbitrary (data-driven) shape extraction, which simply looks for indications of the presence of an unspecified shape in the image (i.e., there is minimal top-down model fitting). Indications of the object's presence may be colinearly aligned clusters of edge pixels, or spatially adjacent short line segments with similar orientations. By assuming only a single object, data-driven extraction techniques can group such pixels and line segments into an appropriate contour, while ignoring pixels which result from the ground or the imaging process (noise). Further, because no knowledge of the geometry of the shape is used, data-driven methods can extract any shape from an image, irrespective of its position, scale or orientation. In practice, however, it is often convenient to provide some minimal *a priori* knowledge by specifying whether the shape is closed, smooth, has concave corners, etc. Under these conditions, shapes can be extracted from images even in the presence of noise. However, excessive noise (especially if not evenly distributed) or occlusion of the object can mislead these techniques. Because they are only working with pixel information, they cannot distinguish whether a given pixel is part of the target/desired shape or not.

Figure 1 shows the processing stages required for shape extraction from images. We assume that the shape will be the dominant feature in some binary edge image, such that pixel density will be greatest along the outline. A deformable contour is used, attracted toward these dense pixel regions and forming a closed path which should then describe the object's outline. To produce the required edge image from an original gray-level image, more extensive preprocessing than simple thresholding is necessary—for example, classic edge-detection algorithms [23], retinoptic-like operations [24] or specialized techniques designed for some particular application (e.g., [25]).

[Figure 1 about here.]

When several transverse scans (slices) are available, the implicit 3D information can be used to extract 3D shape or assist in the recovery of 2D shape. This situation is typical of many medical applications, where multislice methods such as magnetic resonance imaging (MRI), computer tomography and ultrasound imaging make 3D shape reconstruction possible. Such reconstruction would be a useful aid to diagnosis and surgery planning. Indeed, some of the most advanced techniques for shape extraction are currently in the medical domain [26].

Existing methods for 3D shape reconstruction include use of the Hough transform to extract arteries from (non-invasive) ultrasound images [27]. The method locates the artery wall approximately from duplex ultrasound (B-)scans of the region: The actual boundary wall is then characterized as deviations from the best-fit circle. Extraction is performed on each individual slice, allowing a subsequent voxel-based visualization of the artery. However, because the extraction from any one slice is independent of all other slices, no real notion of 'volume' is utilized. A better approach would use the 3D shape information implicit in the B-scan sequence, by applying the Hough transform to all the images simultaneously. Adding another dimension to the voting space allows the image sequence to be considered as a whole, increasing noise tolerance through the consideration of volume [28]. Even so, the technique will only work effectively on image sequences where the cross-sectional shape is consistent and similar—large changes in geometry will defeat the basic Hough transform (and the generalized transform too). Clearly, the problem is inherently three-dimensional and it is only sensible to include any 3D information that is available. The two approaches one could take to this are surface and volume reconstruction.

True surface reconstruction methods, however, introduce three new problems [20]:

- *the correspondence problem*—finding the associations between vertices in one layer with vertices in adjacent layer(s).
- *the tiling problem*—once the correspondences have been assigned, the implicit regions must be tiled to form a surface. The tiling should be complete (no gaps), and the tiles should not intersect.
- *the branching problem*—if the object contains splits or merges, a contour in one slice can correspond to two contours in an adjacent slice. The resulting split in the reconstructed surface can be dealt with in several ways, depending upon the application.

Most surface reconstruction techniques treat one or two of these problems only, yet tackling all three simultaneously can produce impressive results [20].

Volume reconstruction from planar data is described in detail by Boissonnat [29], where 2D Delaunay triangulations of each slice are used as the basis of tetrahedralization. This results in a triangular-faceted solid whose exterior represents the surface of the object between the two slices. The method then concatenates the biplanar solids to produce the full volumetric model. It can tolerate multiple objects in the scene and cases where the number of contours varies from one cross-section to the next. Additionally, the external Voronoi skeleton from the adjacent slice can be incorporated in the Delaunay triangulation of the current slice, permitting more complex branches to be handled [30]. However, formation of the noiseless complete contours was performed in a separate stage (usually manually), indicating that this approach is aimed solely at reconstruction.

There are alternative surface/volume reconstruction methods. These include incremental volume growing [31], deformable superquadrics [32] and elastic meshes [33]. The latter yield an explicit parameterization of the fitted model, making them suitable for recognition.

3 Semi-Localized Deformable Models

Since shape extraction is an inherently geometric problem, it makes sense to exploit the topology of the situation. A deformable contour is a natural choice for extracting shape: Part of the extraction problem is already solved since the contour is itself a primitive model of the target object. Contrast this with an incremental pixel-based edge-linker, which starts at a seed pixel, and navigates along connected pixels deciding which to link next. Such a method has no idea of a region, closure, or smoothness, and will have problems if the outline is fragmented. Attention has also been paid to the neural-network paradigm, which offers important benefits in terms of noise tolerance and generalization abilities. Any technique which incorporates such abilities should be well-suited to the noisy problems found in real-world applications.

The strengths of these two paradigms can be exploited by combining them in a network featuring many simple processing units (the 'neural' component) each connected to neighboring units through elastic connections (the deformable-contour component). Each unit has a multidimensional activation vector. In the 2D case, this can be interpreted as a spatial position. The network ascribes only a very simple role to each unit, but when combined into larger cooperative groups, more complex behaviors can emerge. It is this emergent behavior which allows these techniques to solve geometric optimization problems such as edge linking and the TSP. Such approaches rest on analogies from several disciplines: notably geometry, physics and approximation theory [26]. Their geometry allows deformable contours to represent shape with ease, while the physics of the system governs the permissible deformations and the approximation theory provides the rules for mapping the contour to the data.

Active contours (snakes) [16, 17] are one commonly-used member of this class of techniques. The units move towards nearby image features by virtue of a simple ('external') attraction function. Further, a complex ('internal') regularization term tries to keep the contour smooth, allowing inherent noise tolerance and an implicit model of the known shape to be encoded. Differing attraction functions can propel the snake toward lines, line terminals, or edge points.

The snake's popularity in many image-processing tasks, including medical applications [23, 26, 34], can be attributed to two properties. First, its regularization term avoids sharp corners and favors smooth contours. (Many natural shapes, such as the outlines of body organs, are smooth and have few discontinuous second derivatives.) Second, the various attraction terms are of a local nature—each unit will only be influenced by data within a small spatial neighborhood. However, these distinctive properties can bring their own problems in certain cases. Applications where the shape contains concave corners are problematic because the regularization term discourages formation of tight corners unless sufficient *a priori* knowledge exists to override this by local modification. A consequence of the local property is that the snake has to be seeded near desired features, otherwise it may converge on an undesired feature or diverge altogether. This generally requires human intervention. By contrast, a global technique would not need feature-dependent initialization, and could automatically extract the intended shape irrespective of its position in the image (cf. template convolution [23]). Complications then arise because it is necessary to discriminate between those data which are important (signal), and those which are not (noise).

Such a global technique is typified by the elastic net of Durbin and Willshaw [9], yet it has a fundamental flaw in general image-processing applications. It was designed to find quick, near-optimal solutions to the traveling salesman problem, but the formulation of the TSP inherently assumes all cities are correctly identified and positioned without error. Hence, the elastic net solution has no tolerance to noise, so that it is inapplicable to problems where noise is present. Despite this important restriction, the elastic net still has many desirable qualities as an edge-linker—the technique is global and guarantees that all cities will be linked in a closed path [35]. It is deformable, iteratively distorting a circular path until it passes through every city, enforcing the heuristic that the shortest tour will not cross over itself. Finally, in our experience, the technique is not particularly sensitive to its initial parameters.

Hence, in an edge-linking application, the elastic net will attempt to link all edge points, including noise, so contaminating the final solution. This intolerance to noise is a by-product of the technique's global nature. A better technique would exhibit the same emergent behavior as the elastic net (a global technique), but have a degree of noise tolerance (a local method like the snake). Thus, we arrive at the notion of a *semi-localized* elastic net, to be described in detail shortly. First, however, we describe the basic elastic net.

3.1 Durbin and Willshaw's elastic net

The elastic net consists of a 1D matrix of M units ('neurons'), $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M]^T$, each defined by an ordered number pair formally equivalent to the unit's 'activation' in conventional neural networks. We interpret the activation as the unit's position within a (1.0×1.0) square, so that the 1D matrix corresponds to a closed path in the plane. The image contains a set of N edge points, $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N]^T$, each defined by a coordinate. The aim is to map each of the M units onto the N edge points. Durbin and

Willshaw recommend $M = 2.5N$ to account for the likely uneven distribution of edge points in the image.

During mapping, an appropriate ordering of the units should be maintained: Neighboring units should be spatially nearby in the image. To achieve this, each unit is subjected to two forces. The first allows each edge point to attract nearby units, while the second pulls it towards the mid-point of its neighboring units (elastic tension). The net operates by iteratively updating the activation of the jth unit according to:

$$\Delta \mathbf{u}_j = \alpha \sum_{i=1}^{N} w_{ij} \left(\mathbf{p}_i - \mathbf{u}_j\right) + K\beta \left(\mathbf{u}_{j+1} - 2\mathbf{u}_j + \mathbf{u}_{j-1}\right) \tag{1}$$

where α and β are factors governing the influence of the two components of the update rule, and K is a simulated-annealing term which is gradually reduced during iteration to some convergence criterion.

The attraction force exerted on a unit by an edge point is given by the summation of all weighted distances of that unit to all edge points, controlled by the step size α. The weights are calculated as:

$$w_{ij} = \frac{\phi\left(||\mathbf{p}_i - \mathbf{u}_j||, \sqrt{2}K\right)}{\sum_{k=1}^{M} \phi\left(||\mathbf{p}_i - \mathbf{u}_k||, \sqrt{2}K\right)} \tag{2}$$

where $\phi(d,\ \sigma) = \exp\left(\frac{-d^2}{\sigma^2}\right)$ is a Gaussian distribution function, with standard deviation σ.

[Figure 2 about here.]

The effect of the w_{ij} weighting is more complex than simply normalizing attraction forces. The denominator is based on the summation of distances from every unit to the $\mathbf{p}_i$th edge point, rather than the summation between the $\mathbf{u}_j$th unit and every edge point. The difference is illustrated in Figure 2. Although unit $\mathbf{u}_2$ is closest to edge point $\mathbf{p}_1$, it will get a stronger attraction from point $\mathbf{p}_2$. Thus, the $\mathbf{p}_i$th edge point maximally attracts its nearest unit so that even remote points will have a significant influence on the net (since all points, however remote, will always have a nearest unit). Therefore, the elastic net's global nature is a consequence of the dynamically-scaled Gaussians in equation (2), and removing the denominator will reveal the basic snake technique.

After every n iterations of equation (1), the value of K is reduced, and the units make a gradual transition from being attracted toward groups of edge points (exploration of the data space) to being attracted towards a few edge points (ultimately a single one). Note that for low K, every contour is guaranteed to be a valid TSP tour [35]. Summing the Euclidean distances, d_{ij}, between successive units yields the tour cost.

3.2 Noise-tolerance through semi-localization

It is the global nature of the elastic net method which makes it intolerant to noise. Removing this global behavior will introduce noise tolerance, but will also alter the net's emergent linking properties. This is undesirable, since the global nature guarantees valid tours [35, 36], which translates to automatic location of features when considering image processing. Consequently, removing the net's global aspect introduces the requirement for expert seeding in order to extract desired features. This dichotomy cannot be resolved, so there will always be a trade-off between noise tolerance and initialization sensitivity. To effect an appropriate compromise, we modify equation (2) by replacing the numerator with an α-cut Gaussian weighting term:

$$w_{ij} = \frac{\phi_{\alpha\text{cut}}\left(||\mathbf{p}_i - \mathbf{u}_j||, \sqrt{\frac{3}{4}}K\right)}{\sum_{k=1}^{M} \phi\left(||\mathbf{p}_i - \mathbf{u}_k||, \sqrt{2}K\right)}$$

where:

$$\phi_{\alpha\text{cut}}(d, \sigma) = \begin{cases} \phi\,(d, \sigma) & \text{if } \phi(d, \sigma) \geq \alpha\text{cut}, \quad 0 < \alpha\text{cut} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, any distances greater than $\sqrt{-\sigma^2 \log_e(\alpha\text{cut})}$ are given zero weighting and effectively ignored. Further, the use of a smaller multiplier for K in the numerator narrows the width of the Gaussian, and increases its height, leading to increased localization.

By slowly increasing the value of the α-cut term from a base value during convergence, the technique can initially behave globally, but become more local as the temperature drops. In this way, the semi-localized net still retains its global properties during the early stages of convergence when they are needed most. Through this global behavior, all units are initially influenced by all the edge points and the net is able to position itself appropriately, avoiding the need for manual seeding. However, once an approximate solution has been formed with respect to the input features, the net should then become more localized, allowing specialization onto a solution.

Furthermore, through the use of fewer units than edge points (i.e., $M < N$), the net is forced to generalize by linking dense regions of edge points, rendering the attraction from sparse regions insignificant. Consequently, it forms a contour through the regions of the image with the greatest density of edge points. According to our requirements in Section 2 (that the target shape is the dominant feature in the image), those regions should correspond to the shape to be extracted. An unfortunate consequence of a data-driven extraction method is that the net will also link prominent undesired image features, since it is unable to distinguish between shape and non-shape pixels.

4 Inclusion of Volumetric Information

In the common situation where multiple transverse slices of the object are available, volumetric information can be exploited to increase the robustness of contour extraction, especially in the presence of noise. The basic idea is that the nets are applied as in the 2D case described above, but cooperatively—with additional tension forces employed between adjacent slices—to extract pseudo-3D shape. Thus, movement of units in one slice affects, and is affected by, movement in neighboring slices. We emphasize that each unit's movement is constrained within its respective slice (i.e., movement of units is *not* allowed into the region between slices). This is essentially because the computational cost of allowing full three-dimensional movement would be very high, since the summation in the denominator of equation (2) becomes a double summation. The new vertical tension forces, in combination with the existing horizontal tension forces (Figure 3), promote smoothness in all three dimensions.

[Figure 3 about here.]

This arrangement gives rise to the notion of an 'elastic balloon' encompassing the object [37]. Imagine that air is sucked out of the balloon, causing it to 'shrink wrap' around the object, resulting in a 3D coordinate surface description which can easily be resampled to produce multiple 2D contours. Alternatively, it can be used directly for visualization. However, not all classes of 3D shape can be modeled like this. The technique cannot extract concavities along the scanning axis, since the balloon is ultimately comprised of planar sections.

The new tension forces are incorporated using a two-step update rule, based on equation (1). This allows attraction and horizontal tension forces to act upon each unit (Step 1) independently of vertical forces (Step 2):

$$\begin{aligned}
\text{Step 1:} \qquad \Delta \mathbf{u}_{j,l}^{t+1a} &= \alpha \textstyle\sum_{i=1}^{N} w_{ij} \left(\mathbf{p}_{i,l} - \mathbf{u}_{j,l}^{t} \right) + K\beta \left(\mathbf{u}_{j+1,l}^{t} - 2\mathbf{u}_{j,l}^{t} + \mathbf{u}_{j-1,l}^{t} \right) \\
\text{Step 2:} \qquad \Delta \mathbf{u}_{j,l}^{t+1} &= K\gamma \left\{ \mathbf{u}_{j,l+1}^{t+1a} - 2\mathbf{u}_{j,l}^{t+1a} + \mathbf{u}_{j,l-1}^{t+1a} \right\}
\end{aligned}$$

where a new index, l, is introduced denoting the current layer of the multilayered data, and γ determines the strength of the vertical tension. Splitting the application of forces in this way prevents direct competition, which often results in instability. (The nets would oscillate and then diverge.)

As in the 2D case, the nets are all initialized as large feature-encompassing circles (forming a cylinder in 3D space), and using similar parameters for attraction strength and temperature reduction, the nets then converge cooperatively until the temperature reaches a predetermined value. The coordinates of each of the units represent multiple 2D contours, or can be used as polygon vertices for plotting giving a 3D rendering of the object. However, because of the interdependence of each net on its neighbors, extraction is not performed solely on a per-slice basis: Missing data or excessive noise on one slice can be compensated for by the nets in neighboring slices. Thus, improved tolerance to noise is gained at negligible computational cost, and the limitations of the 2D net described earlier are overcome. Also, since a preformed 3D surface is mapped onto the target surface, the tiling and correspondence problems (see Section 2) are solved intrinsically (contrast with directly constructing surfaces between slices). Hence, 2D contour extraction is more robust and exploits 3D information appropriately.

Since the vertical tension is elastic, changes in size and shape of the imaged object through successive slices can be extracted. The method does not assume identical outlines on each slice. However, there is a limit to the extent of difference that can be tolerated by the nets. If an outline changes too much between successive slices, then the cooperative nets will be unable to follow the outline correctly because of the smoothness incurred by the vertical tension. Such a situation can arise where the imaged object exhibits abrupt changes in shape, or if the spatial resolution of the imaging method is too coarse.

Additionally, since the technique is still only attracted to dense regions of pixels, it is possible for the nets to link noise if it happens to be consistent through several slices. Generally, this is unlikely, but the situation may arise in specific applications as an artifact of the data or the preprocessing stage. Again, this is an inherent problem of a data-driven extraction technique. However, even under these circumstances, the technique will not simply fail, but rather degrade gracefully. This is an important attribute, since even in the presence of considerable signal corruption there should still enough information to perform some very coarse extraction.

5 Example Application—3D Lung Modeling

In connection with other work [25, 38], we require 3D shape models of the lungs, given multislice (256×256) magnetic resonance (MR) scans of the human thorax. Currently, a semi-automated method is used for each slice, in which an expert medical physicist seeds a pixel-based region-growing technique and then hand-edits the final outlines to produce closed, thin contours. This process clearly suffers problems of time efficiency and repeatability. To provide a more fully-automatic alternative, we have previously explored (supervised) segmentation procedures which use a multilayer perceptron (MLP) to classify image pixels as lung-boundary or non-lung-boundary [25, 38]. However, because the classified images contain noise (misclassifications) and the boundaries are fragmented, further processing of the edge image is required. The semi-localized elastic net seems ideally suited to complete the extraction process, since the MLP-segmented data broadly satisfy the conditions set out in Section 2 (first paragraph).

5.1 Classification of Edge Points

The first stage of the segmentation process is the classification of the (256×256) MR image into lung-boundary and non-lung-boundary pixels. This is performed by an MLP with one hidden layer of 30 nodes, and an input layer of 49 nodes which represents the grey-scale value of the MR image from a (7×7) window placed at predetermined points. A single output node is used to signify whether the window is centered on a lung boundary or not. All lung boundary-centered windows from the designated training slices are used for initial training, along with a slightly greater number of randomly selected non-boundary-centered windows. Subsequently, a selection of misclassified lung boundary pixels is added to the training set, and the MLP is retrained. The semi-automatically extracted outlines mentioned earlier provide the target patterns (see Figure 1). The classified images typically contain about 300–600 (binary) edge points.

[Figure 4 about here.]

The complete MRI dataset consists of around 30 slices from each of 12 subjects. Figure 4(a) shows a typical such slice. We present here representative results when the MLP is trained on some 7 or 8 selected slices from each of 4 subjects. Once trained, the MLP classifies an entire dataset from seen and unseen subjects, outputting values which are then thresholded to produce a binary edge image, as depicted in Figure 4(b). Although the lung boundaries are indeed the dominant feature in the output image, the pixel classification is imperfect: The images are contaminated with noise (false positives) and fragmented lung boundaries (false negatives). We stress that the MLP is doing much more than simple edge detection, since boundaries other than the lungs are suppressed.

We now require a second stage to link the fragmented boundary segments together, while ignoring the noise. This second stage of the segmentation process (extraction) is performed by the semi-localized elastic net.

5.2 Extraction by Semi-Localized Net

The semi-localized elastic net technique in its 2D form (Section 3) has previously been applied to the segmented MR data on a per-slice basis [39]. While the results were encouraging, the technique was often defeated by the variable quality of the MLP's classification, and the differing geometry of the lung boundaries found in both intrasubject and intersubject slices. Better robustness, therefore, could be achieved through incorporating the information implicit in the spatial sequence of MR data, using the pseudo-3D approach.

First, the edge-point images are divided vertically to separate the left and right lungs: The technique is applied to each of these separately. From this point on, the term 'image' refers to a half-image containing only one lung boundary, with operations being performed identically on both halves of the MLP output. Next, the balloon is initialized as a large feature-encompassing cylinder and, using empirically-derived values for the parameters in equation (1), converges onto a solution.

[Figure 5 about here.]

Figure 5(a) shows the semi-automatically extracted outline for the lung slice shown in Figure 4(a). The semi-automatically derived outlines act as the target data for training the MLP (and for the determination of effectiveness—see below). Figure 5(b) depicts the outline for the same slice as found by the semi-localized elastic net. Comparing this with Figure 4(b), it is clear that the elastic net has very successfully linked the edge points in the latter image, at the same time ignoring most of the noisy points, to produce a smooth, connected outline of the lungs.

To quantify the performance of the technique, we use the *effectiveness* measure [40]. Since the target data consist solely of binary edge points, no explicit contour information is available. Hence, pixel-based methods must be used for evaluation. Since the target data are themselves imperfect (see Figure 5(a)), the fitted contour should approximate and smooth the target: A perfect fit is unnecessary and undesirable. Thus, a method which accumulates scores of zero for every 'result' pixel (i.e., pixel on the path corresponding to the converged net) which is not coincident with a target pixel will be unduly pessimistic. We prefer, therefore, a method which scores area differences between the target and result contours.

Thus, we count N_{tar}, the number of pixels inside the target lung outline (left and right lungs in each slice are treated separately), and the number of 'result' pixels inside the lung produced by the modified elastic net (N_{res}). The intersection of these two regions is the set of pixels in both regions, and the ratio of the number of intersection pixels, N_{int}, to the number of target and result pixels forms two measures—precision (P) and recall (R) [40]:

$$
\begin{aligned}
P &= \frac{N_{\text{int}}}{N_{\text{res}}} = \frac{\textit{correct\_positives}}{\textit{correct\_positives} \ + \ \textit{false\_positives}} \qquad 0 \leq P \leq 1 \\
R &= \frac{N_{\text{int}}}{N_{\text{tar}}} = \frac{\textit{correct\_positives}}{\textit{correct\_positives} \ + \ \textit{false\_negatives}} \qquad 0 \leq R \leq 1
\end{aligned}
$$

The effectiveness is then:

$$E = 1 - \frac{PR}{\alpha R + (1-\alpha)P} \qquad 0 \leq E \leq 1, \ 0 \leq \alpha \leq 1$$

which combines both types of error (false positives and false negatives) into a single measure. Note that E represents an inverse measure of goodness: Both P and R must be high for E to be low. The parameter α can be used to bias E in favor of precision or recall. Since these are equally important in this application, we use $\alpha = 0.5$.

Results are presented here for an MLP trained using 7 or 8 selected slices (semi-automatically segmented) from each of the three subjects AC, CB and LP. All slices were then classified, lung boundaries extracted, and E values determined for each slice. We also present typical results for a single 'unseen' subject, NH. That is, the MLP was trained on semi-automatically segmented data from AC, CB and LP but tested on NH's data. Summary values quantifying the performance are listed in Table 1. Although $E = 0$ signifies a perfect match, very low values are unlikely because of the approximating nature of the deformable contour. To give a baseline for assessing pseudo-3D shapes extracted from the fully-automatically segmented images, the outlines extracted from the semi-automatically segmented (and therefore noise-free) target slices yielded an average E-score of 0.06.

[Table 1 about here.]

Compared to this baseline, the performance of our method is reduced (average E increases to 0.1) mostly because of difficulty in coping with the dramatic deformation around the liver in the lower slices. This usually results from poorly-classified boundary segments on these slices, where there are too few edge points along the true boundary to deform the balloon appropriately.

[Figure 6 about here.]

A notable exception is subject AC, where the basic lung shape was extracted almost flawlessly. Figure 6 shows two views of wire-frame projections of the extracted pseudo-3D outlines of the two lungs from 24 slices taken from this subject. The deformation caused by the liver can be seen toward the bottom of the rightmost image. The balloon's parameters were: $\alpha = 0.05$, $\beta = \gamma = 1.0$, $K(0) = 0.15$ reduced by 1% per epoch (6 iterations per epoch), and $\alpha - \text{cut} = \{K(0) - K(epoch)\}$. For the lower slices where the lungs are deformed around the liver, we set $\beta = \gamma = 0.5$. The lower strength of the tension forces permit greater deformations in the balloon. We used $M = 50$ units per net, and the algorithm was allowed to run for 160 epochs.

5.3 Problems with the Technique

Two problems arise in applications such as this. First, while the incorporation of 3D information intrinsically increases tolerance to speckle (uncorrelated) noise, it can decrease tolerance to (correlated) noise consistent across several slices. For example, the MLP often misclassifies small parts of the torso boundary and blood vessel walls which naturally range over several slices. These can be erroneously linked by the balloon if they are sufficiently represented in successive images. Second, large changes in boundary shape or position between adjacent slices can mislead the technique, because of the physical smoothness of the balloon. A manifestation of this can be seen in the lower MRI slices where the lung's basically circular cross-section rapidly becomes crescent-shaped as the liver intrudes. Consequently, the effectiveness measure for these slices is poor.

We have experimented somewhat with preforming the balloon on the target data of one subject, relaxing the Gaussian regions of attraction, and applying the deformed balloon again to the MLP's output for the unseen subjects. This has produced improved average E-scores of around 0.08 (since the liver intrusion is now less of a problem). Alternatively, the first problem could be alleviated through better initial classification. Also, increasing the spatial resolution in the medial axis (currently 8 mm between slices) would make boundary deformations smoother and easier to track. Pragmatically, however, increased resolution means increased scanning time (which already takes 20 minutes for 30 slices) and a dramatic increase in data—most of which would be redundant in areas of low change.

6 Conclusions and Discussion

This paper has detailed the implementation of a semi-localized elastic net for data-driven shape extraction from binary edge point images. A desirable property of the original Durbin-Willshaw elastic net, on which this new work is based, is its use of the entire dataset when forming solutions. The elastic net is also geometrically flexible and produces a guaranteed solution with very little *a priori* information needed. Unfortunately, its global nature also makes the technique inherently unsuitable for handling noise. The semi-localized net provides a compromise between retaining global behavior and introducing noise tolerance, allowing the technique to extract 2D shapes from noisy images.

Where a spatial sequence of images (slices) of the object is available, a 3D version of the semi-localized net ('balloon') can be used—consisting of one 2D net per slice, acting cooperatively. By exploiting 3D information in this way, increased noise tolerance can be realized since each net is utilizing the additional knowledge implicitly stored in adjacent nets. The technique now represents a surface reconstruction method which intrinsically solves the correspondence and tiling problems [20]. However, the elastic properties of the balloon which provide these solutions also introduce a new difficulty—extracting rapidly changing boundaries. While this situation can be avoided by increasing the spatial resolution between slices, in general this may not be an option. Extraction under these conditions will always present difficulty for physics-based deformable models in a noisy domain. Non-physics-based methods could cope, but generally would not intrinsically solve the correspondence and tiling problems. At this stage, there is no provision for multiple objects, and so no solution to the branching problem is available.

A further problem with extracting shape from multiple 2D slices is the occurrence of noisy pixels across successive slices. While tolerance to slice-uncorrelated noise is introduced through the cooperative application of a per-slice technique, slice-correlated noise still presents a difficulty. The semi-localized elastic net is susceptible to such noise: This is inherent in any data-driven method. However, with suitable preprocessing of the slices, we are often successful in extracting the desired object.

An example of lung shape extraction from MR scans of the human torso was given, showing a typical application of our technique and identifying the problems that arise (rapidly changing boundaries, and noise correlated through successive slices). Despite this, the net copes well, producing results that we deem already usable for the intended application.

Perhaps the main competitor to the semi-localized elastic net is the snake, which has been intensively developed recently for applications just like this (e.g., [34, 41, 42, 43]). We have not performed any empirical comparison of the methods here: It would in any case be difficult to decide precisely what form of snake should be used. However, it is worthwhile making some comparative remarks. Relative to our method, snakes have lower computational complexity and are therefore likely to be faster. They also have better noise tolerance than even the semi-localized form of the elastic net. However, for anything other than very simple shapes, the snake's regularization term will need local adjustment to cope with concavities, discontinuities etc. Also, the price paid for the snake's good noise tolerance is its requirement for manual seeding. If seeded too far away from the target shape, the snake may well diverge away from it.

In future work, we intend to concentrate on the preforming extension briefly mentioned in Section 5.2. This seems to overcome many of the difficulties arising from poorly classified boundaries, while still retaining the generality of the approach. Additionally, we intend to perform an empirical comparison with snakes (the semi-localized net's closest competitor) for a range of images and objects.

References

[1] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, New York, NY, 1979.

[2] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. *The Travelling Salesman Problem*. Wiley, Chichester, UK, 1985.

[3] J. J. Hopfield and D. W. Tank. Neural computation of decisions in optimization. *Biological Cybernetics*, 52(3):141–152, 1985.

[4] G. V. Wilson and G. S. Pawley. On the stability of the travelling salesman problem algorithm of Hopfield and Tank. *Biological Cybernetics*, 58(1):63–70, 1988.

[5] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Reading, MA, 1991.

[6] D. E. Van den Bout and T. K. Miller. Improving the performance of the Hopfield-Tank neural network through normalization and annealing. *Biological Cybernetics*, 62(2):129–139, 1989.

[7] M. K. Peng, N. K. Gupta, and A. F. Armitage. An investigation into the improvement of local minima of the Hopfield network. *Neural Networks*, 9(7):1241–1253, 1996.

[8] M. Martín-Valdivia, A. Ruiz-Sepúlveda, and F. Triguero-Ruiz. Improving local minima of Hopfield networks with augmented Lagrange multipliers for large scale TSPs. *Neural Networks*, 13(3):283–285, 2000.

[9] R. Durbin and D. Willshaw. An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, 326(6114):689–691, 1987.

[10] P. D. Simic. Statistical mechanics as the underlying theory of 'elastic' and 'neural' optimisations. *Network: Computation in Neural Systems*, 1(1):89–103, 1990.

[11] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

[12] T. Kohonen. The self-organising map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[13] J. C. Fort. Solving a combinatorial problem via self-organizing process – An application of the Kohonen algorithm to the traveling salesman problem. *Biological Cybernetics*, 59(1):33–40, 1988.

[14] F. Favata and R. Walker. A study of the application of Kohonen-type neural networks to the traveling salesman problem. *Biological Cybernetics*, 64(6):463–468, 1991.

[15] N. Aras, B. J. Oommen, and İ. K. Altınel. The Kohonen network incorporating explicit statistics and its application to the travelling salesman problem. *Neural Networks*, 12(9):1273–1284, 1999.

[16] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[17] A. Blake and M. Isard. *Active Contours: The Application of Techniques from Graphics, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer, London, UK, 1998.

[18] A. J. Abrantes and J. S. Marques. Unified approach to snakes, elastic nets, and Kohonen maps. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '95*, volume 5, pages 3427–3430, Detroit, MI, 1995.

[19] M. M. Trivedi and A. Rosenfeld. On making computers "see". *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1333–1335, 1989.

[20] C. L. Bajaj, E. J. Coyle, and K.-N. Lin. Arbitrary topology shape reconstruction from planar cross sections. *Graphical Models and Image Processing*, 58(6):524–543, 1996.

[21] S. J. Gilson and R. I. Damper. An empirical comparison of neural techniques for edge linking of images. *Neural Computing and Applications*, 6(2):64–78, 1997.

[22] S. Webb, editor. *The Physics of Medical Imaging*. Adam Hilger, Bristol, UK, 1988.

[23] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman and Hall, London, UK, 1993.

[24] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, New York, NY, 1982.

[25] I. Middleton and R. I. Damper. Segmentation of magnetic resonance images of the thorax by back-propagation. In *IEEE International Conference on Neural Networks*, volume 5, pages 2490–2494, Perth, Western Australia, 1995.

[26] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[27] M. S. Nixon and T. K. Hames. New technique for 3D artery modeling by noninvasive ultrasound. *IEE Proceedings, Part I–Communications, Speech and Vision*, 140(1):86–94, 1993.

[28] J. M. Nash, J. N. Carter, and M. S. Nixon. Dynamic feature extraction via the velocity Hough transform. *Pattern Recognition Letters*, 18(10):1035–1047, 1997.

[29] J.-D. Boissonnat. Shape reconstruction from planar cross-sections. *Computer Vision, Graphics, and Image Processing*, 44(1):1–29, 1988.

[30] B. Geiger. *Three-Dimensional Modeling of Human Organs and its Application to Diagnosis and Surgical Planning*. PhD thesis, INRIA, Sophia Antipolis, France, 1993.

[31] W. E. Lorensen and H. E. Cline. Marching cubes: A high resoution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

[32] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):703–714, 1991.

[33] Y. F. Wang and J.-F. Wang. Surface reconstruction using deformable models with interior and boundary constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(5):572–579, 1992.

[34] T. McInerney and D. Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4(2):73–91, 2000.

[35] R. Durbin, R. Szeliski, and A. Yuille. An analysis of the elastic net approach to the travelling salesman problem. *Neural Computation*, 1(3):348–358, 1989.

[36] M. W. Simmen. Parameter sensitivity of the elastic net approach to the traveling salesman problem. *Neural Computation*, 3(3):363–374, 1991.

[37] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[38] I. Middleton. *Segmentation of Magnetic Resonance Images Using Artificial Neural Networks*. PhD thesis, Department of Electronics and Computer Science, University of Southampton, Southampton, UK, 1998.

[39] S. J. Gilson, I. Middleton, and R. I. Damper. Neural techniques for outlining the lungs in MR images of the thorax. In *International Symposium on Soft Computing, World Automation Congress*, pages ISSCI–072.1–ISSCI–072.6, Anchorage, AK, 1998.

[40] C. J. van Rijsbergen. *Information Retrieval (2nd Edition)*. Butterworth, London, 1979.

[41] N. E. Davison, H. Eviatar, and R. L. Somarjai. Snakes simplified. *Pattern Recognition*, 33(10):1651–1664, 2000.

[42] W. P. Choi, K. M. Lam, and W. C. Siu. An adaptive active contour model for highly irregular boundaries. *Pattern Recognition*, 34(2):323–331, 2001.

[43] X. M. Pardo, M. J. Carreira, A. Mosquera, and D. Cabello. A snake for CT image segmentation integrating region and edge information. *Image and Vision Computing*, 19(7):461–475, 2001.

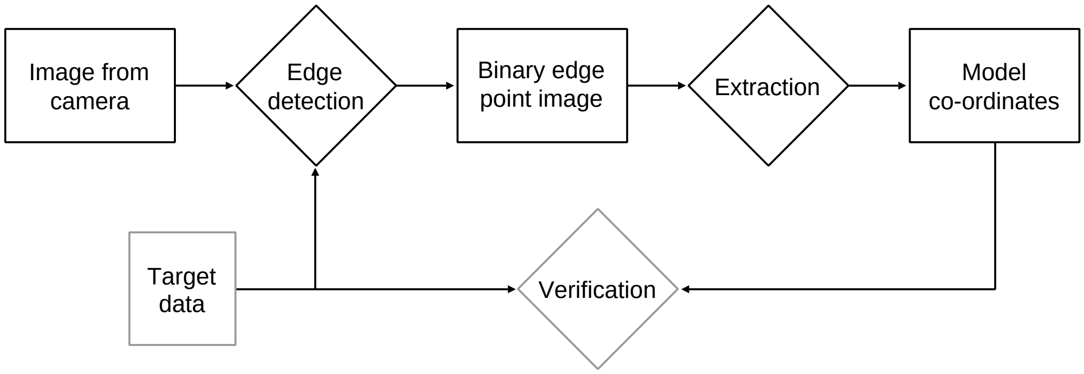


Figure 1: Processing stages required for shape extraction from images. The target data (shown in gray) are used both for training the edge detector and in evaluation of the final contours.

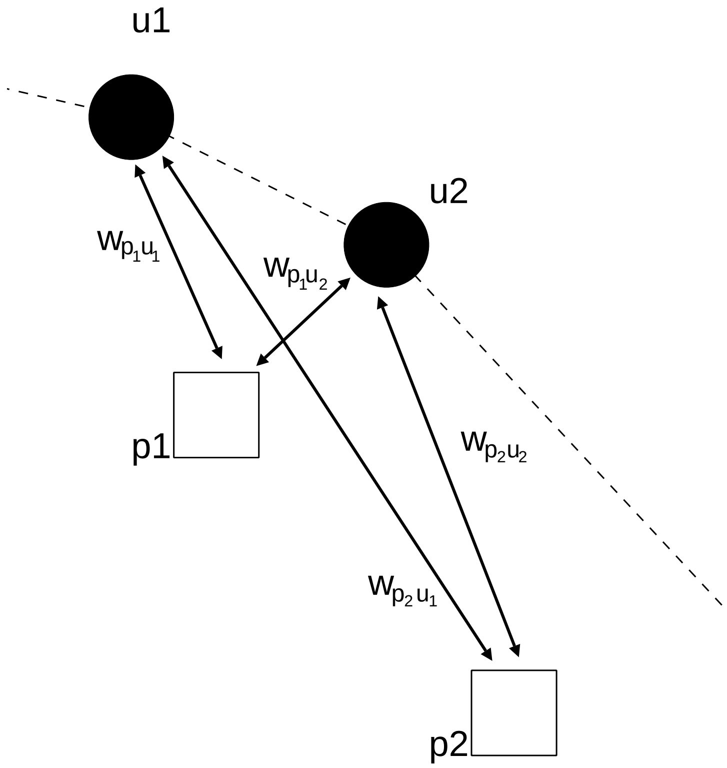


Figure 2: Deformable network model, showing the units (black circles) and their connectivity (dotted lines). Some attraction forces are also shown between the units and edge points.

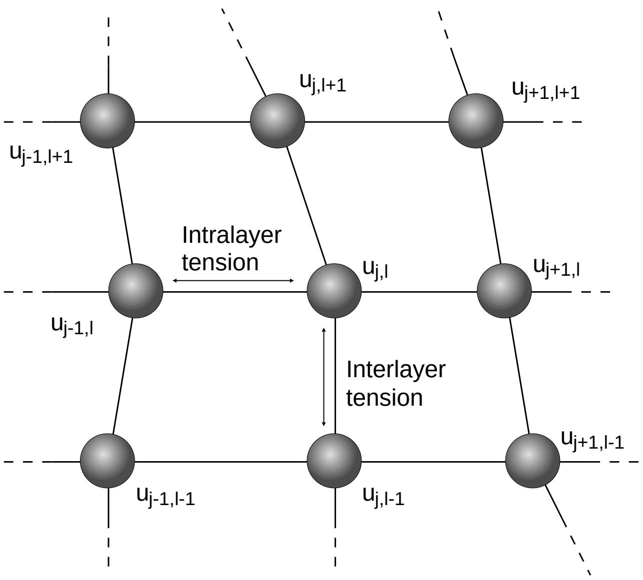
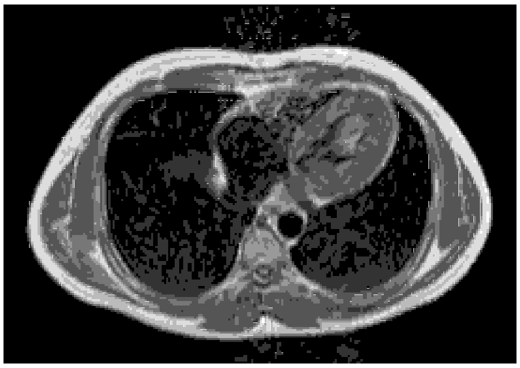
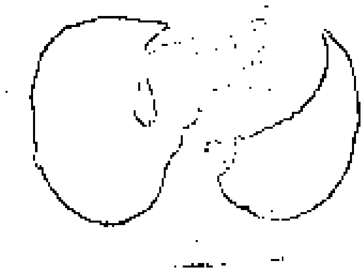


Figure 3: The topology of the 3D surface reconstruction method, showing the tension forces acting on each unit. Note that each unit is still constrained to a horizontal plane.

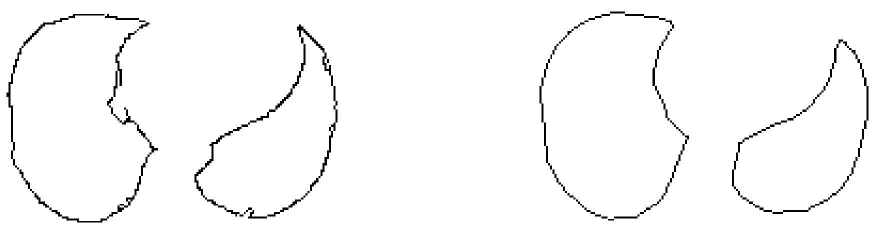


(a)



(b)

Figure 4: (a) A typical MR scan of the human thorax, and (b) the output of the MLP. Note the contamination by torso and internal organ boundaries.



(a) (b)

Figure 5: (a) Semi-automatically extracted outline for the lung slice shown in Figure 4(a). (b) Outlines for the same slice after postprocessing the edge-point image in Figure 4(b) by the semi-localized elastic net.

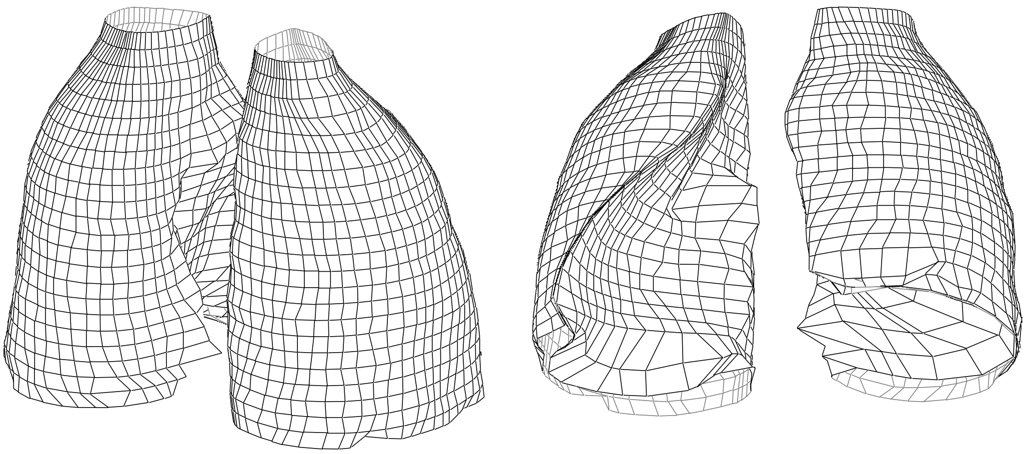


Figure 6: Two views of wire-frame projections of the extracted lung outlines from 24 slices taken from subject AC. The deformation caused by the liver can be seen toward the bottom of the rightmost image.

Subject	N_{tar}	N_{res}	N_{int}	P	R	E
AC	5531	5442	5080	0.92	0.91	0.088
CB	7127	8166	6702	0.83	0.92	0.136
LP	9613	9843	9030	0.89	0.92	0.1
NH	6709	7502	6476	0.86	0.96	0.095

Table 1: Segmentation performance for four subjects (values averaged over 24 slices). See text for definitions of the measures tabulated. Data from subjects AC, CB and LP were used to train the MLP. However, performance for 'unseen' subject NH is not noticeably poorer.