

Published in final edited form as:

*IEEE Trans Image Process.* 2005 September ; 14(9): 1314–1323.

## An Energy-Based Three-Dimensional Segmentation Approach for the Quantitative Interpretation of Electron Tomograms

**Alberto Bartesaghi [Student Member, IEEE] and Guillermo Sapiro [Senior Member, IEEE]**

*Electrical and Computer Engineering Department, University of Minnesota, Minneapolis MN USA 55455 (e-mail: abarte@umn.edu; guille@umn.edu).*

**Sriram Subramaniam**

*National Cancer Institute, National Institutes of Health, Bethesda, MD 20892 USA (e-mail: subramas@mail.nih.gov).*

### Abstract

Electron tomography allows for the determination of the three-dimensional structures of cells and tissues at resolutions significantly higher than that which is possible with optical microscopy. Electron tomograms contain, in principle, vast amounts of information on the locations and architectures of large numbers of subcellular assemblies and organelles. The development of reliable quantitative approaches for the analysis of features in tomograms is an important problem, and a challenging prospect due to the low signal-to-noise ratios that are inherent to biological electron microscopic images. This is, in part, a consequence of the tremendous complexity of biological specimens. We report on a new method for the automated segmentation of HIV particles and selected cellular compartments in electron tomograms recorded from fixed, plastic-embedded sections derived from HIV-infected human macrophages. Individual features in the tomogram are segmented using a novel robust algorithm that finds their boundaries as global minimal surfaces in a metric space defined by image features. The optimization is carried out in a transformed spherical domain with the center an interior point of the particle of interest, providing a proper setting for the fast and accurate minimization of the segmentation energy. This method provides tools for the semi-automated detection and statistical evaluation of HIV particles at different stages of assembly in the cells and presents opportunities for correlation with biochemical markers of HIV infection. The segmentation algorithm developed here forms the basis of the automated analysis of electron tomograms and will be especially useful given the rapid increases in the rate of data acquisition. It could also enable studies of much larger data sets, such as those which might be obtained from the tomographic analysis of HIV-infected cells from studies of large populations.

### Index Terms

Distance functions; electron tomography; energy-based segmentation; geodesics; high resolution; HIV; minimal surfaces; volume segmentation

### I. Introduction

TRANSMISSION electron microscopes have conventionally been used in biomedical research to obtain two-dimensional (2-D) projection images of thin objects such as molecules, cells, and tissues. Such images can be recorded in most modern electron microscopes at

---

This work was supported in part by the Office of Naval Research, the National Science Foundation, the National Institutes of Health, and the National Geospatial-Intelligence Agency. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erik Meijering.

magnifications ranging from  $\sim 100\times$  to  $\sim 1000\,000\times$ . The use of electron microscopes, is, however, not limited to imaging in 2-D. Using emerging methods in electron tomography (see [1] for a recent review), it is now also possible to routinely determine three-dimensional (3-D) structures using principles that are very similar to those used in technologies such as computerized axial tomography. Thus, one can record a series of images of a given object over a wide range of tilt angles and combine them using back projection algorithms to generate a 3-D volume of the imaged object.

A key problem in biological electron tomography is that the images obtained are at relatively low signal-to-noise ratios (SNRs). In part, this is because of the tremendous complexity of biological specimens; for example, a single human cell can contain thousands of copies of tens of thousands of proteins packaged in a variety of multi-protein complexes and organelles of differing shapes and sizes. A second factor comes from the potential of electrons to damage organic matter, which necessitates the use of electron doses that are high enough to obtain measurable contrast, but low enough to minimize structural damage. The rapid, quantitative interpretation of the vast amount of data in tomograms of cells and tissues, therefore, poses a challenging problem. This issue is becoming increasingly important now because of the rapid advances in instrument automation that have led to dramatic enhancements in the speed of data collection. We are interested in developing approaches for 3-D segmentation of features in cellular tomograms that can work robustly and rapidly despite the low SNRs. This is a fundamental step in the automatic analysis of large amounts of data for statistical inference.

As a test case, we use tomograms recorded from human macrophages infected with HIV, although clearly the computational techniques here developed are not limited to this (important) example. The cells were fixed, embedded in plastic in the presence of uranyl acetate and sectioned in an ultra-microtome to produce sections with thicknesses in the range of 100–150 nm. These sections were placed on an electron microscopic grid coated with a thin carbon film, and imaged in a Tecnai 12 electron microscope operating at 120 kV equipped with a LaB<sub>6</sub> filament. Tomograms were constructed using standard back-projection algorithms as implemented in the IMOD reconstruction package [2].

Fig. 1 shows slices from a tomogram recorded from a small region of cells infected with HIV. Within this slice, there are several identifiable features which bear a resemblance to the slice of either an assembled virion or enclosed membranous entities with varying interior density relative to the cytoplasmic medium. Our goal is to detect these structures with minimal user bias, analyze them, and establish a correlation of the nature and extent of these features with progression of viral infection.

Here, we deal specifically with solving the segmentation problem, which is of fundamental importance for the subsequent analysis of the data and is usually the first step required in many such applications. As recent advances in image acquisition technology allow analysis at the scale of large populations (see [3] for some of our efforts on the automatization of data collection), availability of robust and computationally efficient segmentation techniques is of paramount importance.

The segmentation of individual features is formulated as the computation of surfaces of minimal energy on a metric space that depends on image features. Such techniques have been extensively used in the literature both for 2-D and 3-D object segmentation. The idea is to first design an energy functional (combining image driven and regularizing terms) that is minimized at the object of interest. Thereafter, the problem becomes one of nonconvex optimization usually solved by following a gradient descent flow that gives a local minimizer of the energy. Initialization of the flow is a key step that will allow recovery of the desired object provided the initial guess is close to the correct minima. Availability of robust optimization techniques

is then very important, not only because they guarantee finding the correct minima, but also because they allow us to concentrate on the design of the segmentation energy which will ultimately determine the performance of the algorithm.

In this paper, we propose a novel algorithm to obtain a surface estimate that is already the segmenting boundary or else sufficiently close to it, so only a few steps of the gradient descent flow will be required to get the optimal surface. The only requirement of the algorithm is to know the position of a point inside each volume of interest. Full tomograms are segmented in a semi-automatic fashion: A single point inside each cell is first specified by the user selecting only those structures that correspond to simple closed boundaries (this is done by inspection of all slices in the 3-D volume, other cells are ignored at this early stage). For each selected point, we automatically segment the surrounding structure using the proposed algorithm.<sup>1</sup>

The paper is organized as follows. In Section II, we review the energy-based segmentation framework in the planar (2-D) and volumetric (3-D) settings, motivating the need of our technique. In Section III, we describe the proposed 3-D segmentation algorithm. Section IV shows various segmentation results and presents preliminary statistical analysis of two tomograph sets. We conclude with a discussion in Section V.

## II. Previous Work and Background on Energy-Based Segmentation

Electron tomographic images generally display very low SNRs, and the task of segmenting features of interest is a very challenging one. The landscape of the segmentation energy will present a number of local minima demanding the application of robust optimization techniques. In this section, we review the energy-based segmentation framework as presented, for example, in [4].

### A. Energy-Based Segmentation

Let  $\Gamma$  be a contour embedded in  $\mathcal{R}^n$  that represents the boundary of interest (a curve for  $n = 2$ , a surface for  $n = 3$ ). Its intrinsic length/area is

$$\int_{\Gamma} g(\Gamma) d\lambda \quad (1)$$

where  $g : \mathcal{R}^n \rightarrow (0, \infty]$  is the image derived metric that takes small values at boundary points (thin dark regions in the case of our data) and large values elsewhere in the image. By minimizing the quantity in (1),  $\Gamma$  is encouraged to go through areas of small cost (corresponding to boundaries) yielding the desired segmentation. Using calculus of variations, the corresponding gradient descent flow is

$$\Gamma_t = (g\kappa - \nabla g \cdot \vec{N})\vec{N}$$

where  $\vec{N}$  is the normal to the contour and  $\kappa$  is the curvature ( $n = 2$ ) or mean curvature ( $n = 3$ ). This evolution can be implemented in the level set framework [5], by embedding the contour  $\Gamma$  as a level set of a higher dimensional function  $\phi$ . The evolution for  $\phi$  that embeds the motion of  $\Gamma$  is

$$\phi_t = g\kappa \left| \nabla \phi \right| + \nabla g \cdot \nabla \phi. \quad (2)$$

<sup>1</sup>The detection of this single point inside the 3-D shape of interest is independent of the search for the minimal surface. Thereby, we limit our discussion to the segmentation itself, while the automatic finding of the inner point will be reported elsewhere. Even if the selection of this point is left fully manual, the computation speedup and reduced user intervention are extremely significant when compared with the fully manual procedures currently employed for this data.

The scalar metric has typically the form  $g = f(I) + w$ , where  $f$  depends on the input image  $I$ , and  $w$  is a constant that can control the smoothness of the minimizing contour. Given an initial contour  $\Gamma_0$ , the minimizer is obtained as the steady-state solution of the gradient descent flow in (2). Although this technique has proven to be very successful and is considered state-of-the-art for many medical imaging disciplines (it is for example part of the ITK initiative, [www.itk.com](http://www.itk.com)), it strongly depends on the choice of the initial contour  $\Gamma_0$  which has to be close to the desired minima. Many algorithms have been proposed in the literature that provide different mechanisms trying to drive the surface toward the desired minima, e.g., [6]–[8]. Although they offer improved performance they can still get trapped in unwanted minima, producing a surface which does not correspond to the desired segmentation.

Graph-based algorithms have also been applied to solve the minimal surface problem [9], [10]. In this setting, boundary conditions are specified as two sets—object and background—and the surface is found as the optimal cut across the graph separating the two sets. Edge costs can be derived from the continuous metric but accuracy is dependent on the neighborhood system. Although metrification artifacts can be reduced by increasing the number of directions, computational complexity grows rapidly. The level of accuracy attainable with tractable complexity is not enough for the examples presented here given the large size of volumes and the significant amount of noise.

Similarly, in [11], the authors address the computation of optimal surfaces simulating a continuous maximal flow between a source and sink set. The algorithm has to be initialized with discrete maximal flow algorithm and uses a multiscale approach to accelerate convergence of the associated flow. Although increased accuracy can be achieved with this approach, computational complexity is still an issue (at least for the size of volumes we handle in this paper).

In general, limited 3-D evaluation makes it unclear how well these algorithms behave in the presence of significant noise and very weak boundaries typical of tomography images. Efficiently solving and adapting the energy-based segmentation framework presented above to our data, 3-D electron tomograms, is part of the goal of this work.

## B. Closed Geodesic Curves in the Plane

Of particular relevance to our work is the approach in [12], where the authors propose an algorithm for computing planar geodesics between two given points using a noniterative procedure. In a two-step process, they first compute the intrinsic distance function from one (user selected) end point to the rest of the domain, and a back propagation procedure from the second (user selected) point gives the actual geodesic joining both points. Intrinsic distances in the first step are efficiently computed using the fast marching algorithm [13]–[15].

The above technique can actually be applied for the computation of closed contours by adding the restriction of an interior point. That is, we now look for curves of minimal length that have  $p_0$  as an interior point. A nice construction to enforce this constraint was introduced in [16], where a discontinuity (“image cut”) is introduced in the image domain and the image metric scaled to be  $g/\rho$ , where  $\rho$  is the distance to  $p_0$  [see Fig. 2, left]. Since curves cannot go across the discontinuity, this becomes now a problem of computing periodic geodesics between the two sides of the cut. This is reminiscent of the computation of geodesics between points [12] described above, only that the periodicity constraint has to be enforced. Scaling the metric by  $1/\rho$  eliminates the bias toward small circular curves around  $p_0$ .

The above construction can be related to polar-transformed segmentation techniques [17], [18], where the image is first transformed into a polar domain, to then search for periodic

minimal paths between sides of the polar grid<sup>2</sup> (see Fig. 2, right). Minimal paths are then converted back to the original Cartesian grid and a closed contour is guaranteed.

Although these two approaches can be related to each other, they are not mathematically equivalent (a rigorous analysis of the precise relationship is beyond the scope of this paper, we refer the reader to work on that direction [20]). However, let us provide a simple qualitative argument in favor of carrying computations in the polar domain which is consistent with results systematically obtained on real tomographic images. Assuming the resolution of both grids is comparable in some sense, the geometry of the polar grid seems to be more compatible with our particular problem as it imposes a stronger bias toward geodesics with circular trajectories when compared to the equivalent Cartesian implementation. This effect becomes significant when images are very noisy and edge information is excessively weak, a situation frequently found on real tomograms, as shown in Fig. 3.

Regarding the location of the interior point, both constructions will produce the closed contour surrounding  $p_0$  with the lowest possible energy as defined in (1) on the corresponding domain. Given the intrinsic dependence on  $p_0$ , the resulting minimizers will be determined to some extent by the particular location of the interior point. For example, in the case of uniform  $g = 1$ , geodesics will be concentric circles centered in  $p_0$  (i.e., determining completely the location of minimizers). For nonuniform high-contrast  $g$ ,  $p_0$  becomes only a geometric constraint and geodesics will be guided mainly by image content. Fig. 4 illustrates this behavior in real examples.

Note that the previous discussion is only valid for curves that cross the discontinuity/cut only once; for our goal, we do not need to handle the case with multiple crossings as done in [16].

### C. Minimal Surfaces From Cross Sections

Another group of techniques (related to our approach) solve the 3-D minimal surface problem by sectioning the 3-D domain with 2-D planes and finding geodesics restricted to the cutting sections. If we assume that each such 2-D geodesic corresponds to the intersection of the 3-D minimal surface with the slicing plane, we can recover the surface as the collection of 2-D geodesic curves. In general, geodesics restricted to volume cross sections are not necessarily contained in the minimal surface. For example, consider the problem of finding the minimal surface within two parallel rings (boundary condition) with homogeneous metric throughout the volume ( $g = 1$ ) (see Fig. 5, top). It is well known that the solution is the catenoid surface provided the spacing between the rings is appropriate. If we slice the volume with planes in the vertical direction, geodesics restricted to these planes will be straight lines (because the metric is homogeneous), in clear disagreement with the 2-D cross sections of the catenoid (catenary curves). If the metric is ideally heterogeneous, e.g., uniform in the background and has a small value  $g = \epsilon$  on the dark curved wall (see Fig. 5, bottom), planar geodesics coincide with minimal surface cross sections. Of course, for real data, obtaining ideally heterogeneous metrics is not possible. It is then just intuitive that the minimal surface will try to stick to low  $g$  values (in an effort to minimize its intrinsic area) and that 2-D geodesics will have a better chance to coincide with cross sections of the minimal surface.

This is the implicit assumption in 3-D reconstruction algorithms from planar cross sections [21]–[23], that recover the surface as a collection of boundary curves from individual slices.

<sup>2</sup>The main problem when computing periodic geodesics is to find the endpoint position, say, in the first column of the polar domain, in optimal time. To do so, we first compute the distance of all points in the last column to the first one (this can be done in a single fast marching computation setting all points in the first column to zero distance). Similarly, we also compute the distance of all points in the first column to the last one. We select as endpoint the row position that minimizes the sum of both distances computed above. This is a fast approximate solution (see [17]–[19] for alternative techniques).

As we are computing sections originating from a single surface, it is reasonable to assume that curves in consecutive planes are close to each other (provided consecutive slices are close together). An intuitive way to force this condition is by first integrating image information across sections and then use the accumulated values to compute geodesics at individual sections, as done in [24] and [25]. The authors in [25] propose an algorithm that finds the minimal surface given a pair of curves as boundary conditions. Intrinsic distances are first computed from one of the curves (to the rest of the space) and geodesics from points in the second curve are projected into planes around a symmetry axis that joins the centroids of both curves. The surface is then recovered by interpolating between the individual geodesic curves. A similar approach applied to a problem with slightly different geometry is presented in [24] for solving the stereo correspondence problem. In this case, the minimal surface cuts horizontally through a rectangular domain, see Fig. 6. The metric is first accumulated on each vertical slice (from back to front) resulting in an intermediate distance volume which is then used as a new metric to obtain geodesics in the orthogonal vertical direction (from left to right). Both techniques resemble the two step procedure described at the beginning of Section II-B.

Integrating image information across sections by no means guarantees recovery of a smooth surface, i.e., geodesics in consecutive slices (computed with accumulated image values as the metric, as done in [24] and [25]) may be arbitrarily far from each other. Moreover, they may also be poorly localized because local image information is lost after performing the integration. This is illustrated in Fig. 7 where we show geodesics obtained using raw image values as the metric, compared to geodesics obtained using accumulated image values (from back-to-front slices) as the metric. Additional regularity constraints must then be enforced, e.g., [24], where geodesic curves are restricted to a band of width  $b$  around the geodesic in the previous slice. But the choice of  $b$  becomes critical, as too small  $b$ s will result in over smoothed surfaces that poorly follow the minimal surface (the object boundary), and bigger  $b$ s will not properly enforce the smoothness constraint. We address these issues in Section III-A.

### III. Computing Minimal Surfaces With an Internal Point

Motivated by the robustness of techniques described in Section II-B for the 2-D case, we will apply similar methods for the segmentation of volumetric data. The equivalent formulation in three dimensions is to find a closed surface  $S$  that minimizes the intrinsic area given by (1), with the restriction that a given point  $p_0$  is interior to the surface. Unlike planar geodesics around a point, there are no constructive procedures to find the closed minimal surface in this case. However, approximate methods can be designed that give at least a good initial guess assuming the metric  $g$  is well designed and heterogeneous “enough” throughout the surface of interest, i.e.,  $g$  vanishes on  $S$ . If this is the case, cross sections of the minimal surface will coincide with planar geodesics on the cutting sections (as discussed in Section II-C), and the surface can be recovered as a collection of 2-D geodesic curves.

We then consider a spherical coordinate system centered in  $p_0$ .<sup>3</sup> The transformed domain will be a volume with  $\rho \in [0, \rho_{\max}]$  spanning the vertical axis,  $\theta \in [0, 2\pi]$  and  $\phi \in [0, \pi]$  spanning the two horizontal axis as shown in Fig. 8.

Such a construction has a number of desirable features.

- The surface of star-shaped objects can be completely covered by two pencils of parallel vertical planes (this will be fundamental for the design of the minimization algorithm).

<sup>3</sup>This follows the equivalent construction in the polar domain described for the planar case.



- Similar to the planar case, the spherical change of coordinates introduces a scaling that eliminates the global minima of zero energy. If using an homogeneous metric  $g = 1$ , all spheres centered in  $p_0$  (horizontal planes in spherical space) will have the same intrinsic area eliminating the bias toward small area surfaces.
- Boundary conditions in the transformed domain are straightforward to enforce (this will be used extensively in the search algorithm).

This construction restricts the class of objects we can deal with to star-shaped surfaces (this is good enough for the application at hand). That is, given the interior point  $p_0$ , the algorithm can compute all surfaces that accept a parametrization of the form  $\rho = f(\varphi, \theta)$  where  $\rho$ ,  $\varphi$  and  $\theta$  are the spherical coordinates centered in  $p_0$  (as before), and  $f$  is uniquely defined for all pairs  $(\varphi, \theta)$ . Note that this includes convex and nonconvex surfaces provided  $p_0$  is appropriately chosen (see Fig. 9). So, in general, the position of the interior point is important and will determine the surface that we get as a result of the algorithm. On the other hand, the 3-D structures we deal with in this paper belong to a much more restricted class of surfaces (usually close to convex) so the actual position of the point is unimportant as long as it is inside the structure of interest. Following the discussion at the end of Section II-B, it is the strength of image boundaries what determines the actual degree of dependence of the resulting surface with respect to the location of  $p_0$ .

Volumetric reconstructions obtained by electron tomography have significantly better horizontal resolution compared to the vertical one. For this reason, inspection of the raw data for selecting the interior points is usually done in the horizontal direction. Therefore, once an interior point is specified, we first compute a planar geodesic in the selected horizontal slice that will serve as boundary condition for the minimal surface problem. The implicit assumption is that boundaries in the selected slice are strong enough (in the sense that a user was able to detect the presence of a feature of interest) for the 2-D segmentation techniques to get the correct result. The automatic detection of this interior point  $p_0$  should also be based on this and possible directions will be reported elsewhere.

Once we obtained the segmentation curve in the horizontal slice, we show how to find its correspondent through the spherical change of coordinates. As the origin of the transformation was selected to be the interior point, this initial curve will be on the  $z = 0$  plane (see Fig. 8). The crossings of the curve with the  $y$  axis will map through the spherical transformation into horizontal straight lines at the planes  $\varphi = 0$  and  $\varphi = \pi$  (see Fig. 8, right). If we split the referred curve (at the crossings with the  $y$  axis) in two portions, one will correspond to the  $\theta = \pi$  plane and the other to the  $\theta = 0$  plane, and because we are dealing with closed surfaces, we must force the curve at  $\theta = 2\pi$  to be the same as the one at  $\theta = 0$  to satisfy the periodicity constraint. We then have an easy way to enforce boundary conditions in the spherical transformed domain, namely, that geodesics in both vertical directions will have their endpoints fixed at the curves we just constructed.

### A. Computing the Minimal Surface

Similar to the techniques in Section II-C, we adopt a slice by slice approach that implicitly enforces the constraints associated to the closed minimal surface problem.

The algorithm is based in the following two observations.

1. As pointed out in Section II-C, geodesics restricted to planar sections of the domain will approximate minimal surface cross sections provided their intrinsic length is small ( $g$  is properly designed). Intuitively, the minimal surface will stick to low  $g$  regions in an effort to locally minimize its intrinsic area.

2. As we are dealing with a 2-D manifold, we want geodesics in the two orthogonal directions to be in agreement with each other. Since planes in both vertical directions ( $\theta_i$  and  $\phi_j$ ) will be sectioning a single surface, geodesics in two orthogonal cutting planes should go through a common point in three dimensions (the triple intersection of the two planes and the surface), see Fig. 10.

The individual steps of the algorithm are designed to benefit from the above observations and also to satisfy the boundary conditions.

1. For each slice,  $\theta_i$  and  $\phi_j$  of the  $g$  image compute the intrinsic geodesic curve restricted to that slice joining the two fixed endpoints (obtained as the intersection of the current slice with the boundary curves on the spherical transformed domain). Assign to the current slice a cost value  $v = (\int_C g(C)dl) / (\int_C dl)$  i.e., the ratio of the intrinsic to the Euclidean length of the geodesic curve. Intuitively, this cost will be low for curves that go through regions of low  $g$  and higher otherwise.
2. Repeat until all slices in the volume have been processed.
3. Select the unprocessed slice with the lowest cost  $v$ .
4. For  $g$  restricted to that slice, compute the geodesic curve between the corresponding fixed endpoints.
5. Overwrite  $g$  at the current slice with a constant *background* value (e.g., the brightest value in the original image) except at the positions along the geodesic, see Fig. 11. By doing this, we are encouraging geodesics in the orthogonal direction to be in agreement with already computed ones.

By first processing slices with lower cost values, we are relying on areas where boundary information is strong. As we continue to process sections of increasing cost, we start gradually enforcing the restriction that geodesics in both directions should be in agreement with each other. This is done by overwriting processed volume sections with new image values that encourage geodesics in the other direction to go through the already computed positions (see Fig. 11. Sections with higher costs will increasingly see the modified values (that enforce the surface restriction) in substitution of the original image values (see Fig. 12). Although the procedure just described captures the overall location of the minimal surface, it may not guarantee (depending on the metric) smoothness of the surface everywhere. We then perform a second sweep throughout the volume (along slices in both directions in no particular order) and compute geodesic curves using the modified image values. Even though evolution with the gradient descent flow in (2) will do the job, it will be computationally inefficient (may require computation of significant number of iterations) because some portions of the produced surface approximation may be far away from their optimal location.

Observe that unlike the techniques discussed in Section II-C, the minimal surface will be accurately located (because we are not using accumulated image values across sections to compute geodesics), and the smoothness constraint is being implicitly enforced without the need to deal with bands of fixed width or other adhoc constraints.<sup>4</sup>

## B. Image Derived Metric

The image derived metric  $g$  highlights specific features of interest, e.g., edges, depending on the particular application. Its design process requires choosing the particular shape of the  $g$  function and usually the selection of a scale parameter. For the case at hand, and considering the robustness of the algorithm we have developed, we simply use  $g = I + w$ , since the envelope

<sup>4</sup>The smoothness of the resulting surface is implicitly controlled by the constant  $w$  in the metric, as described in Section II-A.



of features to be segmented is stained more darkly than the background (see Fig. 1). Therefore, the raw volumetric data can drive the segmentation algorithm directly and the minimal surface will be attracted toward the dark boundaries. Given the complexity of the images (noise level, proximity of features, missing boundaries, background clutter, fiducial markers, etc.), we found that any kind of regularization or selection of scale parameter will inevitably hinder the subsequent segmentation. This is in contrast with the metric we used in [26] that can enhance the boundaries of interest, but requires computation of derivatives on a regularized version of the image. Again, given the noisy nature of the images, we have found that any amount of regularization may discard valuable boundary information, e.g., causing neighboring boundaries to merge, weak boundaries to disappear, and other common problems of this type. The segmentation technique we reported in [26] required the use of the more sophisticated image metric cited above. Now, because of the improved robustness of the one presented here, we can use the raw image directly and get the best possible (unbiased) localization of boundaries.

#### IV. Three-Dimensional Tomogram Segmentation Results

Full tomograms are processed by first manually selecting the interior points at individual features and then automatically segmenting each structure using the technique described above. The 16-bit volumetric data of a single tomogram has a horizontal resolution of  $2048 \times 2048$  pixels in 120 vertical slices (requiring about a gigabyte of storage only for the raw data), and each tomogram contains more than a hundred individual vesicular features. The segmentation for each 3-D structure is run independently and in a serial fashion. Individual subvolumes are cropped out in regions  $256 \times 256 \times 120$  in size. The resolution of spherical transformed volumes is such that the number of voxels remains roughly the same as in the cropped region and is about  $128 \times 360 \times 180$  ( $\rho$ ,  $\theta$ , and  $\phi$ , respectively) voxels big. As an optional refinement step, we can use the surface estimate as initial condition for the gradient descent of (2) and usually very few iterations are required to achieve the steady-state solution. For all the examples in this paper, only two iterations were needed. The running time for each subvolume is less than 1 min in a 2.1-MHz laptop computer including the data cropping, back-and-forth coordinate transformation, minimal surface search and refinement with the gradient descent flow. The processing of entire tomograms takes then less than 2 h and can well be optimized or parallelized to significantly reduce computation time.

In Fig. 13, we show surface contours on top of slices of unprocessed tomogram for a representative feature of interest. The segmented surface shows an excellent fit to the boundaries of the vesicular features. Several other examples are shown in Fig. 14.

We also show that the segmentation algorithm can be used to classify the volumes in terms of the mean internal density. We show classified 3-D reconstructions in Fig. 15 and results for full tomograms in Fig. 16. Regions are automatically classified (represented with different colors) based on differing internal average grey values, which presumably represent different stages of virus assembly.

Once all relevant structures are segmented, we can also perform some simple statistical analysis on the results. As each volume is obtained as an implicit function, geometry computations (e.g., size, average gray value, shape, etc.) are easily obtained. In Fig. 17, we show histograms of the average gray level (density) distributions inside the selected volumes in two different tomograms. Note that we can clearly classify structures into two groups: the ones with a filled (dark) interior and the empty (brighter) ones. Furthermore, looking at the spatial distribution of gray values inside each of them, more sophisticated criteria can be devised to classify into the different types presented in Section I.

Fig. 17 also shows the size distribution of the segmented vesicles within each tomogram. Note that as overall thickness of the tomogram is under 150 nm, and the virion/vesicular entities are ~100 nm wide, only a few objects are captured completely inside the volume. Acquisition of serial tomograms from successive sections that are stitched together computationally should allow analysis of larger volumes, therefore providing even more reliable statistics on the segmented volumes.

## V. Discussion

In this paper, we report a new algorithm specifically designed for the semi-automated segmentation of critical features in cellular tomograms obtained by electron microscopy. This work then addresses the fundamental challenge of translating the high volume of image information contained in 3-D tomograms into useful quantitative terms.

The proposed technique has a number of advantages/improvements over existing ones.

- The geometry of the spherical transformed domain is particularly suitable for the computation of star-shaped surfaces around a point, even when boundary image information is very weak or has significant gaps.
- Using this construction, well established techniques for the efficient computation of planar geodesics can be used as building blocks to process full 3-D images by proper manipulation of volumetric image information so all geodesics correspond to sections of the same surface.
- The segmentation is directly determined by raw image values, eliminating the need of intermediate edge indicator functions  $g$ , that may be expensive to compute, sensitive to the choice of parameters, and also introduce unnecessary bias in the segmentation.
- Computational efficiency is unmatched by existing 3-D segmentation techniques allowing for the first time processing of large volumes of data.

Moreover, the combination of all these features gives a technique that allows accurate analysis of multiple tomograms as demanded by current applications in electron tomography, a task hardly achievable with the mostly manual segmentation techniques currently used in the field [27], [28].

Three-dimensional reconstructions show the robustness of the technique to the various artifacts characteristic of electron microscopy images, and are in clear agreement with features in the volumetric image data. The results we show cannot be obtained with existing techniques which either yield poorly localized contours, fail to recover the correct minima of the energy, or have demanding computational requirements. Conducting extensive testing in over two hundred individual structures, we have proven the feasibility of quantitative analysis techniques for the processing of large data sets in tractable computational time. Classification based on size or average gray value are just examples of what can be accomplished once the segmentation stage has been completed. The significance of this kind of analysis at this level of resolution may have important biological implications.

Note that the detection of interior points can be done entirely automatically, for example, by detecting singularities of the distance function to robust edges [29]. However, the structure of the tomograms is so complex (membranes merged together, boundaries with large gaps, presence of neighboring structural elements, etc.), that it will require intensive post processing to get rid of false positives and to account for false negatives. Besides, a trained user can select all particles in a tomogram relatively fast (about 5 to 10 min), which is already a significant improvement compared to fully manual processing. We should also note that, at the present

stage of the investigation, our goal is to select as many features as possible in order to minimize user bias in particle selection.

We also demonstrate that our methods are applicable for the quantitative analysis of subcellular features in HIV-infected macrophages. HIV particles in cells are generally present in a variety of maturation states. The immature form of the virus has a clear interior but a wide outer band of uncleaved Gag protein that is darkly stained, while the mature form of the virus has a dense core in the middle and a thin outer boundary corresponding to the bilayer membrane. Intermediate states where the Gag coat has been partially cleaved are also frequently detectable. In addition, infected macrophages can accumulate vesicular compartments whose composition is not well understood, that seem to have neither a visible Gag shell nor an internal core. The shapes, sizes, and textures of these different types of particles encountered in infected cells represents a complex continuum of structures, and the methods we have established here provide tools to make quantitative assessments of the relative frequencies and variations in distributions at different stages of infection. In addition to the overall definition of the types of particles, we also expect that information on their location in the cell will provide new insights into the trafficking of the virus inside the cell, since many efforts at the development of HIV vaccines are aimed at disrupting the assembly of the virus. Thus, the automated segmentation and classification of particle features could provide tools for quantization of the relative efficiencies of drugs aimed at inhibiting different steps in the viral maturation pathway.

As explained above, there is a wide spectrum of viral shapes and sizes that are present in infected cell, and, for the purposes of developing the algorithm, we have used a simplified approach to distinguish between particles that have detectable central density which would likely correspond to a mature virus, and those with a clear and an unstained interior that are likely to correspond to either immature viruses or empty vesicular compartments. As the resolution of the images improves, we expect to add other discriminators to classify the variation in viral morphology. For example, in recent unpublished work, we have shown that improved imaging conditions allow us to delineate the boundary of the internal core in a mature virus. Similarly, we are now able to clearly determine the inner and outer boundaries of the shell of the immature virus. We, therefore, expect that, as the images improve in quality, it will be possible to use more sophisticated methods to discriminate among the different types of viral states observed in the cell.

In addition to simple statistics, such as average gray value and volume of particles, we plan to carry out shape analysis and classification of the segmented regions using novel computational approaches for shape statistics being developed in the computer vision literature.



**Sriram Subramaniam** was born in Madras, India, on November 26, 1960. He received the M.Sc. degree in chemistry in 1981, the Ph.D. degree in physical chemistry from Stanford University, Stanford, CA, in 1987, and carried out postdoctoral work in the Departments of Chemistry and Biology, the Massachusetts Institute of Technology, Cambridge. In 1992, he joined the faculty at the Johns Hopkins University School of Medicine, Baltimore, MD, as an Assistant Professor, and was promoted to Associate Professor in 1998. In 1998, he joined the National Cancer Institute, National Institutes of Health, Bethesda, MD, as a Senior Investigator. The long-term mission of his research program is to develop advanced nanoscale imaging tools to discover the structural signatures that define the differences between healthy and diseased tissue, and to apply them to the effective diagnosis and treatment of diseases such as cancer and AIDS.



**Guillermo Sapiro** (SM'94) was born in Montevideo, Uruguay, on April 3, 1966. He received the B.Sc. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Department of Electrical Engineering, Israel Institute of Technology, Technion, in 1989, 1991, and 1993, respectively.

After postdoctoral research at the Massachusetts Institute of Technology, Cambridge, he became a Member of the Technical Staff at the research facilities of HP Labs, Palo Alto, CA. He is currently with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. He works on differential geometry and geometric partial differential equations, both in theory and applications in computer vision, computer graphics, medical imaging, and image analysis. He recently co-edited a special issue on image analysis in the *Journal of Visual Communication and Image Representation*. He has authored and co-authored numerous papers in image analysis and he has written a book.

Dr. Sapiro is a member of SIAM. He recently co-edited a special issue of IEEE TRANSACTIONS ON IMAGE PROCESSING. He was awarded the Gutwirth Scholarship for Special Excellence in Graduate Studies in 1991, the Ollendorff Fellowship for Excellence in Vision and Image Understanding Work in 1992, the Rothschild Fellowship for Postdoctoral Studies in 1993, the Office of Naval Research Young Investigator Award in 1998, the Presidential Early Career Award for Scientists and Engineers (PECASE) in 1998, and the National Science Foundation Career Award in 1999.



**Alberto Bartesaghi** (S'01) was born in Montevideo, Uruguay, on June 24, 1974. He received the B.Sc. and M.Sc. degrees from the Instituto de Ingeniera Electrica, Universidad de la Republica, Montevideo, in 1999 and 2001, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, in 2005.

He is currently a Postdoctoral fellow at the Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD. He works on geometric computational techniques for applications in computer graphics, computer vision, medical, and biological imaging.

#### Acknowledgements

The authors would like to thank Dr. S. Wahl and Dr. J. Orenstein for providing sections from HIV-infected macrophages, and J. Lefman and S. Lee for their assistance with the tomography.

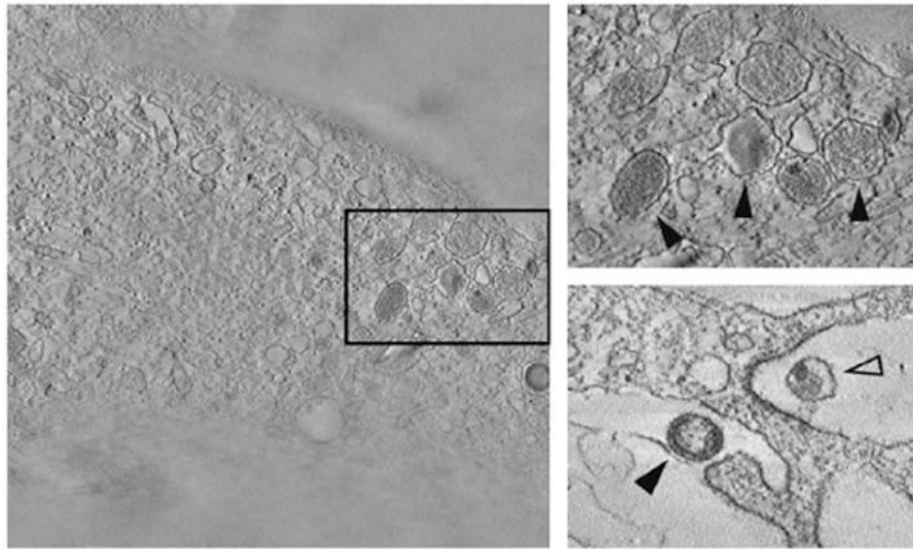
#### References

1. Subramaniam S, Milne JLS. Three-dimensional electron microscopy at molecular resolution. *Annu Rev Biophys Biomol Struct* Jun 2004;33:141–155. [PubMed: 15139808]

2. Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 1996;116(1):71–76. [PubMed: 8742726]
3. Peijun Zhang JLSM, Beatty A, Subramaniam S. Automated data collection with a Tecnai 12 electron microscope: Applications for molecular imaging by cryomicroscopy. *J Struct Biol* Sep 2001;135(3): 251–261.
4. Sapiro, G. Cambridge, U.K.: Cambridge Univ. Press; 2001. *Geometric Partial Differential Equations and Image Processing*.
5. Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J Comput Phys* 1988;79:12–49.
6. Cohen LD. On active contour models and balloons. *Comput Vis, Graph, Image Process* 1991;53(2): 211–218.
7. Siddiqi K, Lauziere Y, Tannenbaum A, Zucker S. Area and length minimizing flows for shape segmentation. *IEEE Trans Image Process* Mar 1998;7(3):433–443.
8. Proc IEEE Conf Computer Vision and Pattern Recognition. Los Alamitos, CA: Jun. 1997 “Gradient vector flow: a new external force for snakes,”; p. 66–71.
9. Boykov, Y.; Kolmogorov, V. , . “Computing geodesics and minimal surfaces via graph cuts,”; *Int Conf Computer Vision*, Nice; France. Oct. 2003 p. 26–33.
10. Hu, TC. *Integer Programming and Network Flows*. Reading, MA; Addison-Wesley: 1969.
11. Appleton, B.; Talbot, H. , . “Globally optimal surfaces by continuous maximal flows,”; *Proc Digital Image Computing: Techniques and Applications*, 7th APRS Conf; Sydney, Australia. Dec. 2003; p. 987–996.
12. Cohen LD, Kimmel R. Global minimum for active contour models: a minimal path approach. *Int J Comput Vis* Aug 1997;24(1):57–78.
13. Sethian JA. A fast marching level set method for monotonically advancing fronts. *Proc Nat Acad Sci* 1996;93(4):1591–1595. [PubMed: 11607632]
14. Tsitsiklis JN. Efficient algorithms for globally optimal trajectories. *IEEE Trans Autom Control* Sep 1995;40(9):1528–1538.
15. Helmsen J, Puckett EG, Colella P, Dorr M. Two new methods for simulating photolithography development in three dimensions. *SPIE Optical/Laser Microlithography IX* 1996;2726:253–261.
16. Appleton B, Talbot H. Globally optimal geodesic active contours. *J Math Imag Vis*. to be published.
17. Sun C, Pallottino S. Circular shortest path in images. *Pattern Recognit* Mar 2003;36(3):711–721.
18. Appleton B, Sun C. Circular shortest paths by branch and bound. *Pattern Recognit* Nov 2003;36(11): 2513–2520.
19. Farin, D.; Pfeffer, M.; de With, PHN.; Effelsberg, W. , . “Corridor scissors: A semi-automatic segmentation tool employing minimum-cost circular paths,”; *Proc Int Conf Image Processing*; 2004. p. 1177–1180.
20. Alkhalifah T, Fomel S. Implementing the fast marching eikonal solver: Spherical versus Cartesian coordinates. *Geophys Prospecting* Mar 2001;49(2):165–178.
21. Burton, BP. Dept Comp Sci. Texas A&M Univ.: College Station; 1999. “Automated 3D reconstruction of neuronal structures from serial sections,” Ph.D. dissertation.
22. Branzan-Albu, A.; Schwartz, JM.; Laurendeau, D.; Moisan, C. , . “Integrating geometric and biomechanical models of a liver tumour for cryosurgery simulation,”; *Proc IS4TM: Int Symp Surgery Simulation and Soft Tissue Modeling*, Juan-Les-Pins; France. Jun. 12–13, 2003; p. 121–131.
23. Treece GM, Prager RW, Gee AH, Berman L. Surface interpolation from sparse cross-sections using region correspondence. *IEEE Trans Med Imag* Nov 2000;19(11):1106–1114.
24. Sun C. Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. *Int J Comput Vis* May 2002;47(123):99–117.
25. “Fast constrained surface extraction by minimal paths,”. In: Faugeras, O.; Paragios, N., editors. *Proc 2nd IEEE Workshop Variational, Geometric and Level Set Methods in Computer Vision*; Nice, France. Oct. 2003; p. 233–240.
26. Bartesaghi, A.; Sapiro, G.; Lee, S.; Lefman, J.; Wahl, S.; Orenstein, J.; Subramaniam, S. , . “A new approach for 3D segmentation of cellular tomograms obtained using three-dimensional electron

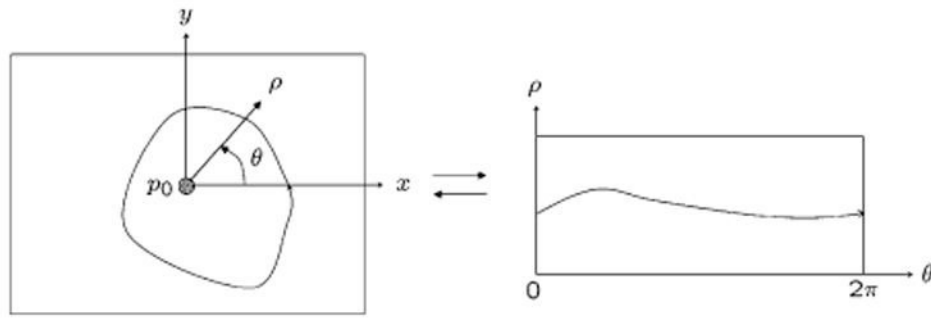


- microscopy,”; Proc IEEE Int Symp Biomedical Imaging, Special Session on PDEs in Biomedical Image Analysis; Apr. 2004; p. 5-8.
27. Lefman J, Zhang P, Hirai T, Weis R, Juliani J, Bliss D, Kessel M, Bos E, Peters P, Subramaniam S. Three-dimensional electron microscopic imaging of membrane invaginations in *Escherichia coli* overproducing the Chemotaxis Receptor Tsr. *J Bacteriol* Aug 2004;186(15):5052–5061. [PubMed: 15262942]
  28. Forster F, Medalia O, Zauberman N, Baumeister W, Pass D. Retrovirus envelope protein complex structure in situ studied by cryoelectron tomography. *Proc Natl Acad Sci* Mar 2005;102(13):4729–4734.
  29. Yu Z, Bajaj C. Picking circular and rectangular particles based on geometric feature detection in electron micrographs. *J Struct Biol* Jan 2004;145(1–2):168–180. [PubMed: 15065684]



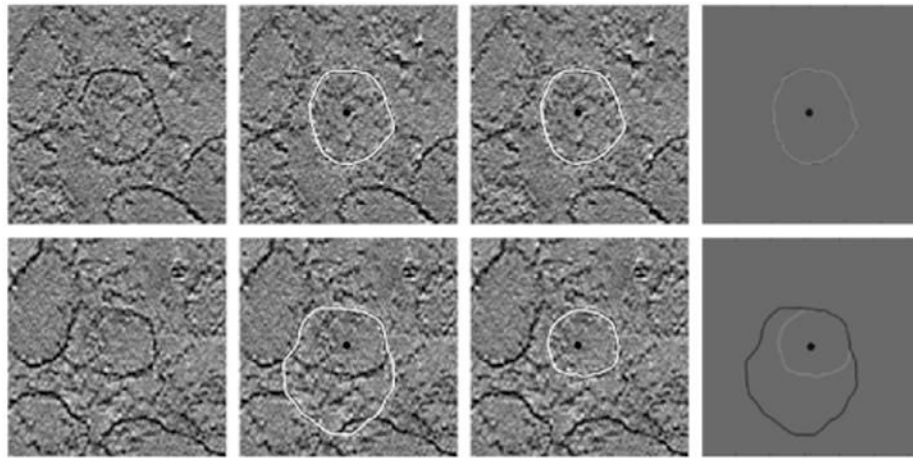
**Fig. 1.**

Tomograms recorded from fixed, plastic-embedded, and stained sections of HIV-infected macrophages. The left panel shows a 2.5- $\mu$ -wide slice from a tomogram and highlights the rich variety of structural detail in the images. The upper-right expanded view corresponds to the box indicated in the left panel. An additional view (obtained from a different tomogram) is shown in the lower right panel. The features identified in the upper right panel are most likely primary lysosomes, while those identified in the lower right panel represent budding (open arrowhead) and mature (closed arrowhead) viral particles.



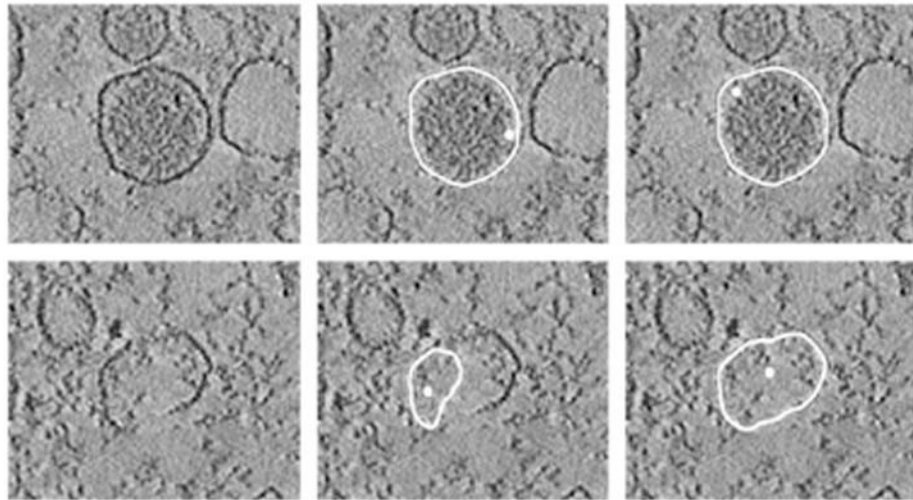
**Fig. 2.**

*Left:* Geometric construction for computing closed geodesics with an interior point as presented in [16]. A discontinuity is introduced at the positive  $x$  axis, and periodic minimal paths are computed between the two sides of the cut. *Right:* Computing closed geodesics in a polar transformed domain. The input image (left) is first converted to polar coordinates (right), where a search is done for periodic minimal paths between the sides  $\theta = 0$  and  $\theta = 2\pi$  of the polar grid. The resulting curve is transformed back to the Cartesian domain guaranteeing a closed curve.



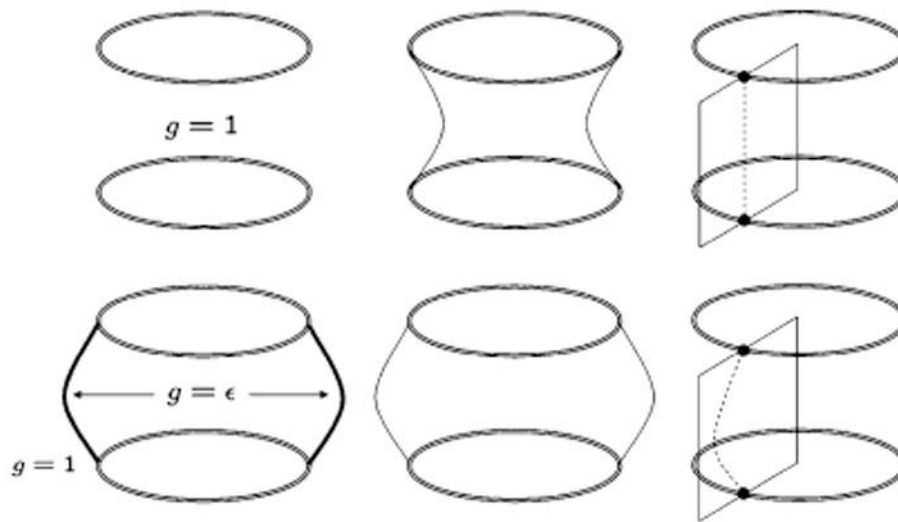
**Fig. 3.**

Computation of closed geodesics around a point in the Cartesian and polar domains. Each row shows (from left to right) the input image, contour computed in the Cartesian domain, contour computed in the polar domain, and comparison between the two. *Top*: If boundaries are strong enough, both methods give similar results. *Bottom*: When dealing with noisy or weaker boundaries, geodesics computed in the Cartesian domain begin to be poorly localized compared to those obtained in the polar domain. This gives an empirical argument in favor of using polar coordinates for the computation of closed geodesics (a rigorous quantitative analysis is not provided here and will be reported elsewhere; see also [20]).



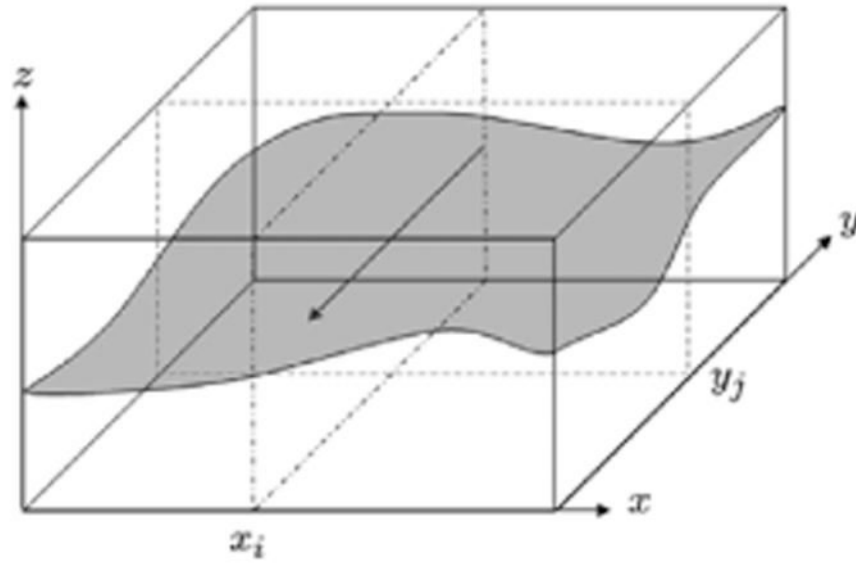
**Fig. 4.**

Effect of interior point location when computing closed contours. Each row shows the raw input image and contours computed using two different point locations. *Top*: If boundaries are strong, the segmentation is not affected by the actual position of the point as long as it falls inside the object of interest. *Bottom*: When dealing with noisy or weak boundaries, very poor localization of the interior point can introduce significant bias on the segmentation result. Users often and naturally avoid such poor localization of the point.

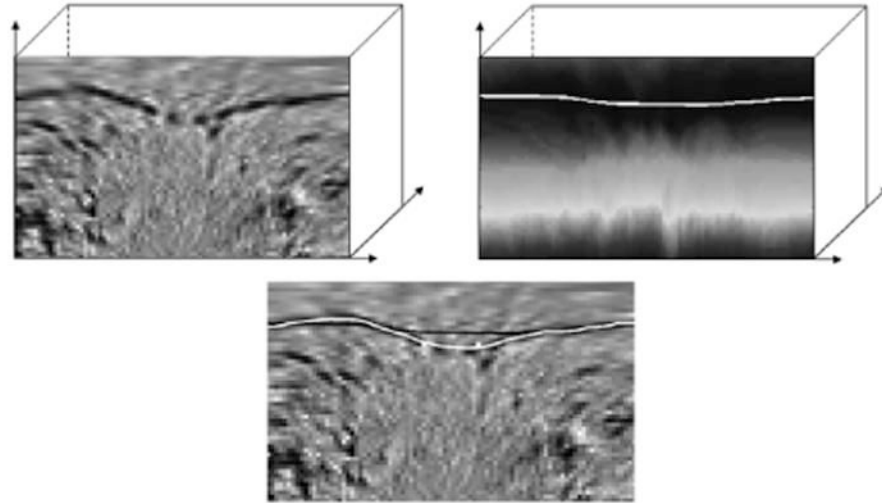
**Fig. 5.**

Geodesics in volume cross sections do not necessarily coincide with cross sections of the minimal surface. The two rings illustrate the boundary conditions. *Top*: Homogeneous metric case, (from left to right) metric function  $g$ , the catenoid surface, 2-D geodesics restricted to volume cross sections do not correspond to catenary curves. *Bottom*: Ideal heterogeneous case, (from left to right) the metric has a small  $\epsilon$  value around the dark curved wall, the intrinsic minimal surface “sticks” to low  $g$  values, planar geodesics intuitively coincide with surface cross sections.



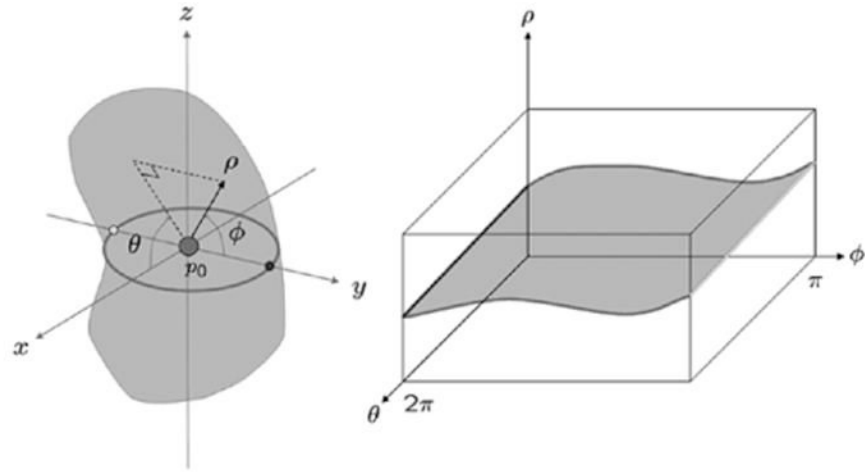
**Fig. 6.**

In the stereo correspondence problem, we search for a minimal surface that cuts horizontally through the rectangular 3-D domain. For each vertical  $x_i$  slice, image values are accumulated from the back of the volume to the front. The volume thus obtained is used as a new metric to compute geodesics now restricted to  $y_j$  sections.



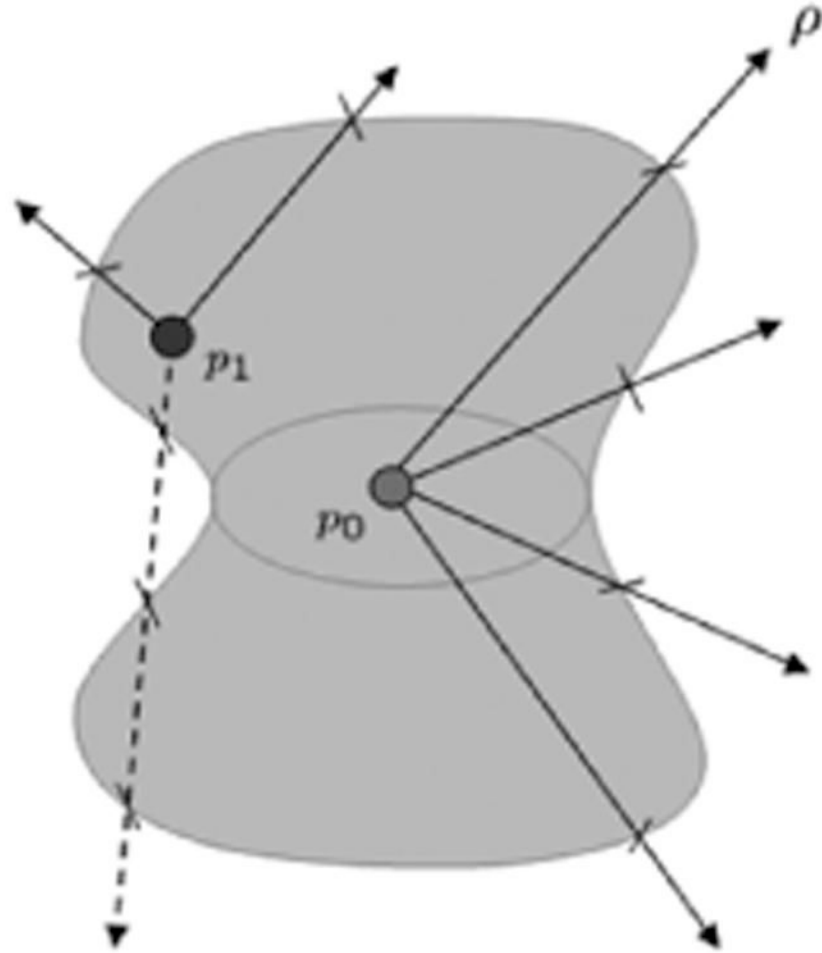
**Fig. 7.**

*Top:* Single-volume slice of raw data in the vertical direction (left), and corresponding accumulated distances in back-to-front vertical slices as computed in [24] (right). *Bottom:* Note how the geodesic curve obtained from the accumulated distances (red curve) is poorly localized compared to the one using the raw data itself (green curve).



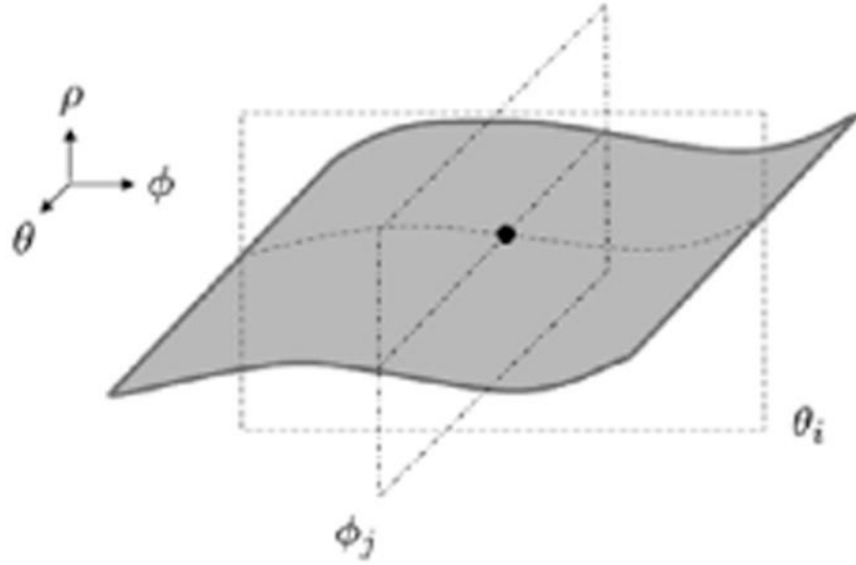
**Fig. 8.**

*Left:* A spherical coordinate system with center in  $p_0$  is used to describe the volumetric data. The spherical transformation is computed as:  $x = \rho \cos \theta \sin \phi$ ,  $y = \rho \sin \theta \sin \phi$ , and  $z = \rho \cos \phi$ . This specific configuration minimizes the distortion introduced near the poles of the spherical transform. *Right:* The transformed domain spans  $\rho \in [0, \rho_{\max}]$ ,  $\theta \in [0, 2\pi]$  and  $\phi \in [0, \pi]$  values.



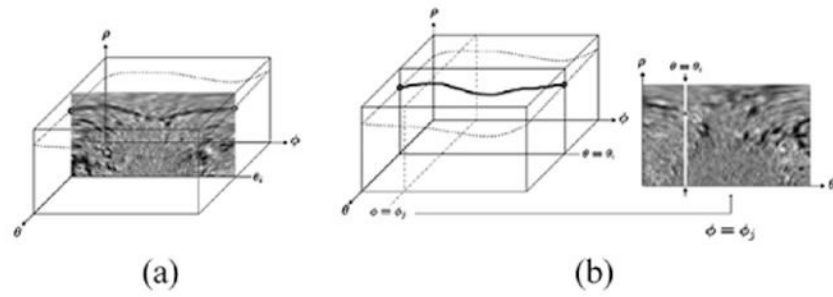
**Fig. 9.**

Class of admissible surfaces that can be obtained by the algorithm. Depending on the position of the interior point, both convex and nonconvex surfaces can be successfully segmented. The surface shown above can be segmented provided the interior point is appropriately chosen;  $p_0$  would be a suitable choice, but not  $p_1$ .



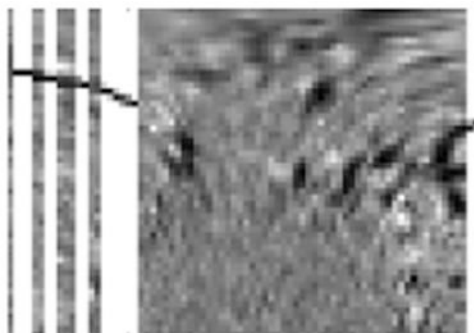
**Fig. 10.**

Forcing planar geodesics to be in agreement with each other. Assuming individual geodesic curves restricted to the  $\theta_i$  and  $\phi_j$  planes come from slicing the same surface, they must intersect with each other at points on the surface.

**Fig. 11.**

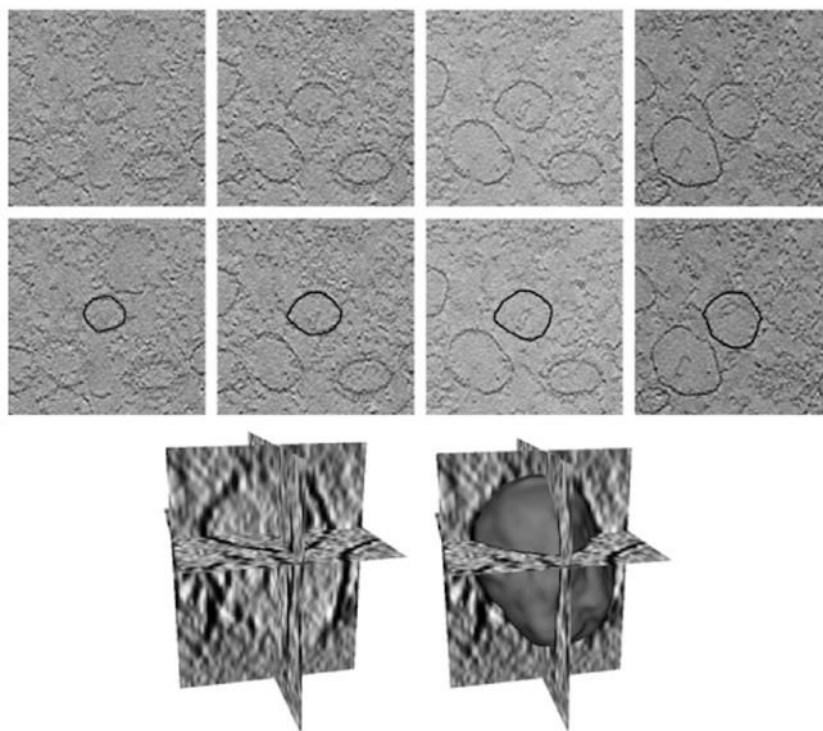
(a) Slice  $\theta_i$  is to be processed. The geodesic between boundary points (left and right dots) is computed using image values, (b) Slice  $\theta_i$  is overwritten with a constant value (white) except at positions on the geodesic curve. When we look at slices  $\phi = \phi_j$  the modified metric will be encouraging geodesics in that direction to go through the already computed surface points.





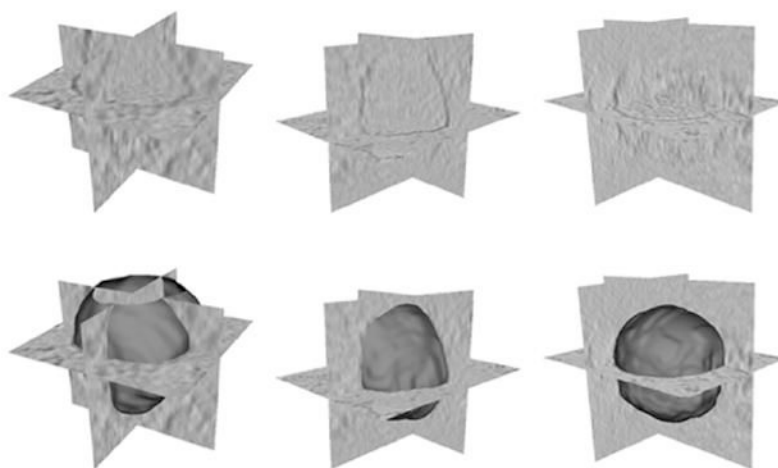
**Fig. 12.**

As the algorithm advances, more sections are being replaced with the modified metric so geodesics are encouraged to go through the already computed surface points.

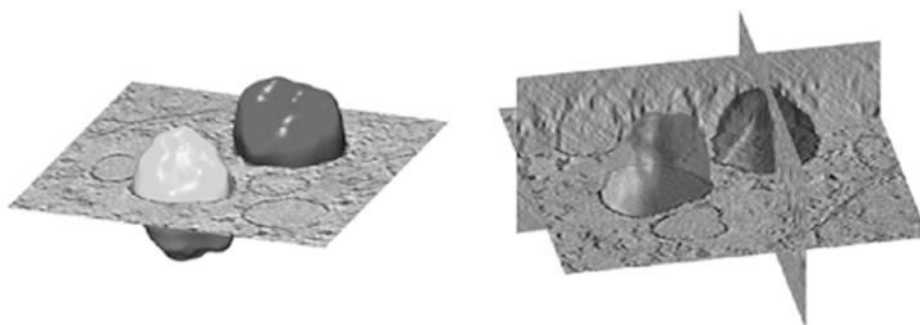


**Fig. 13.**

*Top:* Raw image slices and the corresponding surface contours. Note how even the weakest boundaries are nicely captured by the algorithm. *Bottom:* 3-D view of the extracted 3-D surface.

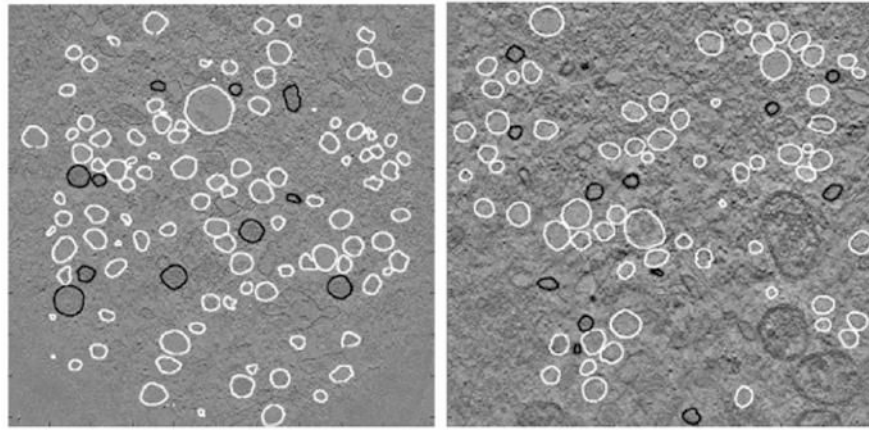


**Fig. 14.**  
Additional examples of 3-D segmented vesicular structures.



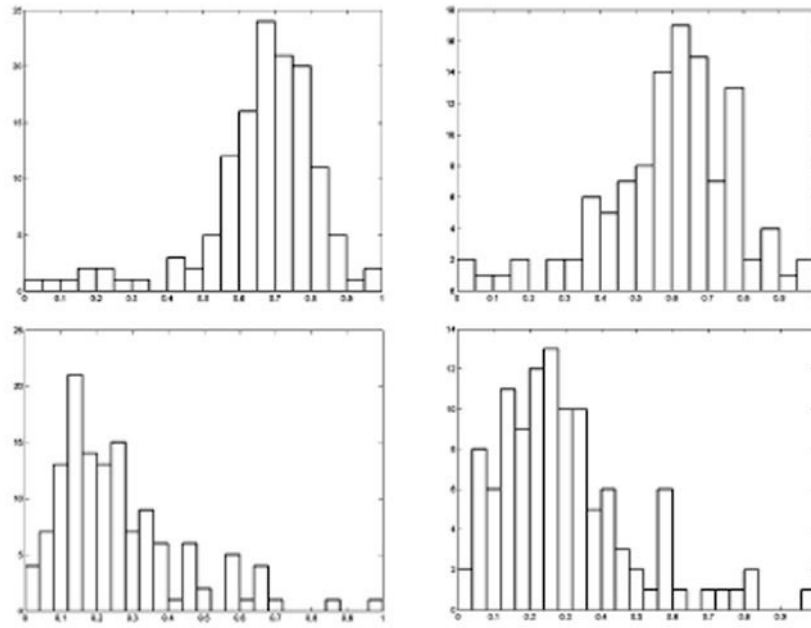
**Fig. 15.**

Two density classes are shown (in different colors) automatically classified according to the average gray level inside the volumes.



**Fig. 16.**

Classification results for full tomograms. Each tomogram contains more than a hundred individual features that can be automatically classified once the segmentation is available. We show 2-D slices of two different tomograms with the corresponding surface contours.



**Fig. 17.**

*Top:* Histogram distribution of average gray levels inside segmented regions in two different tomograms (left and right) from different regions of the infected cell. Horizontal axis shows normalized average gray values. *Bottom:* Histogram distribution of volume size in segmented regions within two tomograms (left and right) normalized with reference to the largest volume in the set. Horizontal axis shows normalized volume.