AN ENERGY-BASED THREE DIMENSIONAL SEGMENTATION APPROACH FOR THE QUANTITATIVE INTERPRETATION OF ELECTRON TOMOGRAMS

By

Alberto Bartesaghi

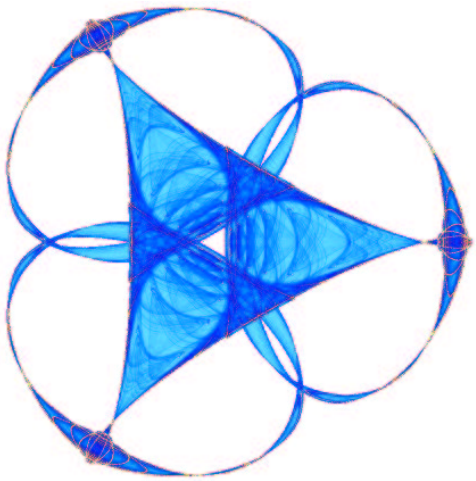
Guillermo Sapiro

and

Sriram Subramaniam

IMA Preprint Series # 2009

(December 2004)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA
514 Vincent Hall
206 Church Street S.E.
Minneapolis, Minnesota 55455–0436

Phone: 612/624-6066 Fax: 612/626-7370
URL: http://www.ima.umn.edu

An energy-based three dimensional segmentation approach for the quantitative interpretation of electron tomograms

Alberto Bartesaghi, Guillermo Sapiro, and Sriram Subramaniam

Abstract

Electron tomography allows determination of the three-dimensional structures of cells and tissues at resolutions significantly higher than 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 center the 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 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 tomographic analysis of HIV-infected cells from studies of large populations.

A. Bartesaghi and G. Sapiro are with the Electrical and Computer Engineering Department, University of Minnesota, abarte,guille@ece.umn.edu.

S. Subramaniam is with the National Cancer Institute, National Institutes of Health, subramas@mail.nih.gov.

Index Terms

Volume segmentation, energy based segmentation, distance functions, geodesics, minimal surfaces, electron tomography, high resolution, HIV.

An energy-based three dimensional segmentation approach for the quantitative interpretation of electron tomograms

I. Introduction

Transmission electron microscopes have conventionally been used in biomedical research to obtain two-dimensional projection images of thin objects such as molecules, cells and tissues. Such images can be recorded in most modern electron microscopes at magnifications ranging from $\sim$100x to $\sim$1,000,000x. The use of electron microscopes, is, however, not limited to imaging in 2D. Using emerging methods in electron tomography (see [1] for a recent review), it is now also possible to routinely determine three-dimensional 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 three-dimensional 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. 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 3D segmentation of features in cellular tomograms that can work robustly and rapidly despite the low signal-to-noise ratios. 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 ultramicrotome to produce sections with thicknesses in the range of 100nm to 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_6 filament. Tomograms were constructed using standard back-projection algorithms as implemented in the IMOD reconstruction package [2].

Figure 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 correlation of the nature and extent of these features with progression of viral infection.

In this paper 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 description at http://hrem.nci.nih.gov/html/research4.html 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 2D and 3D 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 non-convex 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 an initial surface estimate that is already

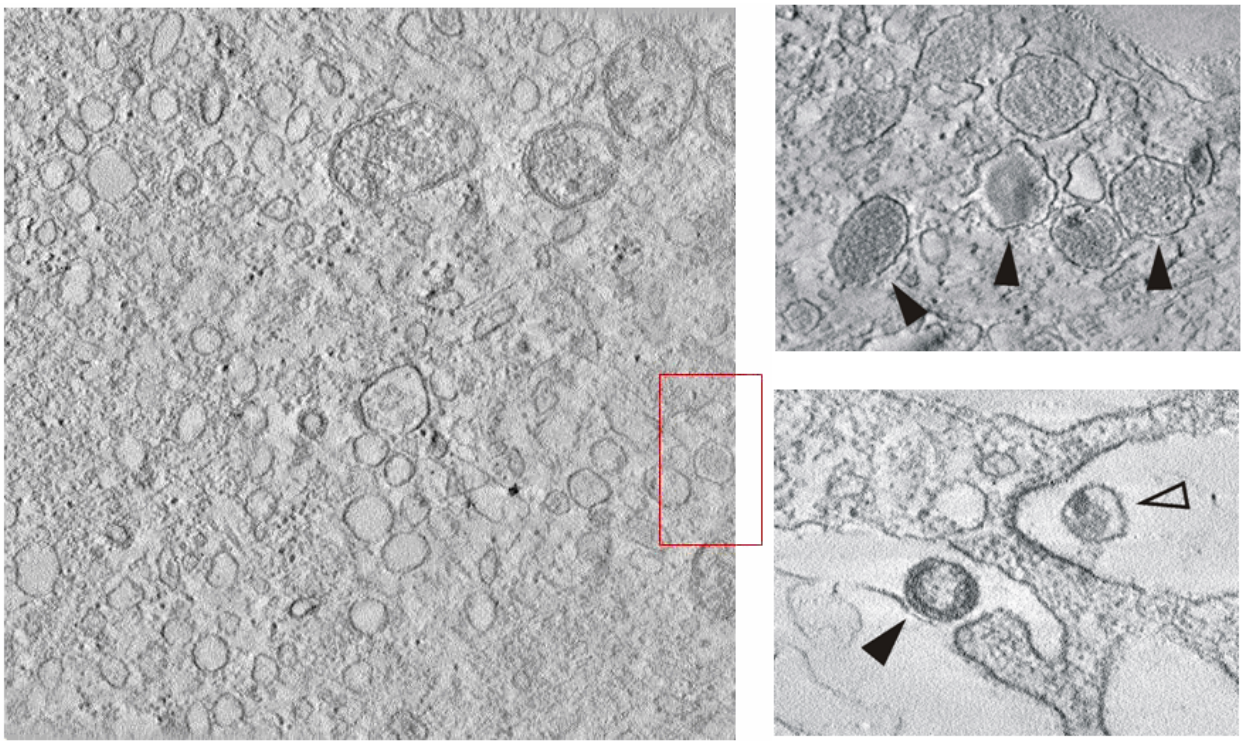


Fig. 1. Tomograms recorded from fixed, plastic-embedded and stained sections of HIV-infected macrophages. The left panel shows a 2.5 micron wide slice from the tomogram and highlights the rich variety of structural detail in the images. More expanded views from regions such as those indicated in the region boxed in red are shown in the right two panels. 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.

the segmenting boundary or sufficiently close to the correct minima so only a few steps of the gradient descent flow will be enough to achieve steady state. 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 3D volume, other cells are ignored at this early stage). For each selected point we automatically segment the surrounding structure using the proposed algorithm.[1]

[1]The detection of this single point inside the 3D 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 speed-up and reduced user intervention are extremely significant when compared with the fully manual procedures currently employed for this data.

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

II. Previous work and background on energy based segmentation

Electron tomographic images generally display very low signal-to-noise ratios, 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 the next section we review the energy based segmentation framework as presented, for example, in [3].

A. Energy based segmentation

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

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

where $g : \Re^n \rightarrow (0, \infty]$ is the image derived metric that takes small values at boundary points (thin dark regions in the case of our data) and bigger values elsewhere in the image. By minimizing the quantity in Equation (1), Γ 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

$$\Gamma_t = \left(g\,\kappa - \nabla g \cdot \vec{\mathcal{N}} \right) \vec{\mathcal{N}}$$

where $\vec{\mathcal{N}}$ is the normal to the contour and κ is the curvature ($n = 2$) or mean curvature ($n = 3$). This evolution can be implemented in the level set framework [4], by embedding the contour Γ as a level set of a higher dimensional function ϕ. The evolution for ϕ that embeds the motion of Γ is

$$\phi_t = g\,\kappa |\nabla \phi| + \nabla g \cdot \nabla \phi \tag{2}$$

The 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 Γ_0, the solution to the segmentation problem is given by the steady state of the gradient descent flow in Equation (2). Although this technique has proven to be very successful and state-of-the-art for many medical imaging disciplines (it is for example part of the ITK initiative, www.itk.com), it strongly depends on the choice of the initial contour Γ_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 towards the desired minima, *e.g.* [5]–[7]. Although they offer improved performance, they can still get trapped in unwanted minima.

Graph based algorithms have also been applied to solve the minimal surface problem [8]. 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 are derived from the continuous metric but accuracy is dependent in the neighborhood system. Although metrification artifacts can be reduced by increasing the number of directions, computational complexity grows rapidly.

Efficiently solving and adapting this energy-based segmentation framework to our data, 3D 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 [9] where the authors propose an algorithm for computing planar geodesics between two given points using a non-iterative 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 [10]–[12].

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 [13], where a discontinuity ("image cut") is introduced in the image domain and the image metric scaled to be g/ρ, where ρ is the distance to p_0, see Figure 2. 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 [9] described above, only that the periodicity constraint has to be enforced. Scaling the metric by $1/\rho$ eliminates the bias towards small circular curves around p_0.

The above construction is equivalent to polar-transformed segmentation techniques [14], [15] where the image is first transformed into a polar domain, to then search for periodic minimal paths between sides of the polar grid[2], see Figure 3. Minimal paths are then converted back to the original Cartesian grid and a closed contour is guaranteed. Although keeping the computation

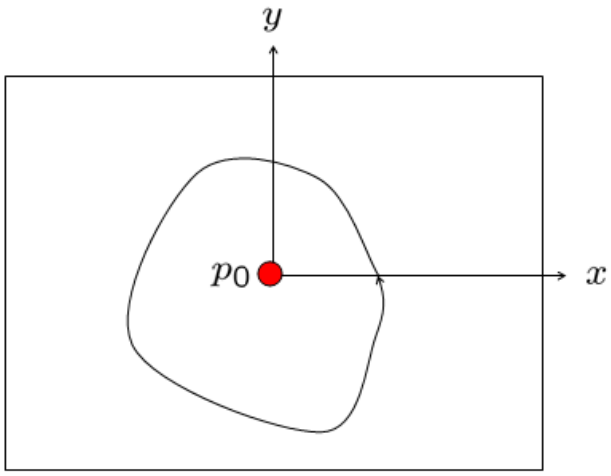


Fig. 2. Geometric construction for computing closed geodesics with an interior point as presented in [13]. A discontinuity is introduced at the positive x axis, and periodic minimal paths are computed between the two sides of the cut.

in the Cartesian grid is highly convenient, we observed that the geometry of the domain (with the discontinuity in place) introduces considerable numerical bias and can result in very poor localization of geodesics, as shown next. Consider the computation of intrinsic distances for the case $g = 1$ in both the Cartesian domain (with the cut) and the polar domain. Figure 4 shows distances computed from one side of the cut to the rest of the image, and the equivalent configuration in the polar domain. Observe the noticeable numeric distortion in the first case. The reason for this is that fast marching errors are larger (accumulate) for big values of ρ since

[2]The 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 [14]–[16] for alternative techniques.

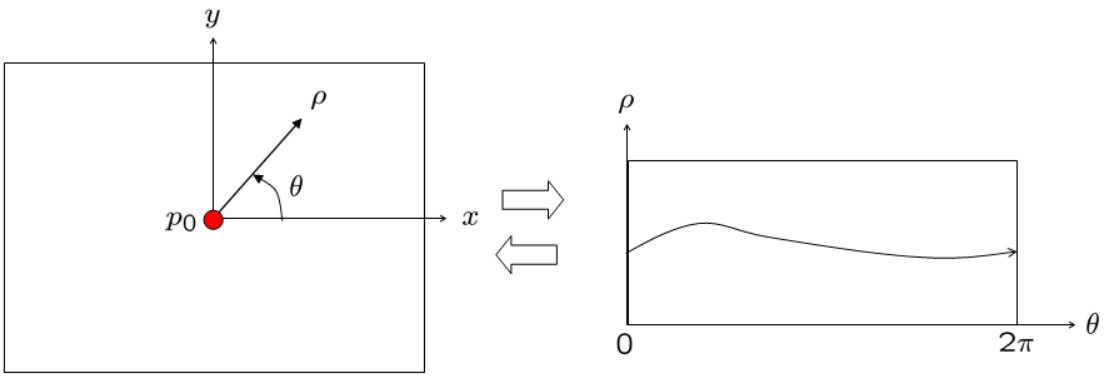


Fig. 3. Computing closed geodesics in a polar transformed domain. The image (left) is first converted to polar coordinates (right), where a search is done for periodic minimal paths between the sides $\theta = 0$ and $\theta = 2\pi$ of the polar grid. The resulting curve is transformed back to the Cartesian domain guaranteeing a closed curve.

more grid points have to be visited compared with low values of ρ. When computations are carried out in the polar domain all points share exactly the same grid geometry so distance values are identical to each other. This gives for the first time a clear justification for the use of polar coordinates for computing closed geodesics. We show the effect on some real examples in Figure 5.

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 [13].

C. Minimal surfaces from cross sections

Another group of techniques (related to our approach) solve the 3D minimal surface problem by sectioning the 3D domain with 2D planes and finding geodesics restricted to the cutting sections. If we assume that each such 2D geodesic corresponds to the intersection of the 3D minimal surface with the slicing plane, we can recover the surface as the collection of 2D 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 isotropic metric throughout the volume ($g = 1$), see Figure 6 (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

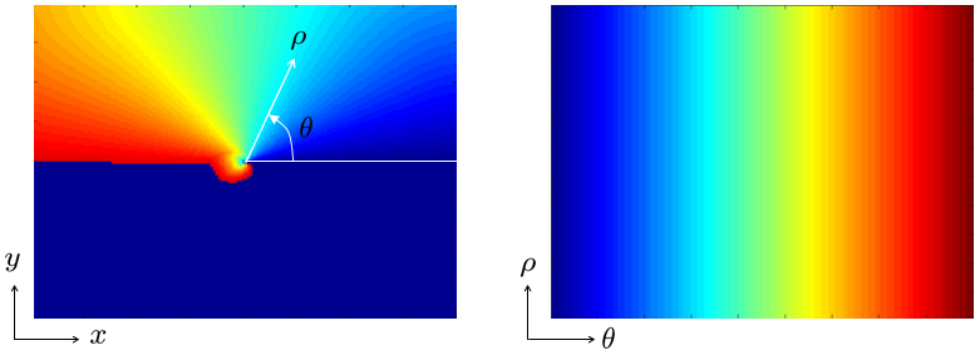


Fig. 4. Effect of numerical errors in distance computations using the Cartesian and polar transformed representations. All computations are done with fast marching using 8-neighbors [17]. *Left*: Cartesian grid with the discontinuity: we show computed distances to the upper side of the cut using uniform metric $g = 1/\rho$. The computation was stopped when the first point in the other (lower) side of the cut was reached. Color code represents increasing distance values ranging from blue (zero distance) to red. Theoretically, distance values should be constant along rays through the center of the image. Instead, observe the significant distortion due to the different number of grid points that have to be visited for small and big values of ρ. *Right*: Polar transformed domain: we show distances from the $\theta = 0$ side to the rest of the domain using the equivalent isotropic metric $g = 1$. As the geometry of the grid does not change for different values of ρ, all level sets of the distance function are vertical as expected.

is isotropic), in clear disagreement with the 2D cross sections of the catenoid (catenary curves). If the metric is ideally anisotropic, *e.g.* uniform in the background and has a small value $g = \epsilon$ on the dark curved wall, see Figure 6 (bottom), planar geodesics coincide with minimal surface cross sections. Of course, for real data obtaining ideally anisotropic metrics is not possible. For anisotropic metrics on real data 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 2D geodesics will have a better chance to coincide with cross sections of the minimal surface.

This is the implicit assumption in 3D reconstruction algorithms from planar cross-sections [18]–[20], that recover the surface as a collection of boundary curves from individual slices. 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 apart). 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 [21], [22]. The authors in [22] 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

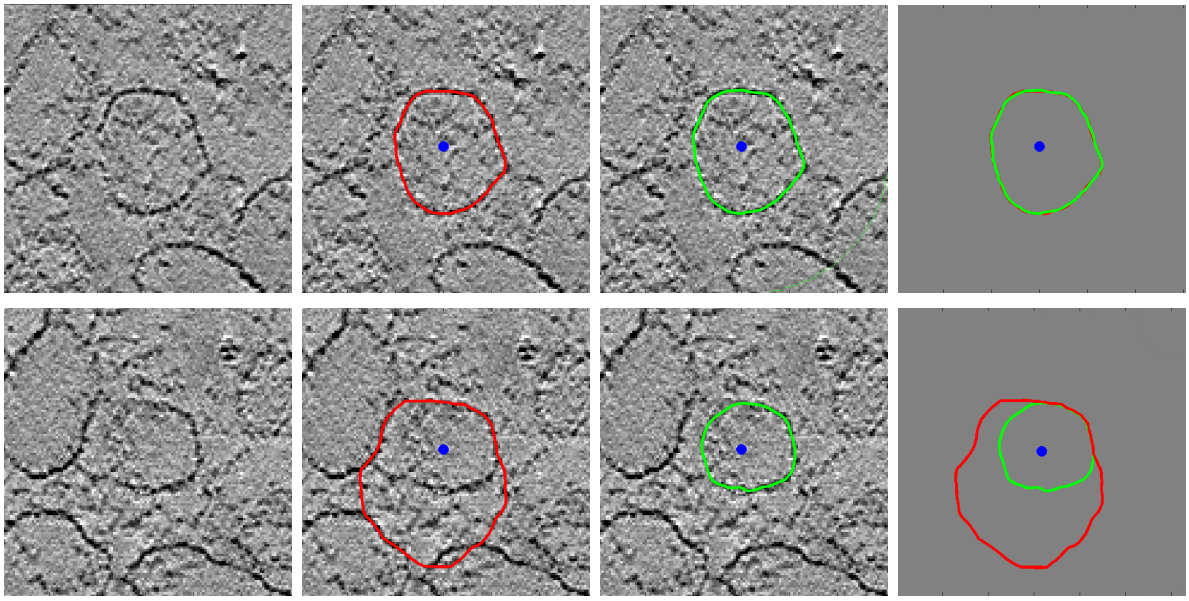


Fig. 5. Effect of numerical errors when computing closed contours in the Cartesian vs. polar domains. Each row shows (from left to right) the input image, contour computed in the Cartesian domain, contour computer in the polar domain and comparison between the two. *Top*: if boundaries are strong enough both methods give similar results. *Bottom*: when dealing with noisy or weaker boundaries, the bias of distance computations in the Cartesian domain becomes significative and poorly localized geodesics are obtained. This example and the description provided in the text bring a novel explanation of the advantage of polar coordinates when computing closed geodesics.

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 [21] for solving the stereo correspondence problem. In this case, the minimal surface cuts horizontally through a rectangular domain, see Figure 7. 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. Moreover, geodesics computed from accumulated image values are poorly localized because local image information is lost after performing the integration, see Figure 8. Additional regularity constraints must then be enforced, *e.g.* [21], 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,

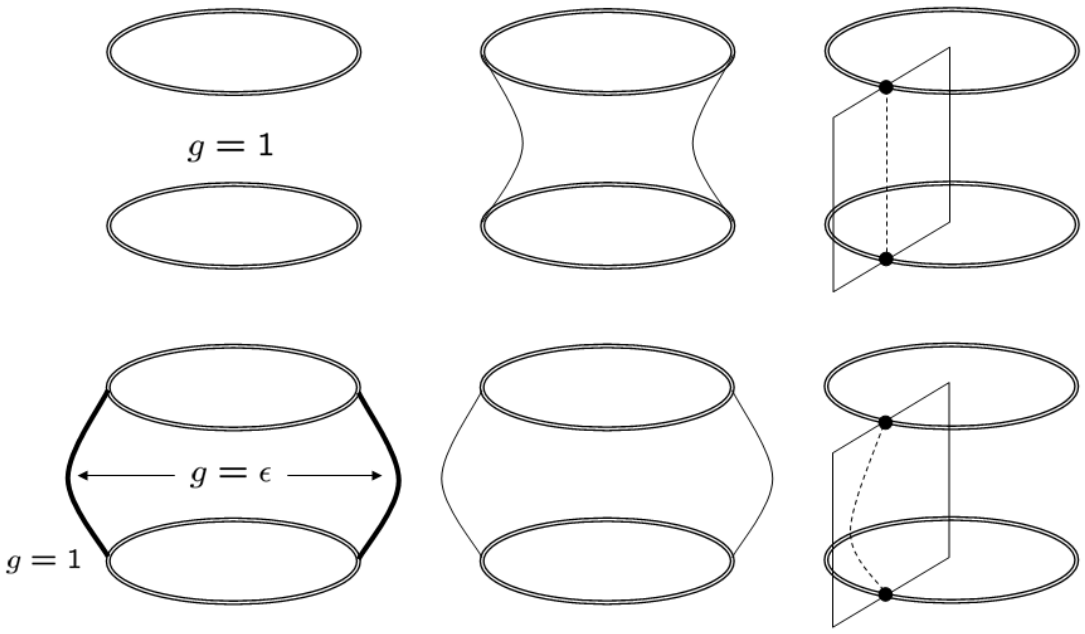


Fig. 6. Geodesics in volume cross sections do not necessarily coincide with cross sections of the minimal surface. The two rings illustrate the boundary conditions. *Top*: isotropic metric case, (from left to right) metric function g, the catenoid surface, 2D geodesics restricted to volume cross sections do not correspond to catenary curves. *Bottom:* ideal anisotropic case, (from left to right) the metric has a small ϵ value around the dark curved wall, the intrinsic minimal surface "sticks" to low g values, planar geodesics intuitively coincide with surface cross sections.

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 two dimensional case, we will apply similar methods for the segmentation of volumetric data. The equivalent formulation in three dimensions is to find a closed surface $\mathcal{S}$ that minimizes the intrinsic area given by Equation (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 anisotropic "enough" throughout the surface of interest, *i.e.* g vanishes on $\mathcal{S}$. If this is the case, cross sections of the minimal surface

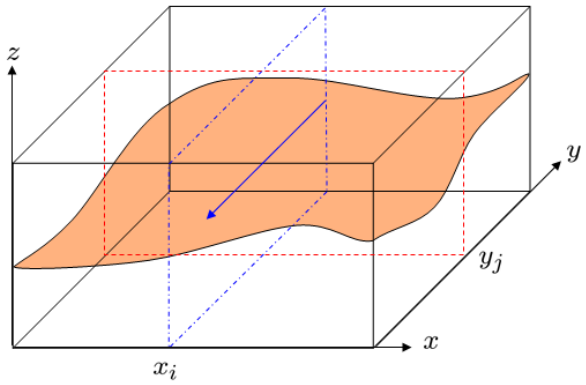


Fig. 7. In the stereo correspondence problem we search for a minimal surface that cuts horizontally through the rectangular 3D domain. For each vertical x_i slice, image values are accumulated from the back of the volume to the front. The volume thus obtained is used as a new metric to compute geodesics now restricted to y_j sections.

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 2D geodesic curves.

We then consider a spherical coordinate system centered in p_0.[3] 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 Figure 9.

Such a construction has a number of desirable features:

- The geometry of the transformed domain is suitable for the accurate computation of intrinsic distances as discussed at the end of Section II-B.
- Similar to the planar case, the spherical change of coordinates now introduces a scaling of the metric g/ρ^2 that eliminates the global minima of zero energy as g becomes unbounded at p_0. For isotropic metrics, all spheres centered in p_0 will have the same minimal energy.
- Boundary conditions in the transformed domain are straightforward to enforce, see below.
- The surface of simple closed objects can be completely covered by two pencils of parallel vertical planes.

Although such a technique cannot deal with arbitrary closed surfaces, the class of admissible surfaces is wide enough (at least for the application at hand) as discussed next. Given the

[3]This follows the previously given explanation on the advantages of using this coordinate system over Cartesian ones as further explained next.

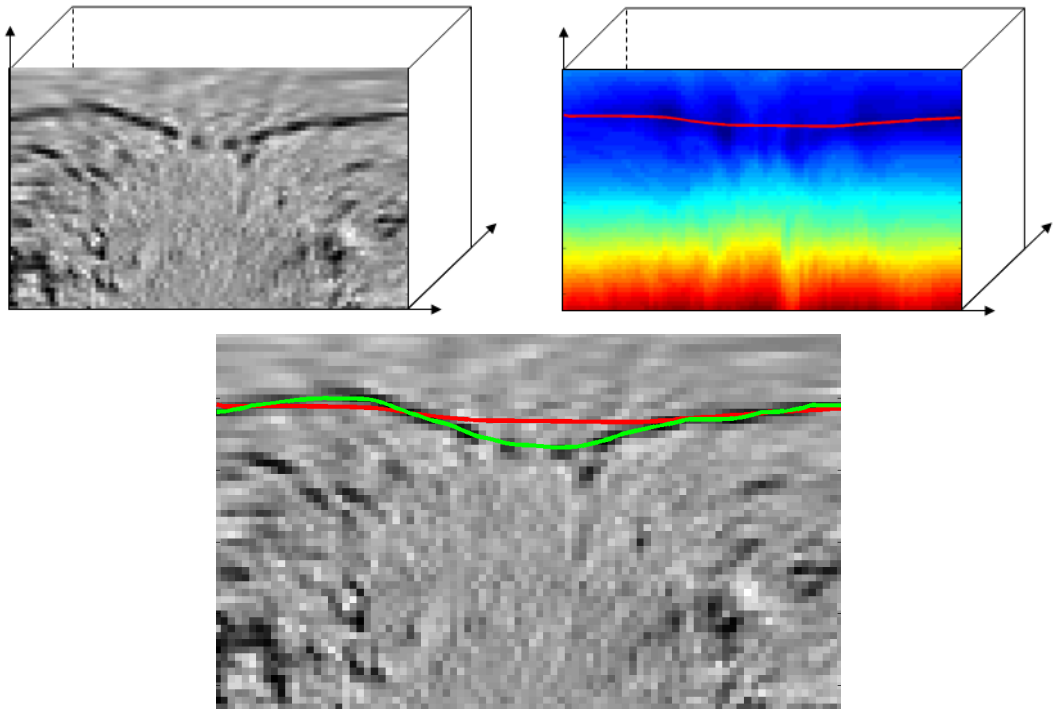


Fig. 8. *Top*: Single volume slice of our data in the vertical direction (left), and corresponding accumulated distances in back-to-front slices as computed in [21] (right). *Bottom*: Note how the geodesic curve obtained from the accumulated distances (red curve) is poorly localized compared to the one using the volume slice itself (green curve).

interior point p_0, the algorithm can compute all surfaces that accept a parametrization of the form $\rho = f(\phi, \theta)$ where ρ, ϕ and θ are the spherical coordinates centered in p_0 (as before), and f is uniquely defined for all pairs (ϕ, θ). These are so called star-shaped surfaces. Note that this includes convex and non-convex surfaces provided p_0 is appropriately chosen, see Figure 10. 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 3D 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 and we only require it to be inside the structure of interest. We discuss this issue further in Section V.

The 3D imaging system at hand has 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

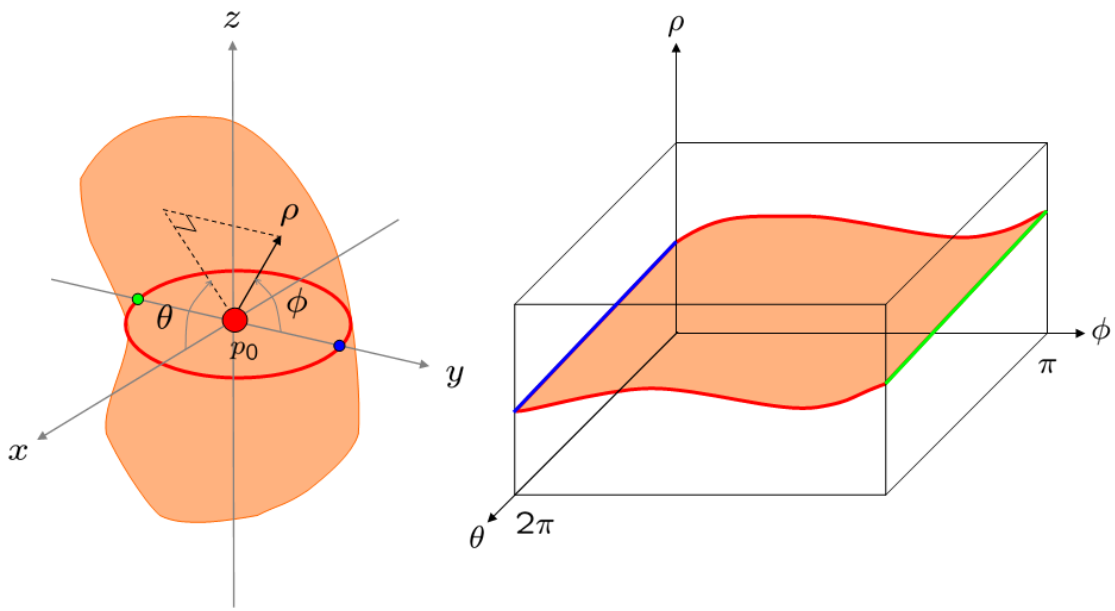


Fig. 9. *Left*: A spherical coordinate system with center in p_0 is used to describe the volumetric data. *Right*: The transformed domain spans $\rho \in [0, \rho_{max}]$, $\theta \in [0, 2\pi]$ and $\phi \in [0, \pi]$ values.

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 2D 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 (represented in red in Figure 9). The crossings of the curve with the y axis will map through the spherical transformation into horizontal straight lines at the planes $\phi = 0$ and $\phi = \pi$ (shown in blue and green respectively in Figure 9). If we split the red curve (at the crossings with the y axis) in two portions, one will correspond to the $\theta = 0$ plane and the other to the $\theta = \pi$ plane. And as we are dealing with closed surfaces, we must force the curve at $\theta = 2\pi$ to be the same as the one in $\theta = 0$ 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.

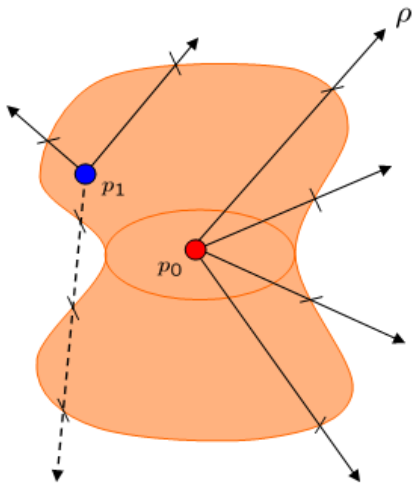


Fig. 10. Class of admissible surfaces that can be obtained by the algorithm. Depending on the position of the interior point, both convex and non-convex surfaces can be successfully segmented. The surface shown above can be segmented provided the interior point is appropriately chosen; p_0 would be a right choice but not p_1.

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-dimensional manifold, we want geodesics in the two orthogonal directions to be in agreement with each other. Since planes in both vertical directions (θ_i and ϕ_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 Figure 11.

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 θ_i and ϕ_j of the g image compute the intrinsic geodesic curve restricted to that slice joining the two fix endpoints (obtained as the intersection of the current slice

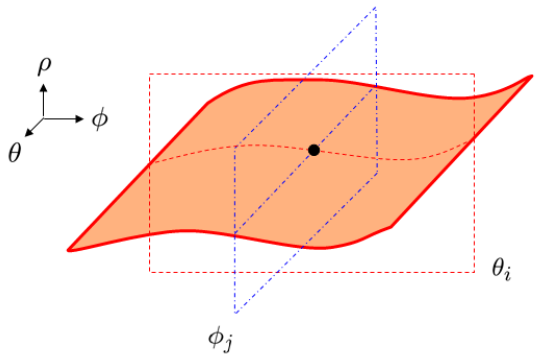


Fig. 11. Forcing planar geodesics to be in agreement with each other. Assuming individual geodesic curves restricted to the θ_i and ϕ_j planes come from slicing the same surface, they must intersect with each other at points on the surface.

with the boundary curves on the spherical transformed domain). Assign to the current slice a cost value $\nu = \frac{\int_{\mathcal{C}} g(\mathcal{C}) dl}{\int_{\mathcal{C}} dl}$, *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:
   a) Select the unprocessed slice with the lowest cost ν.
   b) For g restricted to that slice, compute the geodesic curve between the corresponding fixed endpoints.
   c) Overwrite g at the current slice with a constant *background* value except at the positions along the geodesic, see Figure 12. 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 the metric is strongly anisotropic. 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 a new metric that encourages geodesics in the other direction to go through the already computed positions, see Figure 12. Sections with higher costs will increasingly see the modified metric (that enforce the surface restriction) in substitution of the original image values, see Figure 13. Although the procedure just described captures the location of the minimal surface, it may not guarantee (depending

on the metric) smoothness of the surface everywhere. We then need to do a second sweep across slices in the volume (the order is not important now) computing geodesic curves with the modified metric resulting from the procedure above (which contains the information on the location of the minimal surface). Even though evolution with the gradient descent flow Equation 2 will do the job, it is computationally less efficient than performing this second sweep.

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.[4].

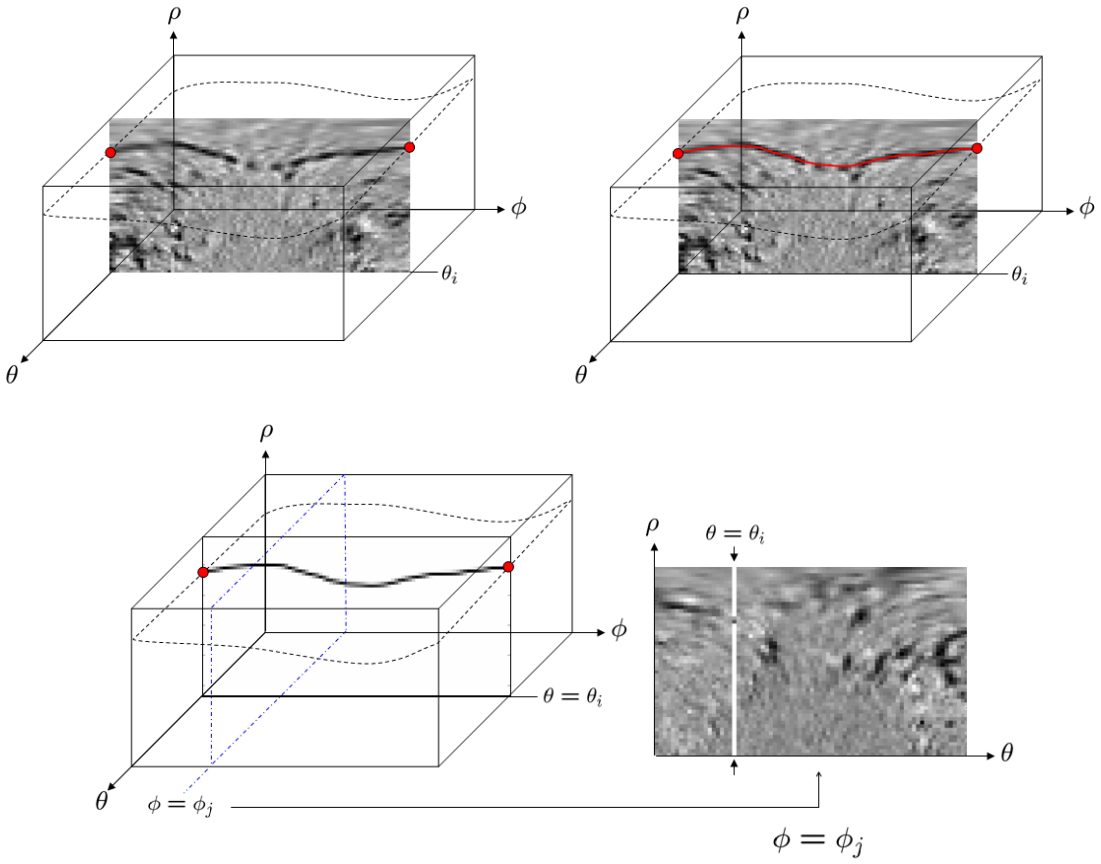


Fig. 12. *Top*: Slice θ_i is to be processed. The geodesic between boundary points (red dots) is computed using image values. *Bottom*: Slice θ_i is overwritten with a constant value (white) except at positions on the geodesic curve. When we look at slices $\phi = \phi_j$ the modified metric will be encouraging geodesics in that direction to go through the already computed surface points.

[4]The smoothness of the resulting surface is implicitly controlled by the constant w in the metric, as described in Section II-A.

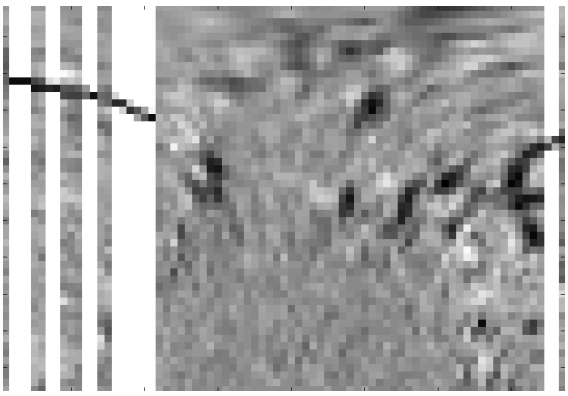


Fig. 13. As the algorithm advances, more sections are being replaced with the modified metric so geodesics are encouraged to go through the already computed surface points.

B. The 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 as described above, we simply use $g = I + w$ since the envelope of features to be segmented is stained more darkly than the background, see Figure 1. Therefore, the raw volumetric data can drive the segmentation algorithm directly and the minimal surface will be attracted towards 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 [23] 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 [23] required the use of the more sophisticated image metric referred 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. 3D 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 2048x2048 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 3D structure is run independently and in a serial fashion. The running time for each of them is less than 20 seconds in a 1.2 Mhz laptop computer. The processing of the whole volume can well be parallelized to significantly reduce the computation time when processing full tomograms. As an optional refinement step, we can use the surface estimate as initial condition for the gradient descent Equation (2) and usually very few iterations are required to achieve convergence. For all the examples in this paper only two iterations were needed.

A. Results

In Figure 14 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 show in Figure 15.

We also show that the segmentation algorithm can be used to classify the volumes in terms of the mean internal density, we show classified 3D reconstructions in Figure 16 and results for full tomograms in Figure 17. Regions are automatically classified (red and green) 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 Figure 18 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 the filled interior and the empty ones. Furthermore, looking at the spatial distribution of gray values inside each of them, more sophisticated criteria can be devised to classify in the different types presented in Section I.

Figure 18 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 $\sim$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 have reported 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 3D tomograms into useful quantitative terms.

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 that are either poorly localized or fail to recover the correct minima of the energy.

Although the detection of interior points can be done automatically (*e.g.*, from singularities of the distance function to robust edges [24]), the structure of the tomograms is so complex (membranes merged together, boundaries with large gaps, presence of secondary structural elements, etc.), that it will require intensive post-processing to eliminate irrelevant objects that are not meaningful for the subsequent analysis. Results in this direction will be reported elsewhere. 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. The segmentation result is fairly independent of the specific position of the interior point as long as it is inside the feature of interest. The degree of robustness depends largely on the strength of image boundaries, and weaker edges will exhibit more variability as the minima of the energy will not be as strong.

We also demonstrate that our methods are applicable for the quantitative analysis of subcellular features in HIV-infected macrophages. We are currently acquiring new data to study statistical aspects of viral progression at different stages of infection. In addition to simple statistics such as average gray value (representing the density inside the region) and volume, 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.

ACKNOWLEDGMENT

We thank Drs. Sharon Wahl and J. Orenstein for providing sections from HIV-infected macrophages, and Jonathan Lefman and Stanton Lee for assistance with the tomography. This work is partially supported by the Office of Naval Research, the National Science Foundation, the National Institutes of Health, and the National Geospatial-Intelligence Agency.

REFERENCES

[1] S. Subramaniam and J. L. S. Milne, "Three-dimensional electron microscopy at molecular resolution," *Ann. Rev. Biophys. and Biomol. Struct.*, vol. 33, June 2004, in press.

[2] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, "Computer visualization of three-dimensional image data using IMOD," *Journal Of Structural Biology*, , no. 116, pp. 71–76, 1996.

[3] G. Sapiro, *Geometric Partial Differential Equations and Image Processing*, Cambridge University Press, New York, 2001.

[4] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[5] L. D. Cohen, "On active contour models and balloons," *Computer Vision, Graphics and Image Processing.*, vol. 53, no. 2, pp. 211–218, 1991.

[6] K. Siddiqi, Y.B. Lauziere, A. Tannenbaum, and S.W. Zucker, "Area and length minimizing flows for shape segmentation," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 433–443, March 1998.

[7] C. Xu and J.L. Prince, "Gradient vector flow: A new external force for snakes," *Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR)*, vol. Los Alamitos: Comp. Soc. Press, pp. 66–71, June 1997.

[8] Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," in *International Conference On Computer Vision*, Nice, France, October 2003.

[9] L. D. Cohen and R. Kimmel, "Global minimum for active contour models: A minimal path approach," *IJCV*, vol. 24, no. 1, pp. 57–78, August 1997.

[10] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proc. Nat. Acad. Sci.*, vol. 93, no. 4, pp. 1591–1595, 1996.

[11] J. N. Tsitsiklis, "Efficient algorithms for globally optimal trajectories," *IEEE Transactions on Automatic Control*, , no. 40, pp. 1528–1538, 1995.

[12] J. Helmsen, E. G. Puckett, P. Colella, and M. Dorr, "Two new methods for simulating photolithography development in three dimensions," *SPIE Optical/Laser Microlithography IX*, , no. 2726, pp. 253–261, 1996.

[13] B. Appleton and H. Talbot, "Globally optimal geodesic active contours," *Journal of Mathematical Imaging and Vision, submitted*, 2002.

[14] C. Sun and S. Pallottino, "Circular shortest path in images," *Pattern Recognition*, vol. 36, no. 3, pp. 711–721, March 2003.

[15] B. Appleton and C. Sun, "Circular shortest paths by branch and bound," *Pattern Recognition*, vol. 36, no. 11, pp. 2513–2520, November 2003.

[16] Dirk Farin, Magnus Pfeffer, Peter H. N. de With, and Wolfgang Effelsberg, "Corrisor scissors: A semi-automatic segmentation tool employing minimum-cost circular paths," in *International Conference on Image Processing (ICIP)*, 2004.

[17] Q. Lin, *Enhancement, Extraction, and Visualization of 3D Volume Data*, Dissertations no. 824, isbn 91-7373-657-0, Linkping University, Sweden, 2002.

[18] B. P. Burton, *Automated 3D Reconstruction Of Neuronal Structures From Serial Sections*, Ph.D. thesis, Texas University, 1999.

[19] A. Branzan-Albu, J. M. Schwartz, D. Laurendeau, and C. Moisan, "Integrating geometric and biomechanical models of a liver tumour for cryosurgery simulation," in *IS4TM : International Symposium on Surgery Simulation and Soft Tissue Modelling*, Juan-Les-Pins, France, June 12-13 2003.

[20] G. M. Treece, R. W. Prager, A. H. Gee, and L. Berman, "Surface interpolation from sparse cross-sections using region correspondence," *Transactions on Medical Imaging*, vol. 19, no. 11, pp. 1106–1114, November 2000.

[21] C. Sun, "Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques," *International Journal on Computer Vision*, vol. 47, no. 1/2/3, pp. 99–117, May 2002.

[22] R. Ardon and L. D. Cohen, "Fast constrained surface extraction by minimal paths," in *2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, Olivier Faugeras and Nikos Paragios, Eds., Acropolis, Nice, France, October 2003, pp. 233–240.

[23] A. Bartesaghi, G. Sapiro, S. Lee, J. Lefman, S. Wahl, J. Orenstein, and S. Subramaniam, "A new approach for 3d segmentation of cellular tomograms obtained using three-dimensional electron microscopy," in *IEEE International Symposium on Biomedical Imaging, Special Session on PDEs in biomedical image analysis*, April 2004.

[24] Z. Yu and C. Bajaj, "Picking circular and rectangular particles based on geometric feature detection in electron micrographs," *Journal of Structural Biology*, vol. 145, no. 1-2, pp. 168–180, January 2004.

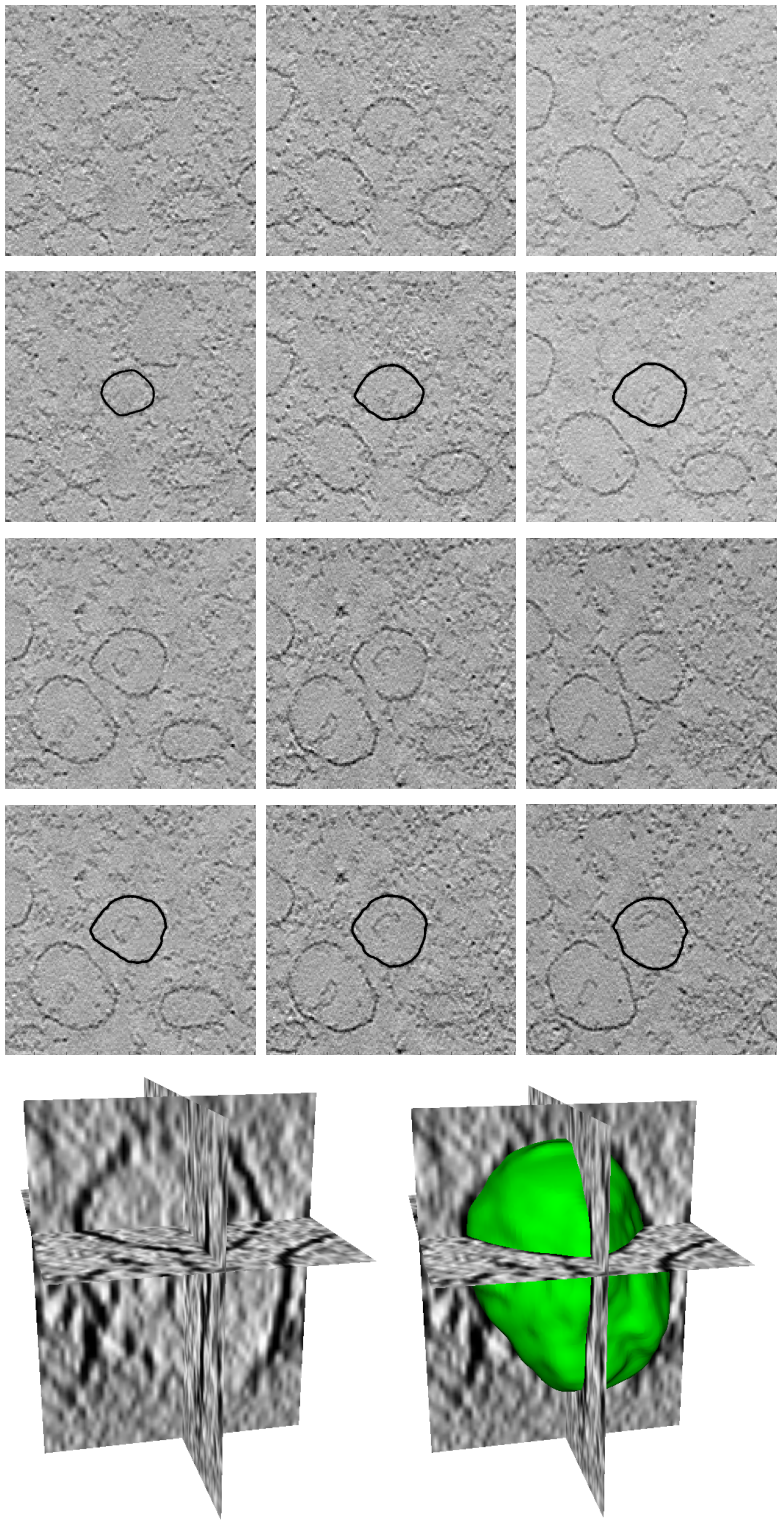


Fig. 14. *Top*: Raw image slices and the corresponding surface contours. Note how even the weakest boundaries are nicely captured by the algorithm. *Bottom*: Three-dimensional view of the extracted 3D surface.

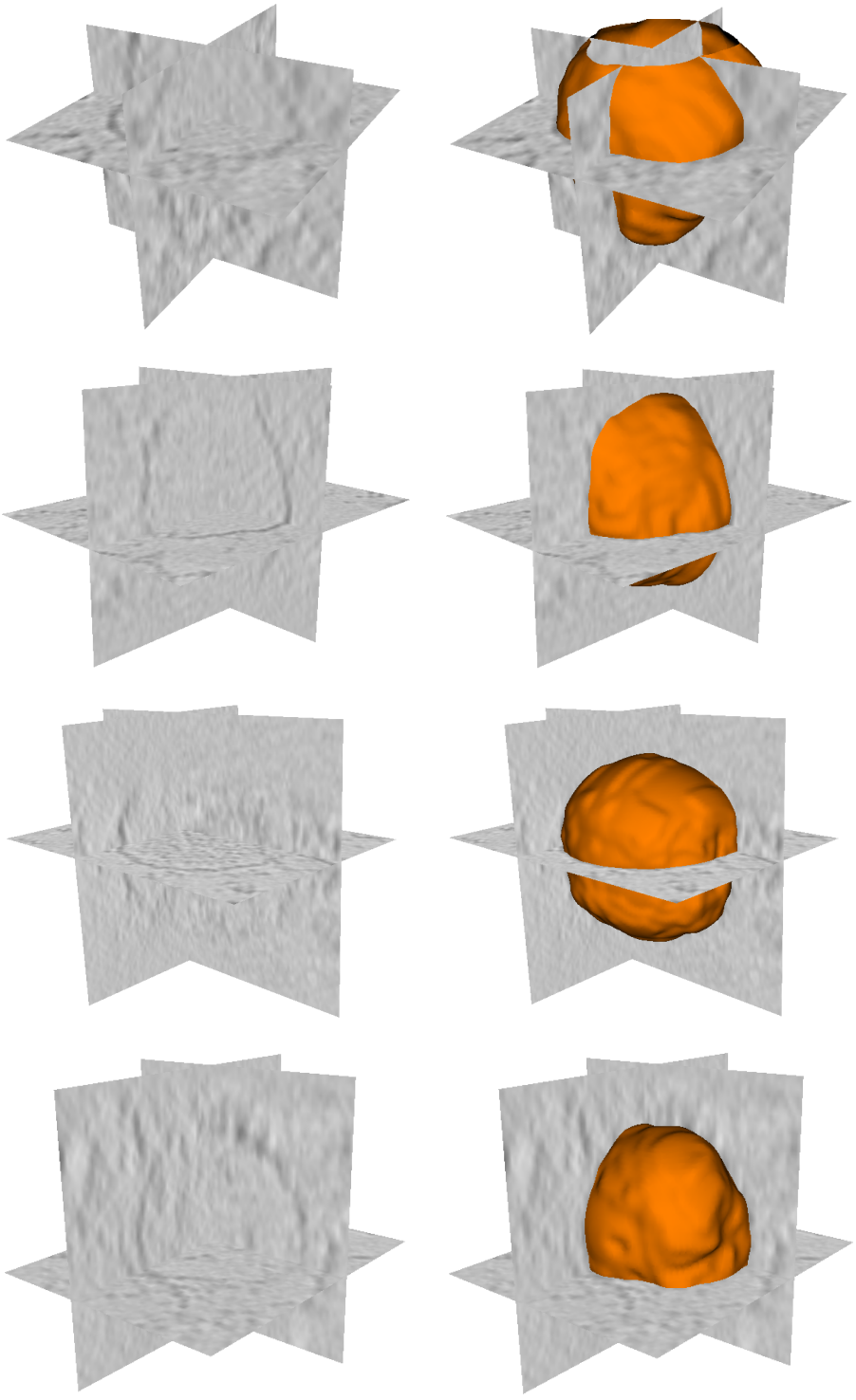


Fig. 15. Additional examples of 3D segmented vesicular structures.

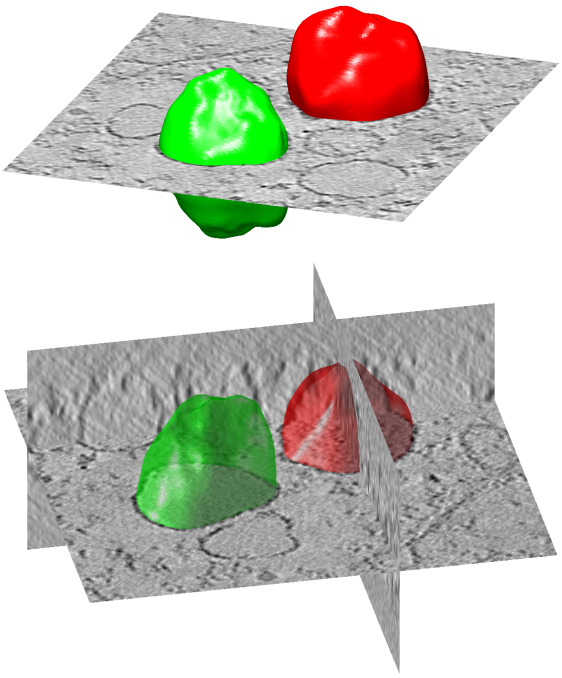


Fig. 16. Two density classes are shown (red and green) automatically classified according to the average gray level inside the volumes.

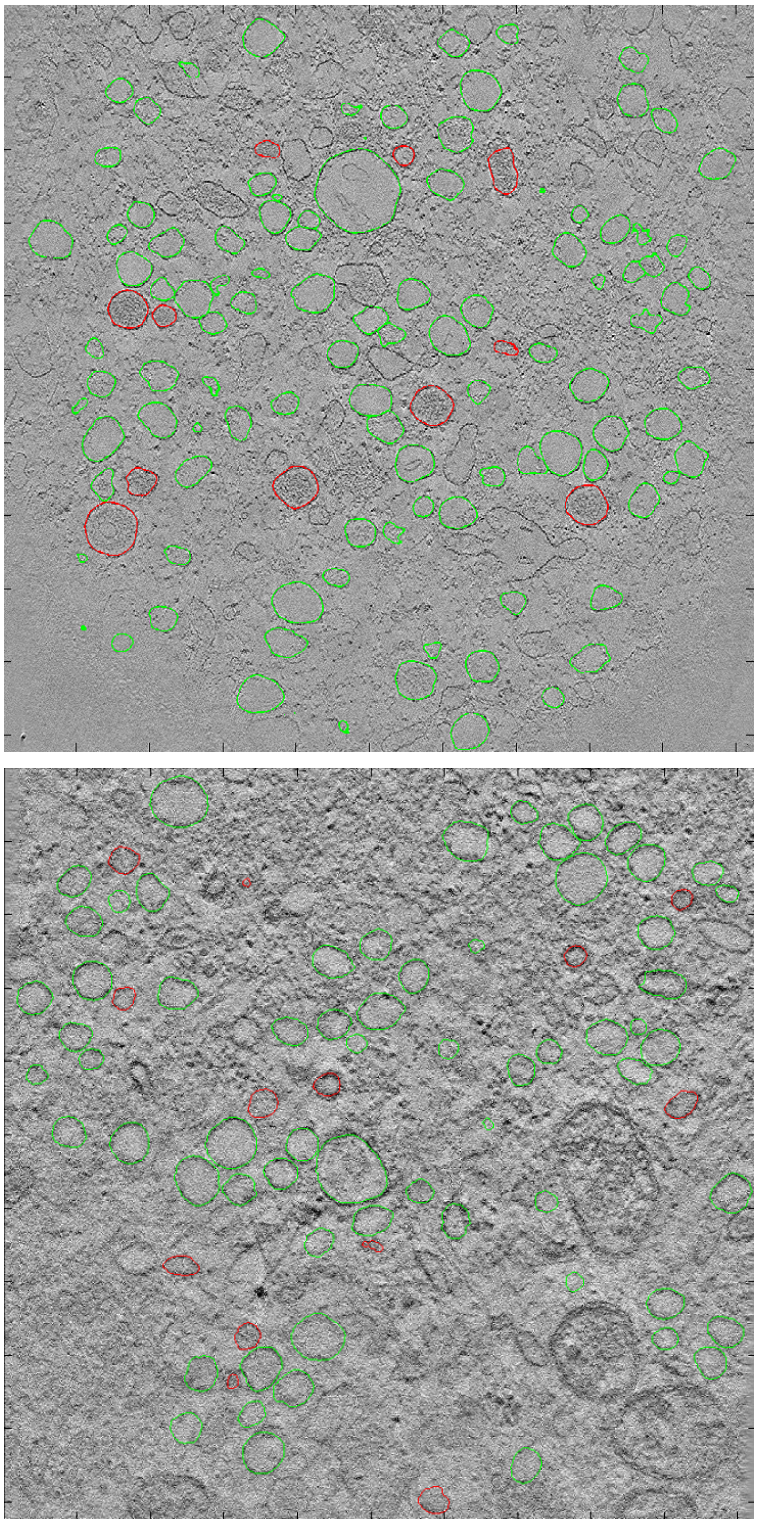


Fig. 17. Classification results for full tomograms. Each tomogram contains more than a hundred individual features that can be automatically classified once the segmentation is available. We show 2D slices of two different tomograms with the corresponding surface contours.

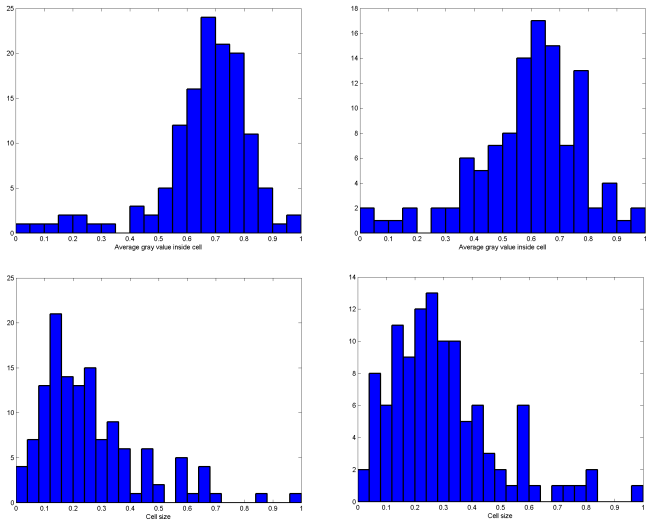


Fig. 18. *Top*: Distribution of average gray levels inside segmented regions in two different tomograms from different regions of the infected cell. *Bottom*: Distribution of size in segmented volumes within a tomogram, normalized with reference to the largest volume in the set.