

Towards automatic visualization of MET biomedical data

Lennart Svensson, Stina Svensson, Ingela Nyström and Ida-Maria Sintorn
Centre for Image Analysis, Uppsala
Email: {lennart,stina,ingela,ida.sintorn}@cb.uu.se

Abstract—This work describes a starting point for estimation of parameters when visualizing Molecular Electron Tomography (MET) volumes. The aim is to instantly and automatically produce an initial rendering of the biological molecules in a MET volume. A one-dimensional transfer function with fixed primitives is suggested to highlight the structure and interaction of the molecules. The correlation between manually set parameters by an expert user for these primitives and measures related to topology, gradient and curvature are explored.

The work is part of the ProViz project aiming at creating better tools for analysis and hapto-visualization of MET volumes of proteins.

I. INTRODUCTION

The proteins are the molecular work horses in any living organism and thus are of interest to study for medical and biological purposes. A protein's structure and flexibility is closely connected to its function, i.e., what other molecules it connects to and how they interact. Usually this is studied using X-ray crystallography, but with Molecular Electron Tomography (MET) [1] it is possible to analyse buffer (in vitro) or tissue (in situ) samples directly, see [2] for a recent example. The output of MET is a 3D scalar field, where scalar densities correspond to the electron opacity of the material. In Figure 1 an MET volume of RNA polymerase II (PolII) macromolecules in buffer is shown.

However, MET volumes are difficult to interpret for a number of reasons: the resolution in a MET volume is relatively low as each protein is represented by a small number of voxels; the contrast is low as electron irradiation destroys biological samples, which means that the total dose used to acquire the MET data must be kept low; the MET volume suffers from missing data due to the acquisition technique; and the grey-levels in the MET volume are relative and not absolute.

For the untrained eye, the volumes often seem to only contain a large number of blobs of varying sizes and shapes. The focus in this paper is to investigate how a transfer function can automatically be generated to produce a visualization that captures the main information in the MET volume. Transfer function generation has been studied in more general settings, see [3] for a good overview. Since then, multi-dimensional transfer functions based on curvature have been introduced [4]. Transfer functions specified as the sum of Gaussians were presented in [5]. Furthermore, schemes based on topology differentiation have been suggested, [6], [7].

We have chosen to address transfer function generation for MET volumes employing a one-dimensional (1D) transfer function approach. Multi-dimensional transfer functions for scalar fields rely on characterizing each point using the local or extended neighbourhood, and combining the information in some way as input to the multi-dimensional function. We have sought to remap each voxel value individually to minimize alteration of the original data.

To generate the transfer function, different measures on the scalar field are explored and correlated to the densities. The different measures investigated are: the density histogram, the number of connected components with a certain size, the mean gradient and the mean curvature. The approach is similar to the contour spectrums [8], but using different measures. It is also taken further by analysing how to set a transfer function from the measures.

II. APPROACH

A MET volume can be described as a function f , not considering the limitations of digital number representation, while a 1D transfer function corresponds to g .

$$f : X \rightarrow Y, X \subset \mathbb{N}^3, Y \subset \mathbb{R} \quad (1)$$

$$g : Y \rightarrow Z, Z \subset \mathbb{R}^4 \quad (2)$$

The four output components of the transfer function are the three color channels, RGB, and the opacity. Note that transfer functions in visualization are different to the classical definition in signal theory, where it describes the response of a linear time-invariant system. In visualization, the same constraints do not apply.

We suggest to render a MET volume with a transfer function built using two primitives, a Gaussian and a piece-wise linear function. See Figure 2 (top). The idea behind the Gaussian is that it should highlight the surface of the molecules of interest. This is denoted the *primary level*. The Gaussian should correspond to the role of the isosurface, but both reveal the shape of objects and give an idea about the blurriness around it, i.e., if the surface is part of a sharp change or a smooth transition. The piece-wise linear function has two roles, partly to reveal some of the inside densities of the molecules and partly to pinpoint outliers. To visualize the first part, a linear ramp with low opacity slope is suggested, to not disturb the surface visualization, while a linear ramp with high slope in a different color is recommended for range of densities corresponding to outliers. To set the positions

for these primitives, the following estimates need to be extracted from the chosen measures:

- Secondary level start (start of low slope ramp)
- Primary level (center of the Gaussian)
- Primary level variance (width of the Gaussian)
- Outlier start level (end of low slope ramp, start of high slope ramp in different color)

In this paper, we study how the different measures relate to the primary level. To establish this, the primary level was manually set by an expert user to give a good visualization of the MET volumes. A correlation analysis was then performed, described in Section III. Primarily, we have investigated direct correlation, i.e., analysis of the different measures separately.

In Figure 2, the measures are shown for the MET volume in 1. A primary thing to identify in these plots is whether they are based on multiple distributions. Such distributions could arise from differences between spurious blobs corresponding to ice crystals, reconstruction artefacts, or detector noise, and the shapes of actual molecules.

The number of bins for calculation of the different measures were set to 256, which is higher than the expected accuracy of the extracted estimates. The different measures calculated for the scalar range are described below. They were implemented in C++ using Insight Segmentation and Registration Toolkit¹ (ITK) routines for applying first and second order Gaussian filters, to acquire the gradient and Hessian. For volume rendering, pre-integrated volume rendering according to [9] was used.

A. Measure 1 - Density histogram

The density histogram gives the amount of volume that corresponds to a certain density range. Let V_k denote the set of points in the volume for the bin with index k . b_k is the lower density bound for the bin and b_{k+1} is the upper bound. The histogram value, $\text{density_hist}(k)$, is the number of points in this set.

$$V_k = \{p : p \in X, b_k \leq f(p) < b_{k+1}\} \quad (3)$$

$$\text{density_hist}(k) = |V_k| \quad (4)$$

B. Measure 2 - Requested Connected Components

The second measure is the number of connected components at each density bin, with size filtering based on prior knowledge about the component size. It is denoted RCC for requested connected components. For determining the components, 26-connectedness is used together with the lower bin bound for determining whether a voxel is part of a component or not. The size window is set to be broad, covering deviations from the expected size of the macromolecule corresponding to discretization effects.

¹<http://www.itk.org/>

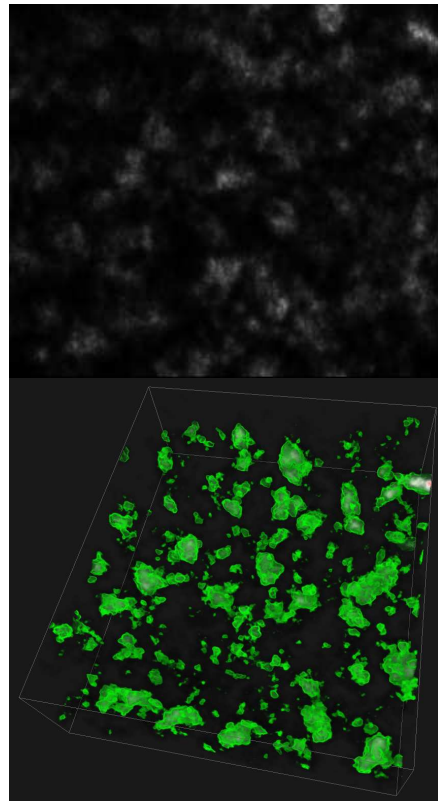


Fig. 1. A slice of a MET volume of a PolII buffer (top), visualized using a manually set transfer function (bottom).

C. Measure 3 - Mean gradient

The third measure is an average gradient measure for one density bin. The gradient is calculated using the first order derivative of a Gaussian kernel with a sigma related to the approximate diameter in voxels of the components of interest in the volumes. The average gradient magnitude value for the set of voxels within the density bin is then calculated, according to

$$M_k = \{|\nabla f(p)| : p \in V_k\} \quad (5)$$

$$\text{mean_gradient}(k) = \overline{M_k} \quad (6)$$

D. Measure 4 - Mean curvature

The mean curvature histogram is based on κ_1 and κ_2 , the curvature values along the principal curvature directions for each voxel. They are calculated following the approach in [4]. The first and second order partial derivatives necessary for the gradient $\mathbf{g}$ and the Hessian matrix $\mathbf{H}$ are calculated using combinations of first and second order derivatives of the Gaussian kernel. Sigma has been set to 1/10 of an estimation of the smallest component diameter. Then the surface normal $\mathbf{n}$ and $\mathbf{P}$, where $\mathbf{I}$ is the identity matrix, are calculated, as

$$\mathbf{n} = \mathbf{g}/|\mathbf{g}| \quad (7)$$

$$\mathbf{P} = \mathbf{I} - \mathbf{n}\mathbf{n}^T \quad (8)$$

Next, the tensor $\mathbf{G}$, referred to in [4] as the *geometry tensor*, is formed and from that the trace T and Frobenius

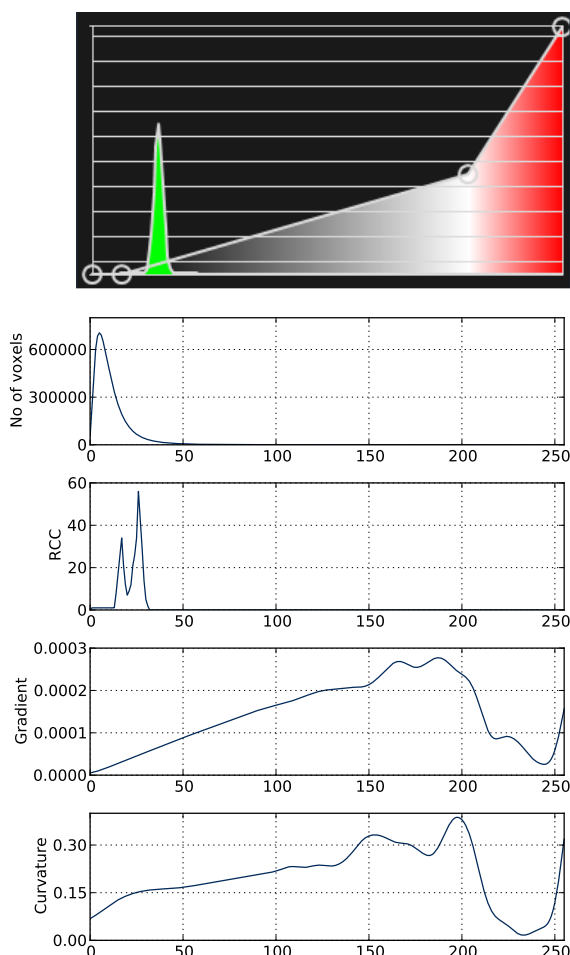


Fig. 2. Example of transfer function (top). Plots of the measures for the MET volume in Figure 1 (bottom).

norm F are calculated. κ_1 and κ_2 are then calculated as

$$\mathbf{G} = -\mathbf{PHP}/|\mathbf{g}| \quad (9)$$

$$\kappa_1 = (T + \sqrt{F^2 - T^2})/2 \quad (10)$$

$$\kappa_2 = (T - \sqrt{2F^2 - T^2})/2 \quad (11)$$

The mean curvature for a single voxel is $(\kappa_1 + \kappa_2)/2$. As in the mean gradient calculation, the total mean curvature for each density bin is calculated to obtain the measure.

III. EVALUATION

We have investigated MET volumes with three types of macromolecules. RNA polymerase II (PolII) is a large macromolecule responsible for mRNA synthesis in eukaryotic cells. The Immunoglobulin G (IgG) antibody is a smaller macromolecule. The carcinoembryonic antigen related cell adhesion molecule 1 (CEACAM1) is even smaller than IgG. Some of the data characteristics can be seen in Table I. All volumes are buffer solutions.

For matching with the primary level, we used two different approaches. For the density histogram, we investigated what histogram percentiles the manually set primary level corresponded to. When features in the

TABLE I
INVESTIGATED MET VOLUMES

Data set	No of investigated volumes	No of molecule instances
IgG	4	~2
PolII	6	~10
CEACAM1	6	~90

measures could be noted, we correlated those features to the manually set primary level. For the RCC measure the maximum was located and analysed. The gradient and curvature measures often level out at the primary level. Therefore, we have looked at the location for the first local minimum of smoothed derivatives of these measures. The correlation plots are shown in Figure 3.

IV. RESULTS AND DISCUSSION

The density histogram is primarily defined by a single distribution. From visual inspection, the distribution is close to a gamma distribution regardless of the kind of molecules in the solution. For the other three measures, there is more variation.

1. *Density histogram*: The primary levels chosen by the expert correspond to percentiles within a small range, but an issue is the thin tail of the distribution, which implies that a 1% difference may correspond to 10–20 density bin levels. Nevertheless, it has been estimated that the primary level should lie between 98–99.8% in the distribution, see Figure 3 (top).

2. *Requested connected components*: RCC is the measure that relies the most on prior knowledge, and also has the highest calculation cost. Intuitively, it should be able to give a robust indicator for the primary level, if the size interval is set correctly and to a narrow range. Here the range has been set broadly, since such accurate prior knowledge cannot be assumed. Because of this, there are variations in our measures, but the overall trend is a linear relation to the primary level.

3. *Gradient*: Probably, there is a linear correlation between the extracted location and the primary level. However, there are too few data samples to establish an exact relationship.

4. *Curvature*: The curvature plateau is also likely to be linearly correlated to the primary level.

When testing on an Intel E5430 2.66GHz CPU, loading and processing a 256^3 volume took approximately 40 seconds (not optimized). In an interactive application, the generation of a transfer function needs to be faster, a few seconds would be preferable, which is possible to reach after code optimization we believe. Building the connected components accounted for more than 25 seconds of the time for the tested data set. It is uncertain what performance gains can be achieved for that calculation, but we expect that the filtering can be faster.

V. CONCLUSION

Our investigated measures are promising for automatic visualization of MET volumes. By analysing more volumes we think that a linear relationship can be established

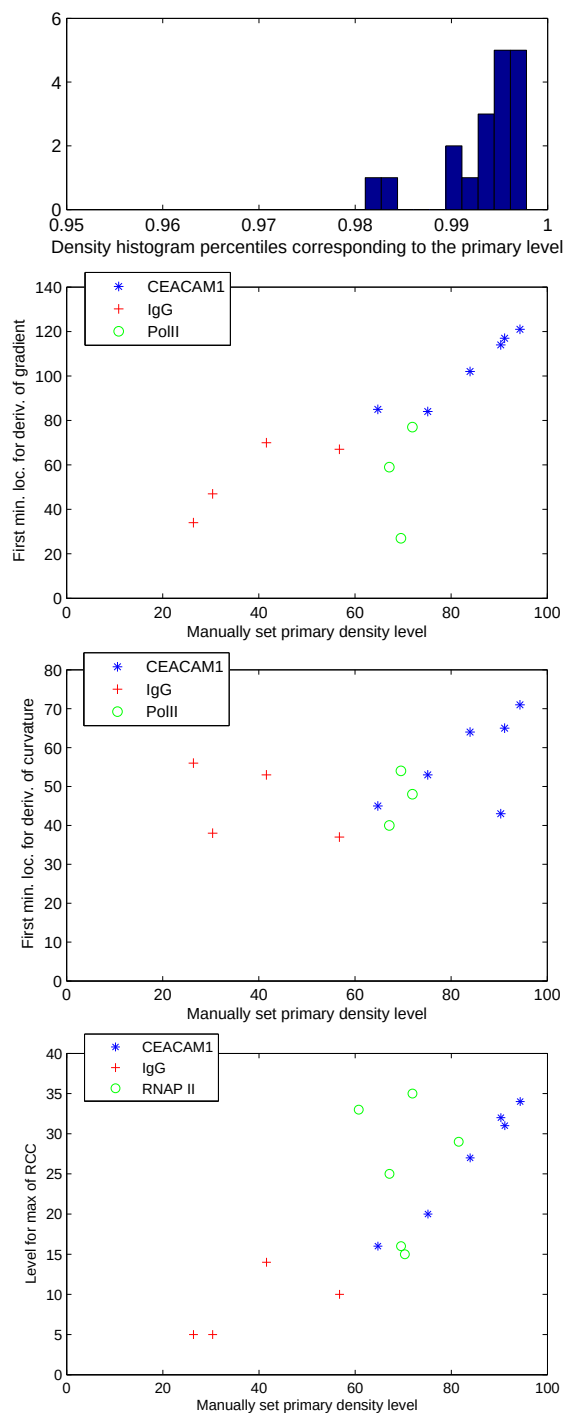


Fig. 3. Histogram percentiles calculated from the manually set primary level (top). Correlation between the user set primary level and different features of measures 2-4 (bottom). The gradient calculation in ITK failed for some of the PolII volumes, therefore these points are missing.

to the user-specified primary level. With that relationship, the primary level can be estimated automatically and the position of the Gaussian primitive in the transfer function can be set thereafter.

It is uncertain whether the performance for the RCC measure can be improved to acceptable levels. The gradient and curvature calculations lie ahead in that aspect. In addition, they are based on separable filtering, which scales well for larger volumes.

A priori information has been used when calculating the gradient and curvature measures, in the form of setting sigma to an adequate level. It would be desirable not to enter such information. However, preliminary tests show that using a standard value for sigma will still give feasible results.

ACKNOWLEDGMENT

This work is funded through the *Visualization Program* by The Knowledge Foundation, Vårdal Foundation, Foundation for Strategic Research, VINNOVA, and Invest in Sweden Agency.

Thanks to: Roger Kornberg laboratory, Dept. of Structural Biology at Stanford University Medical School, USA; Elenor Hauzenberger and Lars-Göran Öfverstedt at Sidec AB, Sweden; Sara Sandin et al [10]; Esther Klaile et al [2], for providing the data and to Daniel Evestedt, SenseGraphics, for partly implementing the pre-integrated volume renderer.

REFERENCES

- [1] U. Skoglund, L.-G. Öfverstedt, R. M. Burnett, and G. Bricogne, "Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: A test application with adenovirus," *Journal of Structural Biology*, vol. 117, pp. 173–188, 1996.
- [2] E. Klaile, O. Vorontsova, K. Sigmundsson, M. M. Müller, B. B. Singer, L.-G. Öfverstedt, S. Svensson, U. Skoglund, and B. Öbrink, "The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters," *Journal of Cell Biology*, vol. 187, no. 4, pp. 553–567, 2009.
- [3] B. Lorensen, C. Bajaj, L. Sobierajski, Machiraju, and J. Lee, "Visualization viewpoints: The transfer function bake-off," *IEEE CGA*, vol. 21, 2001.
- [4] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, "Curvature-based transfer functions for direct volume rendering: methods and applications," in *Proceedings Vis. '03*, 2003, p. 67.
- [5] J. Kniss, S. Premoze, M. Ikits, A. Lefohn, C. Hansen, and E. Praun, "Gaussian transfer functions for multi-field volume visualization," in *Proc. Vis. '03*, 2003, p. 65.
- [6] G. H. Weber, S. E. Dillard, H. Carr, V. Pascucci, and B. Hamann, "Topology-controlled volume rendering," *IEEE TVCG*, vol. 13, no. 2, pp. 330–341, 2007.
- [7] J. Zhou and M. Takatsuka, "Automatic transfer function generation using contour tree controlled residue flow model and color harmonics," *IEEE TVCG*, vol. 15, no. 6, pp. 1481–1488, 2009.
- [8] C. L. Bajaj, V. Pascucci, and D. R. Schikore, "The contour spectrum," in *Proc. Vis. '97*, 1997, pp. 167–173.
- [9] K. Engel, M. Kraus, and T. Ertl, "High-quality pre-integrated volume rendering using hardware-accelerated pixel shading," in *SIGGRAPH/Eurographics Workshop on Graphics Hardware*. ACM, 2001, pp. 9–16.
- [10] S. Sandin, L.-G. Öfverstedt, A.-C. Wikström, O. Wrange, and U. Skoglund, "Structure and flexibility of individual immunoglobulin G molecules in solution," *Structure*, vol. 12, no. 3, pp. 409–415, 2004.