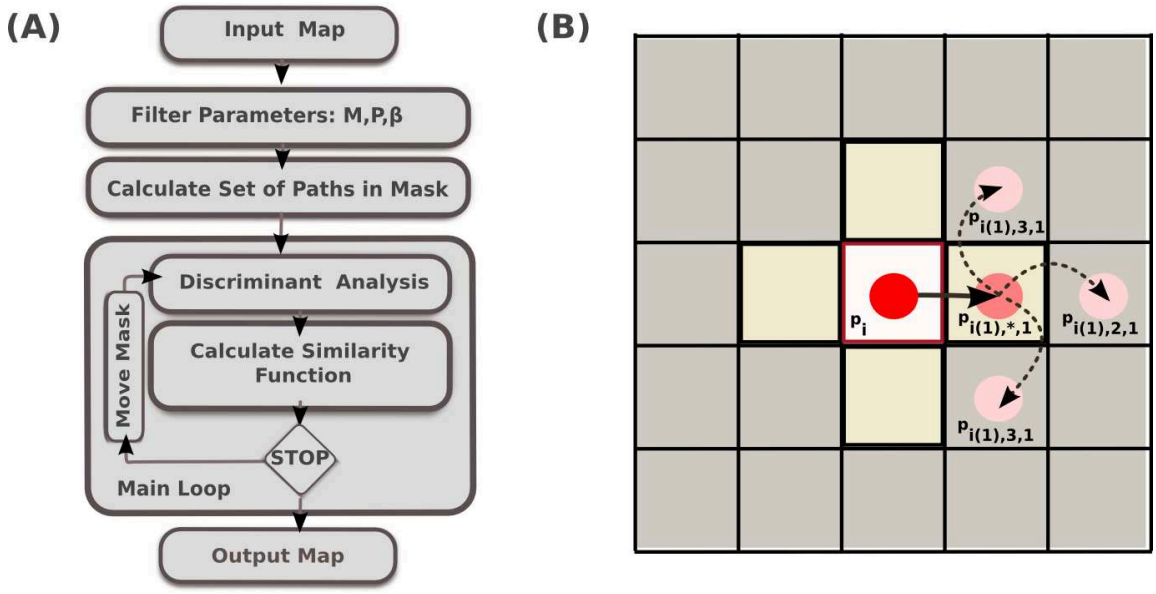
The Digital Paths Supervised Variance (DPSV) Denoising Filter

Supplementary Text (Suppl. Data 1) for
"Automated Tracing of Filaments in 3D Electron Tomography Reconstructions using *Sculptor* and *Situs*" by
Mirabela Rusu, Zbigniew Starosolski, Manuel Wahle, Alexander Rigort, and Willy Wriggers

We describe the Digital Paths Supervised Variance (DPSV) denoising filter as an alternative to the simple Gaussian-weighted averaging used the main text. Free open source implementations of DPSV are available in our *Sculptor* and *Situs* packages at `http://sculptor.biomachina.org` and `http://situs.biomachina.org`. DPSV uses local variance information to control noise in 3D cryo-ET reconstructions in a locally adaptive manner. The method was recently proposed for color image processing (Szczepanski et al., 2004; Smolka, 2008; Szczepanski, 2008), and was adapted for cryo-ET as follows.

The DPSV filter proceeds in three steps. The first step is the generation of digital paths of length P inside a cubic "mask" of (odd-numbered) width M; the second step corresponds to a supervised classification of paths based on a discriminant analysis; and the third step is the computation of the output intensity of the voxel as a kernel-weighted average of the selected paths from the previous steps, where the parameter β defines the size of the exponential kernel. In the *Sculptor* graphics program, version 2.1, the filter can be applied to a map via the menus "Volume" $\rightarrow$ "DPSV Filter" (entering M, P, β parameters in the pop-up dialog box). *Situs* version 2.7 offers a stand-alone denoising program, *volfltr*, that can be run in the UNIX shell. Both implementations of DPSV have been parallelized for multi-core (shared memory) architectures using OPENMP (`http://openmp.org`).

The algorithm for a hypothetical 2D case with $M = 5, P = 2$ (pixel units) is presented in Supplementary Figure 1. S.F. 1A shows how the set of self-avoiding paths in the mask are computed once, after input of the map and the parameters. The mask (and the pre-computed set of paths) are then "moved" across the 3D map during an exhaustive translational scan. The mask size M and path length P are parameters that define the path folding pattern and local reach of the filter. In this work, we used $M = 2P + 1$, which favors straight paths for the detection of linear features such as filaments. S.F. 1B shows paths in a simplified 2D 4-neighborhood model.



Supplementary Figure 1: (A) Schematic overview of the DPSV denoising process. (B) Illustration of a 2D digital self-avoiding walk through one of four nearest neighbors (3 paths of length $P = 2$ are shown).

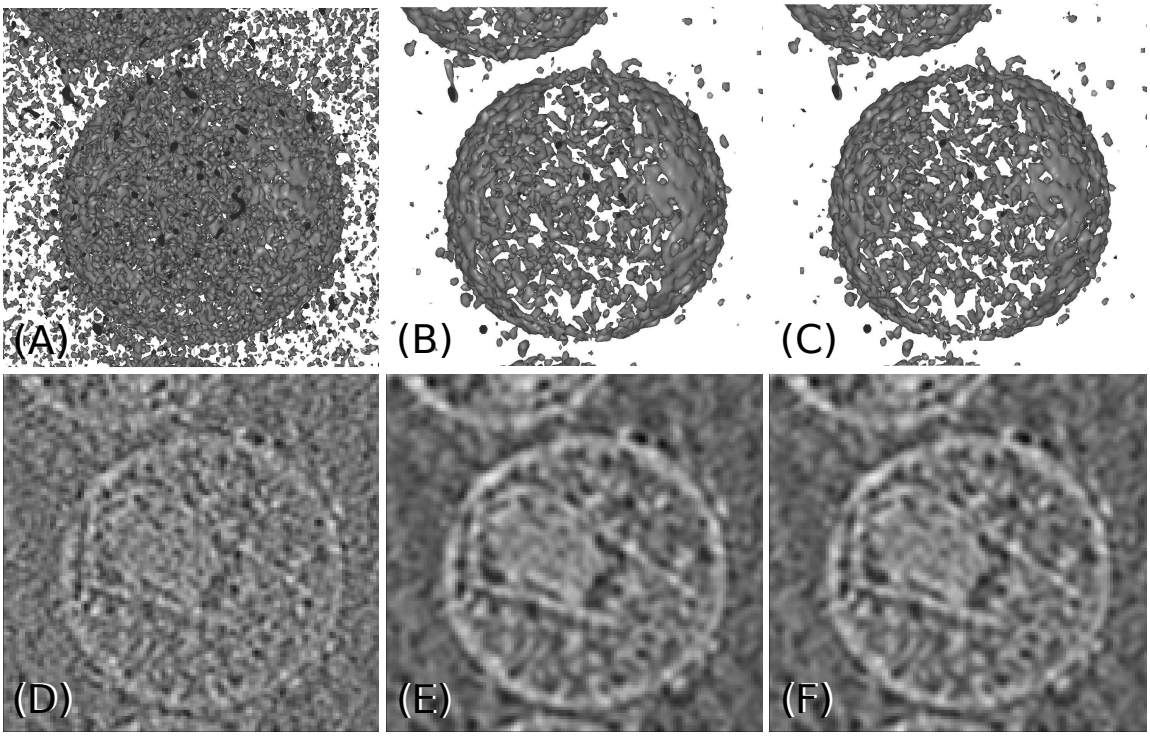
In our case of 3D volume data, we included diagonal connections (26-neighborhood model).

We applied a discriminant analysis (supervised classification) to distinguish between paths affected by noise (which are discarded) and those that ideally include the true signal (Smolka, 2008). For the discriminant analysis, a "connection cost" Λ is computed for all paths. The individual connection cost describes the absolute difference of normalized intensities between a center voxel p_i in the mask and a linked voxel $p_{i(1),l,k}$ (S.F. 1B), divided by their Euclidean distance. The connection cost of a path Λ is then defined as the maximum cost among voxels linked by one path.

Our observation shows that digital paths with high connection cost are usually paths that include noise or cross the edge of the structure. We applied Fisher's discriminant analysis (FDA) to separate the set of paths into two classes (Smolka, 2008). The class with higher connection cost was then excluded from further consideration. This way, the algorithm should ideally preserve only the information that belongs to relatively smooth intensity landscapes and suppress areas affected by noise.

The output intensity of the central pixel of the mask was finally calculated as a cost-weighted mean of the neighboring intensities (termed "Similarity Function" in S.F. 1A). We use an exponential weight for the cost-based averaging, $K(\beta, \Lambda) = e^{-\beta\Lambda}$.

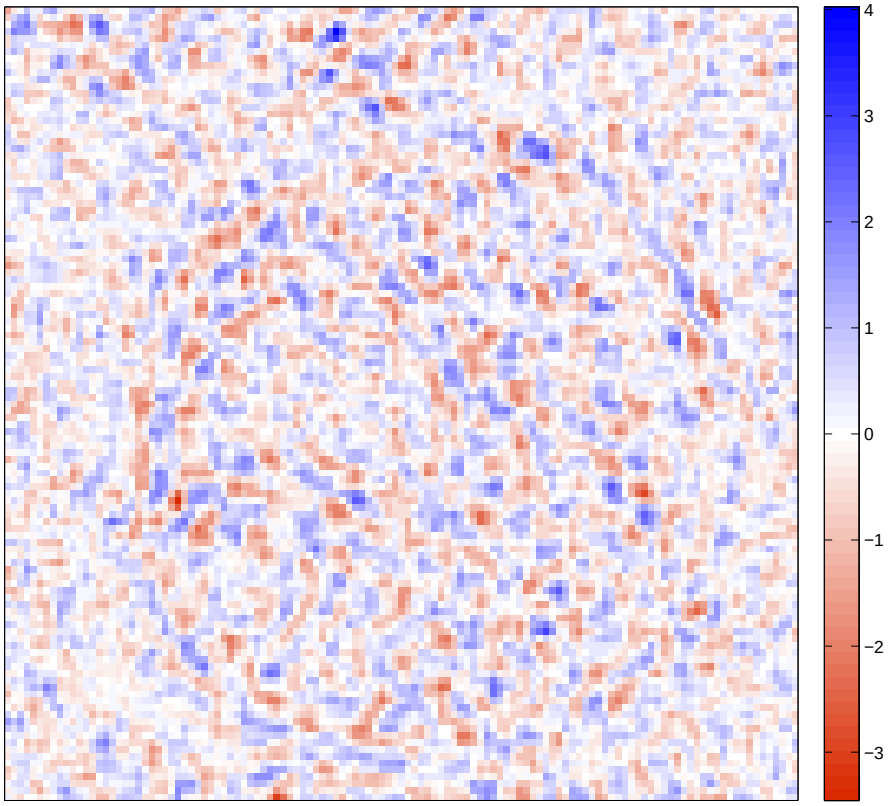
The parameter β defines the "sharpening" of the map, with higher β indicating more sharpening (but due to the cost-weighting it is not a linear relation). Empirical tests suggest useful values (that maximize SNR) in the range of 0.01-0.0001. Although the averaging is performed only over immediately neighboring voxels, the information from more distant voxels is considered indirectly by means of the cost function Λ.



Supplementary Figure 2: Comparison of Gaussian-weighted averaging and DPSV filtration. Shown are iso-surfaces of a 3D cryo-ET reconstruction of unstained HIV-1 virion (EMDB entry 1155, Briggs et al. 2007). (A) Original raw data (cropped; density range: 0-189; iso level: 139.08). (B) Gaussian-weighted average (sigma-1D: 1 voxel; iso level: 134.70) applied to (A). (C) DPSV filter ($M = 5$, $P = 2$, β=0.001; iso level: 136.00) applied to (A). (D) Cross-section of (A). (E) Cross-section of (B). (F) Cross-section of (C). The DPSV filter uses 26-neighborhood model. The molecular graphics were generated with Sculptor (Birmanns et al., 2011).

S.F. 2 compares the results of Gaussian-weighted averaging and DPSV filtration of an HIV-1 virion map. The HIV-1 map is a frequently used test system for denoising (van der Heide et al., 2007; Fernandez, 2009; Wei and Yin, 2010); therefore, our results can be compared to those in the literature. The filtered cross-sections (S.F. 2E,F) clearly show the conical core of the virion, including a region of high density within the core (near the broad end), likely representing the ribonucleopro-

tein complex of the viral genome with the nucleocapsid domain (Briggs et al., 2007). Both filters (S.F. 2E,F) have similar effects on the original raw data and there is little difference discernible in the maps by eye, demonstrating that the DPSV filter ($M = 5$, $P = 2$, β=0.001) is well matched to the Gaussian averaging (sigma-1D: 1 voxel; sigma-1D is the standard deviation of the Gaussian function in 1D, not the 3D standard deviation, $\sqrt{3}$ sigma-1D; the Gaussian was truncated at 3 sigma-1D). We used these matched filter parameters for the validation dataset of the main text.



Supplementary Figure 3: Difference of Gaussian-weighted averaging and DPSV filtration: The DPSV-filtered section of the HIV-1 virion shown in S.F. 2F was subtracted from the Gaussian averaged section shown in S.F. 2E. The Figure was generated with MATLAB 7.9.0 (The MathWorks Inc.).

S.F. 3 shows the difference of Gaussian-weighted averaging and DPSV filtration in more detail. The maximum discrepancy is only 2.1% of the maximum density (189) of the original map in S.F. 2D due to the similarity of the denoising shown above. The

discrepancy map is dominated by "center-surround" patterns that indicate different point spread properties of the filters. Owing to the relatively bigger point spread of the Gaussian, high-density features in the original map (white in S.F. 2D) yield negative (red) centers and positive (blue) surround in the discrepancy map. Likewise, low-density features (dark in S.F. 2D) yield positive centers and negative surround. In other words, DPSV preserves more details than the Gaussian at comparable denoising level.

References

Birmanns, S., Rusu, M., Wriggers, W., 2011. Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. J. Struct. Biol. 173, 428–435.

Briggs, J. A., Grünewald, K., Glass, B., Förster, F., Kräusslich, H. G., Fuller, S. D., 2007. The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. Structure 14, 15–20.

Fernandez, J., 2009. TOMOBFLOW: feature-preserving noise filtering for electron tomography. BMC Bioinformatics 10, 178.

Smolka, B., 2008. Peer group filter for impulsive noise removal in color images. IEEE Trans. Med. Imaging 27, 699–707.

Szczepanski, M., 2008. Fast digital approach spatio-temporal filter. Zeszyty Naukowe Politechiki Slaskiej, seria Automatyka 150, 207–222.

Szczepanski, M., Smolka, B., Plataniotis, K., Venetsanopoulos, A., 2004. On the distance function approach to color image enhancement. Discrete Applied Mathematics 139, 283–305.

van der Heide, P., Xu, X. P., Marsh, B. J., Hanein, D., Volkmann, N., 2007. Efficient automatic noise reduction of electron tomographic reconstructions based on iterative median filtering. J. Struct. Biol. 158, 196–204.

Wei, D. Y., Yin, C. C., 2010. An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data. J. Struct. Biol. 172, 211–218.