Mean Shifting Gradient Vector Flow: An Improved External Force Field for Active Surfaces in Widefield Microscopy.

Margret Keuper*

Chair of Pattern Recognition and Image Processing
Computer Science Department
Centre of Biological Signalling Studies (bioss)
University of Freiburg
keuper@informatik.uni-freiburg.de

May 11, 2010

Abstract

Gradient vector flow based snakes are a very common method in bio-medical image segmentation. The use of gradient vector flow herein brings some major advantages like a large capture range to strong gradients and a good adaption of the snakes in concave regions. In some cases though, the application of gradient vector flow can also have undesired effects, e.g. if only parts of an image are strongly blurred, the remaining weak gradients will be smoothed away. Also, large gradients resulting from small but bright image structures usually have strong impact on the overall result. To tackle this problem, we present an improvement of the gradient vector flow, using the mean shift procedure and show its advantages on the segmentation of 3D cell nuclei.

1   Introduction

Active surfaces are a widely used method in the field of bio-medical image segmentation tasks. Especially when dealing with 3D microscopic data, one has to handle several challenges as there are blurred object edges due to the point spread function (PSF) of the microscope as well as artifacts caused by inhomogeneities in the fluorescent stains. If the data comes from widefield microscopy,

the PSF strongly blurs the recording and can even artificially cause edges in the dataset, that can even become stronger, if we try to linearly deconvolve the data. To make the segmentation robust against those artifacts, one often tries to include as much prior shape knowledge as possible into the model, e.g. [3], [4]. Thus, one tries to weaken the ill-posedness of the segmentation task.

For the segmentation of DAPI-stained nuclei in a *Drosophila* cell line (so called S2 or Schneider cells) recorded with 3D widefield fluorescence microscopy, an active surface model has thus been adapted to the segmentation of sphere-like objects in [4]. In that paper, the force field has been adapted by projecting the gradients onto their radial components such that only the desired gradients have impact on the segmentation result. This indeed leads to better contours but still, artifacts can be observed. One of the major problems is that the resulting active surfaces seem to be elongated and subjected to PSF artifacts compared to what would be expected. See figure 3 for an example of the data in two orthogonal views and the PSF of the used microscopy setting in xz-view. In [4], gradient vector flow (GVF) [6], [5] was used to smooth the gradient field before adapting the active surfaces. This leads to some desired effects like a large capture range for the active surface model as well as smooth gradient fields. On the other hand, the weak but important gradients in upper and lower image regions are smoothed away by the classical GVF.

To diminish the artifacts caused by the dataset blurring, we are presenting a vector diffusion method that not only preserves strong image gradients but also preserves weak gradients if they are dense i.e. if all gradients in a neighborhood have similar direction and length. This method is the result of combining the basic GVF with the mean shift procedure [1], [2].

2 Gradient Vector Flow

The currently best established method to smooth vector fields is the gradient vector flow (GVF)[6], [5]. This diffusion method aims at providing a smooth vector field in regions where no edges are given in the dataset, and in keeping the original gradient information in regions with strong edges. Thus, when the vector field is used as an external force field e.g. for active contours, the capture range is much larger than it would be, if the original gradient information were used.

The three dimensional gradient vector flow field is the vector field $\mathbf{v}(\mathbf{x}) : \mathbb{R}^3 \to \mathbb{R}^3$ that minimizes the energy functional

$$E(M) = \int_{\mathbb{R}^3} \mu(|\nabla \mathbf{v}|^2) + |\nabla M|^2 |\mathbf{v} - \nabla M|^2 \mathrm{d}\mathbf{x}, \tag{1}$$

where M is the edge map of the original dataset and the grad operator ∇ is applied to each component of $\mathbf{v}$ separately. Using the calculus of variations, the following Euler Equation can be found [5]:

$$\mu \nabla^2 \mathbf{v} - (\mathbf{v} - \nabla M) |\nabla M|^2 = 0, \tag{2}$$

where ∇^2 is applied to each component of the vector field $\mathbf{v}$ separately. A solution to this equation is derived by introducing a time variable t and finding the steady-state solution of the partial differential equation (PDE)

$$\frac{\partial \mathbf{v}}{\partial t} = \mu \nabla^2 \mathbf{v} - (\mathbf{v} - \nabla M)\,|\nabla M|^2. \tag{3}$$

For a discrete grid of cubic voxels, the iteration rule resulting from equation (3) is given by:

$$\begin{aligned}
\mathbf{v}_i^{n+1}(x,y,z) =& (1 - |\nabla M|^2)\mathbf{v}_i^n(x,y,z) \\
&+ \mu(\mathbf{v}_i^n(x+1,y,z) + \mathbf{v}_i^n(x-1,y,z) + \mathbf{v}_i^n(x,y+1,z) \\
&+ \mathbf{v}_i^n(x,y-1,z) + \mathbf{v}_i^n(x,y,z+1) + \mathbf{v}_i^n(x,y,z-1) \\
&- 6\mathbf{v}_i^n(x,y,z)) + |\nabla M|^2 \frac{\partial M(x,y,z)}{\partial i},
\end{aligned} \tag{4}$$

where $\mathbf{v}_i^n$ is the i-th component of the vector field $\mathbf{v}$ at iteration step n.

See fig.1 for a toy example. The 2D toy data 1(a)(120×120 pixels) is designed to simulate the blurred recordings of spherical objects with three undesired spots (spot 1-3) that should influence the overall force field as little as possible. The edge image 1(b) is the magnitude of the smoothed gradients of the toy data and the gradient field is displayed in 1(c). The colors indicate the vector direction. In 1(d), the GVF field (300 iterations with $\mu = 0.15$) of the toy data is displayed.

3 Mean Shift Filtering

The mean shift procedure as presented in [1] aims at analyzing the density distribution of arbitrary feature spaces and can be used to detect density maxima or to perform filtering or clustering of feature points. The theoretical basis for the application of the procedure as an image filter has been presented in [2]. The prospect of applying the mean shift procedure on gradient vector fields is to analyze, which gradient vectors correspond to more important structures in the dataset and accordingly, to shift the gradient vectors towards their local density maxima. Thus, gradients corresponding to smaller, less important structures can be filtered out.

The mean shift algorithm is based on the kernel density estimation which is also known as Parzen window technique. Here, we summarize the derivation of the mean shift algorithm as described in [2].

3.1 Kernel Density Estimation

For n independent data points $\mathbf{x}_i$, $i = 1, \ldots, n$ in d-dimensional space $\mathbb{R}^d$, that are generated by an unknown probability distribution f, a *kernel density estimate* is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \tag{5}$$

3

where the d-variate kernel $K(\mathbf{x})$ is a bounded function with compact support, that satisfies

$$\int_{\mathbb{R}^d} K(\mathbf{x})d\mathbf{x} = 1 \quad \lim_{\|x\|\to\infty} \|\mathbf{x}\|^d K(\mathbf{x}) = 0$$

$$\int_{\mathbb{R}^d} \mathbf{x}K(\mathbf{x})d\mathbf{x} = 0 \quad \int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x})d\mathbf{x} = c_K\mathbf{I}, \tag{6}$$

where c_K is a constant. h is the kernel bandwidth parameter. For radially symmetric kernels, the multivariate kernel can be generated by

$$\hat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right), \tag{7}$$

where the function $k(x)$ is the radial profile of the kernel $K(\mathbf{x})$:

$$K(\mathbf{x}) = c_{k,d}k(\|\mathbf{x}\|^2). \tag{8}$$

The function $k(x)$ with $k(x) \geq 0$ only needs to be defined for $x \geq 0$. The constant $c_{k,d}$ with $c_{k,d} > 0$ ensures, that $K(\mathbf{x})$ integrates to one. For complexity reasons, we are using the Epanechnikov kernel (compare [2]), with the profile $k_E(x) = 1 - x$, for $0 \leq x \leq 1$ and 0 for $x > 1$. The radially symmetric kernel is given by

$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\|\mathbf{x}\|^2), & \text{if } \|\mathbf{x}\| \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where c_d is the volume of the d-dimensional unit sphere. k_E is not differentiable for $x = 1$, for $x \neq 1$, $k'_E(x)$ is uniform.

The main step of the mean shift procedure is to determine local maxima in this density distribution of the feature space. Thus, zero positions of the actual density gradient $\nabla f(\mathbf{x}) = 0$ are estimated by gradient ascent of the estimate $\hat{f}_{h,K}(\mathbf{x})$. For differentiable profiles k, the gradient of $\hat{f}_{h,K}(\mathbf{x})$ is given by

$$\nabla \hat{f}_{h,K}(\mathbf{x}) \tag{10}$$
$$= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n}(\mathbf{x}-\mathbf{x}_i)k'\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)$$
$$= \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} k'\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\right] \left[\mathbf{x} - \frac{\sum_{i=1}^{n} \mathbf{x}_i k'\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} k'\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}\right],$$

where the first term is proportional to the density estimate with kernel $G(\mathbf{x}) = c_{g,d}g(\|\mathbf{x}\|^2)$ with $g(x) = -k'(x)$ and the second term is the mean shift

$$\mathbf{m}_{h,G}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}, \tag{11}$$

from which the gradient ascent iteration is derived as:

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)}. \tag{12}$$

With each iteration step, the points are shifted towards the next local density maximum.

3.2 Mean Shift on Gradient Vectors

In [2], the authors show how the mean shift procedure can be employed as a discontinuity preserving image filter. This is done by interpreting the image as a two-dimensional lattice of p-dimensional vectors ($p = 1$ in the gray level case, $p = 3$ for color images). The lattice then forms the spatial domain, the color information forms the range domain of a $d = p + 2$ dimensional feature vector in the joint spatial-range domain. The mean shift is performed using a multivariate kernel

$$K_{h_s, h_r} = \frac{C}{h_s^2 h_r^2} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right), \tag{13}$$

defined as a product of two radially symmetric kernels with the common profile $k(x)$. $\mathbf{x}^s$ is the spatial part, $\mathbf{x}^r$ the range part of the feature vector. The Euclidean metric allows the use of a single bandwidth parameter in each domain, h_s and h_r respectively. These bandwidth parameters control the size of the kernel and thus determine the resolution of the density maximum detection. C is the corresponding normalization constant.

An extension of the algorithm presented in [2] to the three dimensional space is straightforward. In the case of a three dimensional vector field $\nabla M(\mathbf{x}^s)$: $\mathbb{R}^3 \rightarrow \mathbb{R}^3$, where M is a 3D scalar field (the edge map), the mean shift is computed on the $\bar{d} = 3 + 3$ dimensional feature vector, where the first three components contain the spatial information, the last three dimensions contain the corresponding vector information. As in the two dimensional case, the Epanechnikov kernel performs satisfactorily.

When applying this edge preserving mean shift filter to gradient vector fields, the result is quite similar to what we know from 2D color images. Depending on the chosen spatial and range bandwidth parameters, the discontinuities at the dataset gradients are preserved and the vectors converge to equal length and direction in the neighborhood defined by the spatial bandwidth parameter.

In fig. 1(e) we have displayed the result of applying the 3D mean shift filter to our toy example (with the parameters $h_s = 6$, $h_r = 0.4$). Because of the round shape of the object, the gradient vectors converge to the mean direction in their neighborhood, i.e. all vectors are pointing towards the edges. The result is depending on the filter bandwidth. Vectors with $\|\mathbf{v}_i - \mathbf{v}_j\| \leq h_r$ converge to

5

equal length if they are spatially close to one another. Maxima in gradient magnitude are not preserved. Below the actual gradient image in 1(e), we have plotted the profile of the original gradient field (blue) and the mean shifted field (red) along the red line in 1(a).

4 Mean Shift GVF

With the GVF, an isotropic vector diffusion is performed. The regularization only depends on the vector length disregarding the meaning of the vectors in their local neighborhood. In regions, where the gradients are generally weaker, the Laplace term dominates equation 1 and a strong smoothing is performed. In our example of the microscopic recordings of cell nuclei, this is not always desirable. First, we do not want to promote strong gradients, if they are originating from small structures. Second, in the strongly blurred upper and lower dataset regions, we want to keep the gradients pointing towards the cell nuclei even though they are weak.

Generally, we want that large gradients originating from small structures are preserved but do not diffuse as strongly as in GVF, i.e. have less impact on the overall force field. On the other hand, if in a certain region there are many weak gradient vectors pointing into the same direction, these valuable gradients shall not be altered by larger neighboring gradients (as it would happen in GVF), but all gradients in the same neighborhood shall converge to their mean direction and length. This is why we are using the advantages of the mean shift filtering method and the GVF simultaneously, performing a minimization of the energy functional (1) and a maximization of the gradient kernel density estimate (7) at the same time, which we call mean shift gradient vector flow (MSGVF). The new energy functional we are minimizing is given by

$$E(M, \bar{\mathbf{v}}) = \int_{\mathbb{R}^3} \mu \left(|\nabla (\bar{\mathbf{v}})|^2 \right) + |\nabla M|^2 \, |\bar{\mathbf{v}} - \nabla M|^2 \, \mathrm{d}\mathbf{x}, \tag{14}$$

where

$$\bar{\mathbf{v}} = \mathbf{v} + \hat{c} \frac{\nabla \hat{f}_{h,k}(\mathbf{v})}{\hat{f}_{h,g}(\mathbf{v})}. \tag{15}$$

Thus, we are searching for a minimal energy $E(M, \bar{\mathbf{v}})$, constraining every vector to be at its kernel density maximum, i.e. $\mathbf{m}_{h,G}(\mathbf{v}) = 0$. As in [5], this energy functional is minimized by the following Euler-Lagrange equation:

$$\mu \nabla^2 (\mathbf{v} + \mathbf{m}_{h,g}(\mathbf{v})) - (\mathbf{v} + \mathbf{m}_{h,g}(\mathbf{v}) - \nabla M) \, |\nabla M|^2 = 0, \tag{16}$$

with the resulting PDE

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} = {} & \mu \nabla^2 (\mathbf{v} + \mathbf{m}_{h,G}(\mathbf{v})) \\ & - ((\mathbf{v} + \mathbf{m}_{h,G}(\mathbf{v})) - \nabla M) \, |\nabla M|^2. \end{aligned} \tag{17}$$

Unlike for GVF, the optimization cannot be done for each channel separately, because $\mathbf{m}_{h,g}(\mathbf{v})$ is depending of all channels. The new iteration scheme is
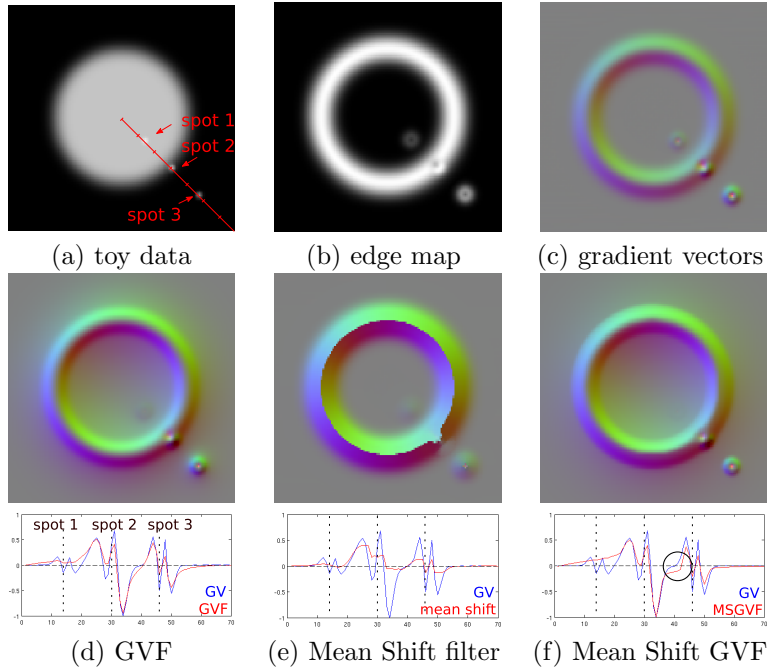
Figure 1: The methods on toy data.

generated be plugging the mean shift criterion (12) into the GVF iteration scheme. This yields

$$
\begin{aligned}
\mathbf{v}_i^{n+1}(x,y,z) =& (1 - |\nabla M|^2)\bar{\mathbf{v}}_i^n(x,y,z) \\
& + \mu(\bar{\mathbf{v}}_i^n(x+1,y,z) + \bar{\mathbf{v}}_i^n(x-1,y,z) + \bar{\mathbf{v}}_i^n(x,y+1,z) \\
& + \bar{\mathbf{v}}_i^n(x,y-1,z) + \bar{\mathbf{v}}_i^n(x,y,z+1) + \bar{\mathbf{v}}_i^n(x,y,z-1) \\
& - 6\bar{\mathbf{v}}_i^n(x,y,z)) + |\nabla M|^2 \frac{\partial M(x,y,z)}{\partial i},
\end{aligned}
\tag{18}
$$

where, at iteration step n,

$$
\bar{\mathbf{v}}^n = \frac{\sum_{j=1}^m \mathbf{v}_j^n g\left(\left\|\frac{\mathbf{v}^n - \mathbf{v}_j^n}{h}\right\|^2\right)}{\sum_{j=1}^m g\left(\left\|\frac{\mathbf{v}^n - \mathbf{v}_j^n}{h}\right\|^2\right)}.
\tag{19}
$$

and $\bar{\mathbf{v}}_i^n$ is the i-th component of the vector field $\bar{\mathbf{v}}$ at iteration step n.

The result of the MSGVF can be seen in fig. 1(f), for 100 iterations with $\mu = 0.15$, $hs = 6$ and $hr = 0.2$. As if applying GVF 1(d), the discontinuities are preserved and long gradients corresponding to object edges are unaltered. As opposed to GVF, gradients resulting from smaller objects have less impact

on the overall result. This can be seen when looking at the magnitude profile in fig. 1(f). The small spot (spot 1) inside the object does not change the vector direction at all. For a visualizition of the vector directions, compare also the streamline plots in fig. 2. When looking at the spot outside the object (spot 3) one can observe, that strong gradients are caused by this structure, that are preserved by both methods. Only the diffusion behavior is different. In the case of MSGVF, the zero crossing, indicating the change of vector direction, is shifted towards the small spot. Thus the capture range for our large object is larger and, even though we have computed less iterations than for GVF, the magnitude of the vectors outside the object are larger.

(a) Streamline plot of the GVF field in the ROI.

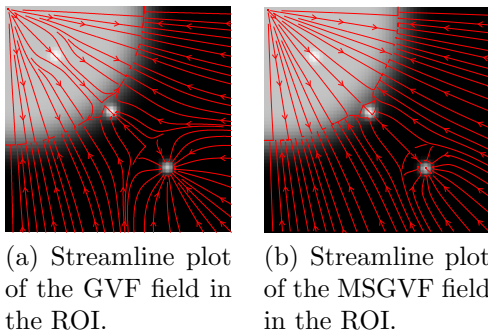(b) Streamline plot of the MSGVF field in the ROI.

Figure 2: Especially when looking at the small spot outside the large object (spot 3), one can see that it has strong impact on the vector field, and this impact is significantly larger, when applying GVF (a), than when applying MSGVF (b). Thus, if we want to segment the main object, our initial surface must fit better with GVF than with MSGVF. Also, when applying MSGVF, spot 1 has no impact on the vector direction at all and even the capture range of spot 2 is less strong. So, both spots will not significantly alter the segmentation result.

5 Application and Results

We have applied the new method to generate force fields for the segmentation of *Drosophila* S2 cell nuclei, see fig. 3(a) for example. In [4], the last step of the external force field generation is applying GVF to the vector field. We are showing that by replacing this step with the presented MSGVF, we can improve the segmentation and diminish the artifacts caused by the PSF. To allow for comparison, all further steps of the active surface procedure are the same as in [4].

We applied our method to a dataset of 393 recordings of *Drosophila* S2 cell nuclei. To measure the susceptibility of the segmentation surface to artifacts, we are locally approximating the surface by spheres and estimate their radii. At the
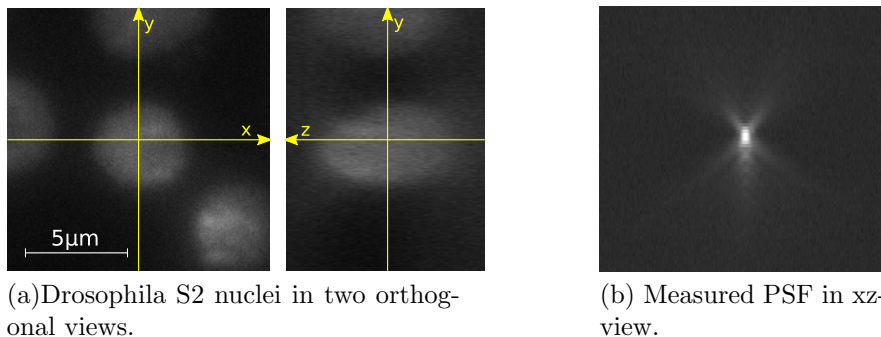
(a)Drosophila S2 nuclei in two orthogonal views.

(b) Measured PSF in xz-view.

Figure 3: Microscopic data from widefield fluorescent microscopy.

best, most radii should be similar to the radius of the cells that ist somewhere near 2μm. The more artifacts we have in a surface, the more very small radii can be found. As we expect to find these artifacts in the lower dataset regions, we compute the measure of the lower 20% of the segmentation mask and compare it to the 20% in the center. The result can be seen in fig. 4. The results show

(a)Histogram of the radii in the center.
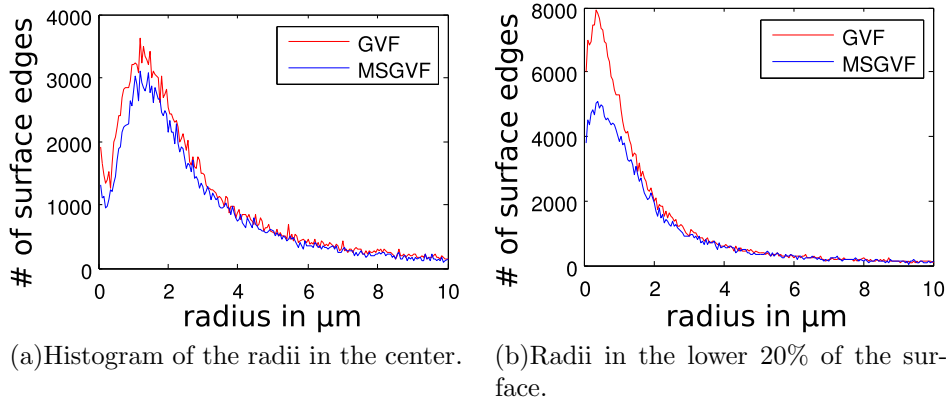

(b)Radii in the lower 20% of the surface.

Figure 4: Fitted radii.

that the surfaces still are more strongly curved in the lower regions than in the center. Though, compared to GVF, the new method works clearly better. For GVF, the ratio of the number of triangles forming an angle in a reasonable range (radii of $1-3\mu$m) and those forming too small angles (radii $< 1\mu$m) is at 0.72, for MSGVF it is at 0.92. Having a closer look at the data (fig. 5, fig. 6 and fig. 7, the red contour has been computed with GVF the green one with MSGVF), one can clearly see that in the central z-slices, the segmentation is very similar. Only in the more difficult lower regions, the result found be MSGVF is smoother and shows less artifacts. Additionally, one can see that most of the resulting

9

surfaces are less elongated than with GVF. This results from the fact that in MSGVF, the vectors in the lower dataset regions that are pointing towards the surface are generally better preserved. Sometimes, when applying GVF, this information is completely lost.

6   Conclusion

We have presented a new method for the generation of external force fields for 2D and 3D active contours. The resulting vector fields are more robust against noise or smaller structures that should not influence the overall force field. Especially small but important gradients in gradient fields from blurred data are preserved. We could show that the segmentation of strongly blurred data from 3D widefield microscopy with an active surface model could be improved and the resulting surfaces show less PSF artifacts.

7   Acklowledgments

We would like to thank Jan Padeken and Patrick Heun from the Max-Planck Institute of Immunobiology in Freiburg who kindly provided the data for this study.

References

[1] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. on PAMI*, 17/8:790–799, 1995.

[2] D. Comanciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on PAMI*, 24/5:603–619, 2002.

[3] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. on PAMI*, 28/8:1262–1273, 2006.

[4] M. Keuper, J. Padeken, P. Heun, H. Burkhardt, and O. Ronneberger. A 3d active surface model for the accurate segmentation of *drosophila* schneider cell nuclei and nucleoli. In *Proc. of the ISVC*. LNCS, 2009.

[5] C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. Imag. Proc.*, 7/3:321–345, 1998.

[6] C. Xu and J. L. Prince. Gradient vector flow: A new external force for snakes. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 66, Washington, DC, USA, 1997. IEEE Computer Society.
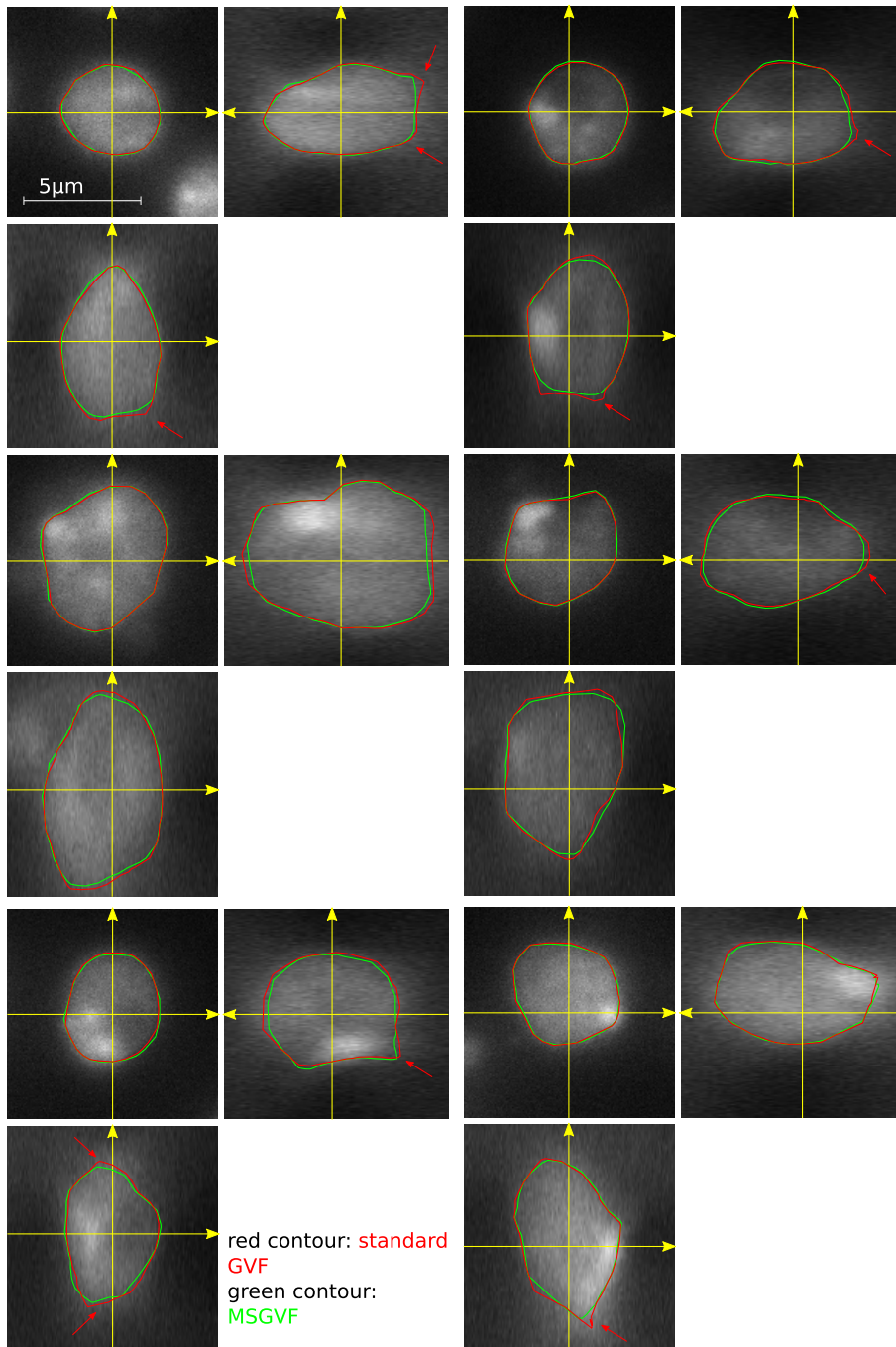
Figure 5: Segmentation results. The red contour has been conputed based on GVF, the green contour based on MSGVF. The red arrow indicate segmentation artifacts.
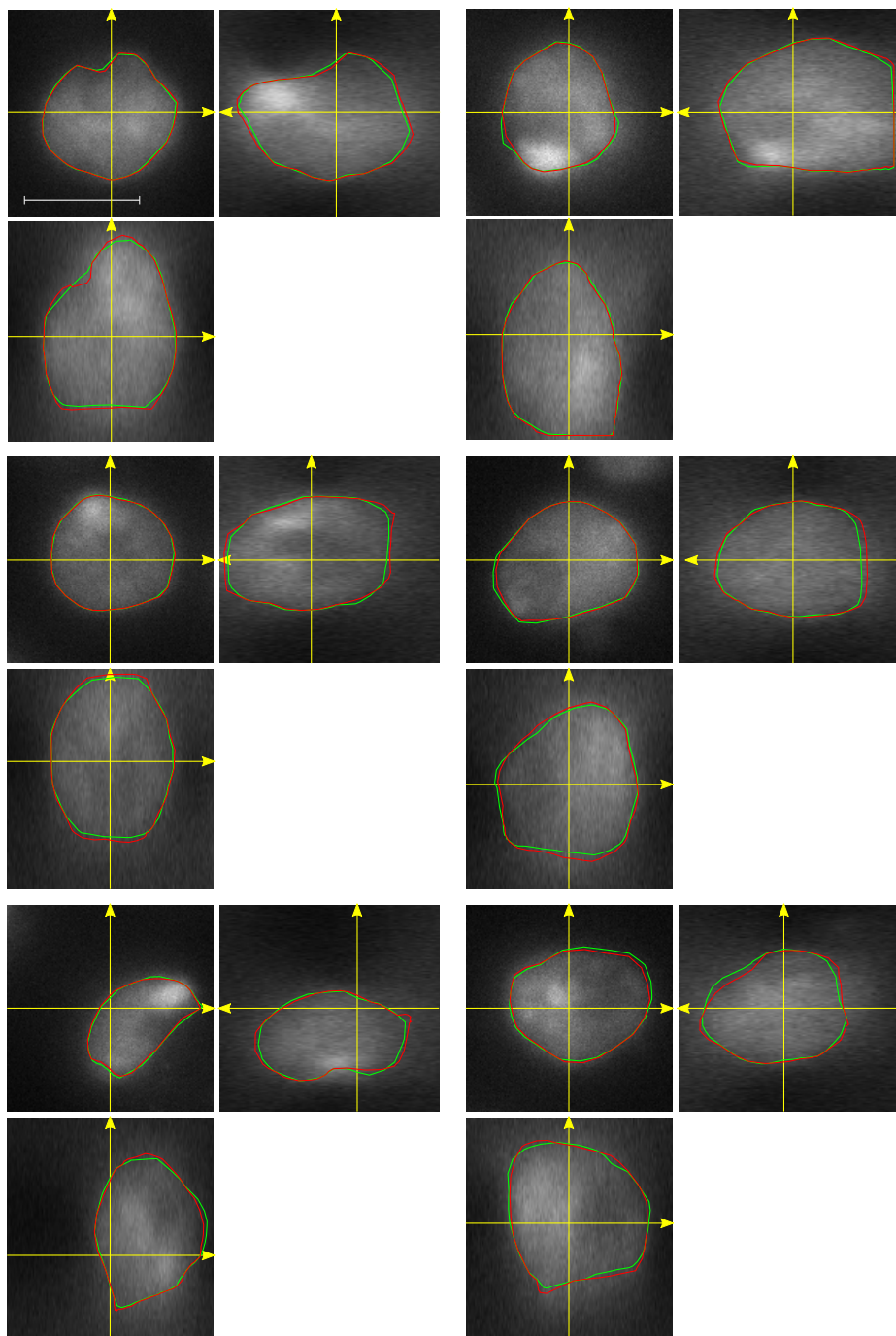
Figure 6: Segmentation results. The red contour has been conputed based on GVF, the green contour based on MSGVF.
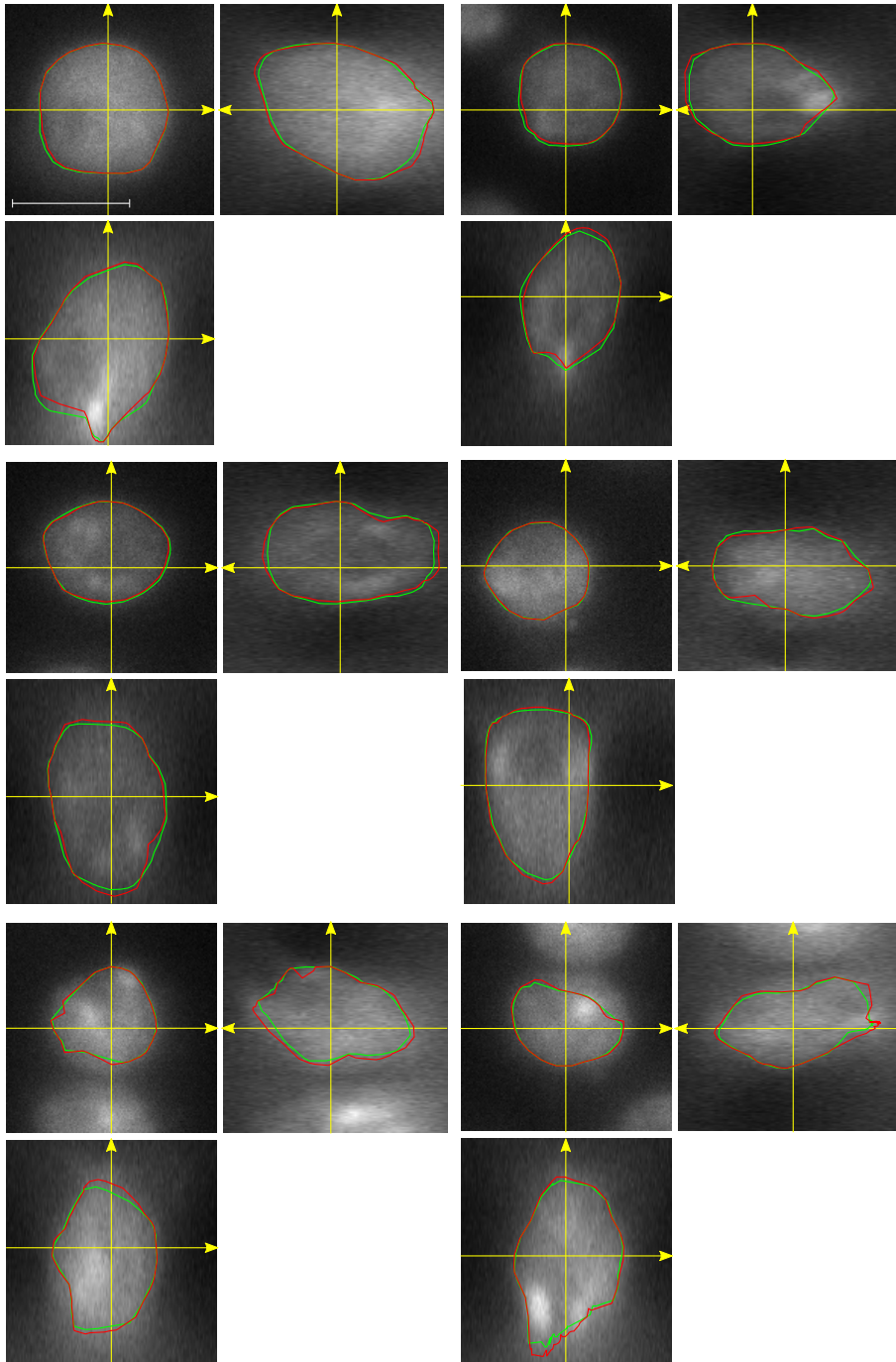
Figure 7: Segmentation results. The red contour has been conputed based on GVF, the green contour based on MSGVF.