

Robust Signal Detection in 3D Fluorescence Microscopy

Amin Allalou,* Amalka Pinidiyaarachchi, Carolina Wählby

The Centre for Image Analysis, Uppsala University, Uppsala, Sweden

Received 3 February 2009; Revision Received 22 June 2009; Accepted 19 August 2009

Grant sponsor: EU-FP6 ENLIGHT (ENhanced LiGase based Histochemical Techniques); Grant number: EU-FP6; Grant sponsors: Swedish Research Council (Collaboration Grant), Swedish International Development Cooperation Agency (SIDA).

Present address of Amalka Pinidiyaarachchi: Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka.

Present address of Carolina Wählby: Broad Institute of MIT and Harvard, Cambridge, MA.

*Correspondence to: Amin Allalou, Centre for Image Analysis, Box 337, 751 05 Uppsala, Sweden.

E-mail: amin@cb.uu.se

Published online 16 September 2009 in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/cyto.a.20795

© 2009 International Society for Advancement of Cytometry

• Abstract

Robust detection and localization of biomolecules inside cells is of great importance to better understand the functions related to them. Fluorescence microscopy and specific staining methods make biomolecules appear as point-like signals on image data, often acquired in 3D. Visual detection of such point-like signals can be time consuming and problematic if the 3D images are large, containing many, sometimes overlapping, signals. This sets a demand for robust automated methods for accurate detection of signals in 3D fluorescence microscopy. We propose a new 3D point-source signal detection method that is based on Fourier series. The method consists of two parts, a detector, which is a cosine filter to enhance the point-like signals, and a verifier, which is a sine filter to validate the result from the detector. Compared to conventional methods, our method shows better robustness to noise and good ability to resolve signals that are spatially close. Tests on image data show that the method has equivalent accuracy in signal detection in comparison to visual detection by experts. The proposed method can be used as an efficient point-like signal detection tool for various types of biological 3D image data. © 2009 International Society for Advancement of Cytometry

• Key terms

3D signal detection; fluorescence microscopy; stable wave detection; Fourier series; DoG; TopHat; multiscale product; segmentation

ROBUST detection and localization of proteins or other biomolecules inside cells enables better understanding of the functions related to them. Fluorescence microscopy together with different types of fluorescent biomarkers makes biomolecules visible as point-like signals in the captured image data. Large experiments of this type produce vast amounts of image data. This has led quantitative analysis to move from manual to semi-automated or fully automated systems (1). This in turn has led to the development of reliable and efficient automated methods for the detection of point-like signals in two-dimensional (2D) images (2–4). As cells and tissue are three-dimensional (3D) structures, exact localization of signals requires detection methods that work in 3D.

A point-like signal in this context can generally be defined as a small object, relatively higher in intensity than the image background. However, image data from cells and tissue often contain point-like signals with a range of varying intensities and sizes and the image background often contains diffuse structures caused by autofluorescence. Signals can also lie in clusters that make separation of individual signals difficult even through visual inspection by an expert. A robust method should, apart from being able to separate clusters of signals, also be capable of distinguishing between a true signal and an intensity maximum caused by noise. An additional consideration when detecting signal localization in 3D is the point spread function (PSF) of the microscope system, which typically affects the axial direction more than the lateral direction (5).

A straightforward method for detecting point-like signals is the difference of Gaussians (DoG) (6), where the image is convolved with two Gaussian kernels

of different sizes; the larger kernel describes the background variation while the smaller kernel describes the signals of interest. Gaussian kernels are also used in (4) where 2D signals from biological samples with varying noise levels are detected using multiscale products (MPs) of sub-band images from a wavelet transform. This MP-based 2D approach has also been extended to 3D for detection of signals from fluorescence microscopy images of biological samples (3).

Another popular method for the detection of signals in fluorescence microscopy is the TopHat transform (2,7,8). Signals are detected/enhanced by morphological opening of the original image, using a selected structuring element followed by taking the difference between the original image and the result of the opening. The TopHat transform has been used for many 2D applications, but also for the detection of signals from 3D images of telomeres (2).

We present a detailed analysis of the performance of a novel 3D point-like signal detection method. The method is an improvement of the 2D stable wave detector (SWD) (9), and is thus referred to as the 3D stable wave detector (3DSWD). The basic idea is to use the Fourier coefficients of the first harmonic in a Fourier series to find signals, seen as local intensity maxima in an image. The method consists of two parts: a detector, which is a cosine filter to enhance the point like signals, and a verifier, which is a sine filter to validate the result from the detector. The 3DSWD was previously presented briefly in (10), an application oriented paper by us. Now we have tested the method in detail for its robustness to noise, accuracy, and resolving power and compared it to the performance of the DoG, the TopHat transform, and MP. The performance of the method is evaluated on artificial 3D data, mimicking 3D data captured by fluorescence microscopy. The different methods are also evaluated using fluorescence microscopy data from a study of protein complex localization during mitosis. Input from two different experts is used as "ground truth" in the performance evaluation.

MATERIALS AND METHODS

Three different methods of detecting small point-source signals in 3D are compared to our proposed method 3DSWD. Characteristics of the different methods as well as their input parameters are discussed and results from applying the different methods to sample images are shown in the result section.

Previous Methods

TopHat transformation. The morphological TopHat transformation (7) is a commonly used method for the detection of signals seen as peaks and local maxima in 2D and 3D images. This method makes use of the difference between morphological opening and the original image. The TopHat transformation to find signals of high intensity is defined as follows:

$$I_{\text{TopHat}} = I - \max_B(\min_B(I)) \quad (1)$$

where I is the input image and $\max_B(I)$ is the maximum in a neighborhood defined by structuring element B , and $\min_B(I)$ is the minimum. The size of the structuring element is impor-

tant for the performance of the method. The structuring element should be slightly larger than the signals that are to be detected. As the PSF of the microscope often smears the signal more along the axial (z) direction than in the lateral (x, y) direction, the structuring element should be larger in the axial direction. The output image contains all the local maxima of the input image that are smaller (in a morphological sense) than the structuring element B and can be intensity thresholded to define individual signals.

Difference of Gaussians. The difference of Gaussians (DoG) is a set of filters that approximate the Mexican hat by subtracting a wide Gaussian from a narrow Gaussian (6). In a technical report by the same authors, a modified 2D DoG is proposed. Based on this, we define 3D DoG as:

$$\text{DoG}_{\gamma, \sigma} = \frac{A_S}{\gamma^3} e^{-\frac{1}{2\gamma^2\sigma^2}(x^2+y^2+\frac{z^2}{r_z^2})} - A_S e^{-\frac{1}{2\sigma^2}(x^2+y^2+\frac{z^2}{r_z^2})} \quad (2)$$

$$A_S = \frac{1}{r_z(2\pi)^{3/2}\sigma^3} \quad (3)$$

$$\sigma \approx \frac{R}{\gamma} \sqrt{\frac{1-\gamma^2}{-6\ln(\gamma)}} \quad (4)$$

$$I_{\text{DoG}} = I \otimes \text{DoG}_{\gamma, \sigma} \quad (5)$$

σ is the standard deviation for the wide Gaussian and γ ($0 < \gamma < 1$) gives the ratio of the standard deviation for the narrow and wide Gaussian. A_S is a normalization factor. R is the distance from the center to the zero crossing of the DoG in the lateral direction and r_z is the ratio of the axial to lateral resolution. The zero crossing of the DoG is of importance when choosing the correct standard deviation for the two Gaussian functions. In addition, the parameter γ can be used to regulate the valleys of the DoG. Narrower valleys will make the filter smaller and consequently better at separating signals that are close in space but on the other hand it will decrease the amount of noise reduction. Setting γ to a high value will tighten the DoG for better resolution. The resulting image I_{DoG} is an image with enhanced signals, which can be used together with a threshold to define the individual signals.

Multiscale product. Vermolen et al. (3) used a 3D extension of a method developed in (4) for detecting signals with MP from the à trous wavelet decomposition of the original image. The method is designed to find signals that resemble a 3D Gaussian intensity profile of a given lateral size. The 3D Gaussian approximates the 3D PSF and is therefore anisotropic. The Gaussian is defined as follows:

$$G(\sigma_i) = \frac{1}{r_z\sigma_i^3(2\pi)^{3/2}} e^{-\frac{1}{2\sigma_i^2}(x^2+y^2+\frac{z^2}{r_z^2})} \quad (6)$$

First, a Gaussian scale space of the image is produced by convolving the image g with Gaussian filters with different widths $G(\sigma_i)$.

$$g_i = g \otimes G(\sigma_i) \quad (7)$$

where $i = 0, 1, 2$ and \otimes is the convolution operator. A scale base, b , is used to define the different widths of σ_i ,

$$\sigma_i = b\sqrt{2}^i \quad (8)$$

The last step consists of multiplying the differences in the Gaussian scale space and producing an image I_{MP} with increased contrast in regions where signals are present. The MP can be thought of as a product of several different DoGs.

$$I_{MP} = (g - g_0)(g_0 - g_1)(g_1 - g_2) \quad (9)$$

The resulting output image I_{MP} can be thresholded to define regions representing individual detected signals.

Comparison of previous methods. For all described methods, a threshold is required to identify signals, and the output is a binary image defining regions of detected signals. As the location of the signals is of importance, we reduced all detected signals to only one pixel representing the approximate center for the signal. This was done by taking the product of the binary image, defining regions of detected signals, and the original image and locating the highest intensity pixel inside each region. This results in a set of pixel positions defining the center of each signal.

To illustrate the different signal detection methods, a 1D example of signals, artifacts, and noise was created. The 1D example can be divided into three segments. Segment 1 consists of a single signal, a pair of signals with the same intensity and a pair of signals with different intensity. Segment 2 consists of a single signal with added Gaussian noise and Segment 3 consists of a sharp edge. The true signals were created with a Gaussian function of $\sigma = 1$. The 1D test image together with input parameters and output from each of the three described methods can be seen in Figure 1. The result of the proposed 3DSWD, described below, is also included.

For Segment 1, all methods give strong output peaks for all true signals. An intensity threshold could separate the five peaks from the image background for all the methods. The DoG and TopHat possess an advantage over the MP regarding peak strength in the output. The relationships between the intensities of the signals in the input data are the same in the output for these two methods, i.e., relative signal intensities are maintained. This makes isolation of the peaks with a single threshold more straight forward. The MP will drown the peak with the lowest intensity, but it is still possible to define all peaks with a single threshold.

Segment 2 consists of a single peak and added Gaussian noise. Here, we can observe how the methods respond to noise. The TopHat does not reduce noise and strong noise may be detected as signals. The DoG will reduce noise, ena-

bling the peak to be identified by a single threshold. Also, the MP reduces noise very well and gives one strong peak for the true signal.

Segment 3 consists of a sharp edge and should not give a response in the output. The only method not responding with a peak is the TopHat. Apart from the proposed SWD, only the DoG produces an output signal where a single intensity threshold can be used to separate all the true signals from the background.

Three-Dimensional Stable Wave Detection

Our proposed method for detecting signals in 3D is based on the idea of the SWD (9). The original version of the SWD was developed for 2D images and we have modified it to work in 3D, therefore referred to as the 3DSWD. The 3DSWD can be divided into two parts: a detector and a verifier/separater. The aim of the detector is to enhance regions with point-like source signals. The verifier/separater has the aim of verifying if a high value from the detector in fact is a true signal. The verifier/separater does this by examining the direction of the slope around the detected point. By looking at the slope around a peak closely lying signals will also be separated.

The SWD uses the Fourier coefficients of the first harmonic wave in a Fourier series to find local intensity maxima or minima in an image. In the 1D case, we assume signals with width $T/2$. The total data of length N is divided into n overlapping frames $i = 1\dots n$ with length T . The frames should overlap more than $T/2$. Sine and cosine Fourier coefficients, b_i and a_i , of the first harmonic wave of the Fourier series are computed using

$$a_i = F_i \cdot C \quad (10)$$

$$b_i = F_i \cdot S \quad (11)$$

$$C(t) = \cos\left(2\pi\frac{t}{T}\right) \quad (12)$$

$$S(t) = \sin\left(2\pi\frac{t}{T}\right) \quad (13)$$

where F_i is the intensity data of frame i , and C and S are discrete cosine and sine functions with period T and $t = 0\dots T - 1$. A frame containing a signal can be defined by some criteria for a_i and b_i . In general, a frame containing a signal will have a negative cosine coefficient with the criteria ($a_i < |b_i|$) and be surrounded by one rising and one falling edge, i.e., a negative sine ($b_i < 0$) and a positive sine ($b_i > 0$).

The presented 3DSWD contains some modifications of the original SWD. As the signals of interest are often small and may be close to each other in space we let the frames overlap by $T - 1$, resulting in a regular convolution of the image with a cosine and sine function. The cosine filtration will work as the detector and the sine function will work as the verifier/separater. When constructing the cosine filters t is set to $0\dots T$, where

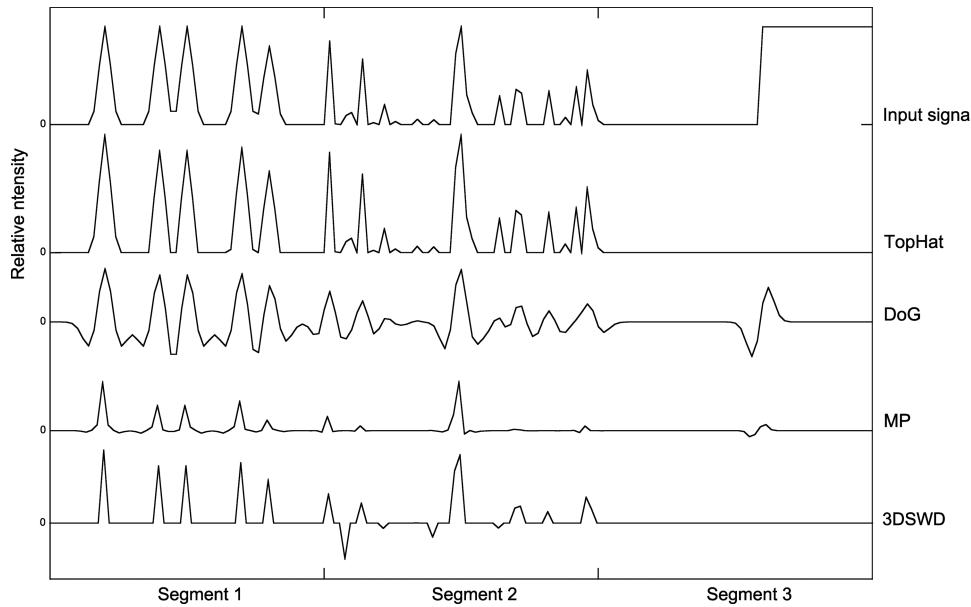


Figure 1. Response curve to an artificial input signal. Segment 1: one single signal, a pair of signals with the same intensity and a pair of signals with different intensities. Segment 2: one input signal with Gaussian noise of variance 0.1. Segment 3: a sharp edge. TopHat: kernel size 7. DoG: $\gamma = 0.9$ and $R = 2.5$. MP: $b = 1.3$ and 3DSWD: $T = 9$.

T is an even number, to make filters symmetric around a center point. We want the filter to sum up to zero, and as values from more than one period are used, cosine values at $t = 0$ and $t = T$ must be divided by 2. For the sine function, the period is half of the cosine to increase the ability to resolve closely lying signals. As a result, the frames of the sine function will be half the size of the frames of the cosine function.

For the SWD to work in 3D, the cosine and sine filters must be extended to 3D. The cosine filter is built up by combining three 1D cosine functions for all three directions x , y , and z into a 3D filter, as illustrated to the left in Figure 2. As we now have three directions, we need to divide the values of the cosine at $t = 0$ and $t = T$ by six to maintain a sum of zero for the filter. Because of the PSF of the signal, the period in the axial direction is made longer than in the lateral direction using the factor r_z which is the ratio of axial to lateral frequency. The equation for constructing a cosine filter C with dimensions is:

$$C(t_x, t_y, t_z) = \begin{cases} \cos(2\pi \frac{t_x}{T}) & \text{for } t_x = 0 \dots T, t_y, t_z = 0 \\ \cos(2\pi \frac{t_y}{T}) & \text{for } t_y = 0 \dots T, t_x, t_z = 0 \\ \cos(2\pi \frac{t_z}{r_z T}) & \text{for } t_z = 0 \dots r_z T, t_x, t_y = 0 \\ 0, & \text{elsewhere} \end{cases} \quad (14)$$

The sine filter works as a derivative filter describing the gradient in the image. It is important to get the orientation of

the gradient in all directions and, therefore, it is not possible to combine the three 1D filters into a single 3D filter. Instead one 1D filter is used in each direction, as illustrated to the right in Figure 2. To increase the specificity of the method, the diagonal directions can also be used. As in the case for the cosine filter, the period in the axial direction is increased by the factor r_z in the lateral direction.

$$S_x(t_x) = \sin(2\pi \frac{t_x}{T_S}); t_x = 0 \dots T_S \quad (15)$$

$$S_y(t_y) = \sin(2\pi \frac{t_y}{T_S}); t_y = 0 \dots T_S \quad (16)$$

$$S_z(t_z) = \sin(2\pi \frac{t_z}{r_z T_S}); t_z = 0 \dots r_z T_S \quad (17)$$

$$T_S = \frac{T}{2} + 1, \quad \text{for odd numbers of } \frac{T}{2} \quad (18)$$

$$T_S = \frac{T}{2}, \quad \text{for even numbers of } \frac{T}{2} \quad (19)$$

The image $I(x, y, z)$ is first convolved with the cosine filter C of period T , producing a_T . A threshold T_r is used on a_T to get a binary result defining potential signals. Next, the image I is convolved with the three different sine filters S_x , S_y and S_z resulting in b_x , b_y , and b_z . For a pixel to be classified as a true signal, the following criteria must be fulfilled.

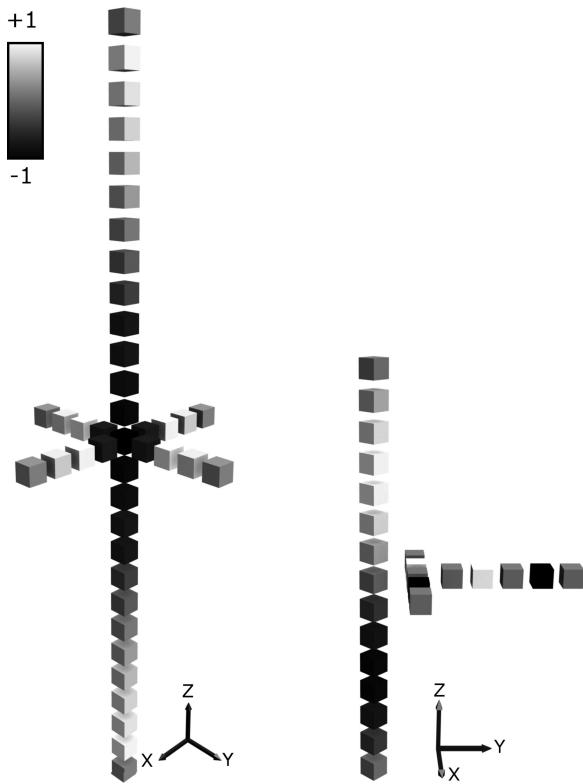


Figure 2. Left: Cosine filter in 3D. Right: Sine filter for three directions.

A pixel is a true signal if

- $-a_T(x,y,z) > T_r$ (threshold on cosine filtered image defining potential peak);
- $b_x(x-1,y,z) < 0$ and $b_x(x+1,y,z) > 0$ (peak surrounded by rising and falling edge in x -direction);
- $b_y(x,y-1,z) < 0$ and $b_y(x,y+1,z) > 0$ (peak surrounded by rising and falling edge in y -direction);
- $b_z(x,y,z-1) < 0$ and $b_z(x,y,z+1) > 0$ (peak surrounded by rising and falling edge in z -direction).

The result of applying a 1D version of the modified 3DSWD on the signal used in "Comparison of Previous Methods" section can be seen in Figure 1. All the true signals of Segment 1 are found, although the relative intensity between the peaks is not maintained. The noise reduction in Segment 2 is comparably good and a clear peak is seen for the true signal. We can also see that the peaks are narrow compared to the result from the other methods. This is due to the verifier/sePARATOR, which will only allow peaks that ful-

fill all the criteria described earlier. In Segment 3, we can see that there is no false response for the edge. The combination of the detector and verifier/sePARATOR makes the detection of the signal more robust than the other methods. It is in this case possible to separate and identify all the true peaks in all segments with a single threshold, without getting any false results. The performance of the method is investigated further in the results section.

Materials: Image Data

To verify the different methods and quantify their accuracy, a ground truth must be available. A 100% reliable ground truth is very difficult to produce for microscopy data. Instead, several simulated images with known signal position were created. Two types of artificial images were produced for two different tests. First, the method's robustness to noise was evaluated with a set of noisy artificial images. Second, detection of closely lying signals in 3D together with sensitivity to intensity was evaluated. Finally, a 3D fluorescence microscopy image was included to verify the performance of the methods. In this case the ground truth consisted of manual, visual detection of signals in 3D.

Evaluation of robustness to noise. To test the robustness of the different methods, an artificial image of size $500 \times 500 \times 20$ pixels with a fixed number of signals of decreasing intensity was created. Each artificial signal was a 3D Gaussian with a ratio of 3:1 for axial to lateral width, the size of the Gaussian in the lateral direction was 5 pixels and $\sigma = 1$. The intensity of the signals decreased from left to right while a constant amount of Poisson noise was added to the image. This resulted in an image with decreasing signal to noise ratio (SNR) from left to right. The definition for SNR used in this article was intensity of the signal, I_{Signal} , divided by the standard deviation of the noise σ_{noise} , eq. (20). The signal to noise ranged from 2.9 to 1, from left to right.

$$\text{SNR} = \frac{I_{\text{Signal}}}{\sigma_{\text{noise}}} \quad (20)$$

A map of the true center for each signal was also produced to simplify evaluation of the signal detection for each of the compared methods. The artificial image was re-created 100 times with randomized noise. Two rows of signals from the test data can be seen in Figure 3.

Evaluation of resolving power and sensitivity to signal intensity. When evaluating the performance of the method's ability to detect signals of varying intensity together with signals lying close in space, we divided the test in two parts: clo-



Figure 3. Central z-slice for two rows of signals from an artificial image with increasing amount of noise from left to right after filtration with an averaging filter of size $3 \times 3 \times 9$.

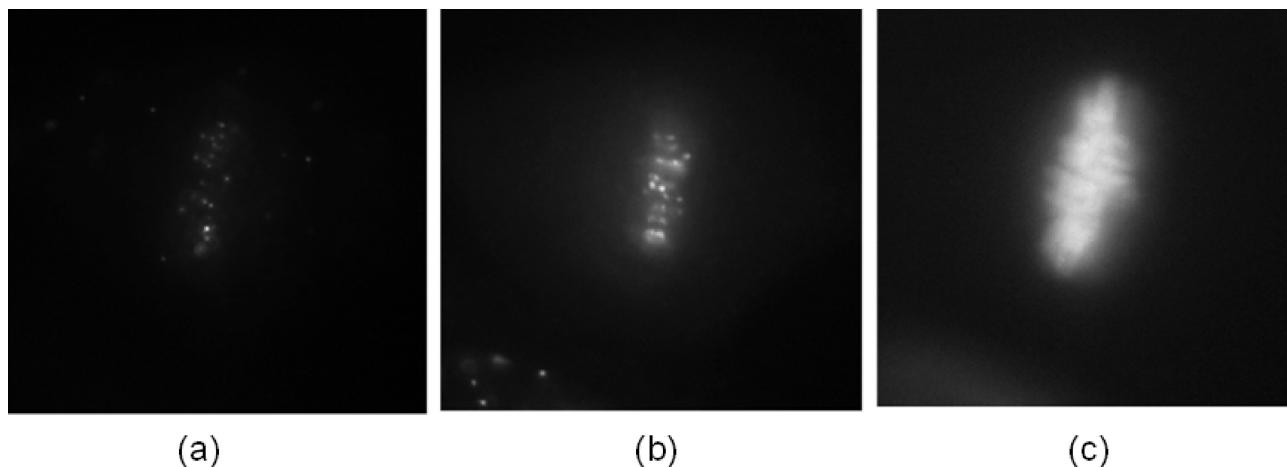


Figure 4. (a) PLA signals. (b) CREST signals. (c) Nuclear stain (DAPI) used as mask for marking regions of interests for analysis.

seness in lateral and axial direction. The artificial signals were created as previously described with a 3D Gaussian of a ratio 3:1 axial to lateral width and $\sigma = 1$. Here we wanted to study the detection resolution of closely laying signals and the effect of the intensity of the signals. Twelve columns and seven rows of pairs of signals were created with increasing intensity left to right. The ratio of the lowest to highest intensity was 1:2. The distance between the pairs ranged from 1 to 7 pixels. Similar images with 12 columns and seven rows were created to test resolution of the methods in the z-direction, with a distance between the centers of the signals ranging from 4 to 10 pixels. A small amount of Poisson noise was also added to the test images. The artificial image was re-created 10 times with randomized noise. SNR ranged from 16 to 33 from left to right. A map of the true center for each signal was also produced to simplify evaluation of the signal detection for each of the compared methods.

Microscope image data. To evaluate the 3DSWD on fluorescence microscopy data, we used images from a study of protein complex localization during mitosis. The aim of the study is to have a closer look (in 3D) of how many AuroraB-Survivin signals co-locate with a kinetochore marker in prometaphase cells versus metaphase cells. The proximity ligation assay (PLA) (11) was used to detect AuroraB-Survivin protein complexes, and CREST antiserum was used as a marker for the kinetochores. PLA converts the recognition of a protein complex by two or more antibodies into a circular DNA molecule. The circular DNA molecule is then amplified using rolling circle amplification and the localized concatameric product is detected using fluorescent probes, enabling visualization of protein complexes as distinct point-like fluorescent signals with high specificity. Figure 4a shows a z-slice from the data set with PLA-Cy3 signals in metaphase. The kinetochore pairs are marked with CREST-FITC, also seen as point-like signals in Figure 4b. Signals were imaged using a Zeiss Axiovert microscope equipped with a 63 \times (NA 1.4) objective, a Hamamatsu Orca-ER camera (Hamamatsu Photonics) and Metamorph imaging software (Universal Imaging), with xy resolution 0.015

$\mu\text{m}/\text{pixel}$ and z-stacks acquired at 0.25 μm z-steps. CREST was labeled with FITC and imaged using a HQ530/50 nm emission filter. Cy3-labeled PLA was imaged using a HQ610/75 nm emission filter. Each data set consists of a total of 49 z-slices of size 322 \times 334 pixels. Original data was compared to deconvolved image data, and as is often the case with weak signals, the deconvolved data contained artifacts created by the deconvolution algorithm. We therefore chose to work with the raw image data, adjusting the signal detection algorithms to handle the difference in axial and lateral image resolution. A simple graphical user interface (GUI) that allows selection of the true signals was created and presented to two experts from the application field in order to create a ground truth for evaluation of the automated signal detection using 3DSWD.

RESULTS

Robustness to Noise

The robustness to noise was compared for the three previous methods and the novel method for 3D signal detection proposed by us. The artificial test data consisted of 3D images with signals of decreasing intensity and random Poisson noise, as described in "Evaluation of Robustness to Noise" section. In each image, the number of true positive (TP) and false positive (FP) signals was counted for each noise level. To be considered as a TP signal, the detected center of a signal had to overlap within a distance of at most 2 pixels from the true center of the signal. FPs were defined as all detected signals that were not true signals. In addition, if two signals were detected inside a region of a true signal then one of them was counted as FP and the other as TP. The ratio of TP and FP signals to the number of true signal is plotted against SNR in Figure 5. The test image was re-created 100 times to estimate the mean and standard deviation of the counts. The artificial images were also filtered with an average filter of size 3 \times 3 \times 9 increasing the performance of all the methods.

The performance of all four methods is affected by a number of input parameters. To get an unbiased estimate of how well the methods perform, we tried to adjust the input parameters and thresholds so that the count of FPs stayed at

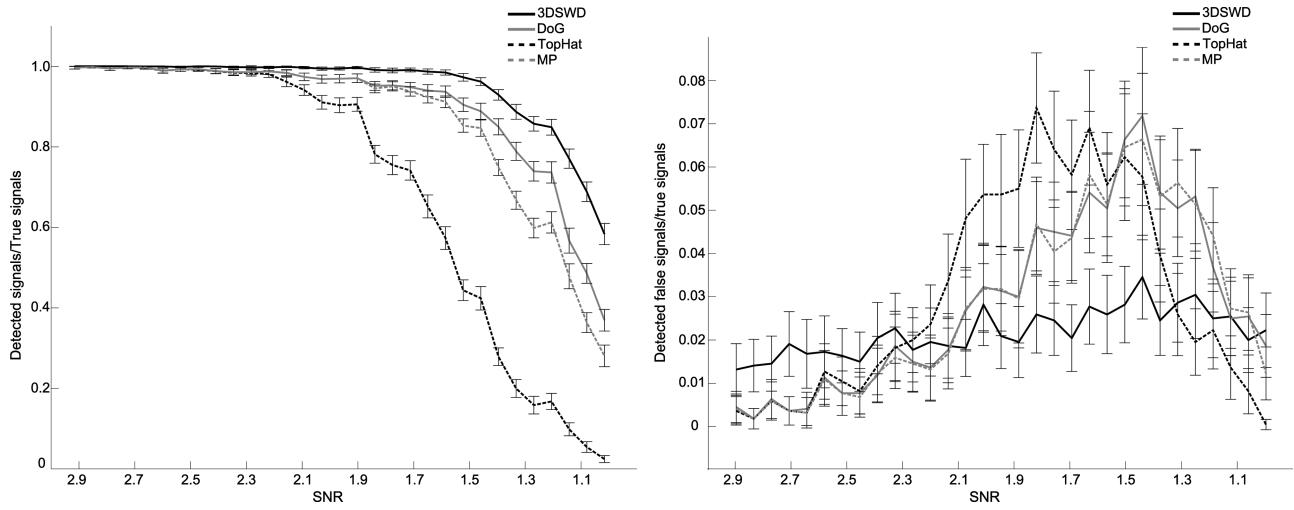


Figure 5. Ratio of TP and FP from robustness to noise test. Error bars represent a 99% confidence interval. Left: Ratio of true detected signals (TP) and true signals. Right: Ratio of false detected (FP) signals and true signals.

approximately the same low level, but above zero, for all methods. If the FP count is zero, we cannot conclude that the number of TPs is at the absolute maximum with that particular method. The thresholds were evaluated individually, and set to include around 3% FP. Once set, the same method specific optimized threshold was used for all noise levels, and all test images. The remaining input parameters were set individually for each method as follows:

TopHat: The structuring element should be slightly larger than the signals that are to be detected. The size of a signal was $5 \times 5 \times 15$ pixels, and a structuring element of $7 \times 7 \times 23$ pixels gave good results.

DoG: The first parameter r_z was set to 3 because the ratio of axial to lateral length of the signal was 3. The second parameter R is the radius of the signal. The diameter of the signals was 5 pixels, and thus R was set to 2.5. γ can range between 0 and 1, a high value will tighten the DoG giving it better resolution. A lower value of γ will make it more robust to noise. We want to have good robustness to noise and still be able to separate fairly closely lying signals, and $\gamma = 0.9$ proved to give good results.

MP: The first parameter r_z was set to 3, same as for the other methods. The value of b was set by trial and error to get the best robustness to noise while still not blurring the image too much, $b = 1.3$.

3DSWD: Once again, r_z was set to 3 as for the other methods. The value for the period T should be double the signal width and an odd number, this value was therefore set to 9.

A summary of the input parameters is presented in Table 1. In addition, the diagonal x - y filter directions of the 3DSWD verifier/separators were included in all experiments.

The result from the test data can be seen in Figure 5. The error bars represent the confidence interval for each noise level for the different methods at significance level 0.01.

At the highest SNR, all methods find almost all TP signals and a few FP. When the SNR reaches around 2.3 the number

of TP starts to decrease and the number of FP starts to increase. From this point and onward TopHat has the fewest number of TP among all the methods tested. The 3DSWD has the highest number of TP and from SNR level 2.9 to 1.6 it finds almost all signals and shows low variance in the number of TP. The 3DSWD also has the lowest total amount of FP among all the methods. The 3DSWD has fewer FP and higher number of TP which leads to the conclusion that the 3DSWD is less sensitive to noise compared to the other methods used in this test.

Resolving Power and Sensitivity to Signal Intensity

To evaluate the different methods regarding resolving power and sensitivity to signal intensity, we constructed an image containing signals at different distances and signal intensities ("Evaluation of Resolving Power and Sensitivity to Signal Intensity" section). All parameters, except for threshold, for each of the methods were kept the same as when testing the sensitivity to noise since we wanted to test the resolving power while maintaining the robustness to noise for each of the methods.

First, we tested the resolving power for each of the methods without any difference in intensity. All the methods, except the DoG, could resolve signals that were separated with 2 pixels in the lateral direction; DoG could only resolve signals that were separated by 3 pixels. In the axial direction, the TopHat could resolve signals separated by 6 pixels while MP and 3DSWD separated signals that were 7 pixels apart and DoG could only separate signals that were 9 pixels apart.

Second, we wanted to test both resolving power and sensitivity to signal intensity. Figure 6 shows an image with increasing signal intensity from left to right and decreasing distance between signals from top to bottom (images have been inverted for better visualization). Here, the objective was to find as many signals as possible in the image. A small amount of Poisson noise is present meaning that all the meth-

Table 1. Parameters used for all methods and all tests in the result section

METHOD	PARAMETER 1	PARAMETER 2	PARAMETER 3	PARAMETER 4
TopHat	$B = 7 \times 7 \times 23$	—	—	—
DoG	—	$r_z = 3$	$\gamma = 0.9$	$R = 2.5$
MP	$b = 1.3$	$r_z = 3$	—	—
SWD3D	$T = 9$	$r_z = 3$	—	—

ods are dependent on an intensity threshold. A low threshold value will cluster signals that are close in space (bottom row of Fig. 6) and a high threshold will miss the weakest signals (left column of Fig. 6) while clusters are separated. If the threshold is increased signals on the left will be lost but the gain will be detection and separation of close signals to the right of Figure 6. A range of threshold values were tested to find the optimal threshold. In this case, the optimal threshold corresponds to the value that will give the highest number of correctly detected signals. The input parameters were optimized to find as many true signals as possible. Since there is some noise present there will be some false positive signals for each method also. In Figure 6 each method has a success region, which contains all the detected pairs of signals. Outside the individual regions, the method in question failed in separating the signals or in detecting any of the signals.

All the methods were tested on 10 images and correctly detected signals, false detected and missed signals were measured. In order to be classified as a TP, the signal had to be detected in more than 5 out of the 10 test images. The FP is an average of the number of falsely detected signals in all test images. FN is the number of total signals minus the number of TP. The results are summarized in Table 2.

In the lateral direction, Figure 6a and Table 2, it can be seen that DoG resolved the fewest number of pairs of signals,

but has no FP at all. The MP and TopHat separate the same number pairs of signals, but the TopHat has a higher number of FP than the MP. The staircase looking shape of the success region shows that the three methods, DoG, MP, and TopHat, lose some of the pairs of signals due to the difference in intensity. The 3DSWD resolved the maximum number of signals due to the dual detector and verifier and has only a few FP in the lateral direction.

The same type of test was performed for the axial direction, Figure 6b. Here the distance between two signals decreases top to bottom and intensity increases from left to right. Each column in Figure 6b shows a y - z slice of the center-point for the signals for better visualization of the results. Here we see that the TopHat and 3DSWD has a slightly higher resolution power than the other methods. The DoG can only resolve a few of the pairs of signals. As in the lateral case, we can clearly see the staircase shape of the DoG, MP, and the TopHat success region. The 3DSWD is not affected by the difference in the intensity of the signals and, therefore, the whole resolving power of the method can be utilized. The 3DSWD and TopHat has quite a few FP signals for the test in the axial direction, while MP only has a few and DoG has none. Because of the 3DSWD low sensitivity to difference in signal strength, it can make full use of its resolving power and therefore finds the highest number of signals in both these tests.

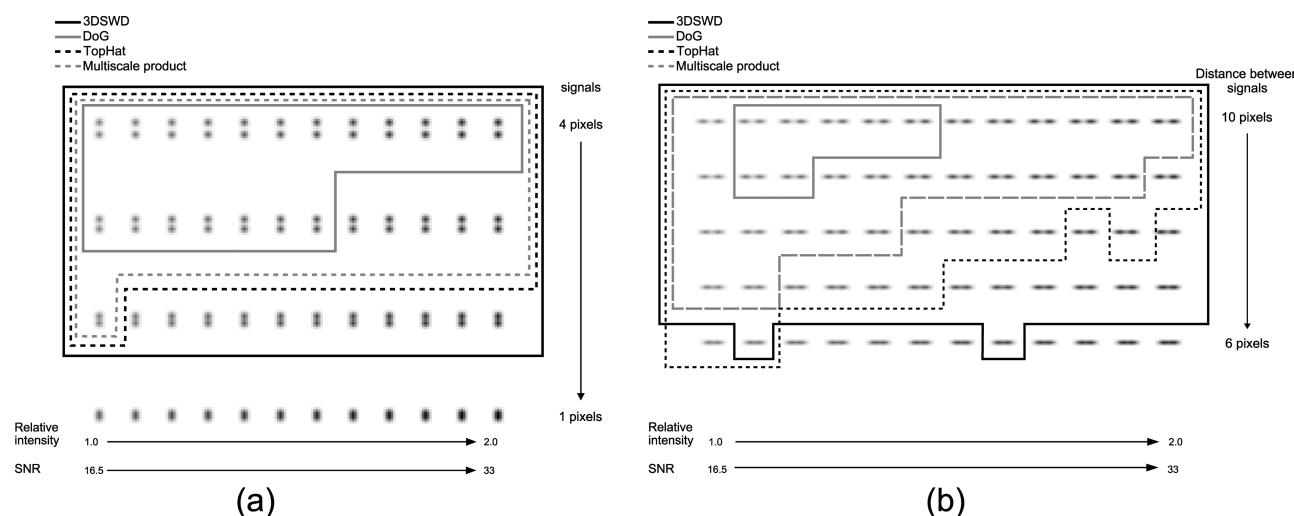


Figure 6. Comparison of resolution power on signals in artificial 3D data. Images show a single central xy- (a) and yz- (b) slice, and have been inverted for better visualization. Signal intensity increases from left to right, and the distance between signals decreases from top to bottom. Left: Lateral resolution power. Right: Axial resolution power. Delineations mark regions where each of the different methods was successful in separating a pair of signals.

Table 2. True positive (TP), false positive (FP), and false negative (FN) signal counts for each of the different methods applied to the image

	3DSWD		DoG		MP		TopHat	
	X-Y	Z	X-Y	Z	X-Y	Z	X-Y	Z
TP	156	129	133	75	141	107	141	120
FP	0.6	8.8	0	0	2.2	3.5	8.3	8.6
FN	0	39	23	93	15	61	15	48

Numbers are averages over 10 images with varying random Poisson noise.

Results on Fluorescence Microscopy Data

A simple GUI allowed experts from the application field to select true signals for creation of a gold standard, or ground truth for evaluation of the automated signal detection using DoG, TopHat, MP, and 3DSWD. Inputs were obtained from two experts (E1 and E2), with each expert doing the selection twice (t_1, t_2), leading to a total of four manual signal detections. The experts defined a signal by setting a mark at the 3D center point of the detected signal. These center points were compared with the center points found by the automated detection methods. The expert results were also compared with each other to measure the interobserver and intraobserver variability.

The expert results are referred to as reference data and the results to which they are compared as test data. A signal in the test data is considered a true positive (TP) if it is within a distance of ≤ 2 pixels from a detected signal in the reference data. A signal is considered a false positive (FP) if the signal is not overlapping with any of the signals in the reference data. In addition, if two signals from the test data are overlapping with only one signal in the reference, one will be classified as TP and the other one will be FP. False negatives (FN) are defined as the signals in the reference data that are not present in the test data. Precision, recall, and *F*-score were used to quantify the performance of the different methods.

$$\text{precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (21)$$

$$\text{recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (22)$$

$$F\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})} \quad (23)$$

The precision measures what fraction of the detected signals are actually true signals while recall measures what fraction of the true signals are detected. The *F*-score is a weighted average of the two measures precision and recall.

As can be seen from Figure 4, PLA signals exist also far away from the nucleus. These signals are not selected by the users as true signals. However, since these signals have the same structure as a true signal, the automated methods will pick them as true signals. To eliminate this bias, the data set

was masked based on the location of the nuclear stain (DAPI) seen in Figure 4c.

As the threshold is of great importance to the results in these types of comparisons, we calculated the *F*-score for a wide range of threshold values for each of the methods with one of the expert results as a reference. The threshold that gave the highest *F*-score was chosen as the threshold used in the comparison. We also chose to decrease and increase the threshold by 20% for all methods to see how that changed the result. The results are presented in Table 3.

When comparing the precision at the optimal threshold we can see that the MP, TopHat, and DoG all have similar and higher precision values than the 3DSWD. Most of the signals found by these methods are TP signals. However, we can see from the Recall data that the 3DSWD outperforms the other methods, meaning that the 3DSWD finds more TP than the other methods. Looking at the *F*-score, which is considered an overall measure of the method performance, we can see that the 3DSWD has the highest value among the tested methods. The MP and DoG have similar, but slightly lower values. TopHat has the lowest *F*-score value of all the tested methods. When looking at the *F*-score values of the expert results, we can see that the inter and intra observer variability is between 0.67 and 0.75. From this we can see that the performance of the 3DSWD is comparable to the results achieved by manual detection of the signals.

Decreasing the threshold by 20% does not affect the precision measure much for most of the methods except for the 3DSWD. The strength of the verifier/separator in the 3DSWD is that it makes it possible to decrease the threshold and still not find false signals. This results in an optimal threshold that is lower than for the other methods. When this threshold is decreased even more, here by 20%, we pick up a lot of the weak signals and noise which causes the precision to decrease significantly. The recall is increased for the 3DSWD when we decrease the threshold by 20% as more TP are picked up. For the other methods, the recall is decreased as signals clump together when the threshold is decreased, causing two true signals to result in one TP and one FN. The verifier/separator restricts this from happening for the 3DSWD.

The decrease in threshold by 20% decreases the *F*-score for all methods but the most significant decrease is seen in the 3DSWD. Yet, it still has the highest average *F*-score of all the methods.

Table 3. Comparison of method performance for DoG, 3DSWD, TopHat, and MP

	Experts				DoG		3DSWD			TopHat			MP			
Precision	E1t1	E1t2	E2t1	E2t2	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%		
E1t1	1.00	0.74	0.70	0.75	0.96	0.96	0.99	0.66	0.93	0.96	0.97	0.97	1.00	0.96	0.96	0.97
E1t2	0.74	1.00	0.67	0.71	1.00	1.00	1.00	0.70	0.96	0.97	1.00	0.99	1.00	0.99	1.00	1.00
E2t1	0.70	0.67	1.00	0.70	0.94	0.95	0.95	0.65	0.92	0.92	0.94	0.94	0.94	0.94	0.94	0.94
E2t2	0.75	0.71	0.70	1.00	0.94	0.94	0.97	0.67	0.93	0.95	0.96	0.97	0.98	0.95	0.96	0.97
Average	0.73	0.71	0.69	0.72	0.96	0.96	0.98	0.67	0.93	0.95	0.97	0.97	0.98	0.96	0.97	0.97
Recall	E1t1	E1t2	E2t1	E2t2	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%		
E1t1	1.00	0.74	0.70	0.75	0.49	0.52	0.51	0.66	0.63	0.62	0.46	0.49	0.46	0.52	0.54	0.54
E1t2	0.74	1.00	0.67	0.71	0.45	0.48	0.46	0.62	0.58	0.56	0.43	0.44	0.41	0.48	0.49	0.49
E2t1	0.70	0.67	1.00	0.70	0.41	0.45	0.43	0.57	0.54	0.52	0.39	0.41	0.37	0.44	0.45	0.45
E2t2	0.75	0.71	0.70	1.00	0.52	0.55	0.55	0.73	0.69	0.66	0.50	0.53	0.49	0.56	0.58	0.58
Average	0.73	0.71	0.69	0.72	0.47	0.50	0.49	0.64	0.61	0.59	0.44	0.47	0.43	0.50	0.52	0.51
F-score	E1t1	E1t2	E2t1	E2t2	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%	-20% OT	+20%		
E1t1	1.00	0.74	0.70	0.75	0.64	0.68	0.68	0.66	0.75	0.75	0.63	0.65	0.63	0.68	0.69	0.69
E1t2	0.74	1.00	0.67	0.71	0.62	0.65	0.63	0.66	0.72	0.71	0.60	0.61	0.58	0.64	0.66	0.66
E2t1	0.70	0.67	1.00	0.70	0.58	0.61	0.59	0.61	0.68	0.66	0.55	0.58	0.53	0.60	0.61	0.60
E2t2	0.75	0.71	0.70	1.00	0.67	0.69	0.70	0.70	0.79	0.78	0.65	0.69	0.65	0.70	0.72	0.73
Average	0.73	0.71	0.69	0.72	0.63	0.66	0.65	0.66	0.74	0.73	0.61	0.63	0.60	0.65	0.67	0.67

The left column shows inter- and intra-observer variation by comparing detection results performed by different experts (E1 and E2) at different times (t1 and t2). An optimal threshold, OT, leading to the maximal F-score, as well as a threshold decreased and increased by 20% of the OT, was tested. Colors reflecting the numerical values have been added for better visualization of the results (red = high, blue = low).

An increase in the threshold by 20% increases the precision for all methods slightly, which seems to follow the logic. Furthermore, the recall is decreased for all methods when the threshold is increased. Since we have a higher threshold now only the really strong signals are picked up causing the number of TP to decrease. The F-score is decreased for all methods except the MP which has the same F-score as when we used the optimal threshold. Also here, the 3DSWD has the highest F-score average among the methods tested.

DISCUSSION AND CONCLUSIONS

We present a method, 3DSWD, for the detection of fluorescent point-like signals. This method has shown better performance than the conventional methods that were used for comparison. The strength lies in the concept of using both a

detector and verifier/sePARATOR that in combination make the signal detection more stable. The detector has the aim of enhancing regions with point-like source signals while the verifier/sePARATOR, by examining the direction of the slope around the point, verifies if a high value from the detector in fact is a true signal. There is a contradiction between having high robustness to noise and high resolutionpower. The 3DSWD has proven to be robust to high levels of noise while the resolution, detection, and separation of closely lying signals is not affected as much as for the other methods used in this comparison. The size of the verifier/sePARATOR to the detector can be altered in order to increase or decrease robustness to noise and hence decrease or increase the resolution power, depending on the image data being analyzed. As a rule of thumb, the period T of the detector should reflect the size of a true signal while the period of the verifier/sePARATOR

should be $T/2$ to give a good balance between sensitivity to noise and ability to resolve closely lying signals.

To evaluate the performance on fluorescence microscopy data and also to demonstrate the usability of the method on biological applications, the method was also tested on fluorescence microscopy images from mitosis research where PLA was used to detect the AuroraB-Survivin protein complexes and CREST antiserum was used as a marker for kinetochores. The purpose of the study is to measure the colocalization of AuroraB-Survivin signals and kinetochore markers in prometaphase cells versus metaphase cells. Two experts have performed manual detection of the signals and this manual detection was compared to the automated signal detection of the 3DSWD and three previous methods. The 3DSWD showed the best performance of the methods tested. The other methods had a slightly higher precision than the 3DSWD but they all had a significantly lower recall. In addition, the *F*-score of the 3DSWD was comparable with the scores of the manual detection of signals. It becomes very clear that a high threshold leads to high precision, while reducing recall, and the other way around. To make the comparison as correct as possible, we tried a wide range of thresholds for all methods and evaluated each to find the optimal threshold as compared to the manual signal detection.

The parameters for the 3DSWD are intuitive to set when the lateral and axial size from the point-spread function of the system is known, making it easy to produce initial good results when tuning input parameters. Signals of different sizes are commonly seen in real images. The methods here have not been tested on finding signals of different size. The 3DSWD targets one size of signals but can be used in a loop over a range of values of T to find signals of different sizes. In a similar way, the parameters for the other 3D signal detection methods may also be varied for finding signals of different sizes.

The tests made in this article assume that signals are correctly sampled during image acquisition. If signals are undersampled, which will be the case when imaging signals smaller than the resolution of the microscope, the performance of all methods will be decreased. It should, however, be possible to detect signals as long as they result in a local maximum in the voxel volume. In addition, very densely packed signals will be difficult to resolve with all of these methods.

All the previous methods for 3D signal detection consist only of a detector as opposed to the 3DSWD which also consists of a verifier/separator. The objective in this study was to evaluate the performance of the 3DSWD in comparison to previously published methods and, therefore, the addition of a verifier/separator was not made to the other methods. An evaluation of how these conventional methods could be improved with the addition of a verifier/separator is currently under investigation. Initial results show that the performance of all these methods can be improved with the addition of a verifier/separator.

In conclusion, the 3DSWD shows superior robustness to noise and signal resolving power as compared to previously described methods for the detection of point-source signals in 3D fluorescent microscopy.

ACKNOWLEDGMENT

The authors thank Mariaana Vuoriluoto, Marko Kallio, and Jeroen Pouwels, at VTT Medical Biotechnology in Turku, Finland, for image data and expert inputs.

LITERATURE CITED

- Glory E, Murphy RF. Automated subcellular location determination and high throughput microscopy. *Dev Cell* 2007;12:7–16.
- Vermolen BJ, Garini Y, Mai S, Mougey V, Fest T, Chuang T C-Y, Chuang A Y-C, Wark L, Young IT. Characterizing the three-dimensional organization of telomeres. *Cytometry Part A* 2005;67A:144–150.
- Vermolen BJ, Garini Y, Young IT, Dirks RW, Raz V. Segmentation and analysis of the three-dimensional redistribution of nuclear components in human mesenchymal stem cells. *Cytometry Part A* 2008;73A:816–824.
- Olivo-Marin J-C. Extraction of spots in biological images using multiscale products. *Pattern Recogn* 2002;35:1989–1996.
- Zhang B, Zerubia J, Olivo-Marin J. Gaussian approximations of fluorescence microscope point-spread function models. *Appl Opt* 2007;46:1819–1829.
- Kruizinga P, Petkov N. Computational model of dotpattern selective cells. *Biol Cybernetics* 2000;83:313–325.
- Meyer F. Iterative image transformations for an automatic screening of cervical smears. *J Histochem Cytochem* 1979;27:128–135.
- Netten H, Young IT, Vanlylt LJ, Tanke HJ, Vrolijk H, Sloos WCR. FISH and chips: Automation of fluorescent dot counting in interphase cell nuclei. *Cytometry* 1997;28:1–10.
- Dupac J, Hlaváč V. Stable wave detector of blobs in Images. In: Proceedings of 28th Annual Symposium of the German Association for Pattern Recognition, LNCS 4174. Berlin, Germany: Springer-Verlag; 2006. pp 760–769.
- Pinidiyarakachchi A, Zieba A, Allalou A, Pardali K, Wählby C. A detailed analysis of 3D subcellular signal localization. *Cytometry Part A* 2009;75A:319–328.
- Söderberg O, Gullberg M, Jarvius M, Ridderstråle K, Leuchowius K-J, Jarvius J, Wester K, Hydbring P, Bahram F, Larsson L-G, Landegren U. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* 2006;3:995–1000.