QUANTITATIVE COMPARISON OF SPOT DETECTION METHODS IN LIVE-CELL FLUORESCENCE MICROSCOPY IMAGING

Ihor Smal[*], *Marco Loog*[‡], *Wiro Niessen*[*], *Erik Meijering*[*]

[*]Biomedical Imaging Group Rotterdam,
Departments of Medical Informatics and Radiology
Erasmus MC – University Medical Center Rotterdam, The Netherlands

[‡]Pattern Recognition Group
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology, The Netherlands

ABSTRACT

In live-cell fluorescence microscopy imaging, quantitative analysis of biological image data generally involves the detection of many subresolution objects, appearing as diffraction-limited spots. Due to acquisition limitations, the signal-to-noise ratio (SNR) can be extremely low, making automated spot detection a very challenging task. In this paper, we quantitatively evaluate the performance of the most frequently used supervised and unsupervised detection methods for this purpose. Experiments on synthetic images of three different types, for which ground truth was available, as well as on real image data sets acquired for two different biological studies, for which we obtained expert manual annotations for comparison, revealed that for very low SNRs ($\approx$2), the supervised (machine learning) methods perform best overall, closely followed by the detectors based on the so-called h-dome transform from mathematical morphology and the multiscale variance-stabilizing transform, which do not require a learning stage. At high SNRs ($>$5), the difference in performance of all considered detectors becomes negligible.

***Index Terms*—** Object detection, noise reduction, image filtering, machine learning, fluorescence microscopy.

1. INTRODUCTION

Quantitative analysis of dynamic processes in living cells using fluorescence microscopy, which is a powerful tool used in biology for visualizing subcellular components [1], involves the detection of many objects of interest. The objects, labeled with fluorescent proteins, appear in the images as bright spots, each occupying only a few pixels. Automated image analysis, which is highly needed for modern high-throughput studies in proteomics, functional genomics and drug screening, is still a great challenge due to limitations in the acquisition process. The signal-to-noise ratio (SNR) is usually very low, as the illumination intensities are kept low during experiments to prevent photobleaching and photodamage [2]. Also, the spatial resolution is rather coarse (on the order of 100 nm) compared to the size of subcellular structures (typically only several nanometers in diameter), resulting in diffraction-limited appearance. As a consequence, it is often difficult, even for expert biologists, to distinguish objects from irrelevant background structures or noise.

In practice, automated detection methods either corrupt the analysis with the presence of nonexistent objects by reporting (too many) false positives, or they bias the analysis towards the (too few) objects that are clearly distinguishable. In time-lapse imaging, where the objects of interest are to be tracked over time to study their dynamics, conventional tracking algorithms, which consist of separate detection (spatial) and linking (temporal) stages [2], also require accurate detectors – poor detection likely causes the linking procedure to yield nonsensical tracks, where correctly detected objects in one frame are connected with false detections in the next (and vice versa), or where tracks are terminated prematurely because no corresponding objects were detected in the next frame(s). Modern tracking approaches, based on Bayesian estimation [3], avoid the hard decision thresholds in the detection stage of conventional approaches. Nevertheless, even in probabilistic tracking frameworks, some form of "deterministic" object detection is still necessary in the track initiation and termination procedures [3], again illustrating the relevance of having a good spot detector.

Several detectors have been proposed in the literature, and the classic, relatively simpler methods have been compared previously for tracking [4,5], but a thorough quantitative comparison including recent, more complex methods is missing. In this paper, we compare seven unsupervised and two supervised methods that are used for object detection in fluorescence microscopy. These range from relatively simple local background subtraction [1], to linear or morphological image filtering [3, 6–8], wavelet-based techniques [9, 10], and machine learning methods [11]. We quantify their performance using both synthetic images and real image data from different biological studies.

2. DETECTION METHODS

The subresolution objects in our studies can be accurately modeled using a Gaussian approximation of the point-spread function (PSF) of the microscope [1, 7]. This involves two parameters: $\sigma_{\max}$ and $\sigma_{\min}$. To model different types of sub-

Financially supported by the Netherlands Organization for Scientific Research (NWO) through grants 639.022.401 and 639.021.611.

cellular structures (round or elongated appearance), we used symmetric ($\sigma_{\max} = \sigma_{\min}$) and asymmetric ($\sigma_{\max} > \sigma_{\min}$) Gaussian intensity profiles [3].

Each detector considered in this paper consists of three steps, but implements them differently. First, the noisy input image $\mathcal{I}$ containing the objects of interest (on a possibly nonuniform background) is preprocessed using noise reduction methods, such as Gaussian smoothing [1], which increase the SNR and improve object visibility. Second, the filtered image $\mathcal{J}$ is transformed into a grayscale classification map $\mathcal{C}$ using a signal processing step that is unique for each detection method and enhances the denoised fluorescent signal only in image regions where the actual objects are, while suppressing background structures. For some methods, these two steps are combined into a single algorithm. Third, to obtain the number of objects and extract position information from the classification map $\mathcal{C}$, hard (binary) decision thresholds need to be applied. The binary classification map $\mathcal{C}_B$ is obtained by applying a threshold to the signal magnitude in the grayscale map $\mathcal{C}$. Additional thresholding on object shape/size in the binary map $\mathcal{C}_B$ may also be performed.

Two types of detectors were considered in our study: unsupervised and supervised. The former implicitly or explicitly assume some object appearance model and contain parameters that need to be adjusted in order to get the best performance for a specific application. Seven methods of this type, frequently used in microscopy image analysis, were included in the evaluation: wavelet multiscale products (WMP) [9], multiscale variance-stabilizing transform (MSVST) [10], top-hat filter (TH) [6], spot-enhancing filter (SEF) [8], morphological grayscale opening top-hat filter (MTH) [1], h-dome based detector (HD) [3], and two types of image-feature based detectors: one that uses curvature information without taking into account spot intensity (IFD_1), and one that combines both sources of information (IFD_2) [7].

Supervised methods, on the other hand, "learn" the object appearance from annotated training data – usually a large number of small image patches containing only the object intensity profiles (positive samples) or irrelevant background structures (negative samples). In order to make our comparison study complete we also included two machine learning (ML) approaches. The first one is the AdaBoost algorithm (AB), which was recently shown to perform well also for spot detection in molecular bioimaging [11]. The second method is Fisher discriminant analysis (FDA) [12], which is a classical and well-known linear classifier, but which (to our knowledge) has not been employed for spot detection in fluorescence microscopy up to now. It uses the same information as AdaBoost but is computationally less expensive.

3. EXPERIMENTAL RESULTS

In the experiments, performed on synthetic as well as real biological image data as described below, connected components found in the binary classification map $\mathcal{C}_B$ produced by each detector for each data set were counted as objects of interest. The position of the detected objects was compared to the ground truth (known exactly in the case of synthetic images and obtained manually by annotation in the case of real biological images). The objects were labeled as true positives (TP) if the correspondence with the ground truth was found and as false positives (FP) otherwise. To compare the algorithms, we considered two common performance measures: the true-positive rate, $\text{TPR} = N_{\text{TP}}/(N_{\text{TP}} + N_{\text{FN}})$, and (because the number of true negatives, N_{TN}, is not defined) the modified false positive rate, $\text{FPR}^* = N_{\text{FP}}/(N_{\text{TP}}+N_{\text{FN}})$. Here, the number of false negatives is defined as $N_{\text{FN}} = N_0 - N_{\text{TP}}$, with N_0 the number of objects in the ground truth. The two measures allowed construction of free-response receiver operation characteristic (FROC) curves to study the sensitivity of the methods to changes in parameter values.

3.1. Experiments with Synthetic Data

Setup: Two types of object appearance were modeled, using 2D Gaussian intensity profiles (GIPs), with $\sigma_{\text{max}} = \sigma_{\text{min}} = 100$ nm for round objects, and $\sigma_{\text{max}} = 250$ nm, $\sigma_{\text{min}} = 100$ nm for elongated objects. Next, three types of images (Type A, B, and C) were created, for each type of object shape and for different levels of Poisson noise in the range of SNR = 2–4 (see Fig. 1). Such SNRs are typical for the real image data acquired in our biological applications and are lower than the critical level of SNR = 4–5, at which several classical detection methods break down [4, 5]. Here, SNR is defined as the difference in intensity between the object and the background, divided by the standard deviation of the object noise [4]. The synthetic images of size 512×512 pixels (with pixel size $\Delta_x = \Delta_y = 50$ nm) contained 256 spots each, placed randomly within the image region with no overlaps in the intensity distributions. Type A images were constructed by adding a background level of 10 to GIPs (similar to previous studies [4]) and applying a Poisson noise generator independently to every pixel of the noise-free image. In the case of Type B images, the background level increased linearly in the horizontal direction (see Fig. 1), from a value of 10 at the left image border to 50 at the right border. Taking into account that the variance of Poisson noise is intensity dependent, we corrected the object intensities accordingly prior to application of the noise generator in order to keep the SNR constant over the whole image. Finally, type C images mimic the intensity distribution in the presence of large (compared to object size) background structures, which are sometimes present in the real image data and can be either larger subcellular structures or acquisition artifacts. In every experiment, the performance of the detection techniques for each object type was evaluated by computing N_{TP}, N_{FP} and N_{FN} for 16 images (each containing 256 ground truth objects) and averaging the results over the 4096 objects.

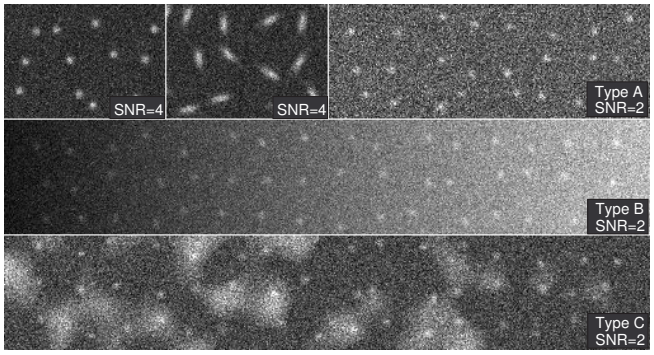

Fig. 1. Examples of synthetic images used in the experiments. The symmetrical GIPs are embedded into uniform (Type A), gradient (Type B), and non-uniform (Type C) backgrounds.

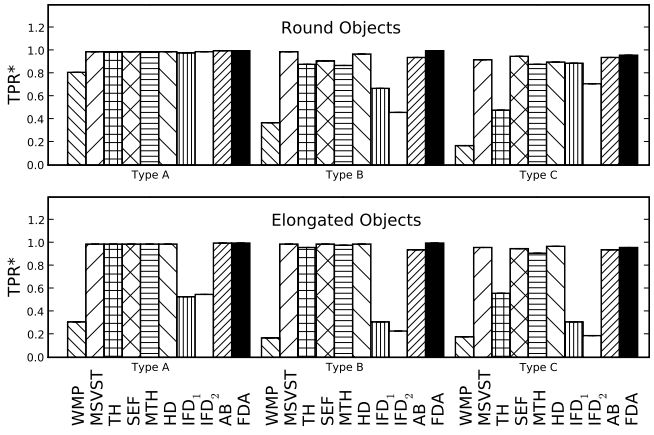

Fig. 2. Maximum detection probabilities (TPR*) at the level FPR* $= 0.01$ for all detectors applied to Type A, B, and C images at SNR $= 2$ in the case of the round and elongated objects.

Results: The performance of all detectors was compared at the level of FPR* $= 0.01$ for the different image data at SNR $= 2$ (see Fig. 2). From the results of the sensitivity analyses (data not shown in this paper due to space limitations), in which we studied the influence of small changes in method parameter values on the behavior of TPR and FPR* for different data types, we conclude that FDA and AB are superior to all other detectors and show the highest TPR* and the lowest sensitivity for all image data (Type A, B and C, SNR $= 2$). The WMP method demonstrated the worst performance and additionally showed high sensitivity to parameter changes, together with the TH detector, which demonstrated high performance only for Type A and B data. The IFDs are quite sensitive to parameter changes and do not have sufficiently high TPR in the case of the elongated objects. MSVST, HD, SEF and MTH demonstrated high TPR* and low parameter sensitivity, but none of these three detectors is better than the other two for *all* types of data. Finally we observed that the difference in performance between the methods decreases when the SNR of the image data increases, and we found that for SNR > 5 all methods perform equally well (TPR $= 1$).

3.2. Experiments with Real Data

Setup: The detection methods were also tested on real fluorescence microscopy image data from several biological studies, where the estimation of important kinematic parameters of subcellular particles in eukaryotic cells was of interest. Two types of representative image data sets were selected for these experiments. The first showed moving microtubule (MT) plus-ends, which have a round or elongated appearance. In the experiments, growing ends of MTs were tagged with so-called plus-end-tracking proteins (+TIP), resulting in typical fluorescent "comet-like" dashes (see Fig. 3(a, b)). A Zeiss LSM-510 confocal laser scanning microscope was used to acquire images of GFP+TIP movements at a rate of 1 frame per 1 or 2 seconds. The image sequences consisted of 30–50 frames of 512×512 pixels of size 75×75 nm^2. The second type of image data showed a variety of GFP-labeled vesicles (Rab6 and peroxisomes), which have a round shape in the images. In this case, HeLa cells and PEX3-GFP fusion were used [3]. Images were acquired on a Zeiss Axiovert 200M inverted microscope at a rate of 0.83 frames per second. The image sequences consisted of 100 frames of 1344×1024 pixels of size 64×64 nm^2 (see Fig. 3(c)).

Results: The parameters of each detection method were fixed to the optimal values following from the experiments on synthetic data. Since the ground truth was not available for the real data, the detection results were analyzed by expert visual inspection and in comparison with manual analysis using the freely available tracking tool MTrackJ for ImageJ (NIH, USA). The FROC curves for all the detection methods applied to two illustrative image data sets showing MTs (each image containing $\approx$ 80–100 spots at SNR $\approx$ 2–4) and one data set showing vesicles (containing ≈ 250 spots at SNR $\approx$ 3–8) are shown in Fig. 3. From these results, it was confirmed that the actual performance of the detection methods depends on the application. For the microtubule data, which contained round or elongated objects of almost identical sizes, we arrived at the same conclusions as in the case of the synthetic image data. For the vesicle data, however, the ranking of the detectors was found to be slightly different. These images have a higher SNR ($\approx$ 3–8) but contain spots of varying sizes. In all cases, the two ML detectors (FDA and AB) and the MSVST and HD detector showed the best overall performance.

4. DISCUSSION AND CONCLUSIONS

In this paper we have evaluated the performance of seven unsupervised and two supervised detection methods that are used in practice for the detection of small spots in fluorescence microscopy images. The results from experiments on synthetic images as well as real image data from two biological studies indicated that no detector outperforms all others in all considered situations. Overall, the supervised (machine learning) methods performed better on the synthetic images as well as on the real image data, but the differences in the

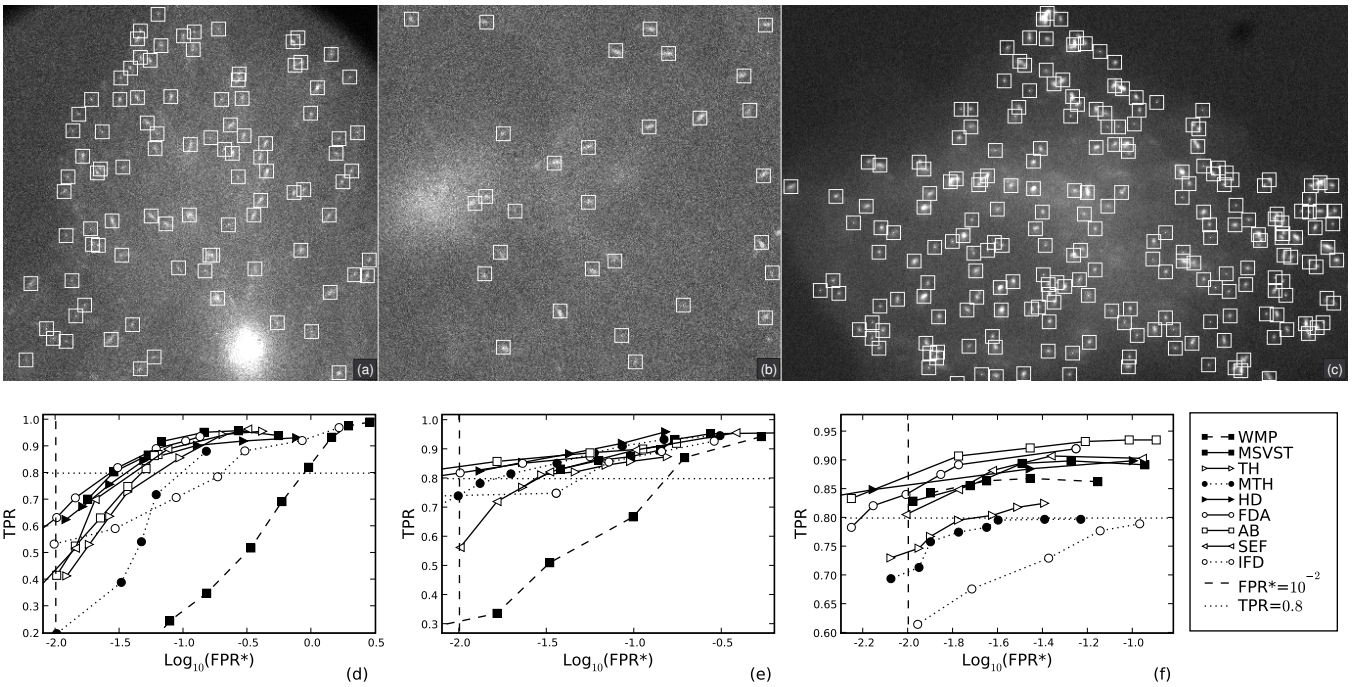

Fig. 3. Examples of real fluorescence microscopy images (a, b, c) with manual spot annotation (white squares) by an expert biologist serving as ground truth (courtesy A. Akhmanova, Erasmus MC). The corresponding FROCs (d, e, f) of all detection methods are shown below the images. In these plots, IFD represents the IFD_1 detector, which in the experiments on synthetic image data performed either similar to or better than the IFD_2 detector (see Fig. 2).

performance were not large compared to some of the unsupervised methods. Based on our extensive experiments, we conclude that when a detector with overall good performance is needed, the supervised AB or FDA detectors or the unsupervised MSVST or HD detectors are to be preferred. The main disadvantage of the supervised methods is that they require a training stage, which involves the extraction of positive and negative samples beforehand. We observed that the training should not be done using only clearly visible spots in image regions with high local SNRs: in order to achieve good classification performance, it must also include many hardly visible objects. Such manual annotation is extremely tedious, time consuming, and observer dependent. Taking this into account, the unsupervised MSVST [10] or HD detector [3] is much easier to use in practice. Finally, when the SNR is sufficiently high (> 5 as a rule of thumb), the other unsupervised detectors perform just as well, and require only minimal adjustment of their parameters to the specific application.

5. REFERENCES

[1] Q. Wu, F. A. Merchant, and K. R. Castleman, *Microscope Image Processing*, Elsevier Academic Press, Burlington, MA, 2008.

[2] E. Meijering, I. Smal, and G. Danuser, "Tracking in molecular bioimaging," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 46–53, 2006.

[3] I. Smal, E. Meijering, K. Draegestein, N. Galjart, I. Grigoriev, A. Akhmanova, M. E. van Royen, A. B. Houtsmuller, and W. Niessen, "Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering," *Med. Image Anal.*, vol. 12, no. 6, pp. 764–777, 2008.

[4] M. K. Cheezum, W. F. Walker, and W. H. Guilford, "Quantitative comparison of algorithms for tracking single fluorescent particles," *Biophys. J.*, vol. 81, no. 4, pp. 2378–2388, 2001.

[5] B. C. Carter, G. T. Shubeita, and S. P. Gross, "Tracking single particles: A user-friendly quantitative evaluation," *Phys. Biol.*, vol. 2, no. 1, pp. 60–72, 2005.

[6] E. J. Breen, G. H. Joss, and K. L. Williams, "Locating objects of interest within biological images: The top hat box filter," *J. Comput. Assist. Microsc.*, vol. 3, no. 2, pp. 97–102, 1991.

[7] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser, "Automatic fluorescent tag detection in 3D with super-resolution: Application to the analysis of chromosome movement," *J. Microsc.*, vol. 208, no. 1, pp. 49–64, 2002.

[8] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser, "Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1372–1383, 2005.

[9] J.-C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognit.*, vol. 35, no. 9, pp. 1989–1996, 2002.

[10] B. Zhang, J. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging," in *Proceedings of the IEEE International Conference on Image Processing*, 2007, vol. 6, pp. VI–233 – VI–236.

[11] S. Jiang, X. Zhou, T. Kirchhausen, and S. T. C. Wong, "Detection of molecular particles in live cells via machine learning," *Cytometry Part A*, vol. 71, no. 8, pp. 563–575, 2007.

[12] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley-Interscience, 2004.