Numerical investigation of integrated dielectric pillars to simplify machine learning classification of cells

A. Lugnan,[1,2] J. Dambre,[3] and P. Bienstman[1,2]

[1]Photonics Research Group, UGent – imec, Technologiepark 15, 9052 Ghent, Belgium
[2]Center for Nano- and Biophotonics (NB-Photonics), Ghent University, Technologiepark 15, 9052 Ghent, Belgium
[3]IDLab, UGent - imec, Technologiepark 15, 9052 Gent, Belgium

The computational power needed to classify interference patterns of biological cells is a major limit to the implementation of fast label-free cell sorting based on digital holographic microscopy and flow cytometry [1]. In this work we discuss some properties of a technique that employs a collection of integrated silica micro-pillars to significantly simplify the classification of raw holograms obtained from cells with different average nucleus size lighted by a laser.

Introduction

Fast sorting of biological cells is crucial for several biomedical applications, like diagnostics, therapeutics and cell biology. However, an accurate classification and separation of different cell types is usually expensive, time consuming and often requires the use of labels, such as fluorescent tags, that may spoil the samples hindering subsequent analyses [1]. Therefore, the development of label-free, high-speed, automated and integrated cell sorting solutions would enable new types of biological investigations and could considerably decrease the costs and duration of existing cell analysis that require selected samples.

Among several options, the integration of digital holographic microscopy in a microfluidic flow cytometry system is a promising candidate. In this technique, the classification is carried out considering the interference pattern (hologram) projected by the cells when lighted by a laser. The hologram is acquired by an image sensor and contains a large amount of information on the 3D refractive index structure of the cells[2], enabling nontrivial analysis and classification. However, the computational cost of reconstructing the image from the hologram acts as a bottleneck to the increase in the cell sorter throughput, e.g. by parallelization of the process.

An important reduction in the required computing power can be achieved by bypassing the reconstruction of the cell image and training a machine learning classifier algorithm on the acquired hologram[2][3]. Moreover, a further increase in classification speed can be obtained by employing a linear machine learning classifier while performing the required nonlinear operations (corresponding to hidden neurons in feedforward neural networks) in the optical domain. In the Extreme Learning Machine (ELM) approach[4], such nonlinear operations can be chosen randomly, optimizing only the linear readout classifier. An implementation of this kind of solution, based on integrated dielectric scatterers and applied to the classification of 2 cell types with different average nucleus sizes or shapes, was numerically demonstrated in [5]. The key concept is that the transfer function linking the phase-encoded optical signal generated by the cell presence to the corresponding intereference pattern acquired by an image sensor is nonlinear, and its complexity can be enhanced by the presence of interposed dielectric pillars (Figure 1).

When a real-life application is considered, a crucial requirement is that the performance of cell classification is robust to the fabrication errors affecting the integrated silica micro-

pillars employed as random scatterers. In this work we investigate this robustness property by showing the classification performances corresponding to the use of 6 different scatterer configurations.

Furthermore, we include the results corresponding to the classification of near field cell holograms, which are indicative of the feasibility of employing an all-optical readout classifier, that could operate at a much higher speed with respect to its electronic counterpart.
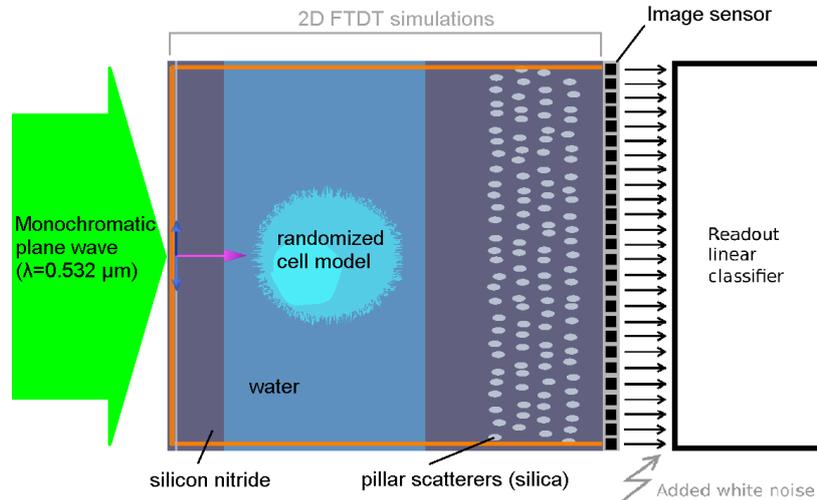

Figure 1. Schematic of the classification process. A monochromatic plane wave impinges on a microfluidic channel containing a cell in water ($n_{H_2O} \sim 1.34$), which has a low refractive index contrast ($n_{cytoplasm}=1.37$, $n_{nucleus}=1.39$); the forward scattered light passes through a collection of silica scatterers ($n_{SiO_2} \sim 1.461$) embedded in silicon nitride ($n_{Si_3N_4} \sim 2.027$) and organized in layers; the radiation intensity is then collected by an 1D image sensor, which is divided into bins (pixels); each pixel value is fed into a trained linear classifier (logistic regression) that consists of weighted sums (1 per class) of the pixel values.

Discussion and results

This work is a continuation of [5], which provided a proof of concept of a novel hardware machine learning technique by processing the results of thousands 2D Finite-difference time-domain (FDTD) optical simulations (Figure 1). The variability of the cells shape and position in the microfluidic channel is simulated using a randomized cell model that makes each simulated cell significantly differ from the others. In this paper, the results obtained from tens of thousands new 2D FDTD simulations are presented, in order to give an insight on important properties of the proposed technique, such as robustness to fabrication errors and employability of near field intensity patterns for a completely on-chip implementation. The simulation and classification design/parameters are the same as described in [5] for the case of nucleus size classification using a green laser, with the exception of:

- the pillar scatterers configuration, 7 different cases are considered in this work;
- in addition to the far field intensity, also the near field intensity at the simulated image sensor was calculated and used as sample for the classification task;
- the number of simulated cell scattering processes was increased from 3200 to 7200 (3600 per class) per scatterer configuration, to slightly reduce the overfitting and the variance of the classifier performance;
- the technique used to validate the classification results is 8-fold cross-validation, instead of bootstrapping.

The considered scatterer configurations (Figure 2) are: (a) 1 layer of scatterers; (b) 2 layers of scatterers; (c) 3 layers of scatterers; (d) 4 layers of scatterers; (e) 5 layers of scatterers; (f) 4 ordered layers of scatterers (no random displacement applied); (g) no scatterers. The spacing between the scatterer layers is 1.846 μm and the inter-scatterer vertical spacing is 1 μm. Except for the last two configurations, the amplitude of the random displacement applied to the scatterers is 0.225 μm.
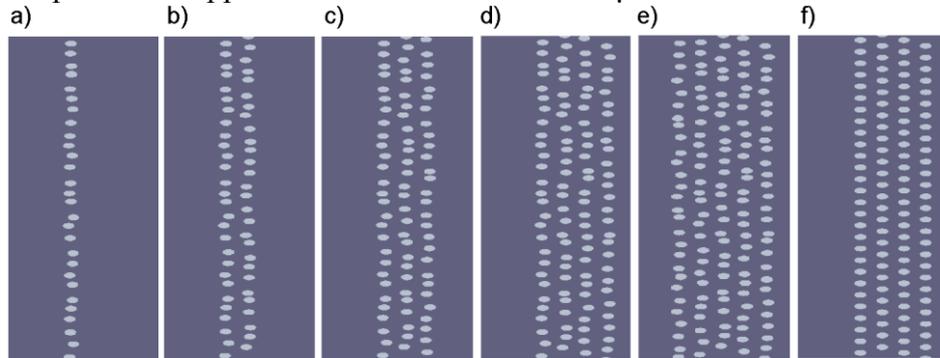

Figure 2. Configurations of silica pillar scatterers as employed in the FDTD simulations (see Figure 1). The 7th configuration, which does not comprise any scatterer, is not shown.

A comparison between the estimated classification error rates obtained using scatterer configurations *d* and *g* shows that the use of 4 layers of scatterers leads to a relative improvement of roughly 75% when the classification is performed on near field interference patterns, while when far field patterns are considered, the improvement is roughly of 50% (Figure 3). This is grosso modo true also for the other investigated scatterer configurations, suggesting that the advantage coming from the use of dielectric pillars is in general significantly larger in the nearfield rather than in the far field.

Considering for example a simulated 1D image sensor of 373 pixels, the obtained classification results show that the improvement in accuracy due to the presence of dielectric scatterers is essentially independent of the differences between the employed configurations (Figure 4).
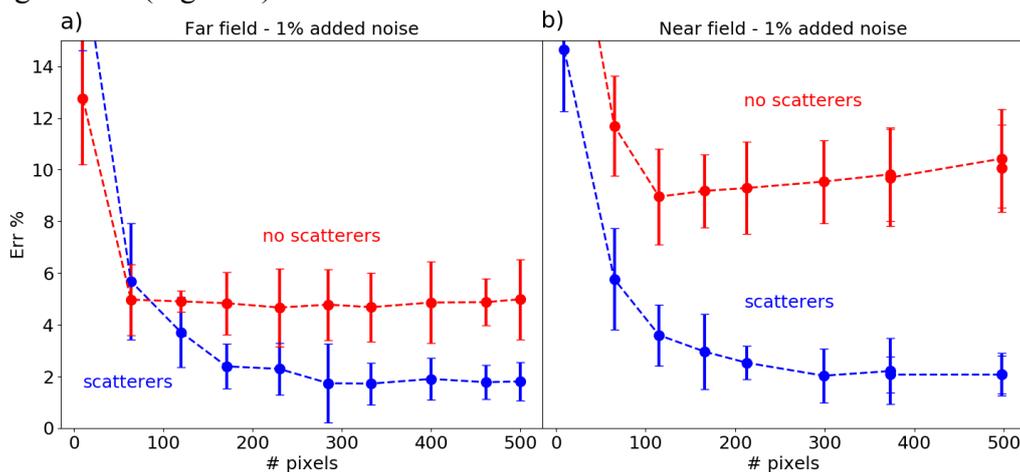

Figure 3. Classification error rate with 2 standard deviation confidence intervals as a function of the pixels number of the simulated 1D image sensor. A white Gaussian noise with standard deviation corresponding to 1% of the average acquired intensity was added to the interference patterns before classification in order to account for experimental noise. Blue and red data respectively correspond to scatterer configurations d and g, i.e. 4 layers of (disordered) scatterers and no scatterers. a) the linear classifier is applied on far field intensity patterns. b) the linear classifier is applied on near field intensity patterns.
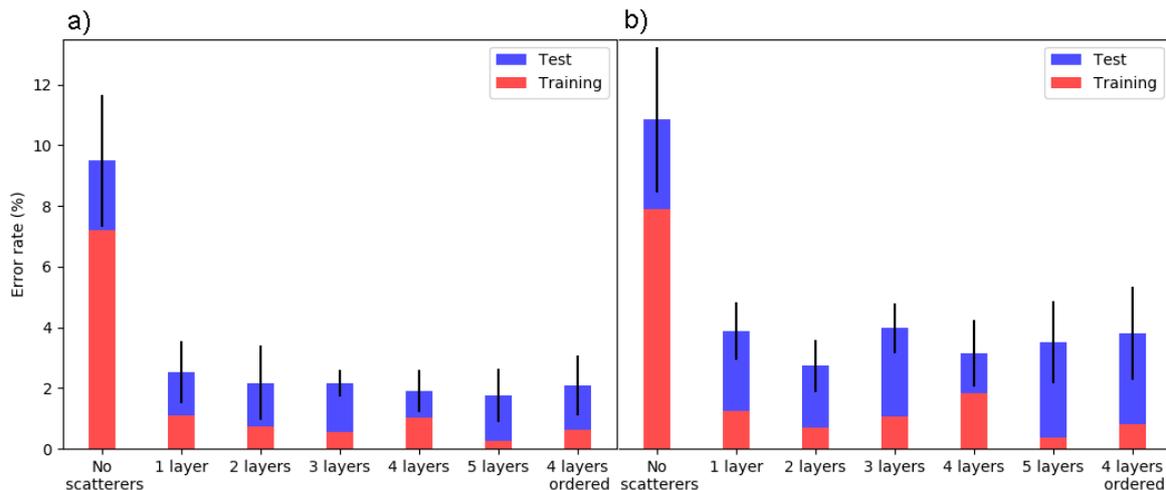
*Figure 4. Training (red) and test (blue) classification error rate with 2 standard deviation confidence intervals (black bars) for different scatterer configurations (Figure 2). A simulated image sensor with 373 pixels was considered. **a)** 1% added white noise level. **b)** 5% added white noise level.*

Conclusions

We discussed some properties of a machine learning technique for fast optical classification of biological cells based on the combination of integrated dielectric scatterers (simulated via FDTD method) and a linear classifier implemented in software. This technique was originally presented and numerically investigated in [5], considering the problem of classifying cells with different nucleus size or shape. In this work, we show that the relative improvement in classification accuracy due to the use of dielectric scatterers is roughly 75% when sample intensity patterns are acquired in the near field. This suggests that the speed of the discussed classification technique could be

further increased by considering an all-optical integrated linear classifier applied on the near field radiation exiting the scatterer stage.

Furthermore, the testing of classification performance for 6 different configurations of dielectric scatterers shows an unexpected robustness to configuration variations such as fabrication errors.

References

[1]   D. R. Gossett, W. M. Weaver, A. J. Mach, S. C. Hur, H. T. K. Tse, W. Lee, H. Amini, and D. D. Carlo: Label-free cell separation and sorting in microfluidic systems, *Anal. Bioanal. Chem.*,vol. 397, no. 8, pp. 3249-3267, 2010.

[2]   L. Lagae, D. Vercruysse, A. Dusa, C. Liu, K. de Wijs, R. Stahl, G. Vanmeerbeeck, B. Majeed, Y. Li, and P. Peumans: High throughput cell sorter based on lensfree imaging of cells, in *Proc. IEEE International Electron Devices Meeting*, pp. 333-336, 2015.

[3]   B. Schneider, G. Vanmeerbeeck, R. Stahl, L. Lagae, J. Dambre, and P. Bienstman: Neural network for blood cell classification in a holographic microscopy system, in *Proc. ICTON*, Budapest, 2015.

[4]   G. Huang, G.-B. Huang, S. Song, and K. You: Trends in extreme learning machines: A review, *Neural Networks*, vol.61, pp. 32-48, 2015.

[5]   Alessio Lugnan, Joni Dambre, and Peter Bienstman: Integrated pillar scatterers for speeding up classification of cell holograms, *Opt. Express*, vol. 25, no. 24, pp. 30526-30538, 2017.

[6]   F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay: Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.