Faculty of Informatics
Masaryk University

Virtual Cell Imaging

Methods and Techniques

Habilitation Thesis

Collection of Articles

Brno, October 2016   David Svoboda

Abstract

Nowadays, simulations are an essential part of the development of new technologies. They are namely used in cases when the observation of some events is technically too complicated, ethically unacceptable or if the repetition of these events is too expensive. This principle is also valid in the field of biomedical image processing where the images represent some specimen acquired using, for example, the optical microscope. Here, we study the visual appearance and behavior of selected biological specimen. The aim is to create a synthetic image or even the whole sequence of synthetic images that imitate all the visual aspects and behaviour of real living specimen. Subsequently, the standard image processing tasks like segmentation and tracking can be validated by applying them on these synthetic data and testing how they perform. Regarding time-lapse image sequences, we are not interested in image processing tasks only. Such computer generated sequences can reveal whether we correctly understand the dynamic processes that are modeled in the utilized simulation frameworks.

The aim of this thesis is an introduction and detailed description of the current methods and principles utilized when generating synthetic image data that imitates the images as acquired by the optical microscope. These include the methods generating the static images as well as time lapse image sequences describing the dynamic processes occurring in living cells. The output of this thesis includes the methods and principles that led to the implementation of the particular simulation toolkits. These toolkits serve as modules in the web-based framework called CytoPacq that is currently capable of generating five types of synthetic cell lines. All the aforementioned method are described in the conference and journal papers listed in this thesis.

Shrnutí

Simulace jsou v dnešní době neodmyslitelnou součástí vývoje nových technologií. Využívají se zejména v situacích, které není možné technicky či morálně realizovat a nebo je jejich opakování natolik nákladné, že by se opětovné sledování daných jevů vůbec nevyplatilo. Nejinak je tomu v oblasti zpracování biomedicínského obrazu pořízeného na optickém mikroskopu. Zde je předmětem simulací snímaný biologický preparát, jehož vizuální stránku popřípadě i pohyby sledujeme a snažíme se je napodobit. Cílem je pak vytvořit umělý obraz popřípadě celou sekvenci syntetických snímků, na nichž se dají následně testovat a ověřovat různé nově vytvářené metody analýzy obrazu. Mezi ně typicky patří segmentace či sledování živých buněk. V případě simulací sekvencí populací živých buněk se nejedná pouze o práci s obrazem. Na takto vygenerované populaci pak můžeme sledovat, zda se uměle vytvořená populace chová jako reálná populace, a tedy zda správně chápeme podstatu biologických procesů, které použitý simulační nástroj implementuje.

Předmětem této práce je představení metod a postupů, které se v současné době používají ke generování umělých obrazových dat imitujících snímky pořízené na optickém mikroskopu. Jedná se jak o statické scény tak i o generování dynamických procesů probíhajících například v živých buňkách nebo takových, které se uplatňují při formování buněčných populací. Výstupem této práce je mimo jiné i několik simulačních nástrojů, které postupně vznikaly jako implementace metod popsaných v jednotlivých publikacích, přiložených k této práci. Společným jmenovatelem všech představených nástrojů je rozhraní CytoPacq, v němž každý jednotlivý simulační nástroj slouží jako modul.

Melius est illuminare quam lucere solum.

Thomas Aquinas

Contents

Part I

Commentary

Chapter 1

Introduction

The advance in computational power of the computers together with the increase in the quality and the acquisition speed of optical microscopes goes hand in hand with the development of new computational methods in the fields such as image segmentation, image restoration, or image recognition. The growing capabilities of the instruments enable the acquisition and storage of high-resolution images in 2D, 3D, 3D+time, or higher dimensions. This requires either the development or at least some small modifications of current image processing methods. The new methods however need to be properly validated and tested before they are published and practically used.

In the field of biomedical image analysis, the validation process requires the use of some benchmark dataset. Such datasets contain real or synthetic image data accompanied with their ground truth. In the case of real data, the ground truth is typically obtained as an expert annotated raw image data. For a purely synthetic case, the image data and also the related ground truth is completely generated in the computer. Both approaches have their pros and cons.

1.1   Focus of the thesis

Due to the lack of publicly available benchmark datasets containing both real and synthetic image data together with their annotations, my research focuses on the computer generated data and its associated challenges.

The simulations have always been of a great importance as they substitute the real processes when those are too expensive to be performed or impossible to be observed. The latter case is typical for optical microscopy. Here, we observe fixed or living cells under assumption, that the optical system and the attached electronic acquisition device does not affect the quality of the original specimen too much. Even though we can measure the most of optical aberrations and estimate the dominant sources of noise, that together cause the final observed image to be blurred and noisy, we are still not able to reveal the original unaffected image how it would appear without any damage.

Although several deconvolution methods are capable of inverting this degradation process they can improve the quality of the data only to some extent.

As there exists no exact knowledge, on how the microscopic specimens look, it is very difficult to evaluate the quality of new emerging segmentation and tracking algorithms that are of a great importance in medicine and biology. The same issue arises when one wants to tune-up their parameters. In the past, the only available quality measurement of the algorithms was an expert's knowledge. The expert either classified the results of selected algorithm or provided an annotation of some real image dataset that was further used for evaluation purposes. Both ways however suffer from two main issues. First, the expert's evaluation is nondeterministic. Second, for higher dimensional data (sequences of 2D or 3D images) the handmade annotation is impractical or even impossible. For this reason, the synthetic data, naturally accompanied by their ground truth, have started to appear. In the very beginning [Pre79], only the basic geometric shapes like spheres or disc without any texture representing the internal structure of the observed cells were employed. Since the late 90s, computer generated images have started to be more complex as the computer capabilities rose and allowed for calculations that required higher performance and extensive memory and disk usage. Namely, in the last 10 years, several simulation frameworks able to generate, for example, cells with detailed description of subcellular components [Mur12], large cell populations [Leh+07; Raj+12; SKS09] and time-lapse image sequences [SU16; Duf+11] emerged.

The goal of this thesis is to summarize the main topics of my research I focused to during my work at the Faculty of Informatics, Masaryk University (CZ) and at the Faculty of Science and Engineering, Manchester Metropolitan University (UK). The results of my research include journal and conference papers accompanied by software packages implementing the ideas and methods described in the papers. The most important software package, which I originally developed in cooperation with my former supervisor Michal Kozubek, is called CytoPacq[1]. During the last few years, it has gradually become the core simulation framework in our group.

1.2   Thesis structure

The following chapters provide a brief overview of the methods and software packages I proposed or analyzed during my research. Chapter 2 is a collection of selected image analysis methods I designed or studied with my colleagues. It explains my interest in validation techniques in this field. Chapter 3 inspects the use of convolution when manipulating with huge multidimensional image data. As convolution plays an important role during the modeling of virtual optical microscope, various optimization techniques are introduced and analyzed in this chapter. In chapter 4, I describe the most common texture descriptors typically used for testing the similarity of the

---

[1]`http://cbia.fi.muni.cz/simulator/`

synthetic and real image data. Finally, the chapter 5 is dedicated to the fundamentals of simulations and the generation of benchmark datasets, the tasks I mainly focused on during my research.

Chapter 2

Image Analysis

Even though the topic of this work is dedicated to simulations and the generation of synthetic image benchmark datasets, I would like first put the reader's attention on the common image analysis methods. One should keep in mind that there would be no need for synthetic data if there were no image processing methods that require the validation and testing. The common tasks in image analysis (segmentation, edge detection, deconvolution, measurement, classification, etc.) were indeed my first research topic that brought me later to the area of simulations. Even after the main topic of my research changed to simulations and modeling I have still kept working or collaborating on projects dedicated exclusively to generic image analysis tasks. I believe that proper understanding of the particular tasks, including their design and implementation, leads to better design of simulation toolkits that should subsequently help with validation and testing of the proposed algorithms. For this reason, I consider the design and further development of various image analysis algorithm utilized in biomedical image analysis to be always a minor but essential part of my research. Here are the individual topics:

Segmentation.   During my Ph.D. studies I collaborated with my former supervisor Pavel Matula and focused on the segmentation tasks. In particular, I aimed my research on the analysis of fully 3D images of human colon tissues. I designed and implemented a method suitable for the segmentation of individual cells that occur inside the image of tight cell cluster. Initially, I co-invented the star-shaped simplex meshes [MS01], a tool suitable for segmentation of clearly separated potato-like objects. In this approach, the analyzed objects (cells) were roughly fitted with a regular parametric mesh, typically in the shape of sphere or ellipsoid, which was further deformed. The deformation iteratively followed the Newtonian law of motion including two principal types of forces: the internal and external one. The former was responsible for keeping the smoothness of the mesh surface while the latter pushed the mesh to the places where the remarkable edges in the inspected image were located. The modified version, called *dual simplex meshes* [SM03] and particularly designed for the segmentation of cells located in tight clusters, introduced two meshes. The inner and

outer one. These two meshes were only allowed to iteratively come near to each other while still keeping the rules given by internal and external forces defined in the original model. This technique was, however, highly sensitive to the initial configuration of the meshes.

Image measurement. The image segmentation is often only a pre-step for some forthcoming image processing operations like, for example, the measurement. In this step, we measure various image or object characteristics. They mostly include the volume of cells, their surface, roundness, distance of some particular spots from the cell membrane, etc. In the study [Ska+07], we segmented the nucleus of each cell by a Chan-Vese segmentation algorithm [CV01] and subsequently we measured the distribution of histones in the individual concentric shells in the cell nucleus [Cre+04]. To validate the proposed method, we artificially generated several types of radial distributions inside a sphere and measured the results of our method.

Edge detection in highly textured microscopy data. During my half-year stay at Manchester Metropolitan University (MMU) in the UK, I collaborated on testing the statistics-based edge detectors. We studied the properties of these filters and tried to identify the most suitable filters for edge detection of highly textured biomedical images [Svo+06; WBS14]. We namely focused on the goodness-of-fit two sample tests including: Fisher test, Student's t-test, Mann Whitey U test, Kolomogorov-Smironov test, χ^2 test, and Difference of boxes. For sufficiently large filter masks utilized during the edge detection, the Kolmogorov-Smirnov and χ^2 tests overcame all the others.

Image restoration. The deconvolution plays an important role in the preprocessing step when manipulating with the image acquired using on optical microscope. In order to speed up the deconvolution process, which is commonly known to be the long-lasting process, we tried to utilize the GPU architecture. For this purpose, we designed a parallel modification of the most common deconvolution algorithms (Wiener, ICTM, EM-MLE) [KKS13].

Image classification. In 2013, we participated the *HEp-2 Cells Classification Contest* associated with International Conference on Pattern Recognition. The aim of this contest was to design and implement the classifier that is able to categorize pre-segmented 2D images of HEp-2 cells into 6 classes to detect autoimmune diseases which correspond to different cell patterns. For this purpose we designed an engine consisting of the following set of image descriptors: Haralick features, Local Binary Patterns, SIFT, surface descriptor and a granulometry-based descriptor. The final classification was based on k-NN classifier [SMS14]. In this contest, we achieved 7th place out of 28. In the following years, we continued in the development of the proposed classification method and extended the classifier to also work with 3D image data.

Articles in collection

- David Svoboda and Pavel Matula. "Tissue Reconstruction Based on Deformation of Dual Simplex Meshes". In: *DGCI*. ed. by S. Svensson I. Nyström G. S. di Baja. Vol. 2886. LNCS. ISBN 3-540-20499-7. Springer – Berlin, Heidelberg, New York, 2003, pp. 524–533

  I invented the method and wrote the paper.

- Magdalena Skalníková et al. "Distinct Patterns of Histone Methylation and Acetylation in Human Interphase Nuclei". In: *Physiological Research* 56.6 (2007), pp. 797–806

  I co-invented the image analysis method applied to the microscopy data acquired and processed in the paper. I edited the paper.

- Roman Stoklasa, Tomáš Majtner, and David Svoboda. "Efficient k-NN Based HEp-2 Cells Classifier". In: *Pattern Recognition* 47.7 (2014), pp. 2409 –2418. ISSN: 0031-3203

  I collaborated on the method design and edited the paper.

- Ian Williams, Nicholas Bowring, and David Svoboda. "A Performance Evaluation of Statistical Tests for Edge Detection in Textured Images". In: *Computer Vision and Image Understanding* 122 (2014), pp. 115 –130. ISSN: 1077-3142

  I analyzed the reviewed methods and wrote the paper.

- Pavel Karas, Michal Kuderjavý, and David Svoboda. "Deconvolution of Huge 3D Images: Parallelization Strategies on a Multi-GPU System". In: *Algorithms and Architectures for Parallel Processing*. Ed. by Joanna Kołodziej et al. Vol. 8285. Lecture Notes in Computer Science. Springer International Publishing, 2013, pp. 279–290. ISBN: 978-3-319-03858-2

  I edited the paper and prepared the data for testing the individual methods.

Chapter 3

Optimization of Convolution

The standard simulation process in the field of biomedical imaging can be split into three consecutive phases: phantom generation, simulation of optical system, and simulation of acquisition device. Even though the convolution is extensively used also in the first phase, when generating the individual objects including their localization within the space, the most clearly the convolution is employed in the second phase, when modeling the transmission of the image through the optical system. As the processed 3D image is typically of a large size ($1024 \times 1024 \times 60$ voxels) and likewise the experimental point spread function (PSF), the standard implementation of convolution would take a long time. Vice versa, the use of fast implementations based on Fourier transform requires extensive memory usage. The research on this topic tries to find some sort of compromise.

3.1 CPU

The combination the *overlap-and-add* and *overlap-and-save* approaches [OS75] together with fast convolution (computed in the Fourier spectrum) was found to be an optimal solution both for time and spatial complexity [Svo11]. In this paper, both image and convolution kernel were split into several regular pieces (tiles) to lower to spatial complexity. The splitting process was however controlled to avoid excessive requirements put on the computational power. The paper shows that if both image and kernel are D-dimensional cubes M^D and N^D, respectively, and the tiling process splits the image into m tiles and kernel into n tiles we need to perform in total

$$(Mn + Nm)^D \left[\frac{9}{2} \log_2 \left(\frac{M}{m} + \frac{N}{n} \right)^D + 1 \right] \tag{3.1}$$

multiplications and the spatial complexity drops to:

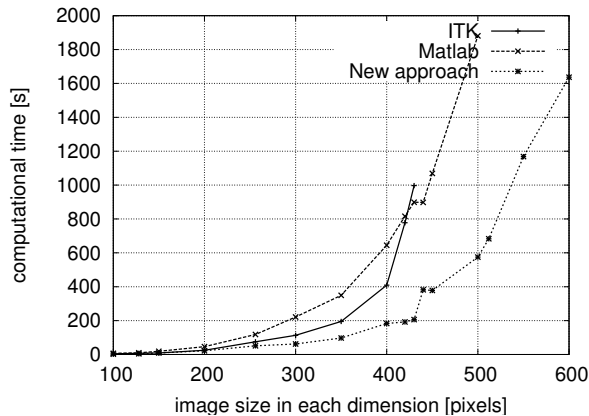$$\left(\frac{M}{m} + \frac{N}{n} \right) \cdot C \tag{3.2}$$

8

Figure 3.1: A graph offering the comparison of the most common implementations of convolution and the new approach. Evaluated over two 3D images of identical size on Intel Xeon QuadCore 2.83 GHz computer with 32 GB RAM. Take note, that ITK and Matlab plots finish earlier as the computation for the images of large dimensions failed due to the lack of memory.

We showed that the optimal solution occurs, when m and n are minimized and equal to each other. The proposed algorithm outperformed the standard implementations of convolution (see Fig. 3.1).

3.2 GPU

In order to further increase the speed of computation of convolution, we focused on GPU programming. Even though the GPU can work in parallel and can perform a large amount of instruction per second, it was not originally designed to handle with large memory blocks. For this purpose, our research was not purely focused on implementing the standard convolution on GPU. We employed the decimation in frequency (DIF) algorithm and carefully manipulated with the memory blocks to eliminate the latency and waiting [KS11; KSZ12]. All the conclusions in terms of optimal design of convolution algorithm both on CPU and GPU are gathered in the book chapter [KS13].

Articles in collection

- David Svoboda. "Efficient Computation of Convolution of Huge Images". In: *Proceedings of the 16th Int. Conference on ICIAP*. ICIAP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 453–462. ISBN: 978-3-642-24084-3

  I am the sole author of this paper.

9

- Pavel Karas and David Svoboda. "Convolution of Large 3D Images on GPU and its Decomposition". In: *EURASIP Journal on Advances in Signal Processing* 2011.1, 120 (2011)

  I co-inveneted the method and wrote the paper.

- Pavel Karas, David Svoboda, and Pavel Zemčík. "GPU Optimization of Convolution for Large 3-D Real Images". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by Jacques Blanc-Talon et al. Vol. 7517. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 59–71. ISBN: 978-3-642-33139-8

  I prepared the image data for testing the method and edited the paper.

Chapter 4

Texture Analysis

The computer generated data together with their annotation (ground truth) help with validating and testing the newly introduced image analysis methods. However, before using such data, one should submit such dataset to the criticism whether it corresponds to some real image dataset. For this reason, both datasets (the real as well as the synthetic one) are commonly submitted to selected set of measures and the results of these measurements are compared to reveal either the difference or the conformity. The most commonly employed measurements are the texture descriptors.

4.1 Texture descriptors

In the years 2009-2010, I supervised the bachelor thesis *Texture descriptors for biomedical image data*, in which the fully 3D biomedical data (two cell lines) were analyzed using 3D Haralick and Zernike descriptors. In master thesis *Application of MPEG-7 descriptors when analyzing 3D biomedical image data*, which I supervised in the years 2012-2013, the student proposed the extension of standard MPEG-7 descriptors to allow the analysis of volumetric data. In parallel to these results, I co-supervised one PhD student, who focused his studies and research on texture-based image descriptors in fluorescence microscopy. Our first team cooperation resulted in the paper [MS12], where we allowed the three standard Tamura features (coarseness, contrast, and directionality) [TMY78] to work also with 3D image data.

Later, we participated in the HEp2-cell classification contest, for which we designed a hand-tailored texture descriptor called RSurf [MSS14]. This descriptor analyzes the gradient changes along given set of directions. It is highly sensitive to small changes in the image texture while less sensitive to image rotation. This property makes it suitable for the analysis of selected cell line without need for rotating each cell into the standard position.

In the last study, we successfully analyzed the classification power of the 3D Gabor descriptor and 3D MPEG-7 edge histogram descriptor when applied on selected microscopy images [MS14].
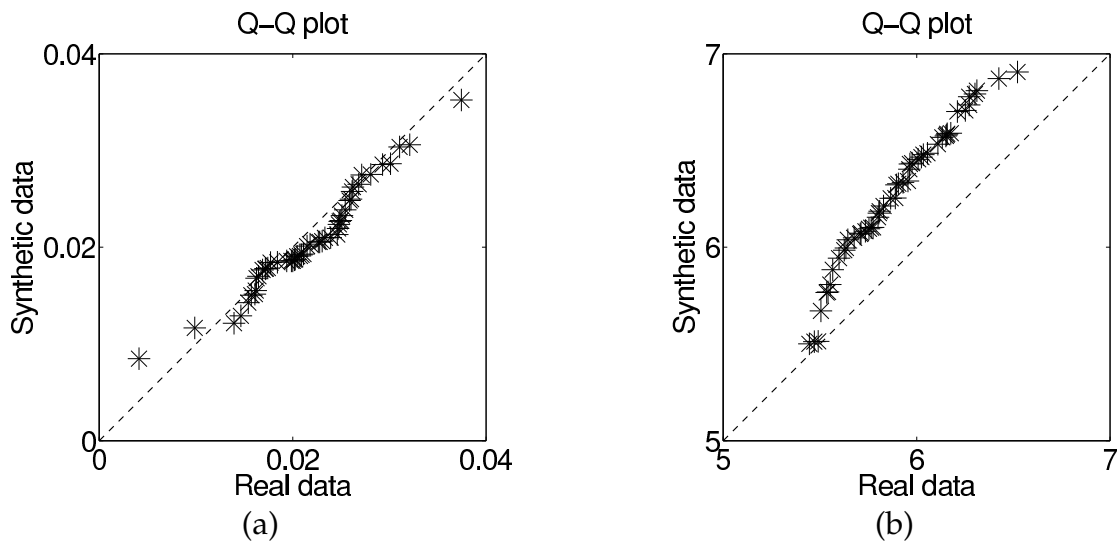
Figure 4.1: Quantile-quantile plots illustrate whether the measured datasets come from populations with similar distributions. If two sample sets come from a population with the same distribution, the points should fall approximately along the reference line $y = x$: (a) sample sets follow nearly the same distribution, (b) sample sets differ.

4.2 Similarity search

All the texture descriptors mentioned above bring an important information about the analyzed images, regardless they were withdrawn form the real or synthetic datasets. However, there are many numbers describing these images (typically grouped into long *feature vectors*) and it is difficult to clearly judge, whether the datasets differ or not. For this purpose, the computed descriptors are further submitted to some statistical methods. These should be able to clearly either approve or reject the similarity of the datasets. The common statistics include *QQ-plots* and *two-sample goodness-of-fit tests*. Both methods analyze the distribution of the inspected sample sets. The first technique shows their similarity by plotting their quantiles against each other in the graph (see Fig. 4.1). The second technique accepts or rejects the hypothesis that the two sample sets come from the same distribution, i.e. the output is one (binary) number. We employed both methods in our research to validate the plausibility of our computer generated data [SKS09; SHS11; SU16].

Articles in collection

- Tomáš Majtner and David Svoboda. "Extension of Tamura Texture Features for 3D Fluorescence Microscopy". In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. 2012, pp. 301–307

  I collaborated on the method design and edited the paper.

- Tomáš Majtner, Roman Stoklasa, and David Svoboda. "RSURF – The Efficient Texture-Based Descriptor for Fluorescence Microscopy Images of HEP-2 Cells". In: *Pattern Recognition (ICPR), 2014 22nd International Conference on*. 2014, pp. 1194–1199

  I collaborated on the method design and edited the paper.

- Tomáš Majtner and David Svoboda. "Texture Analysis Using 3D Gabor Features and 3D MPEG-7 Edge Histogram Descriptor in Fluorescence Microscopy". In: *3D Imaging (IC3D), 2014 International Conference on*. 2014, pp. 1–7

  I collaborated on the method design and edited the paper.

Chapter 5

Simulations

In the last few years, the field of biomedical imaging increasingly utilizes the advantages of simulations. The significance of simulations lies in two main aspects. First, the synthetic, computer generated, data help to measure the quality of newly developed image analysis methods. These methods include, for example, cell segmentation, cell tracking, image deconvolution, object measurements, etc. Even though, the method can be benchmarked with traditionally accessible human-annotated real image data, the synthetic data can be easily generated in larger quantities. Moreover, the synthetic datasets are fully accompanied by the ground truth data, which is not the case of real data. Second, when studying the dynamic processes in living cells, the simulations can help to understand some events, that are difficult to detect or repeat.

In biomedical imaging community, the pivotal study introducing the basic ideas for simulating the large cell populations appeared in 2007 [Leh+07]. This study focused on the main tasks that need to be solved when generating synthetic microscopy image data. The authors explained in detail the principles of generating the cell shape and internal structure. They also imitated the image degradation caused by the imperfections of the optical system. This study was however focused on 2D imaging only.

5.1 Fixed cells

In our paper from the same year [Svo+07], we adopted the main principles from [Leh+07] and created our own simulation framework. Additionally, we introduced the manipulation with fully 3D data and described in detail some selected image degradation phenomena, including uneven illumination, photon shot noise, and noise produced by the signal amplifier. This study was further extended two years later [SKS09] with several principal modifications. First of all, the *three phase* simulation model was introduced (see Fig. 5.1). The first phase called *CytoGen* generates the phantoms of studied cell line. Here, we generated nuclei of HL-60 or granulocytes. This phase creates the image, as it would appear in the objective, if there were no phenomena influencing the final image. The second phase, called *OptiGen* imitates the existing optical system (mi-
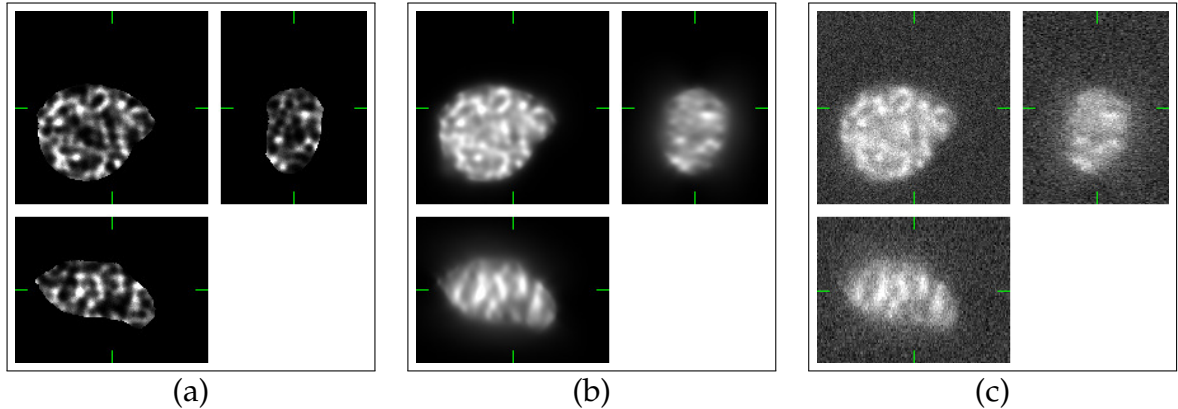
14

Figure 5.1: Three-phase model of simulations: (a) during the first phase the phantom of an object, that should be synthesized, is prepared, (b) the second phase transmits the image of the phantom, generated in the previous step, through the virtual optical system, (c) the transmitted signal is acquired using the virtual acquisition device. Each 3D figure consists of three individual images: the top-left image contains a selected xy-slice, the top-right image corresponds to a selected yz-slice, and the bottom one depicts a selected xz-slice. Three mutually orthogonal slice planes are shown with green ticks.

croscope, objective, excitation and emission filters) and its behavior. Here, the image is typically convolved with some either experimental or theoretical point spread function (PSF) and further modified by simulating the uneven illumination or other important phenomena. Finally, the third phase, called *AcquiGen*, is responsible for simulation of selected acquisition device. In particular, we modeled the behavior of CCD camera, including the standard sources of noise (photon shot noise, dark current noise, and white additive noise produced by an amplifier). The three phase model can be however identified also in other modalities. The camera can be substituted with ultrasound receiver/transmitter or CT scanner, for example. The proposed framework has already been implemented and offered for public use as a web-based service called CytoPacq[1].

In order to extend the capabilities of CytoPacq, we included a new type of cell that can be generated *in silico*. We focused on tightly connected cell colonies – the cells forming the villi of human colon tissues [SHS11] (see Fig. 5.2). When generating the cells forming the colon villi, we initiated the basic structure as a slightly deformed cylinder with randomly generated points on its surface. These points became the seeds for 3D Voronoi regions defining the shape of individual cells. The research group from University of Warwick adopted this principle few year later in [KSR16].

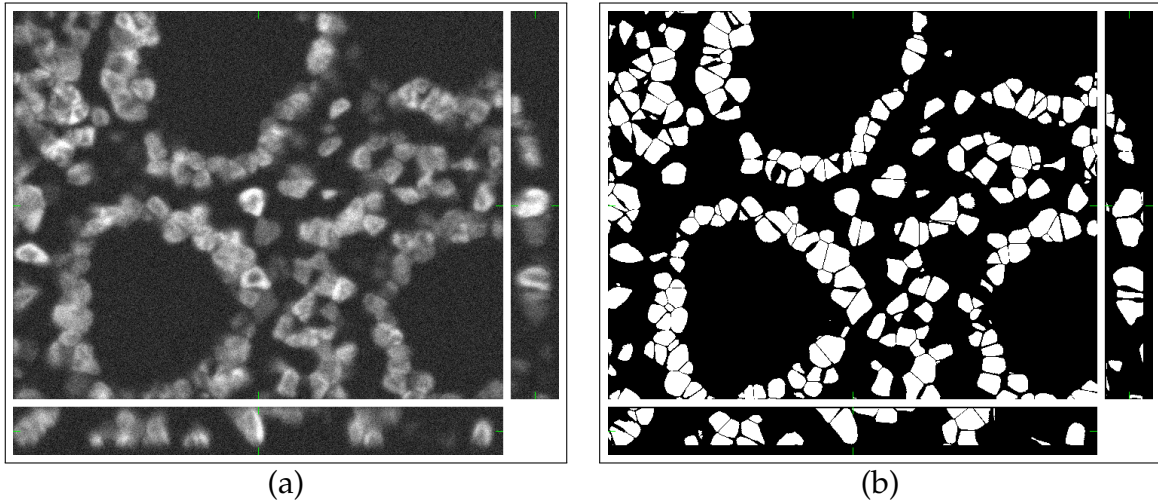
[1]http://cbia.fi.muni.cz/simulator/

Figure 5.2: An example of computer generated fully 3D image of healthy human colon tissue: (a) final synthetic image as it would appear if acquired using real optical microscope, (b) ground truth mask suitable for segmentation purposes.

5.2 Living cells

So far, all the generated synthetic data have represented the fixed (dead) cells. In 2012, we published our first study [SU12] explaining how to simply move with the interphase cells. The motility of such generated cells included simply a rotation, shift and slight deformation. This study became a cornerstone for the activities coming in the following years when we agreed to co-organize the 1st, 2nd, and 3rd edition of Cell Tracking Challenge (CTC) associated with IEEE International Symposium on Biomedical Imaging. As co-organizers, we were responsible for the preparation of the synthetic lapse-lapse image data sequences imitating the life of the whole cell population. Our participation in CTC was concluded in the journal paper [Ma14], where the whole challenge, including the significance of synthetic data, was described in detail.

In order to control the formation of clusters of synthetic cells in the generated dataset and to improve the visual plausibility of the initial cell population, we redesigned the former algorithm randomly distributing cells across the microscope slide already incorporated in CytoPacq. The new algorithm brought a possibility to control the clustering effect (see Fig. 5.3) in the generated cell population [SU13].

The simulated time-lapse sequence did not however contain all the standard visually perceivable artifacts. In the following research, we focused on proper modeling of particular phenomena that influence the visual appearance for the image sequences acquired on the real microscope. In particular, we focused on photobleaching [Svo+14] (also known as fading), which is a photochemical alteration of any dye incapable of permanently and constantly fluoresce. In order to further improve the visual appearance of internal nucleus structure, we studied the biologically motivated models of
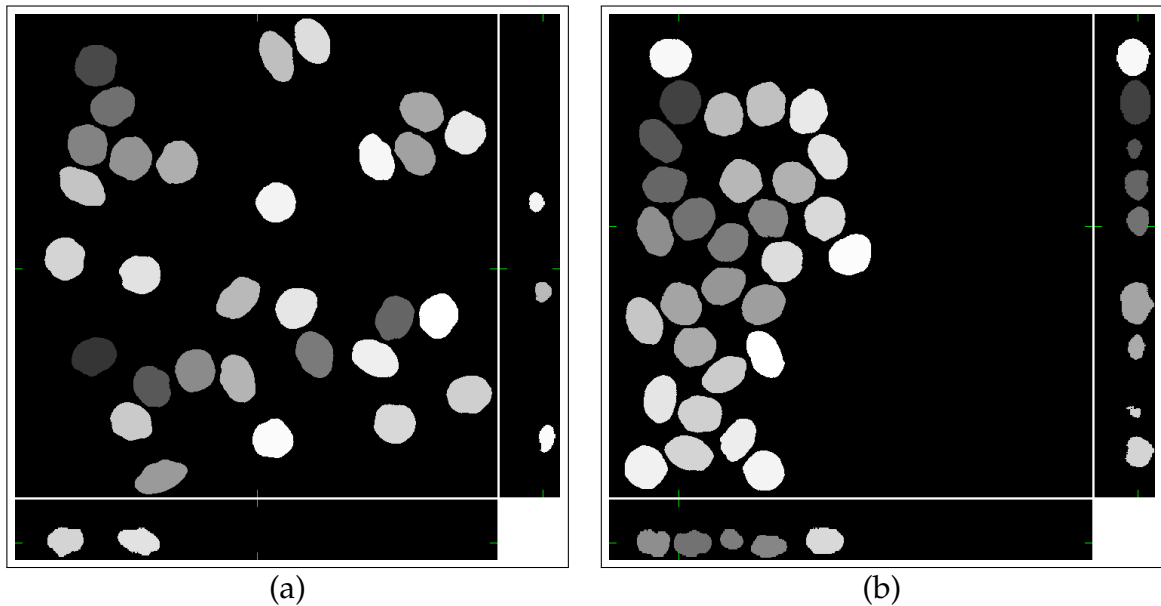
16

Figure 5.3: Different levels of clustering within the initial cell population: (a) 50%, (b) 100%. Each 3D figure consists of three individual images: the top-left image contains a selected xy-slice, the top-right image corresponds to a selected yz-slice, and the bottom one depicts a selected xz-slice. Three mutually orthogonal slice planes are shown with green ticks.

DNA [Str+00] and decided to utilize the so called *free joint chain model* (FJC) [SUP15].

The upcoming simulation of living cells including the mitotic division however required the design of a more complex model. For this reason, we proposed a new model describing also mitotic division (see Fig. 5.4), cell motility, and mutual cell interactions that commonly occur in cell populations [SU16]. The model was realized as a software framework called *MitoGen*[2]. The data produced by MitoGen were included in the benchmark dataset utilized in CTC (2013–2015).

5.3   Living tissues

Initially, our first synthetic images contained only single cells or small clusters. Later on the time-lapse sequences prepared for CTC, for example, have already contained the clusters with the tens of cells. Our ambition was to generate the time-lapse image sequences containing hundreds for cells in order to simulate the would healing or living tissues. The preliminary results in [UOS15; Svo+16] show that this task is feasible.

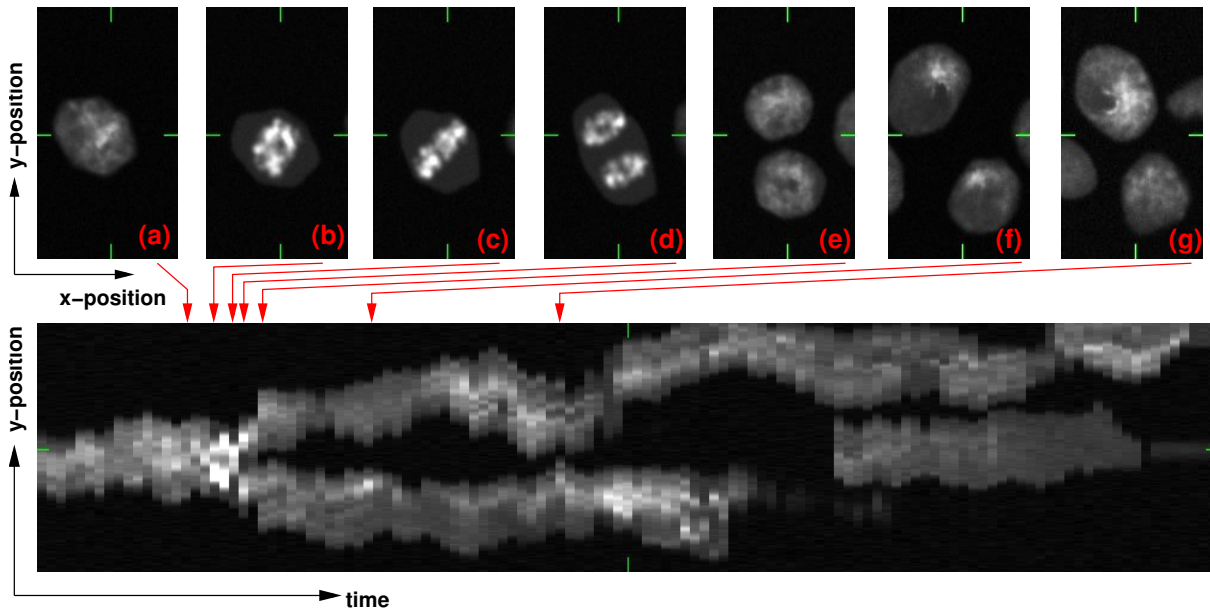[2]http://cbia.fi.muni.cz/projects/mitogen.html

Figure 5.4: One cell cycle of a sample synthetic (computer generated) cell presented using the sequence of 2D z-slices and a spatio-temporal image. The most important phases of the cell cycle are visualized and accordingly marked with red arrows. Following the sequence (left to right), we can recognize the particular phases: **(a)** *G2-phase* in which the chromatin is uncoiled; **(b)** *Prophase* when the chromatin condensates; **(c)** *Metaphase* during which chromosomes form the metaphase plate; **(d)** *Anaphase* when the genetic material is split into two; **(e)** *Telophase* and *Cytokinesis* in which two new daughter cells appear; **(f)** *G1-phase* when the new cell grows; and **(g)** *S-phase* when DNA is replicated.

5.4 Benchmark datasets

The production of synthetic image data as well as annotated real images is not the only objective. The main aim is to establish publicly available and acceptable collection of images along with their ground truth. The research groups need not waste their time by preparing new annotated real and synthetic datasets. They can simply use the data that have already been prepared, validated and accepted by the community.

In the last decade, several benchmark datasets appeared to facilitate the research to those, who develop the image analysis algorithm, including segmentation, tracking, restoration, or classification. These datasets are published on the Internet and freely available. The most common collections include:

- Broad Bioimage Benchmark Collection[3]

---

[3] https://data.broadinstitute.org/bbbc/image_sets.html

- UCSB Bio-Segmentation Benchmark dataset[4]

- Murphy Lab[5]

As a complement to the above mentioned data collections that contain mainly the real annotated data, there are also the collections focused primarily on the synthetic images:

- SIMCEP dataset[6]

- Masaryk University Cell Image Collection (MUCIC)[7].

The last mentioned one is the collection we prepared in Centre for Biomedical Image Analysis (CBIA) at Faculty of Informatics MU. Currently, this collection contains five cellular datasets with the following features:

- HL-60 cell line (fixed cells) . . . This dataset contains 30 synthetic images of nuclei of HL-60 cell line including ground truth (foreground/background) images. Each image set contains 20 cell nuclei with specified probability of clustering (0%, 25%, 50%, and 75%). The two levels of SNR are available. In total, there are 240 3D images.

- Granulocytes (fixed cells) . . . This dataset contains 30 synthetic images of nuclei of granulocytes including ground truth (foreground/background) images. Each image set contains up to 15 cell nuclei. In total, there are 60 3D images.

- Colon tissues (fixed cells) . . . This dataset contains 30 synthetic images of human colon tissue including ground truth (foreground/background) images. In total, there are 60 3D images.

- HL-60 cell line (population of living cells) . . . This dataset contains six computer generated time-lapse image sequences of nuclei of HL-60 cells. The sequences are created as a combination of different levels of noise, different levels of synchronization of cells in the population, the various cell density of the initial cell population, the various number of cells leaving and entering the field of view and the various number of simulated mitotic events, yielding up to 70 cells in the field of view.

- Endothelial cells (living cells) . . . This dataset contains one sequence of frames recording the process called vasculogenesis. The endothelial cells, initially spread across the glass slide, tend to attach each other and form the networks with thin and elongated chords. There is just one sequence of synthetic 2D images.

---

[4]http://bioimage.ucsb.edu/research/bio-segmentation
[5]http://murphylab.cbi.cmu.edu/data/
[6]http://www.cs.tut.fi/sgn/csb/simcep/benchmark/
[7]http://cbai.fi.muni.cz/datasets/

Articles in collection

- David Svoboda et al. "On Simulating 3D Fluorescent Microscope Images." In: *CAIP*. ed. by W. G. Kropatsch, M. Kampel, and A. Hanbury. Vol. 4673. LNCS. Springer, 2007, pp. 309–316

  I co-invented the method and wrote the paper.

- David Svoboda, Michal Kozubek, and Stanislav Stejskal. "Generation of Digital Phantoms of Cell Nuclei and Simulation of Image Formation in 3D Image Cytometry". In: *Cytometry part A* 75A.6 (2009), 494–509. ISSN: 1552-4922

  I co-invented the method and wrote the paper.

- David Svoboda, Ondřej Homola, and Stanislav Stejskal. "Generation of 3D Digital Phantoms of Colon Tissue". In: *Proc. of the 8th Int. Conference on Image Analysis and Recognition, ICIAR 2011*. Vol. 6754. LNCS. Springer, 2011, pp. 31–39

  I co-invented the method and wrote the paper.

- David Svoboda and Vladimír Ulman. "Towards a Realistic Distribution of Cells in Synthetically Generated 3D Cell Populations". In: *Proceedings of 17th International Conference on Image Analysis and Processing*. Vol. 8157. LNCS. Springer, 2013, pp. 429–438

  I co-invented the method and wrote the paper.

- David Svoboda et al. "On Proper Simulation of Phenomena Influencing Image Formation in Fluorescence Microscopy". In: *2014 IEEE International Conference on Image Processing (ICIP)*. 2014, pp. 3944–3948

  I co-invented the method and wrote the paper.

- Martin Maška et al. "A Benchmark for Comparison of Cell Tracking Algorithms". In: *Bioinformatics* 30.11 (2014), pp. 1609–1617

  I prepared the synthetic image data for the challenge and edited the paper.

- David Svoboda, Vladimír Ulman, and Igor Peterlík. "On Proper Simulation of Chromatin Structure in Static Images As Well As in Time-Lapse Sequences in Fluorescence Microscopy". In: *Proceedings of 2015 IEEE International Symposium on Biomedical Imaging*. Stoughton (WI, USA): Engineering in Medicine and Biology Society, 2015, pp. 712–716. ISBN: 978-1-4799-2375-5

  I co-invented the method and wrote the paper.

- David Svoboda et al. "Vascular Network Formation in Silico Using the Extended Cellular Potts Model". In: *2016 IEEE International Conference on Image Processing (ICIP)*. Signal Processing Society, 2016, pp. 3180–3183. ISBN: 978-1-4673-9961-6

  I co-invented the method and wrote the paper.

- David Svoboda and Vladimír Ulman. "MitoGen: A Framework for Generating 3D Synthetic Time-Lapse Sequences of Cell Populations in Fluorescence Microscopy". In: *IEEE Transactions on Medical Imaging* (2016). in press

  I co-invented the method and wrote the paper.

Chapter 6

Conclusion

In this thesis, I have presented my research contribution to the progress within the area of image-based simulations in fluorescence microscopy. I have also mentioned areas of connected work relating to simulations and synthetic data creation. The individual research contributions were described in detail in the selected representative articles I have co-authored, which are also attached to this text[1].

---

[1]The reprints of the articles are excluded from the public version of this thesis to avoid copyright violation.

Bibliography

[Cre+04]   M. Cremer et al. "Three dimensional analysis of histone methylation patterns in normal and tumor cell nuclei". English. In: *European Journal of Histochemistry : EJH* 48.1 (2004). Copyright - Copyright PAGEPress Publications 2004; Last updated - 2014-04-21, pp. 15–28. URL: `http://search.proquest.com/docview/876074331?accountid=16531`.

[CV01]   T. F. Chan and L. A. Vese. "Active Contours Without Edges". In: *Trans. Img. Proc.* 10.2 (Feb. 2001), pp. 266–277. ISSN: 1057-7149. DOI: `10.1109/83.902291`. URL: `http://dx.doi.org/10.1109/83.902291`.

[Duf+11]   A. Dufour et al. "3-D Active Meshes: Fast Discrete Deformable Models for Cell Tracking in 3-D Time-Lapse Microscopy". In: *IEEE Trans. on Image Processing* 20.7 (2011), pp. 1925–1937.

[KKS13]   Pavel Karas, Michal Kuderjavý, and David Svoboda. "Deconvolution of Huge 3D Images: Parallelization Strategies on a Multi-GPU System". In: *Algorithms and Architectures for Parallel Processing*. Ed. by Joanna Kołodziej et al. Vol. 8285. Lecture Notes in Computer Science. Springer International Publishing, 2013, pp. 279–290. ISBN: 978-3-319-03858-2.

[KS11]   Pavel Karas and David Svoboda. "Convolution of Large 3D Images on GPU and its Decomposition". In: *EURASIP Journal on Advances in Signal Processing* 2011.1, 120 (2011).

[KS13]   Pavel Karas and David Svoboda. "Algorithms for Efficient Computation of Convolution". In: *Design and Architectures for Digital Signal Processing*. 1st ed. Rijeka (CRO): InTech, 2013, pp. 179–208. ISBN: 978-953-51-0874-0.

[KSR16]   Violeta N. Kovacheva, David Snead, and Nasir M. Rajpoot. "A model of the spatial tumour heterogeneity in colorectal adenocarcinoma tissue". In: *BMC Bioinformatics* 17.1 (2016), pp. 1–16. ISSN: 1471-2105.

[KSZ12]   Pavel Karas, David Svoboda, and Pavel Zemčík. "GPU Optimization of Convolution for Large 3-D Real Images". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by Jacques Blanc-Talon et al. Vol. 7517. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 59–71. ISBN: 978-3-642-33139-8.

[Leh+07]   Antti Lehmussola et al. "Computational Framework for Simulating Fluorescence Microscope Images With Cell Populations". In: *IEEE Trans. Med. Imaging* 26.7 (2007), pp. 1010–1016.

[Ma14]   Martin Maška et al. "A Benchmark for Comparison of Cell Tracking Algorithms". In: *Bioinformatics* 30.11 (2014), pp. 1609–1617.

[MS01]   Pavel Matula and David Svoboda. "Spherical Object Reconstruction Using Star-Shaped Simplex Meshes". In: *EMMCVPR, 3rd Int. Workshop*. Ed. by Mário Figueiredo, Josiane Zerubia, and Anil K. Jain. Vol. 2134. LNCS. ISBN 3-540-42523-3. Springer, 2001, pp. 608–620.

[MS12]   Tomáš Majtner and David Svoboda. "Extension of Tamura Texture Features for 3D Fluorescence Microscopy". In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. 2012, pp. 301–307.

[MS14]   Tomáš Majtner and David Svoboda. "Texture Analysis Using 3D Gabor Features and 3D MPEG-7 Edge Histogram Descriptor in Fluorescence Microscopy". In: *3D Imaging (IC3D), 2014 International Conference on*. 2014, pp. 1–7.

[MSS14]   Tomáš Majtner, Roman Stoklasa, and David Svoboda. "RSURF – The Efficient Texture-Based Descriptor for Fluorescence Microscopy Images of HEP-2 Cells". In: *Pattern Recognition (ICPR), 2014 22nd International Conference on*. 2014, pp. 1194–1199.

[Mur12]   R.F. Murphy. "CellOrganizer: Image-derived models of subcellular organization and protein distribution." In: *Methods Cell Biol* 110 (2012).

[OS75]   Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice–Hall, 1975.

[Pre79]   Judith Prewitt. "Graphs and grammars for histology: An introduction". In: *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, Washington, DC, 1979, pp. 18–25.

[Raj+12]   Satwik Rajaram et al. "SimuCell: a flexible framework for creating synthetic microscopy images." In: *Nat Methods* 9.7 (2012), pp. 634–5. ISSN: 1548-7105.

[SHS11]   David Svoboda, Ondřej Homola, and Stanislav Stejskal. "Generation of 3D Digital Phantoms of Colon Tissue". In: *Proc. of the 8th Int. Conference on Image Analysis and Recognition, ICIAR 2011*. Vol. 6754. LNCS. Springer, 2011, pp. 31–39.

[Ska+07]   Magdalena Skalníková et al. "Distinct Patterns of Histone Methylation and Acetylation in Human Interphase Nuclei". In: *Physiological Research* 56.6 (2007), pp. 797–806.

[SKS09]   David Svoboda, Michal Kozubek, and Stanislav Stejskal. "Generation of Digital Phantoms of Cell Nuclei and Simulation of Image Formation in 3D Image Cytometry". In: *Cytometry part A* 75A.6 (2009), 494–509. ISSN: 1552-4922.

[SM03]   David Svoboda and Pavel Matula. "Tissue Reconstruction Based on Deformation of Dual Simplex Meshes". In: *DGCI*. Ed. by S. Svensson I. Nyström G. S. di Baja. Vol. 2886. LNCS. ISBN 3-540-20499-7. Springer – Berlin, Heidelberg, New York, 2003, pp. 524–533.

[SMS14]   Roman Stoklasa, Tomáš Majtner, and David Svoboda. "Efficient k-NN Based HEp-2 Cells Classifier". In: *Pattern Recognition* 47.7 (2014), pp. 2409 –2418. ISSN: 0031-3203.

[Str+00] Terence Strick et al. "Twisting and stretching single DNA molecules". In: *Progress in Biophysics and Molecular Biology* 74.1–2 (2000). Single Molecule Biochemistry and Molecular Biology, pp. 115–140. ISSN: 0079-6107.

[SU12]   David Svoboda and Vladimír Ulman. "Generation of Synthetic Image Datasets for Time-Lapse Fluorescence Microscopy". In: *Proceedings of 9th International Conference on Image Analysis and Recognition*. Vol. 7325. LNCS. Springer, 2012, pp. 473–482.

[SU13]   David Svoboda and Vladimír Ulman. "Towards a Realistic Distribution of Cells in Synthetically Generated 3D Cell Populations". In: *Proceedings of 17th International Conference on Image Analysis and Processing*. Vol. 8157. LNCS. Springer, 2013, pp. 429–438.

[SU16]   David Svoboda and Vladimír Ulman. "MitoGen: A Framework for Generating 3D Synthetic Time-Lapse Sequences of Cell Populations in Fluorescence Microscopy". In: *IEEE Transactions on Medical Imaging* (2016). in press.

[SUP15]   David Svoboda, Vladimír Ulman, and Igor Peterlík. "On Proper Simulation of Chromatin Structure in Static Images As Well As in Time-Lapse Sequences in Fluorescence Microscopy". In: *Proceedings of 2015 IEEE International Symposium on Biomedical Imaging*. Stoughton (WI, USA): Engineering in Medicine and Biology Society, 2015, pp. 712–716. ISBN: 978-1-4799-2375-5.

[Svo+06] David Svoboda et al. "Statistical Techniques for Edge Detection in Histological Images". In: *First International Conference on Computer Vision Theory and Applications*. INSTICC Press, 2006, pp. 457–462. ISBN: 972-8865-40-6.

[Svo+07] David Svoboda et al. "On Simulating 3D Fluorescent Microscope Images." In: *CAIP*. Ed. by W. G. Kropatsch, M. Kampel, and A. Hanbury. Vol. 4673. LNCS. Springer, 2007, pp. 309–316.

[Svo+14] David Svoboda et al. "On Proper Simulation of Phenomena Influencing Image Formation in Fluorescence Microscopy". In: *2014 IEEE International Conference on Image Processing (ICIP)*. 2014, pp. 3944–3948.

[Svo+16] David Svoboda et al. "Vascular Network Formation in Silico Using the Extended Cellular Potts Model". In: *2016 IEEE International Conference on Image Processing (ICIP)*. Signal Processing Society, 2016, pp. 3180–3183. ISBN: 978-1-4673-9961-6.

[Svo11] David Svoboda. "Efficient Computation of Convolution of Huge Images". In: *Proceedings of the 16th Int. Conference on ICIAP*. ICIAP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 453–462. ISBN: 978-3-642-24084-3.

[TMY78] H. Tamura, S. Mori, and T. Yamawaki. "Textural Features Corresponding to Visual Perception". In: *IEEE Transactions on Systems, Man, and Cybernetics* 8.6 (1978), pp. 460–473. ISSN: 0018-9472.

[UOS15] Vladimír Ulman, Zoltán Orémuš, and David Svoboda. "TRAgen: A Tool for Generation of Synthetic Time-Lapse Image Sequences of Living Cells". In: *Image Analysis and Processing — ICIAP 2015*. Ed. by Vittorio Murino and Enrico Puppo. Vol. 9279. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 623–634. ISBN: 978-3-319-23230-0.

[WBS14] Ian Williams, Nicholas Bowring, and David Svoboda. "A Performance Evaluation of Statistical Tests for Edge Detection in Textured Images". In: *Computer Vision and Image Understanding* 122 (2014), pp. 115 –130. ISSN: 1077-3142.

Part II

Collection of Articles

This part contains abstracts of journal and conference papers. The full texts are not publicly available to avoid copyright infringements. The comments to all papers and a description of the corresponding research are provided in the previous chapters.

Appendix A

Image Analysis

Conference paper [SM03]

• David Svoboda and Pavel Matula. "Tissue Reconstruction Based on Deformation of Dual Simplex Meshes". In: *DGCI*. ed. by S. Svensson I. Nyström G. S. di Baja. Vol. 2886. LNCS. ISBN 3-540-20499-7. Springer – Berlin, Heidelberg, New York, 2003, pp. 524–533

Abstract. A new semiautomatic method for tissue reconstruction based on deformation of a dual simplex mesh was developed. The method is suitable for specifically-shaped objects. The method consists of three steps: the first step includes searching for object markers, i. e. the approximate centre of each object is localized. The searching procedure is based on careful analysis of object boundaries and on the assumption that the analyzed objects are sphere-like shaped. The first contribution of the method is the possibility to find the markers without choosing the particular objects by hand.

In the next step the surface of each object is reconstructed. The procedure is based on the method for spherical object reconstruction presented in [MS01]. The method was partially changed and was adapted to be more suitable for our purposes. The problem of getting stuck in local minima was solved. In addition, the deformation process was sped up.

The final step concerns quality evaluation: both of the first two steps are nearly automatic, therefore the quality of their results should be measured.

Journal paper [Ska+07]

Abstract.   To study 3D nuclear distributions of epigenetic histone modifications such as H3(K9) acetylation, H3(K4) dimethylation, H3(K9) dimethylation, and H3(K27) trimethylation, and of histone methyltransferase Suv39H1, we used advanced image analysis methods, combined with Nipkow disk confocal microscopy. Total fluorescence intensity and distributions of fluorescently labelled proteins were analyzed in formaldehyde-fixed interphase nuclei. Our data showed reduced fluorescent signals of H3(K9) acetylation and H3(K4) dimethylation (di-me) at the nuclear periphery, while di-meH3(K9) was also abundant in chromatin regions closely associated with the nuclear envelope. Little overlapping (intermingling) was observed for di-meH3(K4) and H3(K27) trimethylation (tri-me), and for di-meH3(K9) and Suv39H1. The histone modifications studied were absent in the nucleolar compartment with the exception of H3(K9) dimethylation that was closely associated with perinucleolar regions which are formed by centromeres of acrocentric chromosomes. Using immunocytochemistry, no di-meH3(K4) but only dense di-meH3(K9) was found for the human acrocentric chromosomes 14 and 22. The active X chromosome was observed to be partially acetylated, while the inactive X was more condensed, located in a very peripheral part of the interphase nuclei, and lacked H3(K9) acetylation. Our results confirmed specific interphase patterns of histone modifications within the interphase nuclei as well as within their chromosome territories.

Journal paper [SMS14]

Abstract. Human Epithelial (HEp-2) cells are commonly used in the Indirect Immunofluorescence (IIF) tests to detect autoimmune diseases. The diagnosis consists of searching and classification to specific patterns created by Anti-Nuclear Antibodies (ANAs) in the patient serum. Evaluation of the IIF test is mostly done by humans, which means that it is highly dependent on the experience and expertise of the physician. Therefore, a significant amount of research has been focused on the development of computer aided diagnostic systems which could help with the analysis of images from microscopes. This work deals with the design and development of HEp-2 cells classifier. The classifier is able to categorize pre-segmented images of HEp-2 cells into 6 classes. The core of this engine consists of the following image descriptors: Haralick features, Local Binary Patterns, SIFT, surface description and a granulometry-based descriptor. These descriptors produce vectors that form metric spaces. k-NN classification is based on aggregated distance function which combines several features together. An extensive set of evaluations was performed on the publicly available MIVIA HEp-2 images dataset which allows a direct comparison of our approach with other solutions. The results show that our approach is one of the leading classifiers when comparing with other participants in the HEp-2 Cells Classification Contest.

Journal paper [WBS14]

Abstract. This work presents an objective performance analysis of statistical tests for edge detection which are suitable for textured or cluttered images. The tests are subdivided into two-sample parametric and non-parametric tests and are applied using a dual-region based edge detector which analyses local image texture difference. Through a series of experimental tests objective results are presented across a comprehensive dataset of images using a Pixel Correspondence Metric (PCM). The results show that statistical tests can in many cases, outperform the Canny edge detection method giving robust edge detection, accurate edge localisation and improved edge connectivity throughout. A visual comparison of the tests is also presented using representative images taken from typical textured histological data sets. The results conclude that the non-parametric Chi-Square (χ^2) and Kolmogorov-Smirnov (KS) statistical tests are the most robust edge detection tests where image statistical properties cannot be assumed a priori or where intensity changes in the image are nonuniform and that the parametric Difference of Boxes (DoB) test and the Student's T-test are the most suitable for intensity based edges. Conclusions and recommendations are finally presented contrasting the tests and giving guidelines for their practical use while finally confirming which situations improved edge detection can be expected.

Conference paper [KKS13]

• Pavel Karas, Michal Kuderjavý, and David Svoboda. "Deconvolution of Huge 3D Images: Parallelization Strategies on a Multi-GPU System". In: *Algorithms and Architectures for Parallel Processing*. Ed. by Joanna Kołodziej et al. Vol. 8285. Lecture Notes in Computer Science. Springer International Publishing, 2013, pp. 279–290. ISBN: 978-3-319-03858-2

Abstract. In this paper, we discuss strategies to parallelize selected deconvolution methods on a multi-GPU system. We provide a comparison of several approaches to split the deconvolution into subtasks while keeping the amount of costly data transfers as low as possible, and propose own implementation of three deconvolution methods which achieves up to $65\times$ speedup over the CPU one. In the experimental part, we analyse how the individual stages of the computation contribute to the overall computation time as well as how the multi-GPU implementation scales in various setups. Finally, we identify bottlenecks of the system.

Appendix B

Optimization of Convolution

Conference paper [Svo11]

- David Svoboda. "Efficient Computation of Convolution of Huge Images". In: *Proceedings of the 16th Int. Conference on ICIAP*. ICIAP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 453–462. ISBN: 978-3-642-24084-3

Abstract. In image processing, convolution is a frequently used operation. It is an important tool for performing basic image enhancement as well as sophisticated analysis. Naturally, due to its necessity and still continually increasing size of processed image data there is a great demand for its efficient implementation. The fact is that the slowest algorithms (that cannot be practically used) implementing the convolution are capable of handling the data of arbitrary dimension and size. On the other hand, the fastest algorithms have huge memory requirements and hence impose image size limits. Regarding the convolution of huge images, which might be the subtask of some more sophisticated algorithm, fast and correct solution is essential. In this paper, we propose a fast algorithm implementing exact computation of the shift invariant convolution over huge multi-dimensional image data.

Journal paper [KS11]

• Pavel Karas and David Svoboda. "Convolution of Large 3D Images on GPU and its Decomposition". In: *EURASIP Journal on Advances in Signal Processing* 2011.1, 120 (2011)

Abstract.   In this article, we propose a method for computing convolution of large 3D images. The convolution is performed in a frequency domain using a convolution theorem. The algorithm is accelerated on a graphic card by means of the CUDA parallel computing model. Convolution is decomposed in a frequency domain using the decimation in frequency algorithm. We pay attention to keeping our approach efficient in terms of both time and memory consumption and also in terms of memory transfers between CPU and GPU which have a significant inuence on overall computational time. We also study the implementation on multiple GPUs and compare the results between the multi-GPU and multi-CPU implementations.

Conference paper [KSZ12]

Abstract. In this paper, we propose a method for computing convolution of large 3D images with respect to real signals. The convolution is performed in a frequency domain using a convolution theorem. Due to properties of real signals, the algorithm can be optimized so that both time and the memory consumption are halved when compared to complex signals of the same size. Convolution is decomposed in a frequency domain using the decimation in frequency (DIF) algorithm. The algorithm is accelerated on a graphics hardware by means of the CUDA parallel computing model, achieving up to $10\times$ speedup with a single GPU over an optimized implementation on a quad-core CPU.

Reference: `http://dx.doi.org/10.1007/978-3-642-33140-4_6`

Appendix C

Texture Analysis

Conference paper [MS12]

- Tomáš Majtner and David Svoboda. "Extension of Tamura Texture Features for 3D Fluorescence Microscopy". In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. 2012, pp. 301–307

Abstract. The image descriptors are a very useful tool in the task of classification. In biomedical image analysis, they may characterize either the shape or the internal structure of studied objects. Both characteristics are very important. When analysing cells, their shape is usually determined first. In the second step, their mask may be used for the selection of the area where the texture descriptor should be applied. In this paper, we are going to focus on the texture-based image descriptors called Tamura features. For their basic properties, they seem to be a very promising tool applicable to the biomedical image data. We will apply them to selected types of cell lines and test how they perform. We will also introduce their extension to higher dimensions and show that they give even better results than in the 2D case.

Conference paper [MSS14]

• Tomáš Majtner, Roman Stoklasa, and David Svoboda. "RSURF – The Efficient Texture-Based Descriptor for Fluorescence Microscopy Images of HEP-2 Cells". In: *Pattern Recognition (ICPR), 2014 22nd International Conference on*. 2014, pp. 1194–1199

Abstract. In biomedical image analysis, object description and classification tasks are very common. Our work relates to the problem of classification of Human Epithelial (HEp-2) cells. Since the crucial part of each classification process is the feature extraction and selection, much attention should be concentrated to the development of proper image descriptors. In this article, we introduce a new efficient texture-based image descriptor for HEp-2 images. We compare proposed descriptor with LBP, Haralick features (GLCM statistics) and Tamura features using the public MIVIA HEp-2 Images Dataset. Our descriptor outperforms all previously mentioned approaches and the kNN classifier based solely on the proposed descriptor achieve the accuracy as high as 91.1%.

Conference paper [MS14]

• Tomáš Majtner and David Svoboda. "Texture Analysis Using 3D Gabor Features and 3D MPEG-7 Edge Histogram Descriptor in Fluorescence Microscopy". In: *3D Imaging (IC3D), 2014 International Conference on*. 2014, pp. 1–7

Abstract.   The recognition of patterns with focus on texture and shape analysis is still very hot topic, especially in biomedical image processing. In this article, we introduce 3D extensions of well-known approaches for this particular area. We focus on the collection of MPEG-7 image descriptors, specifically on the Edge Histogram Descriptor (EHD) and Gabor features, which are the core of the Homogeneous Texture Descriptor (HTD). The proposed extensions are evaluated on the dataset consisting of three classes of 3D volumetric biomedical images. Two different classifiers, namely k-NN and Multi-Class SVM, are used to evaluate the proposed algorithms. According to the presented tests, the proposed 3D extensions clearly outperform their 2D equivalents in the classification tasks.

Appendix D

Simulations

Conference paper [Svo+07]

• David Svoboda et al. "On Simulating 3D Fluorescent Microscope Images." In: *CAIP*. ed. by W. G. Kropatsch, M. Kampel, and A. Hanbury. Vol. 4673. LNCS. Springer, 2007, pp. 309–316

Abstract.   In recent years many various biomedical image segmentation methods have appeared. Though typically presented to be successful the majority of them was not properly tested against ground truth images. The obvious way of testing the quality of new segmentation was based on visual inspection by a specialist in the given field. The novel 3D biomedical image data simulator is presented in this paper. It offers the results of high quality. The comparison of generated synthetic data is compared against real image data using standard similarity techniques.

Reference: `http://dx.doi.org/10.1007/978-3-540-74272-2_39`

Journal paper [SKS09]

Abstract.   Image cytometry still faces the problem of the quality of cell image analysis results. Degradations caused by cell preparation, optics, and electronics considerably affect most 2D and 3D cell image data acquired using optical microscopy. That is why image processing algorithms applied to these data typically offer imprecise and unreliable results. As the ground truth for given image data is not available in most experiments, the outputs of different image analysis methods can be neither verified nor compared to each other. Some papers solve this problem partially with estimates of ground truth by experts in the field (biologists or physicians). However, in many cases, such a ground truth estimate is very subjective and strongly varies between different experts. To overcome these difficulties, we have created a toolbox that can generate 3D digital phantoms of specific cellular components along with their corresponding images degraded by specific optics and electronics. The user can then apply image analysis methods to such simulated image data. The analysis results (such as segmentation or measurement results) can be compared with ground truth derived from input object digital phantoms (or measurements on them). In this way, image analysis methods can be compared with each other and their quality (based on the difference from ground truth) can be computed. We have also evaluated the plausibility of the synthetic images, measured by their similarity to real image data. We have tested several similarity criteria such as visual comparison, intensity histograms, central moments, frequency analysis, entropy, and 3D Haralick features. The results indicate a high degree of similarity between real and simulated image data.

Conference paper [SHS11]

• David Svoboda, Ondřej Homola, and Stanislav Stejskal. "Generation of 3D Digital Phantoms of Colon Tissue". In: *Proc. of the 8th Int. Conference on Image Analysis and Recognition, ICIAR 2011*. Vol. 6754. LNCS. Springer, 2011, pp. 31–39

Abstract. Although segmentation of biomedical image data has been paid a lot of attention for many years, this crucial task still meets the problem of the correctness of the obtained results. Especially in the case of optical microscopy, the ground truth (GT), which is a very important tool for the validation of image processing algorithms, is not available.

We have developed a toolkit that generates fully 3D digital phantoms, that represent the structure of the studied biological objects. While former papers concentrated on the modelling of isolated cells (such as blood cells), this work focuses on a representative of tissue image type, namely human colon tissue. This phantom image can be submitted to the engine that simulates the image acquisition process. Such synthetic image can be further processed, e.g. deconvolved or segmented. The results can be compared with the GT derived from the digital phantom and the quality of the applied algorithm can be measured.

Conference paper [SU13]

• David Svoboda and Vladimír Ulman. "Towards a Realistic Distribution of Cells in Synthetically Generated 3D Cell Populations". In: *Proceedings of 17th International Conference on Image Analysis and Processing*. Vol. 8157. LNCS. Springer, 2013, pp. 429–438

Abstract. In fluorescence microscopy, the proper evaluation of image segmentation algorithms is still an open problem. In the field of cell segmentation, such evaluation can be seen as a study of the given algorithm how well it can discover individual cells as a function of the number of them in an image (size of cell population), their mutual positions (density of cell clusters), and the level of noise. Principally, there are two approaches to the evaluation. One approach requires real input images and an expert that verifies the segmentation results. This is, however, expert dependent and, namely when handling 3D data, very tedious. The second approach uses synthetic images with ground truth data to which the segmentation result is compared objectively. In this paper, we propose a new method for generating synthetic 3D images showing naturally distributed cell populations attached to microscope slide. Cell count and clustering probability are user parameters of the method.

Conference paper [Svo+14]

- David Svoboda et al. "On Proper Simulation of Phenomena Influencing Image Formation in Fluorescence Microscopy". In: *2014 IEEE International Conference on Image Processing (ICIP)*. 2014, pp. 3944–3948

Abstract. The simulation plays an important role in biomedical image analysis as it can inherently provide large collections of image data with absolute ground truth. Compared to real image data typically annotated by some expert, the computer generated data still lacks the authenticity due to the simplifications of various natural phenomena. In this paper, we focus on simulating the photobleaching effect and the uneven illumination, being simplified or even omitted from the majority of present simulation toolkits, to considerably improve the authenticity of computer generated data.

Conference paper [SUP15]

• David Svoboda, Vladimír Ulman, and Igor Peterlík. "On Proper Simulation of Chromatin Structure in Static Images As Well As in Time-Lapse Sequences in Fluorescence Microscopy". In: *Proceedings of 2015 IEEE International Symposium on Biomedical Imaging*. Stoughton (WI, USA): Engineering in Medicine and Biology Society, 2015, pp. 712–716. ISBN: 978-1-4799-2375-5

Abstract. In fluorescence microscopy, where the benchmark datasets for validating the various image analysis methods are difficult to obtain, a great demand is either for manually annotated real image data or for realistic computer generated ones. In the last two decades, the latter case has become more and more accessible due to an increasing computer capabilities. However, the development of elaborate models, especially in the field of fluorescence microscopy imaging, is less progressive. In this paper, we propose a novel approach, based on well established concepts, to properly imitate the structure of chromatin inside the interphase cell nucleus as well as its dynamics. The performance of the approach was quantitatively evaluated against the real data. The results show that the produced images are sufficiently plausible and visually resemble their real counter parts, both for fixed and living cells.

Journal paper [Ma14]

Abstract. Automatic tracking of cells in multidimensional time-lapse fluorescence microscopy is an important task in many biomedical applications. A novel framework for objective evaluation of cell tracking algorithms has been established under the auspices of the IEEE International Symposium on Biomedical Imaging 2013 Cell Tracking Challenge. In this article, we present the logistics, datasets, methods and results of the challenge and lay down the principles for future uses of this benchmark.

The main contributions of the challenge include the creation of a comprehensive video dataset repository and the definition of objective measures for comparison and ranking of the algorithms. With this benchmark, six algorithms covering a variety of segmentation and tracking paradigms have been compared and ranked based on their performance on both synthetic and real datasets. Given the diversity of the datasets, we do not declare a single winner of the challenge. Instead, we present and discuss the results for each individual dataset separately.

Conference paper [Svo+16]

• David Svoboda et al. "Vascular Network Formation in Silico Using the Extended Cellular Potts Model". In: *2016 IEEE International Conference on Image Processing (ICIP)*. Signal Processing Society, 2016, pp. 3180–3183. ISBN: 978-1-4673-9961-6

Abstract. Cardiovascular diseases belong to the most widespread illnesses in the developed countries. Therefore, the regenerative medicine and tissue modeling applications are highly interested in studying the ability of endothelial cells, derived from human stem cells, to form vascular networks. Several characteristics can be measured on images of these networks and hence describe the quality of the endothelial cells. With advances in the image processing, automatic analysis of these complex images becomes increasingly common. In this study, we introduce a new graph structure and additional constraints to the cellular Potts model, a framework commonly utilized in computational biology. Our extension allows to generate visually plausible synthetic image sequences of evolving fluorescently labeled vascular networks with ground truth data. Such generated datasets can be subsequently used for testing and validating methods employed for the analysis and measurement of the images of real vascular networks.

Journal paper [SU16]

• David Svoboda and Vladimír Ulman. "MitoGen: A Framework for Generating 3D Synthetic Time-Lapse Sequences of Cell Populations in Fluorescence Microscopy". In: *IEEE Transactions on Medical Imaging* (2016). in press

Abstract. The proper analysis of biological microscopy images is an important and complex task. Therefore, it requires verification of all steps involved in the process, including image segmentation and tracking algorithms. It is generally better to verify algorithms with computer-generated ground truth datasets, which, compared to manually annotated data, nowadays have reached high quality and can be produced in large quantities even for 3D time-lapse image sequences. Here, we propose a novel framework, called MitoGen, which is capable of generating ground truth datasets with fully 3D time-lapse sequences of synthetic fluorescence-stained cell populations. MitoGen shows biologically justified cell motility, shape and texture changes as well as cell divisions. Standard fluorescence microscopy phenomena such as photobleaching, blur with real point spread function (PSF), and several types of noise, are simulated to obtain realistic images. The MitoGen framework is scalable in both space and time. MitoGen generates visually plausible data that shows good agreement with real data in terms of image descriptors and mean square displacement (MSD) trajectory analysis. Additionally, it is also shown in this paper that four publicly available segmentation and tracking algorithms exhibit similar performance on both real and MitoGen-generated data. The implementation of MitoGen is freely available.