
Abstract

Modern fluorescence microscopes with high-resolution cameras are capable of acquiring large images at a fast rate. Data rates of 1 GB/s are common with CMOS cameras, and the three-dimensional (3D) image volumes acquired by light-sheet microscopy (Huisken et al., 2004) routinely exceed tens of gigabytes per image, and tens of terabytes per time-lapse experiment (Huisken and Stainier, 2007; Preibisch et al., 2010; Schmid et al., 2013). This defines new challenges in handling, storing, and analyzing the image data, as image acquisition outpaces analysis capabilities.

Ideally, the images are analyzed during acquisition with analysis times that are smaller than the time until the next image is acquired. This “real-time” image analysis not only alleviates the data bottleneck, but is also a prerequisite for smart microscopes that optimize the acquisition of the next image based on the contents of the current image (Scherf and Huisken, 2015). Real-time segmentation also enables interactive experiments where, e.g., optical manipulation and tracking become feasible in a developing embryo (Amat et al., 2014).

Real-time, or more precisely acquisition-rate, segmentation of large images is usually hindered by the memory requirements of the image data and the analysis algorithm. Segmenting an image requires about 5 to 10 times more memory than the raw image data (Caselles et al., 1997; Beare and Lehmann, 2006; Al-Kofahi et al., 2010). This means that in order to segment a 30 GB 3D light-sheet microscopy image, one would need a computer with 150 to 300 GB of main memory. Image segmentation at acquisition rate has hence mainly been achieved for smaller images (Stegmaier et al., 2014). For example, segmenting a $2048 \times 2048 \times 400$ pixel image of stained nuclei, which translates to about 3 GB file size at 16 bit depth, required more than 32 GB of main memory (Stegmaier et al., 2014).

Acquisition-rate processing of large images has so far been limited to low-level image processing, such as filtering or blob detection. Pixel-by-pixel low-level processing has been accelerated by Olmedo et al. (2012) using CUDA as a parallel programming tool on graphics processing units (GPUs). In their work, pixel-wise operations are applied to many pixels simultaneously, rather than sequentially looping through pixels. While such GPU acceleration achieves high processing speeds and data rates, it is limited by the size of the

GPU memory, which is in general smaller than the main memory. Another approach is to distribute different images to different computers. In a time-lapse sequence, every image can be sent to a different computer for processing. Using 100 computers, every computer has 100 frames time to finish processing its image, until it receives the next one. While this does not strictly fulfill the definition of acquisition-rate processing (e.g., it would not be useful for a smart microscope), it improves data throughput by pipelining. Galizia et al. (2015) have demonstrated this in the parallel image processing library *GENoa*, which runs on computer clusters using the Message Passing Interface (MPI) to distribute work, but it also runs on GPUs and GPU clusters. This library focuses on low-level image processing. Both GPU acceleration and embarrassingly parallel work-farming approaches are, however, unable to provide acquisition-rate high-level image analysis of single large images or time series comprised of large images.

High-level image analysis in fluorescence microscopy is mostly concerned with image segmentation (Aubert and Kornprobst, 2006; Cremers et al., 2007). In image segmentation, the task is to detect and delineate objects represented in the image. This is a high-level task, which cannot be done in a pixel-independent way. It also cannot be formulated as a shader or filter, rendering it hard to exploit the speed of GPUs. Finally, as outlined above, high-level image analysis of large images quickly exceeds the main memory of a single computer. This memory limitation can be overcome by subsampling the image, for example coarse-graining groups of pixels to *super-pixels*. This has been successfully used for acquisition-rate detection of nuclei and lineage tracking from large 3D images (Amat et al., 2014). The generation of super-pixels only requires low-level operations, where the high-level analysis is done on the reduced data. While this effectively enables acquisition-rate high-level analysis, it does not provide single-pixel resolution and is somewhat limited to the specific application of lineage tracing.

Pixel-accurate high-level analysis of large images can be achieved by splitting each image into smaller sub-images and distributing them across multiple computers or memories, thus distributing the data and the work. The computers then work in parallel, each on its sub-image. They communicate over a network interconnect in order to collectively solve the same high-level image-analysis problem that a single computer would have solved. However, since the data are distributed, the solution is available faster, and arbitrar-

ily large images can be accommodated by distributing across more computers. This is the hallmark of *distributed-memory parallelism*. This strategy enables segmentation of large images, and accelerates segmentation to match the time scale of image acquisition. However, image segmentation constitutes an ill-posed problem. Thus, the segmentation might be uncertain without providing any information about the uncertainty in the solution. Providing confidence intervals and uncertainties of the analysis results would provide more information and enable higher-level reasoning about the solution. Therefore, it is critical to capture the uncertainty of the segmentation before declaring the final decision or storing the results, in order to decide whether an observed difference between samples is real, or a processing artifact. In Bayesian image analysis, the distribution of potential results and a representative collection of them can be estimated by drawing samples from the posterior distribution. Representative samples from the posterior provide additional insight into the robustness of the segmentation. However, the sampling approach, no matter how smart, is computationally expensive and inherently sequential. We explore the distributed-memory approaches that relax these issues for data that do not fit the memory of a single computer.

In this thesis, we address both processing-time and memory issues by developing a distributed parallel framework for segmentation and uncertainty quantification of large fluorescence microscopy images. The method is based on the versatile Discrete Region Competition (Cardinale et al., 2012) algorithm, which has previously proven useful in microscopy image segmentation. The present distributed implementation decomposes the input image into smaller sub-images that are distributed across multiple computers. Using network communication, the computers orchestrate the collective solving of the global segmentation problem. This not only enables segmentation of large images (we test images of up to 10^{10} pixels), but also accelerates segmentation to match the time scale of image acquisition. Such acquisition-rate image segmentation is a prerequisite for the smart microscopes of the future and enables online data inspection and interactive experiments.

We also address both the processing-time and memory issues of Markov-chain Monte Carlo algorithms for assessing the segmentation quality and robustness. The method is based on a novel, efficient, particle-based Metropolis-Hastings algorithm (Cardinale, 2013), called discrete region sampling (DRS). Again, the present dis-

tributed implementation decomposes the input image into smaller sub-images that are distributed across multiple computers. Using network communication, the computers orchestrate the collective sampling from the posterior distribution of the defined segmentations on the observed image in a statistically unbiased manner.

We begin by discussing the relevant background material in Chapter. 2. After motivating the Bayesian formulation of image analysis, we introduce the DRC image-segmentation algorithm. We give a brief introduction to Markov chain theory, followed by a review of previous shape sampling approaches. We provide an introduction to distributed computing and conclude with an overview of the parallel particle mesh library (PPM) as a middleware for distributed particle methods.

We present a distributed-memory parallel design and implementation of the generic image-segmentation algorithm DRC. The present implementation scales to large images. Here, we test images of size up to $8192 \times 8192 \times 256 = 1.7 \cdot 10^{10}$ pixels, corresponding to 32 GB of data per image at 16 bit depth. We show that distributing an image across 128 computers enables acquisition-rate segmentation of large light-sheet microscopy images of *Drosophila* embryos. *Discrete Region Competition* (DRC) (Cardinale et al., 2012) is a general-purpose model-based segmentation method. It is not limited to nucleus detection or any other task, but solves generic image-segmentation problems with pixel accuracy. The method is based on using computational particles to represent image regions. This particle-method character renders the computational cost of the method independent of the image size, since it only depends on the total contour length of the segmentation. Storing the information on particles effectively reduces the problem from 3D to 2D (or from 2D to 1D). Moreover, the particle nature of the method lends itself to distributed parallelism, as particles can be processed concurrently, even if pixels cannot. In terms of computational speed, DRC has been shown competitive with fast discrete methods from computer vision, such as multi-label graph-cuts (Cardinale et al., 2012; Delong et al., 2012). Single-processor DRC has previously been demonstrated on 2D and 3D images using a variety of different image models, including piecewise constant, piecewise smooth, and deconvolving models (Cardinale et al., 2012).

The piecewise constant and piecewise smooth models are also available in the present distributed-memory parallel implementation. This makes available a state-of-the-art generic image segmentation

toolbox for acquisition-rate analysis of large images that do not need to fit the memory of a single computer. The main challenge in parallelizing the DRC algorithm is to ensure global topological constraints on the image regions. These are required in order for regions to remain closed or connected. The main algorithmic contribution of the present work is hence to propose a novel distributed algorithm for the independent-sub-graph problem. Moreover, we present a new parallel algorithm for connected-component labeling in 2D and 3D images. The presented algorithm is both memory and computationally efficient and scales to large numbers of processors.

The algorithmic solutions presented in Chapter 3 ensure that the final result computed is the same that would have been computed on a single computer, and that the network communication overhead is kept to a minimum, hence ensuring scalability to large images.

Since each computer only stores its local sub-image, information needs to be communicated between neighboring sub-images in order to ensure global consistency of the solution. Since DRC is a particle method, we use the Parallel Particle Mesh (PPM) library (Sbalzarini et al., 2006; Awile et al., 2010, 2013) for work distribution and orchestration of the parallel communication.

We then present the main algorithmic contribution that made this possible: the distributed independent-sub-graph algorithm. We demonstrate correctness of the parallel implementation by comparing with the sequential reference implementation of DRC (Cardinale et al., 2012), as available in ITK (Ibanez et al., 2005). We then benchmark the scalability and parallel efficiency of the new parallel implementation on synthetic images, where the correct solution is known. Finally, we showcase the use of the present implementation for acquisition-rate segmentation of light-sheet fluorescence microscopy images.

Accurately and robustly segmenting an image is a challenging task. Usually, the segmentation is not unique and can be locally uncertain. Even using a global approach and providing bounds on the final energy of a solution does not provide information about its quality. This means that many cases require user interaction, for example by iterating over the results and correcting mistakes. This process is usually too time-consuming or does not yield reproducible results, especially at places in the image with low signal-to-noise ratio.

Estimating the uncertainty and robustness of a segmentation can reduce and guide user interaction. It also enables statistical

tests conveying the information about the segmentation reliability. Cardinale (2013) presented a method for sampling from the posterior distribution of explicitly or implicitly defined segmentations, conditioned on the observed image. In this approach, a discrete deformable model evolves, such that the sampled segmentations approximate the posterior distribution of possible segmentations. This allows assessing segmentation robustness. The presented particle-based Metropolis-Hastings algorithm, called *Discrete Region Sampling*, has been compared with a state-of-the-art algorithm by Chang and Fisher (2012) in terms of solution quality and has been shown competitive w.r.t. computation time.

In Chapter 4, we present a distributed-memory parallel extension of this sampling approach. Because of the inherently sequential nature of the sampling approach, parallelization is challenging. The primary challenge in parallelizing the DRS algorithm is to update a large number of particles in a statistically unbiased way and to guarantee the detailed balance, which is a sufficient condition for convergence. The main algorithmic contribution of the present work is to propose and implement a novel parallel distributed algorithm for efficiently sampling from the posterior distribution of defined segmentations on the observed image in a statistically correct manner.

Finally, in Chapter 5, we summarize the present work and the contributions of this thesis. We provide recommendations for extensions and future work pertaining to the developed parallel algorithms for segmentation and uncertainty quantification of fluorescence-microscopy images.