

Non-invasive mouse embryonic stem cells population estimation using machine vision

Nicolas Jaccard¹ - August 27, 2010

Supervisors: Dr. Nicolas Szita² | Dr. Lewis Griffin³

¹Center for Mathematics and Physics in the Life Sciences and Experimental Biology

²Department of Biochemical Engineering

³Department of Computer Sciences

Embryonic stem cells derived from the inner cell-mass of the blastocyst are self-renewable, meaning they can undergo an infinite number of divisions. They are also pluripotent and can give rise to most adult cell types through differentiation, a genetically regulated process during which cells gain specialised phenotypes. These unique properties make them a potentially unlimited source of material for developmental biology, drug discovery and regenerative medicine. However, stem cells-based bioprocesses still are in their infancy and remain inefficient. Cell growth, along with dissolved oxygen and pH, is part of the measurements required to fully characterise a bioprocess. Yet, traditional cell density assessment methods require adherent cells to be detached from their growth surface prior to analysis, thus limiting growth characterisation to the initial and final densities.

The goal of this project was to develop a machine vision-based mouse embryonic stem cells population estimation tool that could operate without addition of dyes or any other interference with the culture itself. This method would enable kinetic studies of cell growth over long periods of time and therefore could be used for stem cells-based bioprocesses optimisation. First, the theoretical feasibility of this approach was evaluated using numerical simulations. Then a flexible and extensible image processing framework was employed to devise the best algorithms for the analysis of large sets of microscopy images. Finally, the method was used to characterise the growth kinetics of two cell lines (CHO K1 and mESC Oct4-GiP) and was validated against conventional counting methods.

The numerical simulations showed that both the projected cell area (termed cells pixel ratio) and the extrapolated cell number (average number of cells counted per image multiplied by the number of images necessary to cover the whole culture surface) could be used to estimate the population of an adherent cell culture. In addition, 20 images were shown to be enough to obtain an estimation within 10% of the ground truth value for cultures in 6 well-plates. The results of the experimental implementation of the cells pixel ratio estimator showed that apriori information about the process was required in order for this estimator to be used as a proxy for cell density. A simple exponential decay model was used to take into account the dynamic changes in cell morphology and resulted in a nearly linear relationship between cell density and cells pixel ratio determined from mESC microscopy images.

Contents

1	Introduction	3
1.1	Embryonic stem cells	3
1.1.1	Embryonic stem cells and fate determination	3
1.1.2	ESC bioprocessing: challenges and opportunities	4
1.2	Machine-vision assisted bioprocessing	6
1.3	Project	7
1.3.1	Context and relevance	7
1.3.2	Structure	9
2	Material and methods	9
2.1	Material	9
2.1.1	Cell culture	9
2.2	Methods	10
2.2.1	Cells maintenance	10
2.2.2	Growth characterisation	11
2.2.3	Analytics	12
2.2.4	Image processing and numerical simulations	12
2.2.5	Statistical analysis	13
3	Results & discussion	14
3.1	Numerical simulations	14
3.1.1	Quantification methods	15
3.1.2	Effects of the cell distribution homogeneity on steady state deviation	18
3.1.3	Limiting the number of images	20
3.1.4	Insights from the simulations	21
3.1.5	Current limitations	22
3.2	Image processing framework	22
3.2.1	Scripting engine, queue system and scoring	23
3.2.2	Framework features of interest	25
3.2.3	Example of applications	27
3.3	Experimental implementation	30
3.3.1	Cells pixel ratio as a cell density estimator	32
3.3.2	Data reconciliation	33
4	Conclusion	34
5	Future work	35
6	Acknowledgements	36
	Bibliography	37

1 Introduction

1.1 Embryonic stem cells

'Holy grail of medicine', 'magic bullet' and more recently 'superheroes' [1]: there is no shortage of metaphors to qualify the promises held by embryonic stem cells (ESC). Their ability to give rise to any adult cell type through differentiation combined with their high proliferative capacity could potentially lead to the emergence of novel therapeutic strategies targeting yet untreatable conditions. But it has yet to be shown whether ESC can live up to their promises in a clinical context. In addition, numerous bioprocessing challenges will need to be overcome if stem cells-based therapies are to become mainstream.

1.1.1 Embryonic stem cells and fate determination

In the end of the eighties, the first mouse embryonic stem cells (mESC) line was derived by Hooper et al. [2]. A decade later, the first successful culture of human embryonic stem cells (hESC) was reported by Thomson and colleagues [3]. In both cases, cells originated from the inner cell mass of a developing embryo at the blastocyst stage of the development. The cells in culture retained characteristics that were until then only observed in the embryo.

Self-renewing division Most of the cells found in an adult organism are locked out of the cell cycle (in phase ' G_0 '), with notable exceptions like fibroblasts, which are able to divide. They have, however, a limited lifespan; most likely as a tumor protection mechanism, cells age and, after a certain amount of divisions (called the Hayflick number), lose proliferation capacity and eventually die, a phenomena known as senescence. In contrast, embryonic stem cells can escape that fate, due to their high telomerase activity [3].

Pluripotency ESC are not yet committed to any specific cell type, they are said to be *pluripotent*. Specific environmental cues can trigger a succession of genetic regulatory events, termed differentiation, during which cells gradually specialise before eventually reaching a terminally differentiated state [4]. ESC can potentially give rise to any cell type of the three germ layers (endoderm, mesoderm and ectoderm)[5].

Fate determination Pluripotency seems to be a requirement for self-renewing division. As cells undergo differentiation, they progressively lose their proliferative capacity. Their fate is mostly determined by the environmental conditions. The fundamental mechanisms underlying fate determination are yet to be fully understood.

Research and clinical applications of ESC Embryonic stem cells have been extensively studied in developmental biology and led to major advances in that field [6]. They also serve as ideal models for drug discovery [7] and toxicology studies [8]. It is, however, in the context of the clinic that stem cells garner the most interest: generating large amount of cells of any type and transplant them into a patient would allow to target a large number of yet untreatable conditions. This regenerative approach uses cells as the therapeutic agents to restore tissue functions, in opposition to traditional therapeutic strategies that use small molecules such as monoclonal antibodies to target individual symptoms. ESC-based therapies have been proposed for neurodegenerative diseases [9], the treatment of brain damages after a stroke [10], cancer [11] and HIV [12]. Recently, an important step was made when the first ESC-based therapy clinical trial was approved for the treatment of spinal cord injuries [13].

Controversy and induced pluripotent stem cells Embryonic stem cells have been surrounded by ethical issues and controversies regarding their origin; in order to derive a cell line, a human embryo has to be destroyed. However, recent breakthroughs might alleviate these concerns: Takahashi and colleagues successfully reprogrammed adult fibroblasts into pluripotent stem cells by expressing a specific cocktail of genes [14]. The resulting induced pluripotent stem cells (iPSC) have been shown to be very similar to embryonic stem cells. In addition, the fact that any adult cell could potentially be reprogrammed paves the way for patient-specific therapies, reducing risks of immune response upon transplantation.

1.1.2 ESC bioprocessing: challenges and opportunities

As noted above, many challenges need to be overcome so the transition from the laboratory to the clinic can happen. Stem cells bioprocessing usually consists in generating a large population of pluripotent cells before triggering their differentiation into the desired cell type (figure 1). It is critical to maintain cells in a pluripotent state during expansion as they will lose their proliferative capacity when differentiation is triggered.

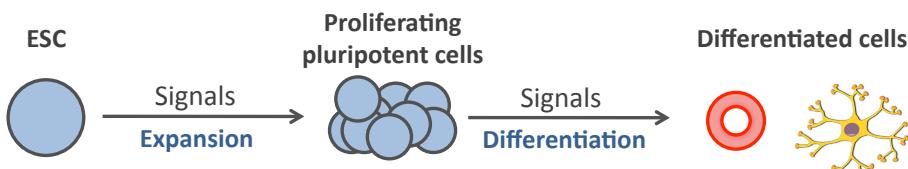


Figure 1: Stem cells bioprocessing in a nutshell: cells are expanded in conditions promoting their proliferation. Once the desire cell density is attained, conditions are changed, triggering differentiation to the desired cell type (e.g. red blood cell or neuron)

Controlling cell fate The behavior of stem cells is regulated by complex combinations of signals. These can be categorised as soluble factors (culture medium formulation), physico-chemical conditions (temperature, pH, osmolarity), extra cellular matrix composition and physical forces [15]. The nature of the signals and their intensity will dictate the fate of the cells. The key to optimise ESC bioprocessing protocols is to be able to both control the input signals and reliably measure cell response (figure 2). Additionally, it is important not to generate undesired signals during culture itself but also during routine maintenance. For instance, it has been shown that pipetting and centrifugation steps could impact the proliferative capacity and the differentiation yield of mESC [16].

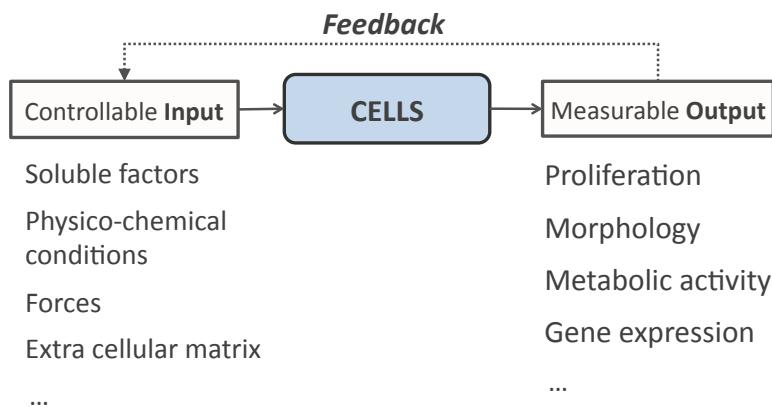


Figure 2: By controlling different inputs, it is possible to measure the cell response and use the information as a feedback to optimise the process

Limitations of current approaches Large scale cell culture is mostly performed using cell lines adapted to suspension. This approach allows attaining high cell densities while maintaining high level of monitoring and control on various culture conditions (e.g. temperature, pH or agitation). Although promising early results for the culture in suspension of pluripotent hESC have recently been reported [17, 18], most of the existing protocols rely on the attachment to a culture surface as cell-to-cell interactions seem to be required to maintain long-term pluripotency [19, 20]. The degree of control and the experimental throughput offered by most 2D culture vessels are very limited and are therefore not suited for bioprocess optimisation. As a middle-ground solution, it has been proposed to culture cells on microcarriers that are themselves in suspension. Preliminary results are encouraging but the additional downstream processing steps required to separate the cells from their carriers have yet to be perfected [21, 22].

Opportunities It is estimated that a transplant therapy would require up to two billion cells to be viable [23]. Although current approaches are not yet capable of generating the

amount of cells necessary to provide such therapies to a large number of patients, there is room for improvements and optimisations [24]. These will most likely require changes in stem cells cultures paradigms; novel bioprocessing platforms such as microfluidics bioreactors offer a high degree of control over the cells microenvironment while drastically improving experimental throughput through miniaturisation and multiplexing [25, 26]. Their ability to closely mimic *in-vivo* conditions, together with appropriate monitoring tools, could potentially lead to a better understanding of the mechanisms underlying stem cells proliferation and differentiation.

1.2 Machine-vision assisted bioprocessing

Be it qualitative or quantitative, microscopy is an integral part of any bioprocess. Indeed, microscopy images are information rich and can be used to infer various characteristics of a culture. For instance, the morphology of the cells depends on the culture conditions and can be considered as a marker for starvation due to a lack of nutrients, physical stress, cell death or even differentiation in the case of stem cells. In addition, the use of fluorescent dyes greatly simplified the imaging of whole cells or sub-cellular components. However, turning this visual information into quantitative data remains a challenge. It is possible to manually process images using packages such as ImageJ¹ but this approach is not practicable when large amount of images need to be analysed or when real-time processing is required. Machine vision approaches have been developed to alleviate these issues and have been applied (with more or less success) to bioprocessing.

Cell detection using fluorescence markers The localisation and identification of cells on microscopy images can be considerably simplified if fluorescent markers are used to stain specific cell structures such as the nucleus, cytoskeleton or membrane. The position of said structures is inferred from the fluorescence signal and is then used to seed segmentation algorithms. Zhou and colleagues used the signal of fluorescently-labeled DNA to determine the centres of a Voronoi diagram, allowing the segmentation of the cells [27]. Similarly, Yu *et al* recently proposed an evolving Voronoi diagram approach that resulted in very accurate segmentation of touching and overlapping stained cells [28]. Chang and Zandstra used stained nuclei to identify mouse embryonic stem cells in culture and colocalize various fluorescently-marked proteins of interest, allowing to determine the effect of different conditions on the expression of said proteins [29]. A similar approach was employed to estimate the population of human embryonic stem cells that were previously stained using DAPI (nucleus marker); the nuclei signal served as a basis for cell segmentation using a watershed algorithm [30].

Label-free cell detection In some cases, it is not possible to use fluorescent markers; staining is disruptive to the culture and the expression of labeled proteins require the

¹<http://rsbweb.nih.gov/ij/>

transfer of exogenous DNA to the cell and its integration, which again introduces a potential bias into the experiment. In addition, marker proteins such as GFP (Green Fluorescent Protein) have been found to be toxic when accumulated in the cell [31]. Label-free cell segmentation is a difficult machine vision problem and has yet to be perfected. However, many approaches offer satisfactory preliminary results. Korzynska and colleagues proposed a combination of texture and contour segmentation techniques for the detection of neutrophils and lymphocytes on brightfield microscopy images [32]. Another major challenge of microscopy image processing is to tell apart cells that are touching each other; one solution is to use edge linking, an approach that's based on the identification of pixels that have a high probability to be part of the cell membrane followed by their association to recreate an abstract representation of the adjacent cells membranes that can be used as a basis for further segmentation [33]. Alternatively, a method termed *snakes* was shown to efficiently segment touching cells by combining intensity gradient analysis and prior knowledge such as the cells curvature and size [34].

Automated machine vision approach for bioprocess optimisation Most of the examples cited above focus on the characterisation of small culture areas. There seems to be little to no attempts described in the literature to use a machine vision approach for the cell density estimation of a culture with dimensions relevant to bioprocess applications. A few commercial products claim to be able to track the number of cells in single colonies overtime but due to their closed nature, it was not possible to verify these assertions [35]. Alternatively, by combining traditional phase contrast microscopy and digital holography, a swedish group was able quantify the number of cells present in culture with an accuracy similar to that of traditional invasive methods. However, the price of the required equipment remains prohibitive [36].

1.3 Project

The aim of this project was to develop a machine vision-based mouse embryonic stem cells population estimation tool that could operate without addition of dyes or any other interference with the culture itself (figure 3). Multiple images of the culture would be acquired using a conventional inverted microscope before being processed without any user intervention. The end result would be an estimation of the cell density accurate enough to be used in a bioprocess optimisation context.

1.3.1 Context and relevance

Cell growth, along with dissolved oxygen and pH, is part of the measurements required to fully characterise a bioprocess. Growth rate is dependant of both the culture conditions (e.g. pH, temperature, gas phase or medium formulation) and the intrinsic properties of the reactor (e.g. material, mass transfer, size or geometry). It can be used as a basis

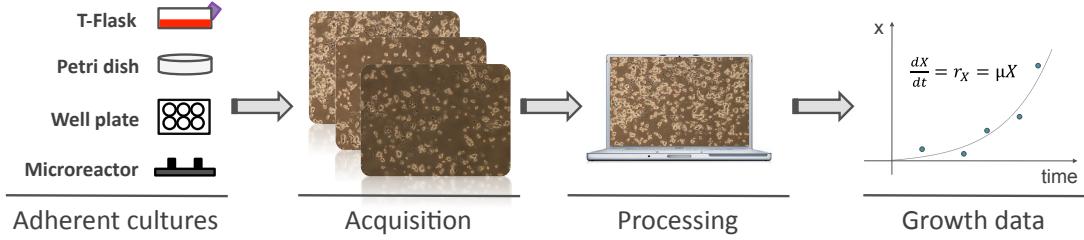


Figure 3: Method for the estimation of adherent cells population using machine vision

for the optimization of a specific process or for the comparison of the performances of different reactors. In the case of cultures in suspension, cell growth can be assessed using on-line methods (e.g. optical density) or at/off-line methods (e.g. manual or automated counting using a hemocytometer, dry weight biomass or packed cell volume [37]). While most of these apply to adherent cell cultures as well, they require cells to be detached from their growth surface and re-suspended prior to analysis. Not only such operations can induce cellular stress and potentially lead to partial differentiation of embryonic stem cells [16], but it also limits cell density determination to the initial and final densities (figure 4.a).

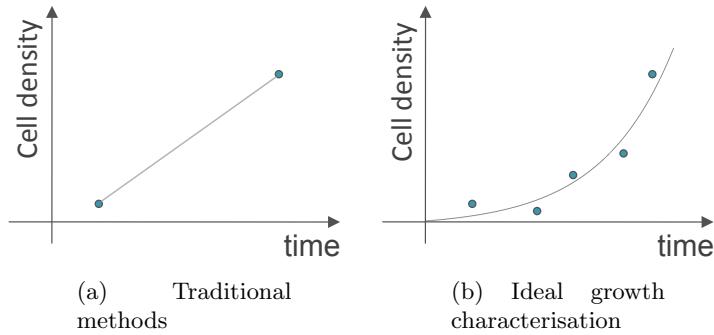


Figure 4: Difference between the cell density characterisation using (a) existing cell counting method and (b) the proposed method

It is clear that two points are not enough to fully characterise any process, let alone ones as complex as stem cells expansion or differentiation. Intermediate datapoints are required for kinetic studies of cell growth, which are of particular importance for expansion optimisation where the goal is to maximise growth rate while maintaining the cells in a pluripotent state. In addition, proliferation and differentiation being two mutually exclusive processes, the decrease in growth rate over time can be used to detect early differentiation events.

The proposed method (figure 3) would alleviate these issues by allowing non-invasive

time-lapse monitoring of the cell density (figure 4.b). The tool would be ideal for the optimisation of bioprocesses involving adherent cells in general and embryonic stem cells in particular. Unlike traditional counting methods that require tens of thousands of cells to generate reliable results, this approach would be applicable as long as cells can be imaged using a microscope. The range of applications spans from large culture flasks down to microfluidic reactors that only contain a few thousands cells and are usually well below the detection limit of conventional counting methods. In addition, the potential for automation makes it ideal for high-throughput bioprocessing, which employs multiple miniaturised reactors (microfluidics-based or otherwise) running in parallel.

1.3.2 Structure

The project consisted of three main parts: numerical simulations, the development of an image processing framework and the experimental implementations of the method.

Numerical simulations Simple yet appropriate numerical simulations were used to evaluate the feasibility of an image processing approach for cell density estimation.

Image processing framework The powerful built-in functions of Mathematica were leveraged into a flexible and extensible image-processing framework that was used to devise the various algorithms used through this project

Experimental implementations The cells pixel ratio method was used to characterise the growth kinetics of chinese hamster ovary cells (CHO) and mouse embryonic stem cells (mESC). The results were then validated against those obtained using conventional counting methods.

2 Material and methods

2.1 Material

2.1.1 Cell culture

Oct4-GiP culture medium DMEM Knockout (Invitrogen, UK) medium supplemented as described in table 1 before filtration (0.45 μm filter bottles, TPP, Switzerland). Leukaemia inhibitory factor was added post-filtration.

CHO K1 culture medium GMEM (Invitrogen, UK) medium supplemented as described in table 2 before filtration (0.45 μm filter bottles, TPP, Switzerland).

Table 1: Oct4-GiP culture medium formulation

Component	Final Concentration
β -mercaptoethanol (VWR, UK)	0.1 mM
Fetal Bovine Serum (Invitrogen, UK)	10% v/v
Non-essential amino acids (Invitrogen, UK)	1% v/v
L-Glutamine (Invitrogen, UK)	2 mM
Sodium pyruvate (Invitrogen, UK)	1 mM
Leukemia inhibitory factor (LIF, Millipore, UK)	$1 \cdot 10^6$ U l ⁻¹

Table 2: CHO K1 culture medium formulation

Component	Final Concentration
Fetal Bovine Serum (Invitrogen, UK)	10% v/v
Non-essential amino acids (Invitrogen, UK)	1% v/v
L-Glutamine (Invitrogen, UK)	2 mM
Sodium pyruvate (Invitrogen, UK)	1 mM

Cell detachment solution The solution used to detach cells from their culture surface was prepared by dissolving the reagents (table 3) in a Dulbecco's phosphate buffered saline solution without Ca²⁺/Mg²⁺ (DPBS, Sigma, UK).

Table 3: Detachement solution formulation

Reagent	Concentration [% v/v]
Trypsin (Invitrogen, UK)	0.04
EDTA (Invitrogen, UK)	0.9
Chick serum (Sigma, UK)	0.9

2.2 Methods

2.2.1 Cells maintenance

Mouse embryonic stem cells During routine maintenance, cells were grown in T-25 flasks (Nunc, USA) pre-coated with 0.1% gelatin in presence of 5 ml culture media. Cells were cultured in an Heraeus Hera-Cell 150 incubator (Jencons-PLS, UK) at 37°C and 5% CO₂. Medium was changed on a daily basis. Cells were passaged every 2 days: supernatant was discarded and cells were washed with 2 ml of DPBS before being incubated for 3 minutes in presence of 500 μ l of detachment solution (see table 3 for the formulation). Cells were then quenched with 2 ml of fresh culture medium, centrifuged for 3 minutes at 300g, re-suspended in growth media and then transferred to a new pre-coated flask.

CHO K1 cells During routine maintenance, cells were grown in T-25 flasks (Nunc, USA) in presence of 5 ml culture media. Cells were cultured in an MCO CO₂ incubator (Sanyo, UK) at 37°C and 5% CO₂. Cells were passaged every 3 days: supernatant was discarded and cells were washed with 2 ml of DPBS before being incubated for 3 minutes in presence of 500 μ l of detachment solution (see table 3 for the formulation). Cells were then quenched with 2 ml of fresh culture medium, centrifuged for 3 minutes at 300g, re-suspended in growth media and then transferred to a new flask.

2.2.2 Growth characterisation

Sacrificial method A sacrificial method was used to characterise the growth kinetics of different cell lines (figure 5). Each datapoint consisted of a 6-well plate with 3 wells seeded with cells at t_0 (each well is a replica). Wells were kept in their controlled environment until analysis.

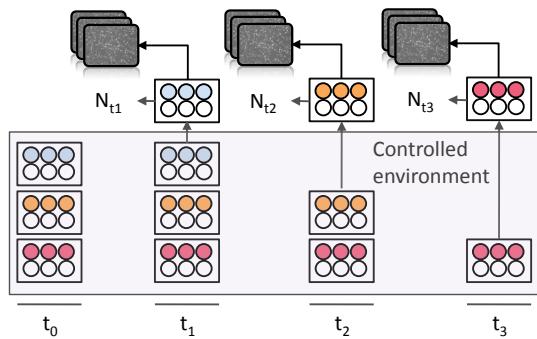


Figure 5: Sacrificial method: cells were seeded at t_0 and the content of the wells was sacrificed in sequence in order to have density information for each time point. Cells were maintained in a controlled environment until analysis

CHO K1 cells growth kinetic characterisation Cultures were performed in 6 well plates (Nunc, USA). Cells were seeded at a density of 50,000 cells cm⁻². Each datapoint was generated using the following procedure: for image analysis purposes, 30 random pictures were taken using a TE2000-U inverted microscope (Nikon UK Ltd. UK) at the specified magnification. Cells were washed with DPBS (Sigma, UK) before being incubated for 5 minutes in presence of a detachment solution (see table 3 for the formulation). Cells were then quenched with fresh culture medium, centrifuged for 3 minutes at 300g, re-suspended in growth media and then their density determined using a Vi-Cell automatic cell counter.

Mouse embryonic stem cells growth kinetic characterisation Cultures were performed in 6 well plates (Nunc, USA). Cells were seeded at a density of 10,000 cells cm⁻² and

culture medium was changed every 24 hours. Each datapoint was generated using the following procedure: for image analysis purposes, 15 random pictures were taken using a TE2000-U inverted microscope (Nikon UK Ltd. UK) at the specified magnification. Cells were washed with DPBS (Sigma, UK) before being incubated for 5 minutes in presence of a detachment solution (see table 3 for the formulation). Cells were then quenched with fresh culture medium, centrifuged for 3 minutes at 300g, re-suspended in growth media and then their density determined using FACS analysis.

2.2.3 Analytics

Cell density determination using FACS Resuspended cell samples were analysed using a fluorescence-activated cell sorting PCA-96 instrument (Merck KGaA, Germany). Cellular debris were gated out by using forward scatter (objects with a FSC-HLog lower than 441 were supposed to be debris). Cell viability was assessed by incubating cells with the ViaCount reagent (Merck KGaA, Germany) for at least 10 minutes at room temperature. Cells were considered GFP positive when the signal (GRN-HLog) was higher than 13 units.

Cell density determination using Vi-Cell Resuspended cell samples were analysed using a Vi-Cell instrument (Beckman Coulter, USA). The software parameters used are detailed in table 4 for CHO K1 cells.

Table 4: Vi-Cell parameters for CHO K1 cells

Paramater Name	Parameter Value
Minimum diameter	6 μ m
Maximum diameter	50 μ m
Number of images	50
Image brightness	85%
Viable cell spot brightness	75%
Declustering degree	Medium

2.2.4 Image processing and numerical simulations

Numerical simulations and the image processing framework were programmed using Mathematica (7.0.1.0, Wolfram Research Inc.).

Scoring metrics The various metrics employed to assess the performances of the image processing algorithms were for the most part taken from an introduction to receiver operating characteristics (ROC) analysis by Tom Fawcett [38]. Let FP be the number

of false positives, TP the number of true positives, P the total number of positives and N the total amount of negatives, then we can define the following metrics:

$$\text{Recall} = \text{True positive rate} \approx \frac{\text{Positives correctly classified}}{\text{Total positives}} = \frac{TP}{P} \quad (1)$$

$$\text{False positive rate} \approx \frac{\text{Negatives incorrectly classified}}{\text{Total negatives}} = \frac{FP}{N} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{F-Score} = \frac{2(\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \quad (4)$$

Cells pixel ratio and data reconciliation Microscopy images were first processed using the steps shown in table 5. The cells pixel ratio was then computed as the ratio of pixels that were detected as positive (pixels_{ON}) to the total number of pixels (equation 5). For data reconciliation, the cells pixel ratio results were corrected using an exponential decay term (equation 6) where α is the proportionality constant and t the culture time.

$$\gamma = \frac{\sum \text{pixels}_{ON}}{\sum \text{pixels}} \quad (5)$$

$$\gamma^* = \gamma \cdot \exp(\alpha t) = \frac{\sum \text{pixels}_{ON}}{\sum \text{pixels}} \cdot \exp(\alpha t) \quad (6)$$

Table 5: Image processing steps for the thresholding of the cells on microscopy images

Operation	Parameters
Rescale image	Size: 1280*768 (width*height)
Color Convert	Grayscale
Extract multiple BIFs	Features 2 to 7

Deviation from groundtruth The deviation from the ground truth (or relative error) was computed as following:

$$\text{Deviation} = \text{Relative Error} = \frac{|x_{groundtruth} - x_{measured}|}{x_{groundtruth}} \quad (7)$$

2.2.5 Statistical analysis

All the statistical analyses presented in this document were computed with a 95% confidence level ($\alpha=0.05$) and corrected for small sample size using Student's t-distribution when necessary.

3 Results & discussion

The results for the numerical simulations, the image processing framework and the experimental implementation are shown and discussed below. An overarching conclusion follows.

3.1 Numerical simulations

Numerical simulations were carried out in order to determine the theoretical feasibility of an image-processing based approach to the estimation of the cell density in the case of adherent cultures. The culture surface was abstracted as a matrix whose elements were positions in the plane that could be either empty or part of the area occupied by a given cell. It was critical to have a scale consistent with what was observed during *in-vitro* experiments; the field of view at different magnifications was determined using a microstructure of known dimensions (table 6). For comparison purposes, the number of

Table 6: Field of view at various magnifications

Magnification	Field of view [μm^2]	Field of view [cm^2]
2x	$1.0031 \cdot 10^7$	$1.003 \cdot 10^{-1}$
4x	$2.5078 \cdot 10^6$	$2.508 \cdot 10^{-2}$
10x	$4.0124 \cdot 10^5$	$4.012 \cdot 10^{-3}$
20x	$1.0031 \cdot 10^5$	$1.003 \cdot 10^{-3}$

images required to cover the whole culture surface of an individual well of a 6 well-plate (the culture vessel used for the different culture experiments in this study) was calculated for various magnifications (table 7). The matrix had fixed dimensions but a scaling

Table 7: Number of images required to cover the entirety of the culture surface of an individual well (9.6 cm^2) of a 6 well-plate culture vessel at various magnifications

Magnification	2x	4x	10x	20x
Number of images	96	383	2393	9570

factor allowed to simulate vessels of various dimensions by varying the resolution of the system. Cells were represented as rectangular subsets of the main matrix with dimensions randomly generated within the limit of specified parameters. The distribution of the cells on the culture surface is rarely homogeneous when doing *in-vitro* experiments. Typically, higher density areas can be observed in the central part of the vessel and around the edges, due to the nature of the seeding procedures. In the simulations, cells were either truly randomly distributed with each position having the same probability to be chosen or a bias could be introduced to force higher cell densities in the central area and on the edges (figure 6). Qualitative observations of mouse embryonic stem cell cultures in 6 well-plates vessels were used to determine dimensions of the key components of the

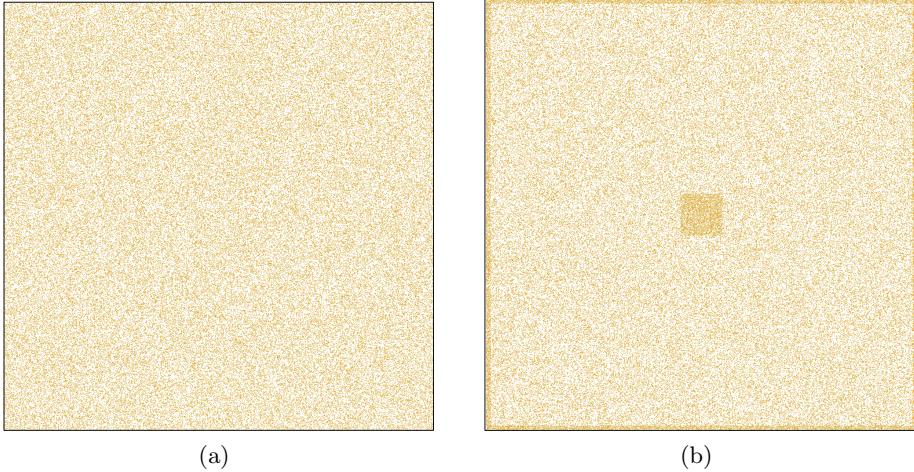


Figure 6: Numerical simulations of cells growing in an individual well of a 6 well-plate culture vessel. The seeding density was set to $13,800 \text{ cells cm}^{-2}$ (a) Homogeneous distribution of cells, the probability for a cell to be placed at a given position was identical for all possible positions; (b) Biased distribution of the cells on the culture surface with edges and central areas containing 3 times more cells than they would in the case of a non-biased distribution

simulation (table 8). Images of the virtual culture were generated by extracting sub-matrices of dimensions corresponding to the different fields of view (figure 7).

Table 8: Dimensions of key simulation components for cells growing in an individual well of a 6 well-plate culture vessel

Feature	Dimension
Culture surface	$30984\mu\text{m} \cdot 30984\mu\text{m}$ $9.6 \cdot 10^8 \mu\text{m}^2$ 9.6cm^2
Cells	min. volume: $625\mu\text{m}^2$ max. volume: $900\mu\text{m}^2$
Central area	$8.1 \cdot 10^7 \mu\text{m}^2$
Edge bands	$100\mu\text{m}$

3.1.1 Quantification methods

The numerical simulations were used to investigate two approaches to the estimation of cell density using machine vision (figure 8). The most intuitive one would be to count the number of distinct cells present in the images. While this was easily doable in the context of *in-silico* simulations, the detection and segmentation of single cells is a major challenge when using real microscopy images. The other quantification method

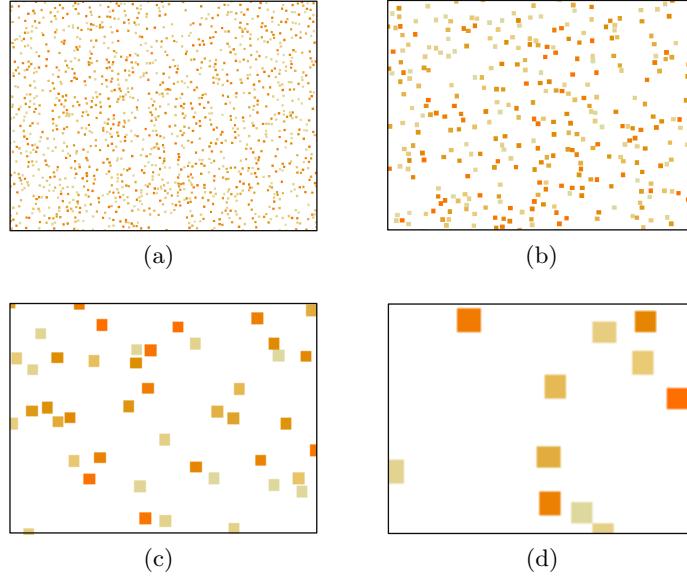


Figure 7: Images of the virtual culture at different magnifications (a) 2x; (b) 4x; (c) 10x; (d) 20x

consisted in detecting the total area occupied by the cells in a given image. A ratio was then computed to determine the fraction of the image occupied by the cells (called cells pixel ratio, see equation 5). In both cases, 100 random images of the culture were generated for all 4 magnifications. The successive results were averaged and the value obtained for the 100th image was considered to be the steady-state. The deviations from the ground truth values were computed as described in equation 7. This measures the systematic bias in the cell density results that would be introduced by the use of an image-analysis method.

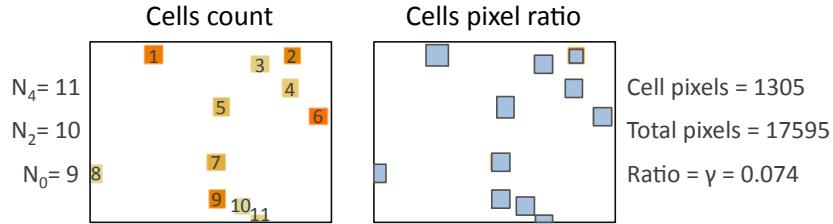


Figure 8: Two approaches to cell density quantification using microscopy images: counting the number of distinct cells present in the image (left) or quantifying the total area occupied by the cells (right). N_4 is the total number of cells, N_2 the count excluding cells on the lower and right borders and N_0 the number of cells excluding all cells in contact with an edge

The first simulations showed that the methodology employed to count cells had a no-

ticeable effect on the results obtained (figure 9). The extrapolated cell number (that is, the average number of cells counted per image multiplied by the number of images required to cover the whole culture) was overestimated when counting cells on the edges of the field of view. Conversely, ignoring all cells that were in contact with an edge of the image led to an underestimation of the overall cell density (figure 9.a). This was easily explained by the fact any given cell could be counted multiple times, even if none of the field of view overlapped. The deviation from the ground truth value was minimised by counting cells that were in contact with two of the edges only (upper and left borders). The same effect was not observed for the cells pixel ratio determination where counting all the cells resulted in a steady state value very close to the ground truth value (figure 9.b).

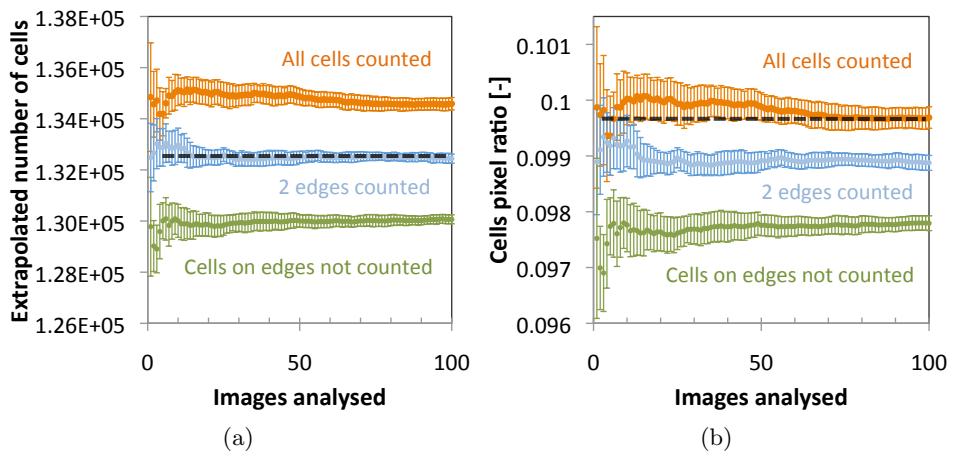


Figure 9: Effect of the counting method on the (a) extrapolated cell number and (b) cells pixel ratio both computed for 100 images at a 2x magnification. Each successive datapoint is the average of the values obtained up to that point. Cells were randomly distributed on the culture surface. The dashed line represents the ground truth value (the exact number of cells and the overall cells pixel ratio for (a) and (b) respectively). The error bars represent the standard error of the mean corrected for small sample sizes ($n=10$, 95% confidence)

In order to find out whether the effects of the counting methodology could be generalised, the deviation of the steady state from the ground truth value was determined for the different magnifications (figure 10). Counting either all or none of the cells on the edges resulted in a high deviation of the extrapolated cell number from the ground truth value. Again, the deviation was minimised for all magnifications when only cells on two of the edges were counted, with a maximum error of **3.6% \pm 1.9%** at 20x magnification.

Partially or completely excluding cells in contact with a border resulted in high deviations from the cells pixel ratio ground truth value. When all cells were considered, the deviations remained under 1.5% with a maximum of **1.4% \pm 0.9%**.

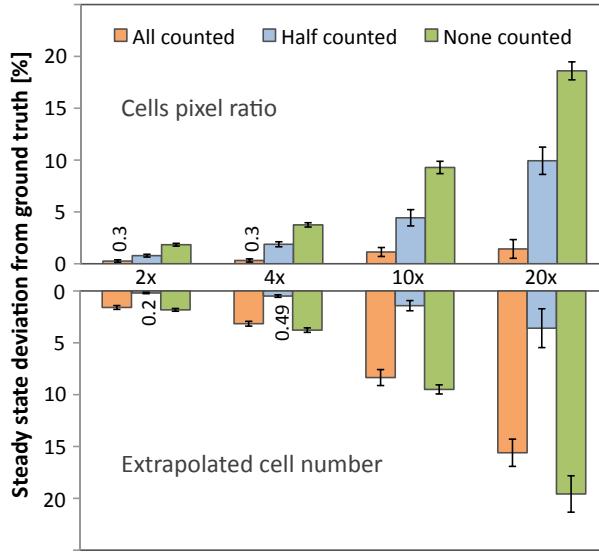


Figure 10: Effect of different counting approaches on the steady state deviation from the ground truth value. The horizontal axis represents the magnification. Cells were randomly distributed on the culture surface. The error bars represent the standard error of the mean corrected for small sample size ($n=10$, 95% confidence)

In all cases, increasing the magnification led to higher deviations from ground truth values. The field of view being inversely proportional to the magnification, the number of cells present in an image decreases as the magnification increases. To compensate, it would necessary to increase the number of images analysed. The deviations would likely be lower for higher magnifications if the amount of images was increased beyond the arbitrary 100 images steady state limit. In addition, it is important to note that in an experimental context, increasing the magnification would also lead to an increased resolution or image quality, simplifying the detection of the cells. This aspect was not taken into account in the simulations.

All extrapolated cell number results to follow were computed by excluding cells in contact with the lower and right edges of a an image while the cells pixel ratio was determined without any prior treatment of the image (all cells were considered).

3.1.2 Effects of the cell distribution homogeneity on steady state deviation

Previous results were based on a truly random distribution of cells on the culture surface. The introduction of a distribution bias to mimic experimental adherent culture conditions significantly affected the steady state values obtained at various seeding densities and magnifications for both cells pixel ratio and extrapolated cell number estimators (figure 11).

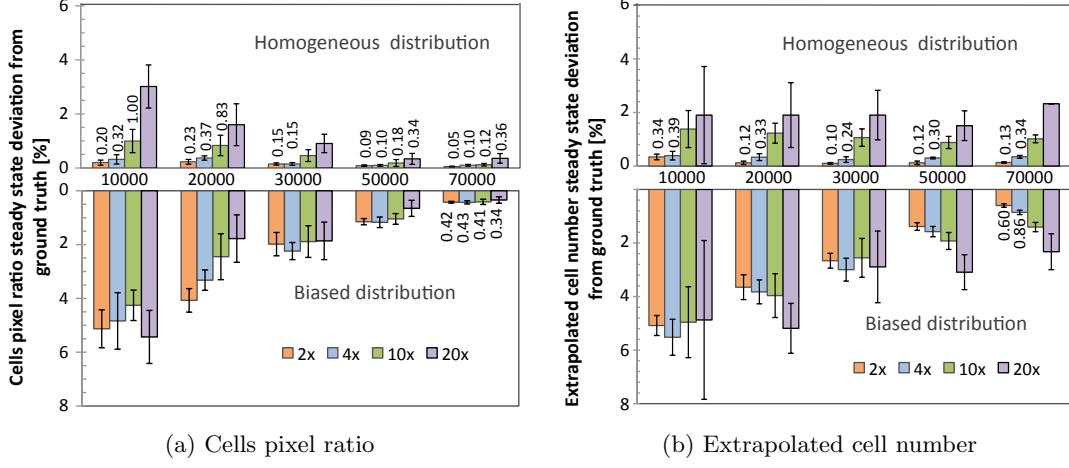


Figure 11: Steady state deviation from ground truth in the cases of homogeneous and biased distributions of cells for (a) cells pixel ratio and (b) extrapolated cell number estimators. The horizontal axis represents the seeding density in cells cm⁻². The error bars represent the standard error of the mean corrected for small sample size (n=10, 95% confidence)

Cells pixel ratio The cells pixel ratio deviations for the homogeneously distributed culture followed a trend similar to that previously observed (figure 11.a); for a given seeding density, the deviation increases with the magnification. In contrast, there was no noticeable pattern observed in the case of cultures with a biased cell distribution. In both cases, there was a decrease of the deviations as seeding density increased. This could be an artefact related to the way the deviation was computed (equation 7), which was basically the relative error on the measured cells pixel ratio. The observed trend is typical of a constant or slowly increasing absolute error. Indeed the absolute error was found to increase with the seeding density in all cases (data not shown), but not enough to produce a constant deviation. Another point to take into account concerns the seemingly sharper decrease for the biased distribution, which could be explained by the mitigation of the bias by the increased cell density; the high density regions as defined in the simulation are of finite dimensions and as more cells get added to the culture, the available area in these regions decreases, forcing additional cells to be placed in other regions. These observations should remain true for *in-vitro* cultures.

Extrapolated cell number The results for the extrapolated cell number approach were similar with a few notable differences (figure 11.b). Unlike what was observed for the cells pixel ratio, the deviations remained relatively stable as density increased for homogeneously distributed cultures. However, the differences between replicas (shown by the error bars) decreased with the number of cells seeded in the culture. When the cells distribution was biased, the extrapolated cell number deviation from ground truth

had a very similar profile to that of the pixels ratio but the decrease at higher densities wasn't as marked. As the seeding densities increased, the deviations converged towards the corresponding homogeneously distributed values, mostly due to the compensation of the bias as discussed for the cells pixel ratio results above. It is also worth noting that the variation between replicas was much higher for the extrapolated cell number values than it was for the cells pixel ratio estimator.

3.1.3 Limiting the number of images

Cells pixel ratio and extrapolated cell number approaches were shown to produce estimations that were very close to their respective ground truth values. However, using 100 images to reach the steady state value is not practicable for the characterisation of *in-vitro* cultures. Simulations were used to determine how accurate the estimations would be if only 20 images were processed (figure 12).

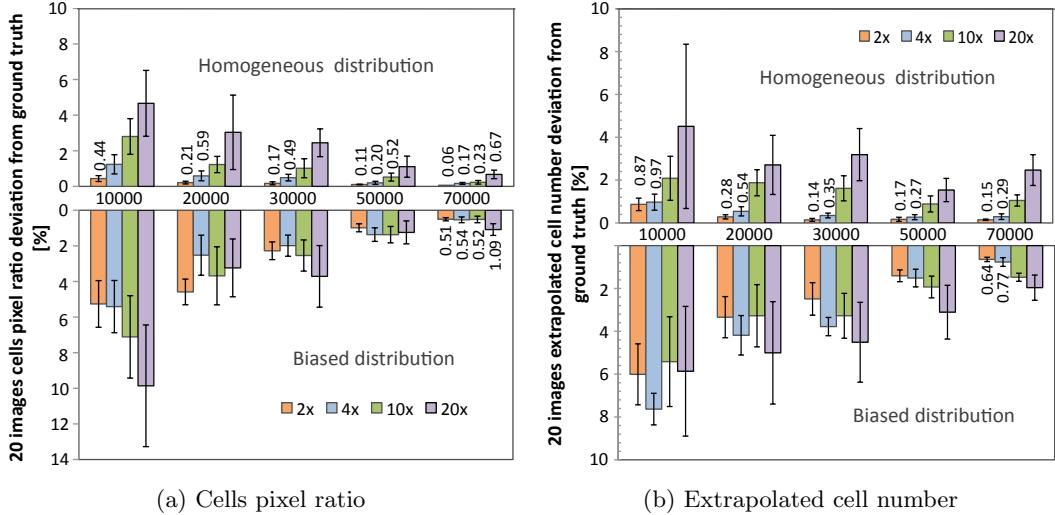


Figure 12: Deviation of the estimator computed using 20 images from the ground truth in the cases of homogeneous and biased distribution of cells for (a) cells pixel ratio and (b) extrapolated cell number estimators. The horizontal axis represents the seeding density in cells cm⁻². The error bars represent the standard error of the mean corrected for small sample size (n=10, 95% confidence)

Cells pixel ratio The deviations of the cells pixel ratio was only moderately higher at low seeding densities when using only 20 images compared to those observed at the steady state (figure 12.a). However, the variation between replicas was much larger, especially for the biased distribution. The overall trend was similar with deviations decreasing as the seeding density increased. At high densities, there was little to no difference between

the deviations observed with the pixels ratio computed with 20 images and those at the steady state.

Extrapolated cell number Similarly to that observed at the steady state, the deviation of the extrapolated cell number computed with 20 images from the ground truth remained relatively constant as the cell density increased when cells were homogeneously distributed (figure 12.b). Again, the deviations for the biased distribution decreased at higher cell densities and converged towards the values observed for the homogeneously distributed culture.

In all cases, linear regressions between the estimator computed using 20 images and the seeding density had coefficients of determination superior to 0.9980 (figure 13), the relationship between the two variables could therefore be considered nearly linear.

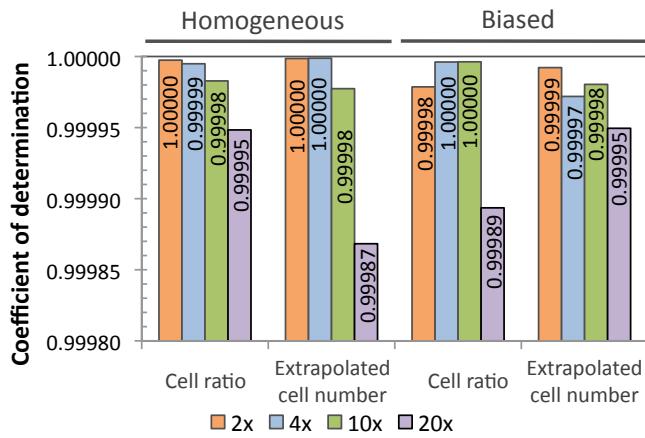


Figure 13: Coefficients of determination for the linear regression between the estimator value (either cells ratio or extrapolate cell number) computed for 20 images and the seeding density. The values used are the mean of 10 experiments

3.1.4 Insights from the simulations

The various simulations highlighted some key aspects of the method as well as issues that might arise at the experimental implementation stage. Be it at the steady state or after the analysis of 20 images, the deviation from the ground truth was inferior to 10% in all cases. In addition, the relationship between the estimators and the seeding density was nearly linear. Thus, it was shown that the estimation of the cell density of an adherent cell culture is, at least in theory, feasible using an image processing approach. Cells in contact with two pre-determined edges should be discarded from the analysis in order to avoid an overestimation of the extrapolated cell number. In the case of the determination of the cells pixel ratio, all cells appearing in the field of view of a given image should be taken into account.

In general, higher magnifications had higher deviations from the ground truth value, which could prove to be problematic in some situations where a high magnification is actually required to detect the cells. Moreover, a biased distribution of the cells, which most likely is a better representation of a real culture, further increased the deviations when compared to an homogeneously distributed culture. Finally, the extrapolated cell number approach showed a systematic error even at high seeding densities whereas the cells pixel ratio deviations decreased as cell density increased and eventually became negligible (inferior to 1%).

3.1.5 Current limitations

While the function was present in the simulation code, time constraints did not allow to try and generate cultures of different dimensions. For instance, it could be used to compare the performances of the different approaches between a conventional scale vessel such as a 6 well-plate and a microfluidic reactor. Another major limitation of these simulations was the fact that no actual image processing was required to compute the estimators value as the input were essentially perfect binary images. It would have been interesting to degrade the quality of the generated images and apply the algorithms that were employed for the experimental implementation. This way, the whole process would be simulated.

3.2 Image processing framework

The choice of a development environment had to be made early on during the project. Although numerous general purpose and microscopy-oriented image processing packages were readily available, none fulfilled all the requirements set for the project. It was therefore decided to build an image processing framework by leveraging the powerful capabilities of Mathematica.

Extensible catalogue of functions Mathematica has a very large catalogue of image processing functions built-in ranging from basic filtering to complex morphological operations. Unlike most specialised packages with a closed architecture, new functions can easily be added using symbolic programming. In addition, a large fraction of the image processing functions are based around matrix operations, an area in which Mathematica excels.

Flexibility It was important to have a flexible environment that allowed to easily define sequences of operations to be applied on images. A simple scripting language was used to define sequences of operations that could be saved and re-used.

Batch optimisation and quantification The primary function of the framework was to rapidly screen image processing algorithms and their associated set of parameters. But finding the best algorithm for a single image would not be very useful. Instead, the framework had to allow the screening of algorithms against large training sets (e.g. all images for a given cell line). The highest scoring algorithm would then be used to quantify experimental images. Again, batch processing capabilities were required as the experimental dataset was made up of thousand of images.

3.2.1 Scripting engine, queue system and scoring

Image processing can be described, at least in the context of this study, as a sequence of operations applied to an input image that will eventually result in the extraction of particular features of interest that can then be then quantified if necessary (figure 14). The main idea behind the scripting language was to create an abstraction layer that

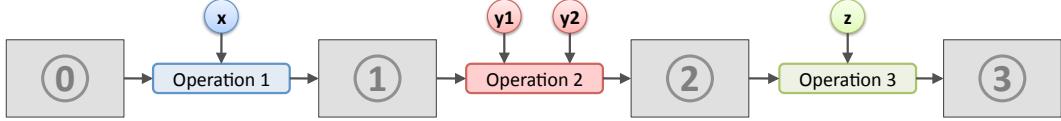


Figure 14: Image processing as used in this study: a sequence of operations applied to an input image (0) producing new intermediate images, eventually leading to an output image (3) only containing the features of interest

would allow any user to define new sequences easily without having to tinker with the code itself. The framework was designed so that any Mathematica function could be accessible through scripting via specific hooks, thus enabling the large catalogue of built-in functions to be leveraged. In addition, this approach allowed to easily extend the catalogue by implementing custom functions (figure 15).

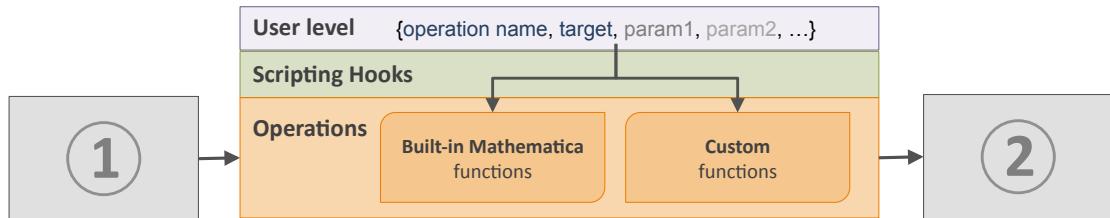


Figure 15: Abstraction levels of the framework. Users write scripts that call built-in or custom functions through scripting hooks. The operation is then applied to the input image (1) to produce the output image (2)

The syntax of the scripting language was as following:

```
{operation name, target, [param1], [param2], ...}
```

where *target* is the image to apply the operation to (0 applies the operation to the last image in the process queue) and *paramX* are the operation parameters. Sequences of operations are defined by using queues, which are lists containing multiple operations defined as shown above. A queue can be saved as a simple text file and be loaded back into the program at a later time.

The main advantage of such an abstraction layer is to allow the generation of queues on the fly. For instance, instead of using a single value for a given parameter, it is possible to supply the different operations with range of values for each parameter:

```
{operation name, target, {min,max,step}, [{min,max,step}], ...}
{operation name, target, {value1,value2,value3}, [{value1,...}], ...}
```

Effectively, the framework will generate a queue for each possible combination of parameters. This is particularly useful when combined with the scoring of an output image by comparison with a ground truth image. Various metrics (see section 2.2.4 for details) are available to determine the performance of a queue. This feature of the framework was employed to perform linear optimisations of the different sequences of operations to be applied to the microscopy images used in this study (figure 16).

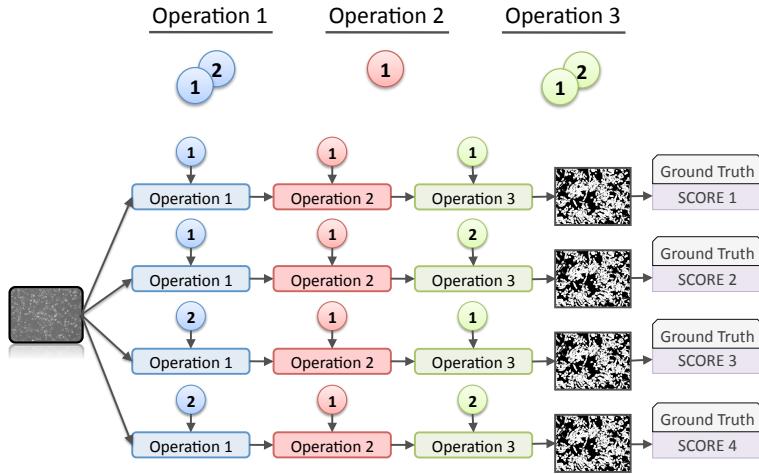


Figure 16: Simplified example illustrating the linear optimisation method made possible by the framework

It is difficult to predict which sequence of operation and which set of parameters will yield the best results for a given image processing problem; the possibility to automate the screening of large numbers of queues against whole batches of images greatly facilitated the work done for this study.

3.2.2 Framework features of interest

Here are some of the features that were implemented in the framework. While not all of these were used in the context of this study, they will most likely be useful for an eventual continuation of this work.

Basic Image Features Recently, Griffin and coworkers described a method based on derivative-of-Gaussian filters to classify all points of an image into 7 different categories, termed basic image features (BIFs), effectively producing a *primal sketch* of the image [39]. Reported applications include texture classification through the use of spatial complexes of BIFs computed at different scales and object categorisation. The method was implemented in the framework with permission (figure 17).

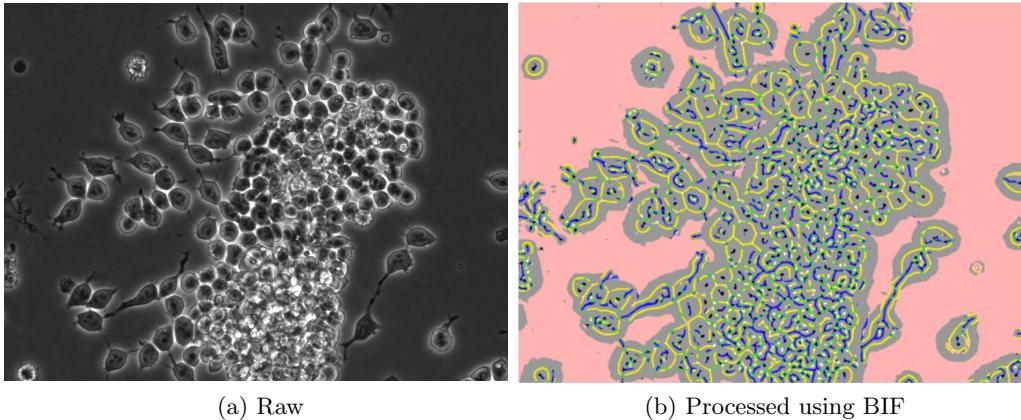


Figure 17: Basic Image Features (BIFs) computed for a mouse embryonic stem cells microscopy image (Oct4-GiP line, acquired at 20x). Each color represents a different feature

Region specific scoring Some of the microscopy images presented multiple features that were very distinct from each others, mostly cells that adopted dramatically different morphologies such as mESC undergoing differentiation (figure 18.a). It was of interest to devise algorithms that could, for example, extract a particular feature while ignoring others. One option would have been to crop the image and analyse the different features separately. However, this approach would have introduced a bias as the algorithm would most likely produce different results when analysing the whole image in an experimental context. The framework was designed to allow region specific scoring; the user could select as many regions as needed (figure 18.b) and while the image was processed as a whole, the scoring step was applied separately for each region.

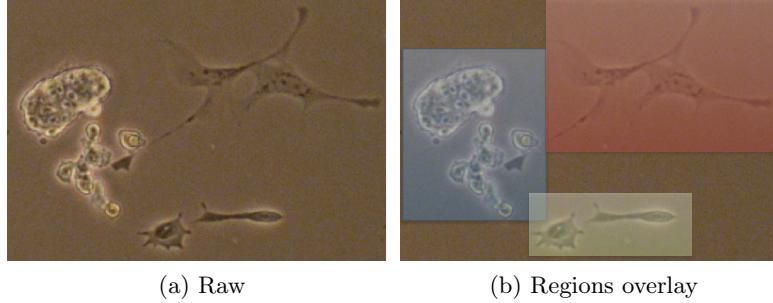


Figure 18: Example of region selection for region-specific scoring (Oct4-GiP mESC image of acquired at a 10x magnification)

Z-stacking In some circumstances, it was not possible to obtain microscopy images with sufficient contrast between the cells and the background. This was mainly the case with images acquired at magnifications that did not allow the use of phase contrast illumination. As proposed by Selinummi and colleagues, it is possible to alleviate this issue by using a technique called Z-stacking[40]. It consists in acquiring multiple brightfield images that are slightly out of focus; the value of the pixels that belong to the cells will vary greatly between the different images while that of the background pixels will not be as affected (figure 19.a). As a result, by extracting pixels with large variations in intensity between the different images, it is possible to reconstruct an image with little to no background noise. This technique was successfully implemented in the framework using standard deviation, median absolute deviation and interquartile range as projection methods (figure 19.b).

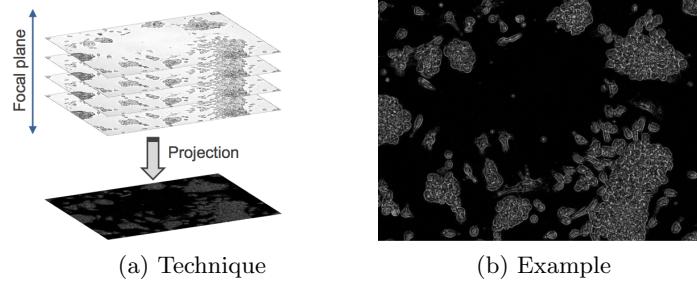


Figure 19: Z-stacking technique to improve contrast of microscopy images. (b) was generated by using 20 slightly out of focus images of Oct4-GiP mESC acquired at a 10x magnification

3.2.3 Example of applications

The framework was used to tackle various image processing problems related to microscopy images. Here some of the possible applications are briefly summarised, mostly taking advantage of basic image features extraction.

Thresholding Thresholding is a very common operation in image processing. It consists in extracting pixels whose intensity are within a defined range. All pixels that lie in this intensity range will be set to the value 1, while the rest will be set to 0, resulting in a binary image. Many thresholding functions were implemented in the framework; one could define a manual range or instead, use automatic thresholding value determination methods based on clustering, entropy, mean, median or minimum error. None of these were found to produce consistent results across different type of images (e.g. different cell lines, magnifications or illuminations). However, the extraction of a specific set of BIFs resulted in good thresholding for all the images screened. This approach was used as a basis for the experimental implementation of the cells pixel ratio estimator.

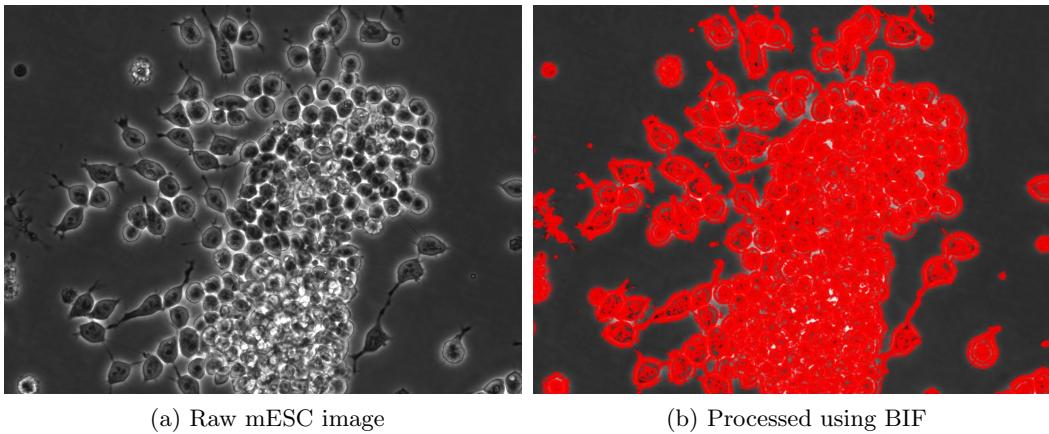


Figure 20: Example of using BIFs for image thresholding (feature 2, $\sigma=1$, $\gamma=0.05$). Image acquired at 20x magnification (Oct4-GiP cell line)

Mouse embryonic stem cells counting As discussed in the introduction, detecting cells as individual objects so they can be counted is a difficult task in image processing, especially when said cells form colonies and are in contact with each other. Different approaches involving multiple steps of filtering and morphological operations were assessed but the results were far from being satisfactory. However, one method involving the extraction of the 'blob' basic image feature yielded promising results for the detection of individual mouse embryonic stem cells (figure 21). Further optimisation work will require the generation of high quality ground truth data, which is difficult as individual

cells are difficult to discern even at high magnification. One solution would be take advantage of the fact that the mES cell line used here (Oct4-GiP) expresses a fluorescent protein. The fluorescence channel could be used as a basis for the cell segmentation of ground truth images.

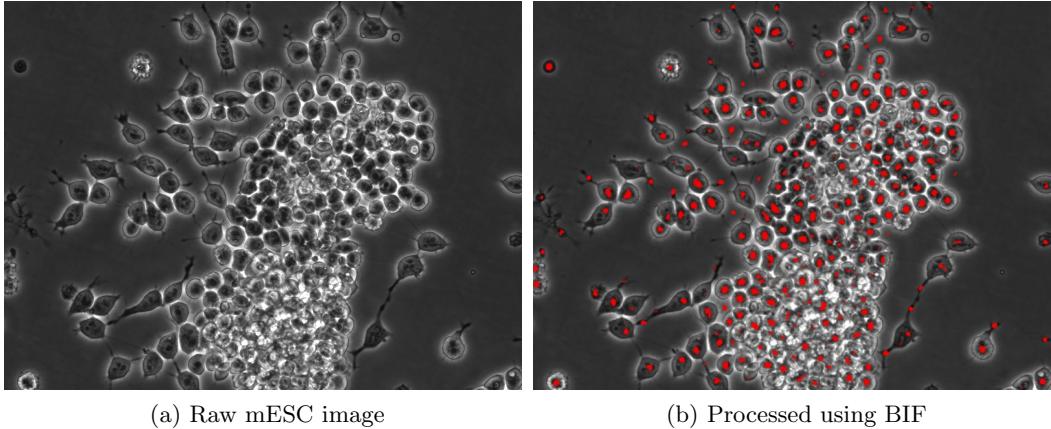


Figure 21: Example of using BIFs for mES cell counting (feature 3, $\sigma=9$, $\gamma=0.1$). Image acquired at 20x magnification (Oct4-GiP cell line)

CHO cells counting Multiple approaches to cell counting found in the literature were tested for the detection of single CHO cells in adherent cultures but none was found to be accurate enough to be used as a reliable alternative to traditional counting methods. Unfortunately, the approach for mouse embryonic stem cells counting described above was not applicable to CHO cultures due to their distinct morphologies. Nevertheless, another BIFs-based approach could be used in this case. Indeed, the feature 4 tracks elongated segments, which are characteristic of CHO microscopy images (figure 22.a). The number of segments present on the processed images (figure 22.b) was then counted using morphological components. Again, more optimisation work would be required to use this approach as a reliable estimator for the cell density but early versions of the algorithm was able to detect 92% of the cells present in an image with a false positive rate of around 7%.

Human embryonic stem cells colony detection All the applications discussed so far have been limited to either chinese hamster ovary cells or mouse embryonic stem cells. While they are interesting models for initial optimisations of the culture conditions, there will eventually be a need to optimise both expansion and differentiation processes for human embryonic stem cells (hESC). There is, however, an additional challenge to overcome when it comes to the processing of hESC microscopy images: most protocols require inactivated mouse fibroblast cells to be co-cultured with the hESC as they provide

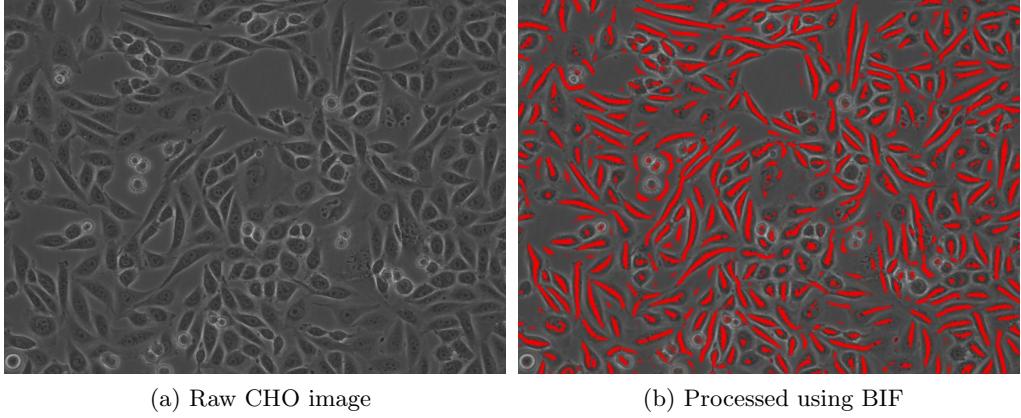


Figure 22: Example of using BIFs for CHO cell counting (feature 5, $\sigma=8$, $\gamma=0.04$). Image acquired at 20x magnification (CHO K1 cell line)

essential signalling molecules [41]. Thus, the image processing algorithm needs to be able to detect hESC colonies while ignoring the fibroblasts that are also present in the field of view (figure 23.a). Encouraging results were obtained by using a combination of BIF extraction (feature 4) and a custom filling algorithm (figure 23.b). This approach could potentially be used to track and characterise the expansion of hESC colonies in real-time.

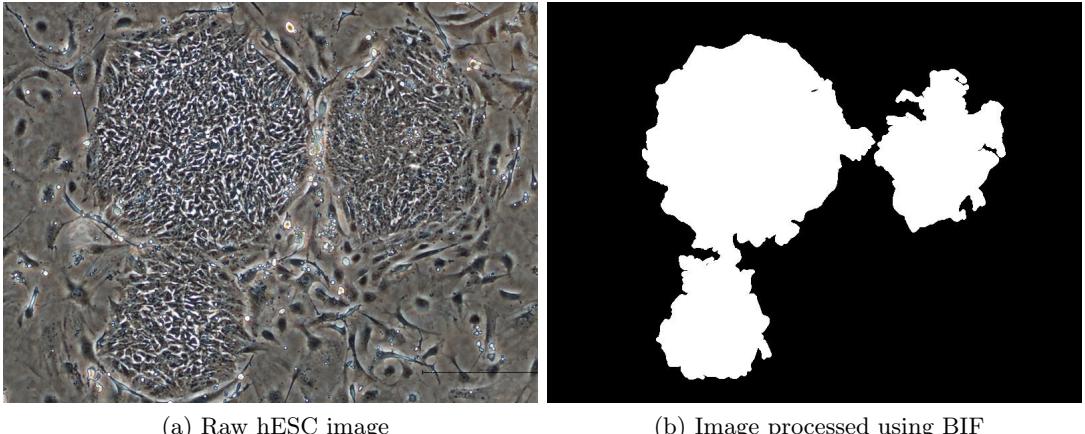


Figure 23: Detection of hESC colonies co-cultured with mouse fibroblasts (feature 1, $\sigma=8$, $\gamma=0.15$). hESC images were obtained from Marcel Reichen (UCL, Department of Biochemical Engineering, Microfluidics group)

3.3 Experimental implementation

Numerical simulations results showed that, at least in theory, cells pixel ratio (a proxy for the projected area) and extrapolated cell number were both good estimators for the cell density of an adherent culture. However, due to time constraints only the former was implemented experimentally. Experiments were carried out using a chinese hamster ovary cell line (CHO K1) and mouse embryonic stem cells (Oct4-GiP). CHO K1 is a well characterised model mammalian cell line commonly used in the industry. During growth, it adopts a typical elongated shape (figure 24.a). Oct4-GiP, the chosen mouse embryonic stem cells line, grows in clumps of round-shaped cells and tend to form colonies early on during the culture (figure 24.c).

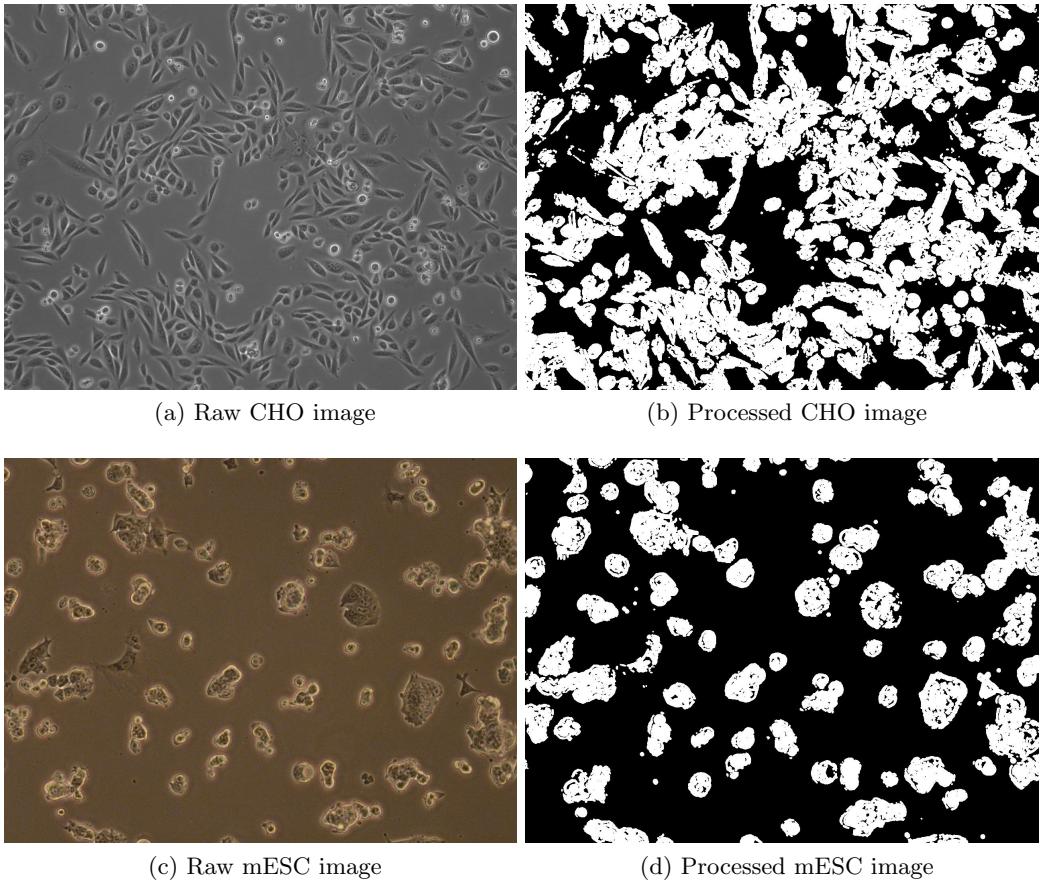


Figure 24: Microscopy images before and after processing

Results validation A sacrificial approach was used to characterise the growth of the cells over time (see 'Sacrificial method' in section 2.2.2). The ground truth dataset was

generated by using an automatic cell counter for CHO cells and FACS analysis for the mESC. 30 phase-contrast images per replica were acquired at two different magnifications (10x and 20x for CHO cells 4x and 10x for mESC) using a regular inverted microscope (see section 2.2.2 for the detailed methodology). The number of datapoints for the image processing results of CHO cells had to be reduced to 4 due to a very large accumulation of dead cells in suspension after about 60 hours of culture.

Image processing algorithm The prototyping framework was used to determine the best method to compute the cells pixel ratio for the set of images that were acquired during this experiment (around 1400 images). Although most of the conventional methods involving extensive morphological operations failed to achieve satisfactory results, a simple, one step method was found to be very effective at thresholding the foreground cells (figure 24). It was based on the extraction of 6 of the basic image features (see table 5 for the the algorithm specifics and section 3.2.2 for details on BIFs). The proposed method scored very high for the precision and moderately high for the recall (see equations 1 and 3 respectively), which means that most of the pixels detected as being part of a cell were also present in the ground truth image but some pixels that were labeled as cells in the ground truth image were not detected, mostly in the central area of the cells. Further optimisation, such as the inclusion a filling pass, could alleviate this issue. Nevertheless, this method proved to be simple to implement and relatively low in term of computational complexity.

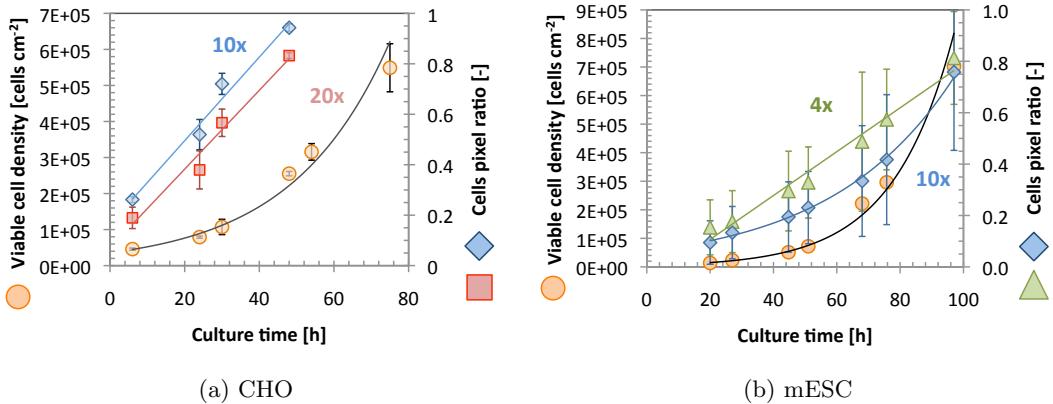


Figure 25: Growth kinetic for (a) CHO cells and (b) mESC. The black lines are the exponential regressions of the viable cell density ($y = 35658 \cdot \exp^{0.0381x}$, $R^2=0.985$ for CHO cells and $y = 6983.2 \cdot \exp^{0.0479x}$, $R^2=0.983$ for mESC). Coloured lines represent linear regressions: $y = 0.0166x + 0.1639$, $R^2= 0.97801$ and $y = 0.0156x + 0.0699$, $R^2=0.97393$ for 10x and 20x magnification respectively for CHO cells, $y = 0.0085x - 0.0636$, $R^2= 0.97622$ and $y = 0.0609 \cdot \exp^{0.0257x}$, $R^2=0.99423$ for 4x and 10x magnification respectively for mESC. Error bars represent the standard error of the mean corrected for small sample size ($n=3$, 95% confidence)

3.3.1 Cells pixel ratio as a cell density estimator

The viable cell density increased exponentially overtime for both cel lines (figure 25). The maximum growth rates were 0.036h^{-1} and 0.06h^{-1} for CHO and mES cells respectively. For the CHO cultures, cells pixel ratio at both magnifications increased linearly as the cultures progressed to reach near confluence after 48 hours. Concerning the mESC cultures, the cells pixel ratio increase was linear at 4x magnification but was nearly exponential for the 10x images. After 48 hours the cells pixel ratio was only between 0.2 and 0.4 and reached 0.8 towards the end of the cultures (96 hours). There was also a notable difference in term of variability between the different replicas (shown by the error bars); mESC cells pixel values had large variations between replicas whereas CHO results had little to none. This was most likely due to the large variations in morphology of mESC while CHO cells have a fairly consistent morphology.

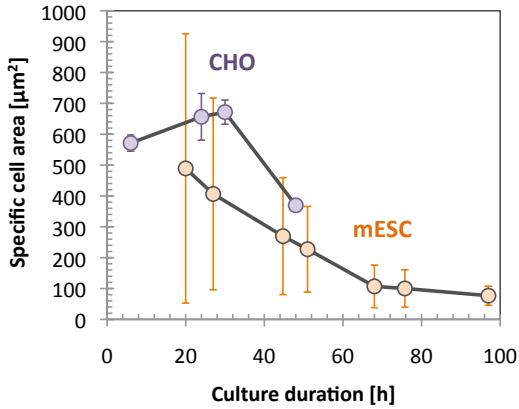


Figure 26: Specific cell area computed by dividing the total area occupied by the cells (inferred from the cells pixel ratio computed with 10x images) divided by the total cell density

The differences between the two cell lines in term of cells pixel ratio increase during the culture showed that CHO cells tend to spread very early to cover as much surface as possible and then shrink to make space for freshly divided cells while mESC appeared to continuously decrease in size over the course of the culture, probably due to the formation of colonies very early on, which imposes strict packaging constraints (figure 26).

According to the numerical simulations, cells pixel ratio should increase linearly with the number of cells present in the culture (figure 13), which was clearly not the case for these experiments. As shown by the specific area results (figure 26) and unlike their representation in the simulations, cells aren't static objects but undergo changes in both shape and size during the course of a culture. This highly dynamic system led to a non-linear relationship between the cell density and the cells pixel ratio (figure 27).

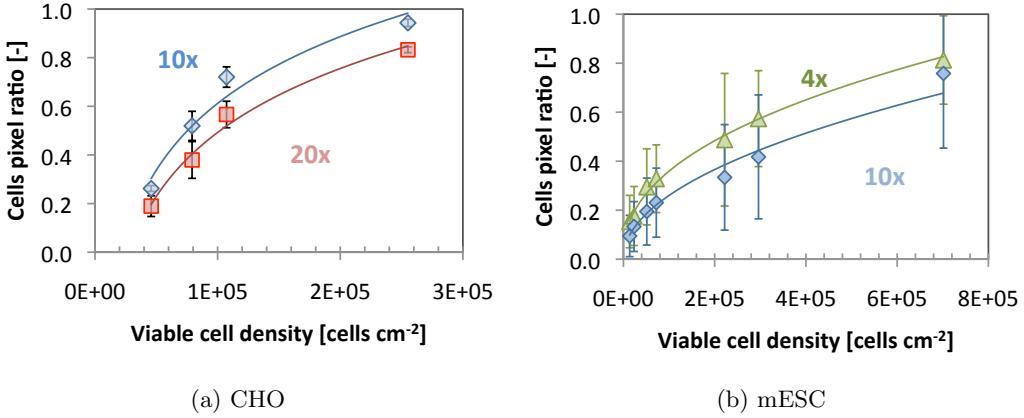


Figure 27: Pixel ratio in function of cell density for (a) CHO and (b) mES cells. Error bars represent the standard error of the mean corrected for small sample size ($n=3$, 95% confidence). (a) Lines are logarithmic regressions: $y = 0.3963 \ln x - 3.9504$, $R^2 = 0.9613$ and $y = 0.3795 \ln x - 3.8762$, $R^2 = 0.9862$ for 10x and 20x magnification respectively; (b) Lines are exponential regressions: $y = 0.0026 \cdot \exp^{0.427x}$, $R^2 = 0.9914$ and $y = 0.0009 \cdot \exp^{0.491x}$, $R^2 = 0.98563$ for 4x and 10x magnifications respectively

3.3.2 Data reconciliation

The end goal of this project being to develop a method that allows the inference of the cell density using image-analysis, the results presented above were clearly not satisfactory. While some differences are expected from a cell line to another, the relationship between cells pixel ratio and cell density wasn't consistent for a given cell line at different magnifications. Nevertheless, it might be possible to approximate the ideal linear relationship between the two variables by applying data reconciliation, a transformation of the data based on apriori information about the studied process, in this case the growth dynamics of adherent cells.

The decrease of the specific cell area of the mESC in function of time (figure 26) can be approximately fitted to an exponential decay function. This information can be used for data reconciliation; by applying an exponential decay factor proportional to the culture time (with the proportionality constant equal to that of the exponential decay function on figure 26) to transform the cells pixel ratio data (see equation 28), the relationship with the cell density became nearly linear (figure 28.a).

The corrected cells pixel ratio in function of time was very similar to the growth curve based on the viable cell density (figure 28.b). Thus, by employing data reconciliation, it was possible to use cells pixel ratio as a good estimator for the overall cell density. This approach is not without limitations; the model used as prior information is simplistic and probably has a limited scope of application. In addition, the proportionality constant

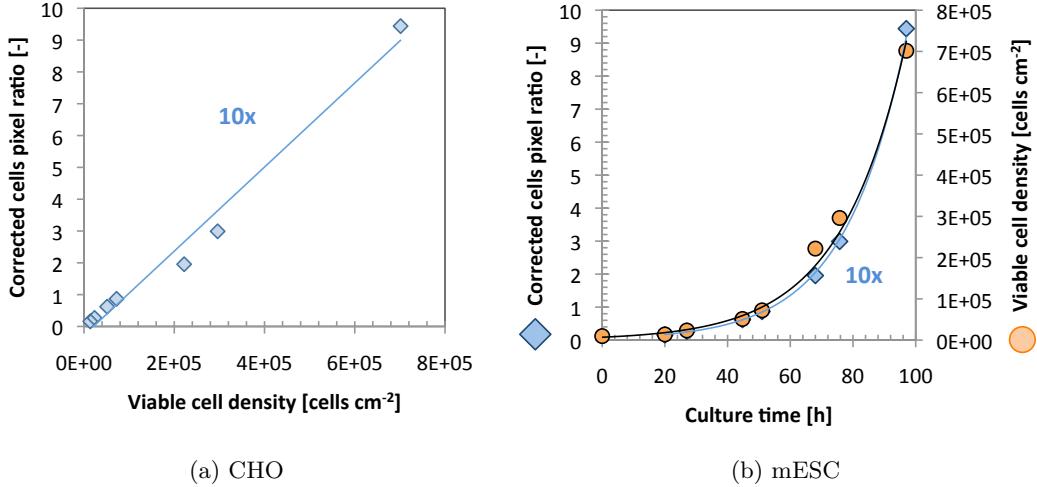


Figure 28: Corrected cell pixel ratio in function of (a) viable cell density and (b) time. (a) The line represents a linear regression ($y = 10^{-5}x - 0.2753$, $R^2=0.97988$); (b) The lines are exponential regression for the cell density ($y = 6983.2 \cdot \exp^{0.0479x}$, $R^2=0.983$) and corrected cells pixel ratio ($y = 0.0609 \cdot \exp^{0.0517x}$, $R^2=0.99857$)

is most likely highly dependant on the studied process. In the best case scenario, it is only affected by a small number of factors such as the cell type or the culture vessel. However, it is likely that every time a culture conditions that effect growth is altered (e.g. temperature or culture medium formulation), a 're-calibration' using other means of cell density determination would be required, defeating the purpose of the approach as a versatile tool for bioprocess optimisation. Additional experiments would be required to determine what conditions effect the proportionality constant. Additionally, the implementation of a robust extrapolated cell number estimator would probably alleviate these issues as it is expected to be directly proportional to the cell density.

4 Conclusion

The main objective of this project was to develop a machine-vision approach to the estimation of adherent cell density. It was shown through numerical simulations that both cells pixel ratio (representing the average fraction of an image occupied by cells) and the extrapolated cell number (the average number of cells counted per image multiplied by the number of images necessary to cover the whole culture area) could be used as estimators for the cell density of the culture. Moreover, even though 96 to 9570 images would be required to cover the whole culture area, 20 random images were shown to be sufficient to accurately infer the cell density in a well from a 6 well-plate culture vessel. Algorithms devised using the image processing framework were then employed to experimentally validate the cells pixel ratio method for the characterisation of CHO

cells and mESC growth kinetics. The relationship between the cell density and the cells pixel ratio was shown to be non-linear, contradicting the results of the simulations. The discrepancies were attributed to the dynamic changes in cell shape and size during growth. The cells pixel ratio data was transformed using a simple exponential decay model to take into account the changes in size of mESC through the culture. The relationship between the corrected cells pixel ratio and the cell density was linear and thus, using this method, cell density could be derived from the analysis of microscopy images.

The goals of this project were mostly met. The feasibility of the approach was investigated through numerical simulations and experimental implementation, paving the way for the development of a set of image processing methods for the non-invasive characterisation of adherent cell cultures.

5 Future work

The work done on the three aspects of this project (numerical simulations, development of an image processing framework and experimental implementations) could serve as a strong basis for further developments. Briefly, a few aspects that should be considered for future work are presented below.

Numerical simulations The simulation program could be improved upon by incorporating the key findings that emerged from the experimental work; instead of treating cells as static objects, both their size and shape should be able to change during the culture. Additionally, it would also be interesting to simulate cell division instead of repopulating the whole simulation with a higher number of cells as it was done for this project. Another aspect that could be improved upon to minimise the discrepancies with experimental results is the quality of the generated images; currently, the simulated images are not representative of their 'real' counterparts in that their quality is perfect, to the point that no actual image processing was required to determine the value of the estimators. The generation of low resolution images with added noise would not only make for a much more credible simulation but also would be a potentially unlimited source of training images for the image processing algorithm optimisation.

Image processing framework The framework currently employs a linear algorithm optimisation approach. The computation time will be directly proportional to the number of combinations to screen. If the set of parameters gets too large, the computation time becomes impractical. It would be interesting to use genetic or machine learning-based approaches to alleviate this issue and streamline the optimisation process. Concerning the cell counting algorithms that could be used for the experimental implementation of the extrapolated cell number estimator, more optimisation work needs to be carried

out to produce consistent and reliable results. Alternatively, a texture recognition approach could be implemented; instead of trying to identify individual objects, the whole image would be characterised (e.g. by using BIFs) and compared to a database that was previously populated with textures associated with a cell density and other culture conditions. This approach could prove to be more reliable if enough quality 'ground truth' textures can be generated.

Experimental implementation The experimental applications described in this study were solely based on cells pixel ratio. In parallel to their optimisation using the image processing framework, individual cell counting algorithms should be implemented experimentally. They might be a better estimator than cells pixel ratio as they wouldn't require any apriori information about the process. The cells pixel ratio approach could also be improved by developing a model for the data reconciliation that does not depend on the culture conditions. The future work suggested for the numerical simulations should help devise such a model.

6 Acknowledgements

I would like to thank my supervisors **Dr. Nicolas Szita** (Department of Biochemical Engineering) and **Dr. Lewis Griffin** (Department of Computer Sciences) for their encouragement, guidance and support through the project.

Marcel Reichen for the human embryonic stem cells images and advices.

Alexandre Super for the fruitful discussions.

Chris Landridge and **Ellie Nye** for their outstanding job as CoMPLEX administrators.

Prof. Andrew Pomiankowski, **Prof. Alan Johnston**, **Prof. Rob Seymour**, **Dr. Guy Moss** and all the staff at CoMPLEX for allowing me to be part of this department.

Bibliography

- [1] Lawrence Burns. "You are our only hope": trading metaphorical "magic bullets" for stem cell "superheroes". *Theor Med Bioeth*, 30(6):427–42, Jan 2009.
- [2] M Hooper, K Hardy, A Handyside, S Hunter, and M Monk. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, 326(6110):292–5, Jan 1987.
- [3] J Thomson, J Itskovitz-Eldor, S Shapiro, M Waknitz, J Swiergiel, V Marshall, and J Jones. Embryonic stem cell lines derived from human blastocysts. *Science*, 282 (5391):1145–7, Nov 1998.
- [4] J Schoorlemmer, L Jonk, S Sanbing, A van Puijenbroek, A Feijen, and W Kruijer. Regulation of Oct-4 gene expression during differentiation of EC cells. *Mol Biol Rep*, 21(3):129–40, 1995.
- [5] J Odorico, D Kaufman, and J Thomson. Multilineage differentiation from human embryonic stem cell lines. *Stem Cells*, 19(3):193–204, 2001.
- [6] Anna M Wobus and Kenneth R Boheler. Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol Rev*, 85(2):635–78, Apr 2005.
- [7] Peter Sartipy, Petter Björquist, Raimund Strehl, and Johan Hyllner. The application of human embryonic stem cell technologies to drug discovery. *Drug Discov Today*, 12(17-18):688–99, Sep 2007.
- [8] Julio C Davila, Gabriela G Cezar, Mark Thiede, Stephen Strom, Toshio Miki, and James Trosko. Use and application of stem cells in toxicology. *Toxicol Sci*, 79(2): 214–23, Jun 2004.
- [9] Steve Goldman. Stem and progenitor cell-based therapy of the human central nervous system. *Nat Biotechnol*, 23(7):862–71, Jul 2005.
- [10] Su-Chun Zhang, Xue-Jun Li, M Austin Johnson, and Matthew T Pankratz. Human embryonic stem cells for brain repair? *Philos Trans R Soc Lond, B, Biol Sci*, 363 (1489):87–99, Jan 2008.
- [11] M Mimeault, R Hauke, and S K Batra. Stem cells: a revolution in therapeutics-recent advances in stem cell biology and their therapeutic applications in regenerative medicine and cancer therapies. *Clin Pharmacol Ther*, 82(3):252–64, Sep 2007.
- [12] Steven G Deeks and Joseph M McCune. Can HIV be cured with stem cell therapy? *Nat Biotechnol*, 28(8):807–10, Aug 2010.
- [13] Joe Alper. Geron gets green light for human trial of ES cell-derived product. *Nat Biotechnol*, 27(3):213–4, Mar 2009.

- [14] K Takahashi and S Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, Aug 2006.
- [15] Gregory H Underhill and Sangeeta N Bhatia. High-throughput analysis of signals regulating stem cell fate and function. *Curr Opin Chem Biol*, 11(4):357–66, Aug 2007.
- [16] F Veraitch, R Scott, J Wong, G Lye, and C Mason. The impact of manual processing on the expansion and directed differentiation of embryonic stem cells. *Biotechnol Bioeng*, 99(5):1216–29, Apr 2008.
- [17] Debora Steiner, Hanita Khaner, Malkiel Cohen, Sharona Even-Ram, Yaniv Gil, Pavel Itsykson, Tikva Turetsky, Maria Idelson, Einat Aizenman, Rita Ram, Yael Berman-Zaken, and Benjamin Reubinoff. Derivation, propagation and controlled differentiation of human embryonic stem cells in suspension. *Nat Biotechnol*, 28(4):361–4, Apr 2010.
- [18] Ruth Olmer, Alexandra Haase, Sylvia Merkert, Wei Cui, Jirí Palecek, Chen Ran, Andreas Kirschning, Thomas Schepers, Silke Glage, Konstantin Miller, Eliza C Curnow, Eric S Hayes, and Ulrich Martin. Long term expansion of undifferentiated human iPS and ES cells in suspension culture using a defined medium. *Stem Cell Res*, 5(1):51–64, Jul 2010.
- [19] Yue Xu, Xiuwen Zhu, Heung Sik Hahm, Wanguo Wei, Ergeng Hao, Alberto Hayek, and Sheng Ding. Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc Natl Acad Sci USA*, 107(18):8129–34, May 2010.
- [20] Raymond C B Wong, Martin F Pera, and Alice Pébay. Role of gap junctions in embryonic and somatic stem cells. *Stem Cell Rev*, 4(4):283–92, Dec 2008.
- [21] James A King and William M Miller. Bioreactor development for stem cell expansion and controlled differentiation. *Curr Opin Chem Biol*, 11(4):394–8, Aug 2007.
- [22] Marti Lecina, Sherwin Ting, Andre Choo, Shaul Reuveny, and Steve Oh. Scalable Platform for hESC Differentiation to Cardiomyocytes in Suspended Microcarrier Cultures. *Tissue engineering Part C, Methods*, Jun 2010.
- [23] Robert Zweigerdt. Large scale production of stem cells and their derivatives. *Adv Biochem Eng Biotechnol*, 114:201–35, Jan 2009.
- [24] Larry A Couture. Scalable pluripotent stem cell culture. *Nature Publishing Group*, 28(6):562–563, Jan 2010.
- [25] Danny van Noort, Siew Min Ong, Chi Zhang, Shufang Zhang, Talha Arooz, and Hanry Yu. Stem cells in microfluidics. *Biotechnol Prog*, 25(1):52–60, Jan 2009.

- [26] Marie Csete. Q&A: what can microfluidics do for stem-cell research? *J Biol*, 9(1):1, Jan 2010.
- [27] Xiaobo Zhou, K Y Liu, P Bradley, N Perrimon, and Stephen T C Wong. Towards automated cellular image segmentation for RNAi genome-wide screening. *Med Image Comput Comput Assist Interv*, 8(Pt 1):885–92, Jan 2005.
- [28] Weimiao Yu, Hwee Kuan Lee, Srivats Hariharan, Wenyu Bu, and Sohail Ahmed. Evolving generalized Voronoi diagrams for accurate cellular image segmentation. *Cytometry A*, 77(4):379–86, Apr 2010.
- [29] Karen H Chang and Peter W Zandstra. Quantitative screening of embryonic stem cell differentiation: endoderm formation as a model. *Biotechnol. Bioeng.*, 88(3):287–98, Nov 2004.
- [30] T Thurnherr, A Choo, I Reading, and S Oh. Population estimation of human embryonic stem cell cultures. *Biotechnol Prog*, Nov 2009.
- [31] H S Liu, M S Jan, C K Chou, P H Chen, and N J Ke. Is green fluorescent protein toxic to the living cells? *Biochem Biophys Res Commun*, 260(3):712–7, Jul 1999.
- [32] A Korzynska, W Strojny, A Hoppe, D Wertheim, and P Hoser. Segmentation of microscope images of living cells. *Pattern Analysis & Applications*, 10(4):301–319, 2007.
- [33] L G Alexopoulos, G R Erickson, and F Guilak. A method for quantifying cell size from differential interference contrast images: validation and application to osmotically stressed chondrocytes. *J Microsc*, 205(Pt 2):125–35, Feb 2002.
- [34] M Tscherepanow, F Zöllner, M Hillebrand, and F Kummert. Automatic segmentation of unstained living cells in bright-field microscope images. *Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry*, pages 158–172, 2008.
- [35] S Narkilahti, K Rajala, H Pihlajamaki, R Suuronen, O Hovatta, and H Skottman. Monitoring and analysis of dynamic growth of human embryonic stem cells: comparison of automated instrumentation and conventional culturing methods. *Biomed Eng Online*, 6:11, 2007.
- [36] A Mölder, M Sebesta, M Gustafsson, L Gisselson, A Gjörloff Wingren, and K Alm. Non-invasive, label-free cell counting and quantitative analysis of adherent cells using digital holography. *J Microsc*, 232(2):240–7, Nov 2008.
- [37] Matthieu Stettler, Nicolas Jaccard, David Hacker, Maria De Jesus, Florian M Wurm, and Martin Jordan. New disposable tubes for rapid and precise biomass assessment for suspension cultures of mammalian cells. *Biotechnol. Bioeng.*, 95(6):1228–33, Dec 2006.

- [38] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [39] L Griffin, M Lillholm, M Crosier, and J van Sande. Basic Image Features (BIFs) Arising from Approximate Symmetry Type. *Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision*, Jan 2009.
- [40] Jyrki Selinummi, Pekka Ruusuvuori, Irina Podolsky, Adrian Ozinsky, Elizabeth Gold, Olli Yli-Harja, Alan Aderem, and Ilya Shmulevich. Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images. *PLoS ONE*, 4(10):e7497, Jan 2009.
- [41] Irina Klimanskaya, Young Chung, Sandy Becker, Shi-Jiang Lu, and Robert Lanza. Derivation of human embryonic stem cells from single blastomeres. *Nat Protoc*, 2(8):1963–72, Jan 2007.