UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Computación

Algoritmos de procesamiento de imágenes para la reconstrucción del desarrollo embrionario del pez cebra

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área de Ciencias de la Computación

Camilo Melani

Director de tesis:	Doctor Alessandro Sarti.
	Doctora Marta Mejail.
Consejero de estudios:	Doctora Marta Mejail.
Lugar de trabajo:	Universidad de Buenos Aires, Argentina
	Univerdidad de Bologna, Italia
	Instituto de sistemas complejos, Paris, Francia

Buenos Aires, 2013

Resumen

Algoritmos de procesamiento de imágenes para la reconstrucción del desarrollo embrionario del pez cebra

Algunos procesos biológicos fundamentales, como el desarrollo embrionario, se han conservado a lo largo de la evolución y son comunes a especies pertenecientes a diferentes posiciones filogenéticas. Aún hoy, esos procesos no son totalmente entendidos. La comprensión de la morfodinámica celular, que produce tejidos u órganos, puede ser alcanzada por la reconstrucción y análisis de la forma y posición celular durante el desarrollo embrionario de un animal.

Esta tesis estudia una serie de algoritmos de procesamiento de imágenes que permiten segmentar y seguir los núcleos y membranas celulares durante el desarrollo embrionario del pez cebra. Este pez ha sido ampliamente estudiado como organismo modelo para comprender el desarrollo embrionario, expresión genética y mecanismos de regeneración y reparación de tejidos en vertebrados.

Para lograr este objetivo, el embrión es previamente marcado con proteínas fluorescentes que se adhieren a los núcleos y membranas celulares. Usando un microscopio láser, se obtiene una serie temporal de imágenes volumétricas del embrión a escala subcelular. La posición de cada célula se obtiene procesando las imágenes por medio de la transformada de Hough. Se segmenta la forma de las membranas y núcleos por medio ecuaciones en derivadas parciales. El movimiento celular se estima usando técnicas de flujo óptico. El seguimiento celular se obtiene combinando información previamente adquirida con restricciones biológicas. Los resultados son manualmente validados y reconstruyen durante 6 horas la formación del cerebro del pez cebra en la fase somítica (7-8) con todas las células seguidas desde la finalización de la fase esfera con un error menor a 2%. Esta reconstrucción abre el camino a una investigación sistemática y cuantitativa del comportamiento celular y origen clonal de los órganos y cerebro.

Palabras claves: Morfogénesis. Organogénesis. Desarrollo del pez cebra. Identificación celular. Segmentación celular. Seguimiento celular. Estimación del movimiento. Procesamiento de imágenes. Microscopio confocal. Transformada de Hough 3D. Superficie subyetiva. Flujo óptico.

Abstract

Image processing algorithms for zebrafish development reconstruction.
Some fundamental biological processes such as embryonic development have been preserved during evolution and are common to species belonging to different phylogenetic positions, but are nowadays largely unknown. The understanding of cell morphodynamics leading to the formation of organized spatial distribution of cells such as tissues and organs can be achieved through the reconstruction of cells shape and position during the development of a live animal embryo.
We design in this work a workflow of image processing methods to automatically segment and track cells nuclei and membranes during the development of a zebrafish embryo, which has been largely validates as model organism to understand vertebrate development, gene function and healing-repair mechanisms in vertebrates. The embryo is previously labeled through the ubiquitous expression of fluorescent proteins addressed to cells nuclei and membranes, and temporal sequences of volumetric images are acquired with laser scanning microscopy. Cells position is detected by processing nuclei images through the Hough transform. Membranes and nuclei shapes are reconstructed by using a PDEs based variational techniques. Cells movement is estimated using optical flows methods. Cells tracking is performed by combining informations previously detected on cells shape and position with biological regularization constraints. Our results are manually validated and reconstruct the formation of zebrafish brain at 7-8 somite stage with all the cells tracked starting from late sphere stage with less than 2% error for at least 6 hours. Our reconstruction opens the way to a quantitative and systematic investigation of cellular behaviors, of clonal origin and clonal complexity of brain organs.
Key words: Morphogenesis, Organogenesis, Zebrafish development, Cell identification, Cell segmentation, Cell tracking, Motion estimation, Image processing, Confocal

microscopy, 3D Hough transform, Surjective surface, Optical flow.

a Ingrid

Agradecimientos

Contents

Introduction

The work done in this thesis is devoted to the cell morphodynamical "reconstruction" of the cell lineage tree underlying the processes of animal embryogenesis. This is a fundamental issue for bio-medical research. Such a goal can be achieved through the automated tracking of cell nuclei and identification of cell divisions from 3D+time in vivo imaging [1]. The full reconstruction of the cell lineage tree from the egg cell to the adult stage has only been achieved for the worm *Caenorhabditis elegans*. However, in that case, the total cell number in the adult is less than one thousand and the cell lineage is largely invariant.

We designed a set of strategies, methods and algorithms to "sequence" the cell lineage tree as a branching process annotated in space and time.

In this work we focus on the zebrafish, a model organisms largely studied by the community with the objective to have a animal development reference. The phylogenetic position, ease of breeding, observation and the possibility of using certain strategies of genetic engineering are taken into account in the choice of model.

The zebrafish has a very large cell number, this challenge has not been taken up so far for vertebrate organisms. Nevertheless, recent advances in imaging strategies open the way to in toto 3D plus time imaging providing data suitable for in vivo cell tracking and cell morphodynamics reconstruction. The zebrafish (*Danio rerio*) is a vertebrate model that has been chosen for its transparency allowing in vivo inspection at the cellular level deep into the tissues by confocal laser scanning microscopy [2]. The zebrafish exhibits typical vertebrate differentiated cell types and has been largely validated for investigations related to humans including cancerogenesis and a number of genetic diseases [3]. Achieving the automated reconstruction of the zebrafish embryo cell morphodynamics is highly relevant for investigating stem cells populations, early steps of cancerogenesis or drug effects in vivo. Such a goal requires engineering live

zebrafish embryos to highlight sub-cellular structures to be imaged by time lapse laser scanning microscopy, designing image processing algorithms and computational methods. We define in this thesis a workflow of image processing algorithms to: filtering the noise in the 3d images; identifying the 3D location of cell nuclei; segmenting the nucleus and membrane of the zebrafish cells; estimate cell movements; identify cell divisions and track the cells. This precious data could be processed to measure the the cell proliferation and cell death rate in time and space; cell density; understand the monoclonal groups of cells and cell differentiation.

The chapters on this thesis focus on the image processing algorithms developed in this research (see figure 1). These algorithms build a workflow (or steps) of image processing algorithms where images are acquired, transformed from raw data coming from the microscope to filtered and enhanced data where is possible to detect and segment each cell nucleus and membrane. Then, consecutive images are transformed to vector field images representing the motion and deformation of sub-cellular structures that allow us to to track the cells. The tracing of cells could be backtracked towards its earlier 3D fate map.

This work is organized as follows:

Chapter 1: We introduce the zebrafish animal, their characteristics, properties and the different embryo stages. This chapter is important to understand the problem of dealing with live embryos and understand the process of morphogenesis and organogenesis.

Chapter 2: We shown the microscopy system used to acquire the 3d + time multichannel images, we explain how the H2B/mCherry fusion protein and farnesylated eGFP allow us to dye the membranes and nuclei with fluorescents colors. Then, we compare the confocal and the multi-photon laser scanning microscope, later then we present the image parameters used and how they determine the image quality.

Chapter 3: We compare multiple denoising algorithms that filter noise intrinsically linked to the acquisition method. These filters favours intra-region smoothing while

inhibiting inter-region smoothing, this imply that edges, like membranes, are preserved and small aberration or non homogeneous distribution of fluorescent labeling are filtered out.

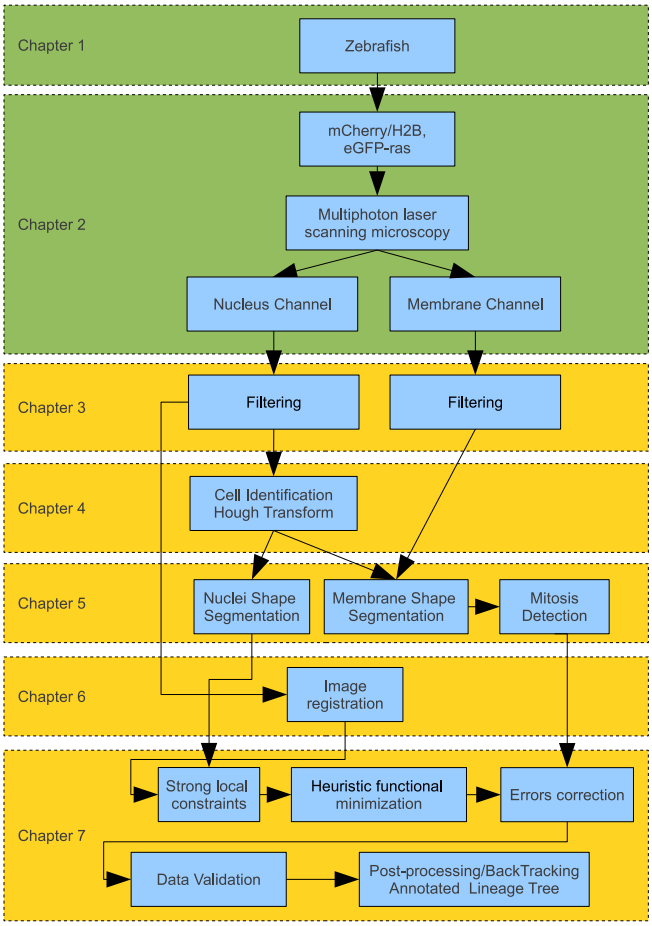


Figure 1: This figure depict a workflow of image processing algorithms that is developed in this thesis. Each subject is explain in a different chapter. Arrows indicate the dataflow, inputs and outputs of each step. In green the introductory chapters that are important to understand the problem. In yellow the actual research developed in this work.

Chapter 4 An algorithm based on the Hough transform is used to detect the sperical shape of cell nuclei. This algorithm is used to identify and count the number of cells in the images. This is a critical step because the segmentation and tracking rely on the position of each cell.

Chapter 5: An algorithm based on Subjective Surfaces perform an fully automatic membranes and nuclei segmentation. Later is studied the stability condition of the algorithm and a preliminary algorithm validation comparing the segmentation results with a gold standard. Also is explain how is possible to detect cells divisions, given that during mitosis two nuclei are inside a single cell membrane.

Chapter 6: Migration of cells is a crucial issue to describe vertebrate development and explore the relations between cell mechanical activities and the formation of macroscopic structures. In this chapter, we studied different methods that calculate the optical flow of consecutive images. Motion estimation using registration is the first step towards the understanding of cells and tissues biomechanics. An evaluation protocol has been set up in order to measure the error produced by a given algorithm. Evaluation relies on a reference, a gold standard of cells trajectories. The difference between this reference and the cell trajectories built by registration is the error measurement. Four non-rigid registration methods used in the biomedical field have been tested using as dataset artificial and real images.

Chapter 7: A cell lineage tree can be interpreted as a binary tree where the first cell of the embryo is represented by the root node and the relationship of mother cells and the two daughter cells are represented by edges. In this chapter we develop an algorithm that link the same cell at consecutive time steps. The cell divisions are identified and new born cells are related with their mother cell. The tracking algorithm follows three phases: the first phase use strong local constraints for restricting the unknown tracking graph; the second phase minimize a heuristic elastic functional, combining information extracted from images with biological regularization constraints to measure the correctness of cell matching at consecutive time steps; the last phase minimize false-positive and false-negative errors looking at the whole biological coherence of the

lineage tree identifying discontinuities in cell trajectories and cells that lived for only a few times. The tracking results are manually validated and errors are classified. The exploitation of the virtual embryo opens the way to in silico experimental embryology to follow the formation and transformation of 4D morphogenetic fields, backtracking a 3D morphogenetic section towards its earlier 3D fate map.

Chapter 1

Zebrafish embryogenesis

Some fundamental biological processes such as metabolism, cell cycle, embryonic development and genes expression have been preserved during evolution and are common to species belonging to different phylogenetic positions. Animal models are widely studied in many scientific fields, such as Developmental Biology, to understand biological mechanisms common also to humans but still widely unknown. The zebrafish (*Danio rerio*) is a small tropical freshwater fish largely validates as powerful experimental system to understand vertebrate development, gene function and healing/repair mechanisms in vertebrates. At early stage of embryogenesis the zebrafish shares indeed with other vertebrates the basic organization of the body [4]. The nervous system and the skin come from the ectoderm, the gut form a tube of endoderm, while the mesoderm forms the musculoskeletal and vascular systems. The observation and the manipulation of zebrafish embryos are easy, because embryos are robust, transparent and develop externally to the mother. The observation is also very fast, if compared with those of others widely studied animal models, e.g. mice. Considering an optimal temperature of incubation, 28.5°C, zebrafish embryo proceeds indeed from the zygote to the larval stage (first and last stage of embryonic development) in only three days. Moreover, embryos can be easily genetically modified to express fluorescent proteins [5, 6], allowing the observation of internal cellular structures. Mutations to express human diseases can be generated using a variety of techniques, to perform fast and low cost tests of medical treatments and drugs [3]. The combination of all these features makes the zebrafish one of animal models most studied in the scientific community. Among other things, its genome has been fully sequenced and a database containing a wide range of information about zebrafish genetics, genomics, and development has been created [7].

1.1 Zebrafish embryo development

In the course of embryonic development a single, undifferentiated egg-cell undergoes characteristic changes - cell division, migration, differentiation and interaction-, giving rise to organized morphogenetic patterns such as tissues and organs. Following Kimmel et al. [8], in the case of zebrafish the embryogenesis can be subdivided in seven broad periods of development: zygote, cleavage, blastula, gastrula, segmentation, pharyngula,

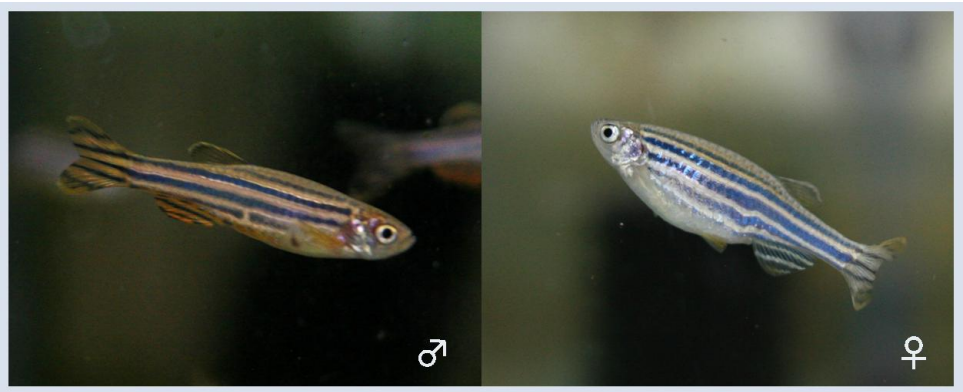


Figure 1.1: Zebrafish in the adult stage. The fish is named for the horizontal strips covering the side of its body. Left image: male animal. Right image: female animal.

and hatching (shown in figure 1.2). Periods of subdivision are related to particular developmental processes occurring during the embryogenesis and can be further divided in stages. Stages are usually named depending on the shape of embryo. In this section we describe the main morphological features of zebrafish embryogenesis periods. Stages are shown in Figures 1.3 – 1.10.

1.1.1 Zygote Period (0-3/4 h)

The zygote is designated as the period of time between the egg fertilisation and the first cell division, that occurs about 40 minutes later. The newly fertilised egg measures about 700μ in diameter and is composed by a disc of blastoderm (blastodisc) on the top of a noncellular mass of nutrients (yolk) (Fig. 1.3). We can distinguish between an animal and a vegetal pole (left and right of Fig. 1.3), which nearly identify the *anteroposterior* axis of the embryo body (from the future head to the future tail).

1.1.2 Cleavage Period (0.7-2.2 h)

As suggested by the name, in the cleavage period the original zygote divides six times, synchronously, about every 15 minutes, at regular orientations and without changes in the total mass. (Fig. 1.5). At the end of the period the cells number has been

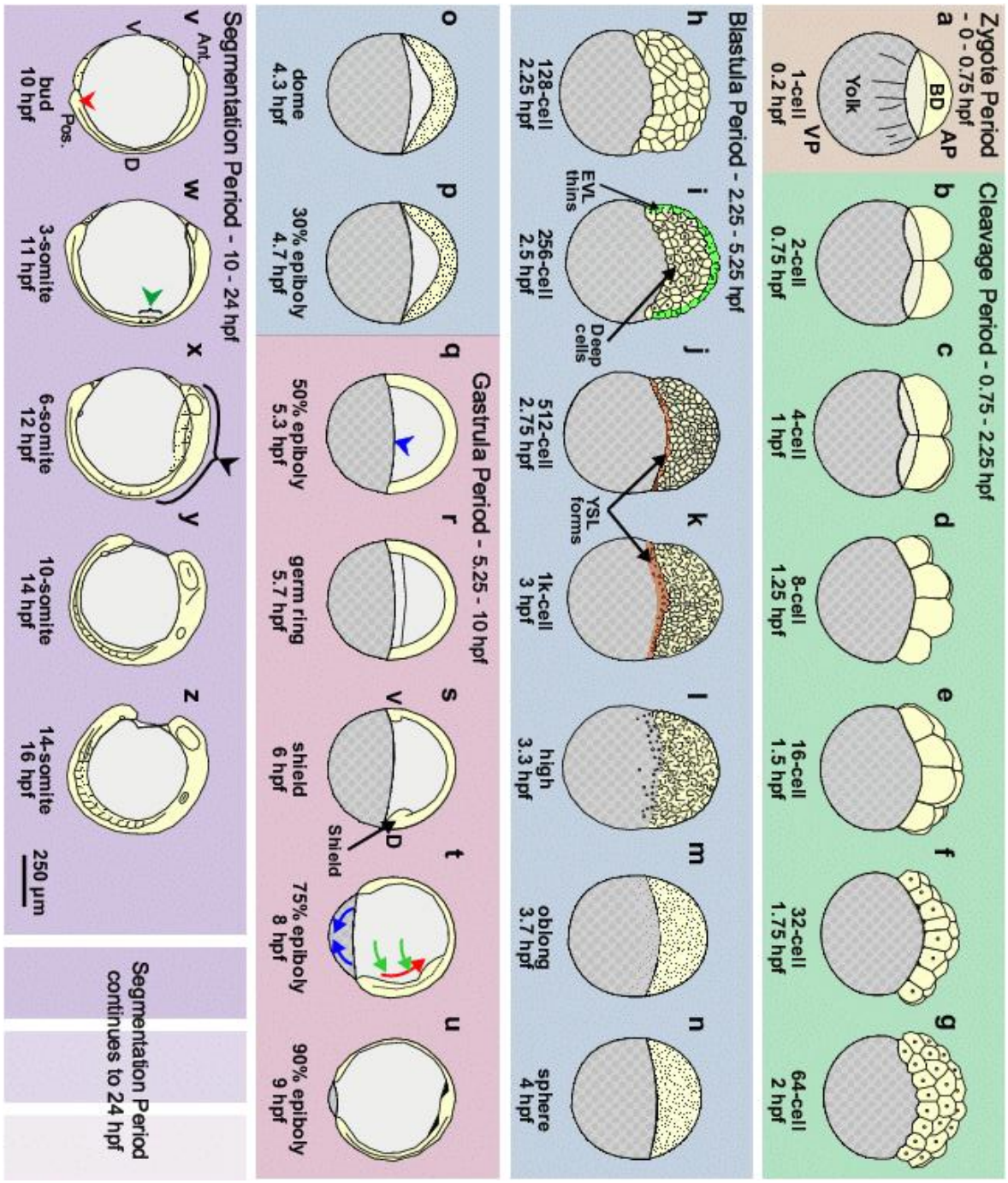


Figure 1.2: The stages differentiated the spatio-temporal evolution with morphological criteria. The figure represent some of these stages of zebrafish [8]

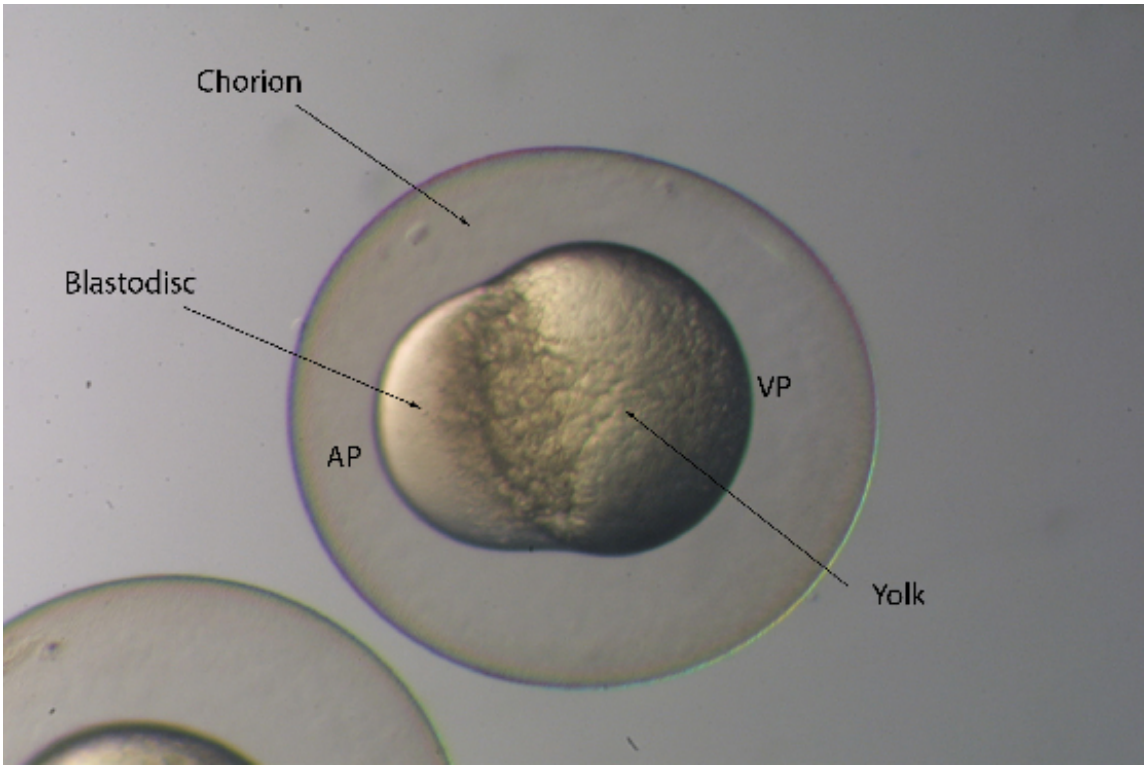


Figure 1.3: Zebrafish embryo in the zygote period at 1 cell stage. Yolk-free cytoplasm are segregated to the animal pole (AP), clearly identified on the left. The opposite side identifies the vegetal pole (VP). Figure modified from Steve Baskauf bioimages page [9].

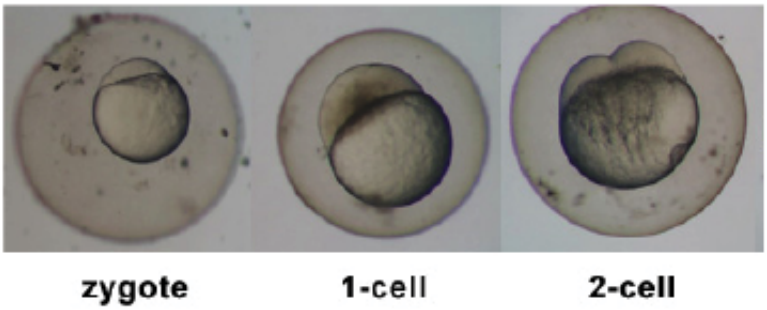


Figure 1.4: The stages of the zebrafish development begins with the zygote period, in which the zygote reaches its morula stage and then reaches the 2 cell stage. Figures from [10]

increased to 64 (see figure 1.5). It is possible to identify an *enveloping layer* (EVL) of outer, marginal cells, so called as they are close to the margin of blastodisc, and nonmarginal, deep cells, completely surrounded by EVL and marginal cells.

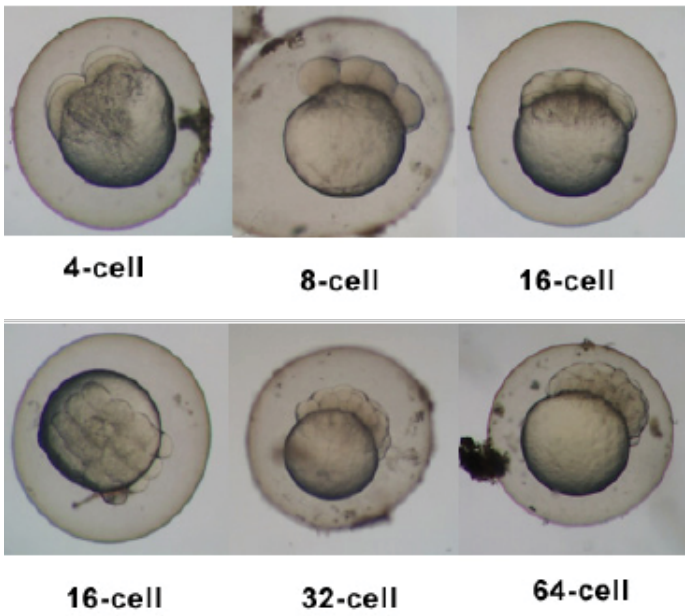


Figure 1.5: Within about 0.75 to 2.25 hours, it reaches the cleveage period where it will divide into the 64 cell stage. Meroblastic cleavage occurs in the zebrafish. Figures from [10]

1.1.3 Blastula Period (2 1/4 - 5 1/4 h)

With the term blastula is designed the period between the seventh cleavage and the beginning of gastrulation (Fig. 1.6). The blastula is marked by three important events. The *midblastula transition* (MBT), the formation of the *yolk syncytial layer* (YSL) and the beginning of *epiboly*. The MBT occurs around the tenth mitotic division (Fig. 1.6) and is characterised by the beginning of genes transcription. As a consequence cells start to divide asynchronous and cell cycle lengthens. The YSL is formed by marginal cells which collapse and release their nuclei into the yolk. Further division of such cells are noncytoplasmic, and the yolk remains a unique, syncytial cell. Both cells

of blastodisc and YSL spread over the yolk, starting the process of epiboly. Epiboly continues during the subsequent period of gastrulation until covering entirely the yolk with a thin cells layer. The outermost, EVL cells, increase considerably their number and change shape, becoming flat, thin and large, organized as epithelial sheet.

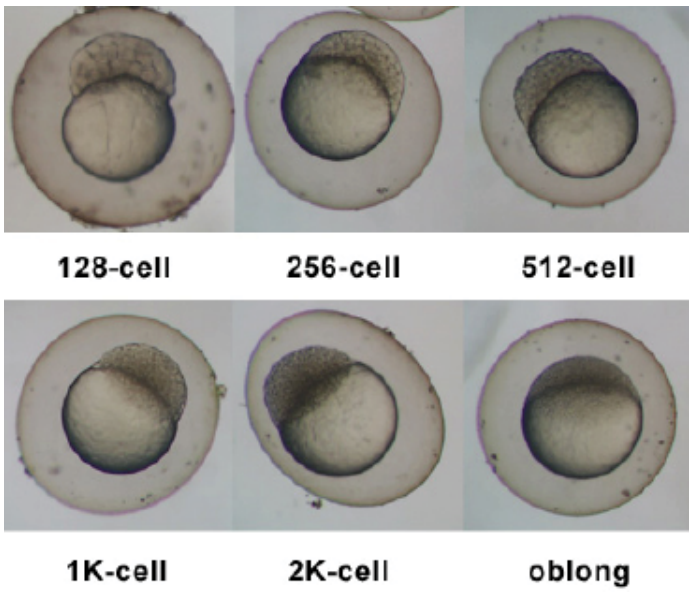


Figure 1.6: Views of embryos during the blastula period. Between 2.25 and 5.25 hours, the embryo reaches the blastula stage. After the mid-blastula transition stage. Figures from [10]

1.1.4 Gastrula Period (5 1/4 - 10 h)

The gastrula period is characterized by complex cells movements giving rise to the primary germ layers. Consequently to movements of DEL cells involution, at about 50% of epiboly appears a thick anular region - named germ ring -, all around the margin of blastodisc (Fig. 1.7). The germ ring is composed by both hypoblastic (close to the yolk) and epiblastic cells. Hypoblast will give rise to endoderm and mesoderm, epiblast to ectoderm. After the germ ring formation epiboly arrests until the formation of the embryonic shield, an accumulation of cells in a particular position along the germ ring, as shown in Fig. 1.7. The *dorsoventral* (from the back to the belly) and *mediolateral*

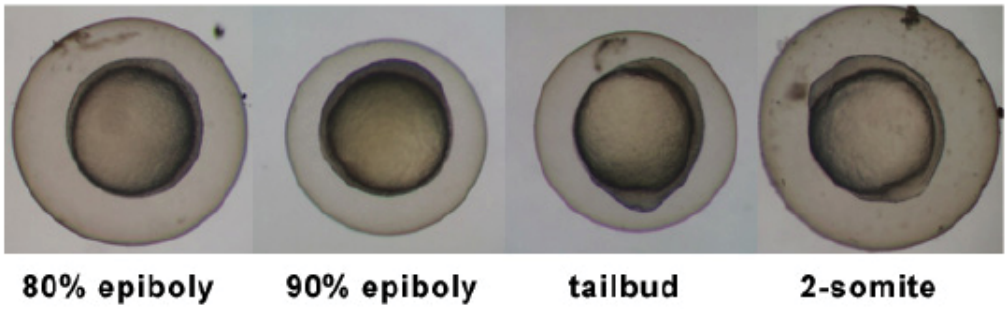


Figure 1.7: Zebrafish embryo development during the gastrula period. Gastrulation begins between 5.25 10.33 hours. The tailbud begins to form and 2 somites can be observed. Figures from [10].

axes are well defined for the first time. Then the epiboly continues until covering completely the yolk. At the same time the anterior axial region of hypoblast migrates towards the animal pole (Fig. 1.7) to form the prechordal plate. General movements of convergence to the dorsal midline make it thicker than the ventral one. The dorsal epiblast thickens anteriorly to form the precursor of brain, the neural plate, while its posterior cells will contribute to form the spinal cord. At 100% of epiboly, on the vegetal pole, appers the tail bud. Hypoblastic cells of future prechordal plate accumulate close the the animal pole to form a bulge of gland cells.

1.1.5 Segmentation Period (10-24 h)

During the segmentation period occur many important events, as cells start to differentiate morphologically and organogenesis begins. Somites, which will give rise to muscles and vertebral cartilage, develop sequentially on the dorsal side of the embryo. The neural plate, the primordium of the central nervous system, thickens to forms the so-called neural keel. Optic primordium become visible (Fig.1.8). At about 16 hours of development (14-19 somites stage) the tail start to stretch such elongating the embryo, which until this stage of development had maintained, in general, a spherical shape with a diameter of $700 - 800\mu m$. It is now possible to distinguish four region of the brain, telencephalon, diencephalon, midbrain and hindbrain. Otic placodes can be seen near to the hindbrain. The neural keel starts to round to form the neural tube and the notochord beneath it, forming the primary longitudinal skeletal axis of the body,

became distinguishable. Then the embryo elongates and the tail become more evident. Otic placodes form the otic vesicle and the neural tube is almost formed along the whole length of trunk. Neurons start to develop axons and blood cells are forming. Near to the trunk somites have developed pioneers of muscles that produces muscular contractions and some body movements appears.

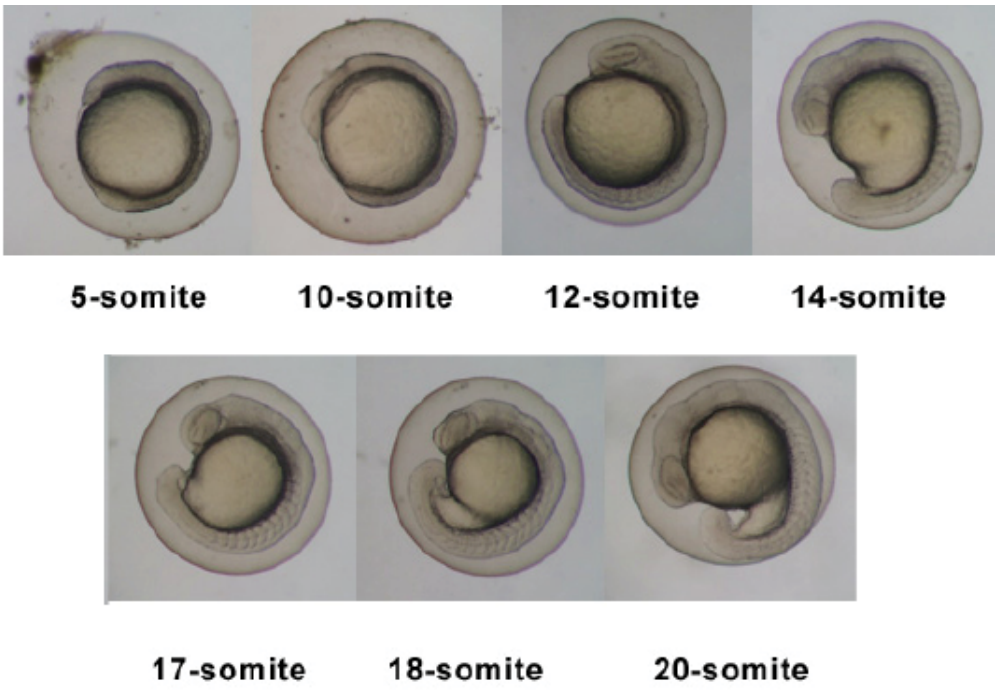


Figure 1.8: Zebrafish embryo development at the segmentation period. Betwen 10.33 and 24 hours, segmentation begins to happen. This the period in which the folding of the embryo begins to occur and the formation of the somites begins to continue. The embryo begins to elongate and the neural cord and notochord continue to develop. During this stage, the dermis, vertebrae and skeletal muscle are formed as well. Figures from [10].

1.1.6 Pharyngula Period (24-48 h)

The pharingula period corresponds to the second day of development of zebrafish embryo (see figure 1.9). The period is named for the formation of primordium of six pharingeal arches. Since the beginning the notochord is already well developed, somites are formed and the brain is divided in 5 clearly recognizable lobes. Pigment cells dif-

ferentiate and pigmentation appears first in pigmented epithelium of retina, then in melanophores on the dorsolateral side of embryo, rapidly progressing and becoming a prominent feature. Anyway, also at the end of period, lateral strips are still not complete. The circulatory system forms. The hearth become visible and start to beat. At first it is visible as a straight tube in the most anterior region of the yolk, then it folds to form the atrium and the ventricle. Blood circulates through carotid artery and cardinal veins and aortic arches appear. The median and pectoral fins form and become gradually more prominent. The tail lengthens, the head shortens and straightens out.

Figure 1.9: Pharyngula period, which occurs between 24-48 hours. During this time period, the notochord is fully developed, the pectoral fins begin development and the cirgulatory system can be observed along with a heart beat. Figures from [10].

1.1.7 Hatching Period (48-72 h)

At the last period of embryonic development the organogenesis is nearly complete except for the gut. The pectoral fins continue to grow and show cartilage and blood vessels. The mouth appears in the pharyngeal region, lying at first between the eyes, in the back of the head. Subsequently, following the formation of the jaw, the mouth moves anteriorwards. In the first and second pharyngeal arch (mandibular and hyoid arch) is indeed possible to observe the formation of cartilage and muscular tissues. The remaining four pharyngeal arches have function of support for the gills and are called branchial arches. Each gill homes an aortic arch, allowing the exchange of oxygen between the blood and the external environment. In this period forms also a fifth branchial arch but it does not give rise to formation of gills and it has only function

of support. The formation of branchial arches in cartilage appears about twelve hours after the first two pharyngeal arches.

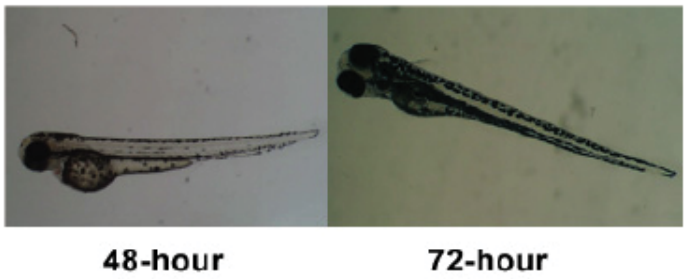


Figure 1.10: The hatching period in which the olfactor palcodes are fully developed, the pectoral fins are elongated and the development of cartilage begins. This stage usually occurs between 48 and 72 hours. Figures from [10].

Chapter 2

Live-Cell imaging methods

Live-cell Imaging techniques are widely used, especially in the field of Biological Imaging, to provide critical insight into the fundamental nature of cells and tissues. Recent advances in both fluorescent proteins and laser-scanning microscopy open the way for studying and understanding mechanism that were previously impossible to observe. *In toto* 3D + time imaging, in which every cell in a living embryo can be traced through space and time during its development, may become a standard technique for small transparent embryos such as zebrafish. After a short introduction on the fluorescent proteins, in this chapter we describe the imaging techniques used for acquiring our experimental dataset, the confocal and the multi-photon laser scanning microscopy. Particular attention is devoted to the main drawbacks of these imaging techniques. They influence and the typology of data to be examined, impose physical limits to the dataset dimensionality and resolution, to the image contrast, the signal-to-noise ratio and the specimen survival.

2.1 Fluorescent proteins

The first fluorescent protein which has been observed was the GFP (green fluorescent protein), so called as it has the property of emitting green light if illuminated with ultraviolet light. Although the GFP has been isolated from the jellyfish *Aequorea victoria* [11] in 1970s, the significant potential as a molecular probe was not realized until several years later, in 1994, when the GFP was used as marker for gene expression in E. coli and C. elegans [12]. Since these early studies, green fluorescent protein has been engineered and modified to produce a vast number of fluorescent proteins, with a fluorescence emission between the blue and the yellow. Others longer wavelength fluorescent proteins, emitting in the orange and red spectral regions, have been developed from the marine anemone, *Discosoma striata*. Fluorescent proteins are nowadays widely used as *in vivo* reporter molecules in biological research, such that the discovery and the development of GFP earned M. Chalfie, O. Shimomura and R. Y. Tsien the 2008 Nobel Prize in Chemistry.

2.2 Imaging

Live-cell imaging can be performed with two similar techniques, the Confocal Laser Scanning Microscopy and the Multi Photon Laser Scanning Microscopy (CLSM and MPLSM, respectively). These imaging methods are based on the detection of fluorescent signal emitted by a sample previously modified to express fluorescent proteins. Both techniques allow *in vivo* 3D + time imaging, providing serial optical sections of intact, living specimen. The current generation of microscopes generally have the capability to simultaneously detect the light emitted at two or more different frequencies, allowing the simultaneous observation of several internal structures with a reduced dose of irradiation absorbed by the sample. Live-cell imaging is nevertheless always characterized by the the best compromise between the quality of images (spatial and temporal resolution) and the damage-survival of sample. A spatial resolution allowing the detection of finest details is *de facto* in contrast with an high temporal resolution able to accurately follow any changes into the specimen. Furthermore, the exposition to an excessive level of illumination irreparably damages the cells making them useless for further observations. We first describe in next paragraphs the main principles of Confocal Laser Scanning Microscopy and the physical constraints of imaging process. We than show how the most part of drawbacks of such technique can be solved by using more than one photon to excite fluorescent proteins in the sample, as done in the MPLSM.

2.2.1 Confocal Laser Scanning Microscopy

In a confocal microscope a spot of laser beam is focused into a volume element (ideally diffraction limited) of a fluorescent sample (fig. 2.1). The light emitted by the focused volume element is then detected and amplified by a photo-multiplier tube. The laser beam is scanned across the specimen and each volume element is associated with a discrete fluorescence intensity and represents one voxel in the resulting image. By scanning along the xy plane and varying the depth in z are obtained several slices that, recombined, provide three-dimensional, volumetric images. Time-lapse sequences of three-dimensional data can be acquired giving rise to tetra-dimensional images. The laser source is emitted at a particular frequency, selected so as to excite the fluorescent proteins in the sample. It is then focused by electronic lens and further filtered by a

dicroic mirror. Interfering intensity from outside the focused volume element is removed by a pinhole located in a conjugate focal plane to the specimen [13]. Excitation sources can be provided by the argon ion laser, which emits at wavelengths of 488nm and 514nm and by the argon/krypton mixed gas laser. It emits instead at wavelengths of 488nm, 568nm and 647nm, that cover many of the commonly used fluorophores.

A series of physical factors limit the spatial and the temporal resolution achievable in a Confocal Laser Scanning System, such as in a Multi Photon Laser Scanning Microscope. First of all, due to typology of acquisition, which requires a scanning along every plane of imaged volume, the temporal resolution depends on the spatial resolution and

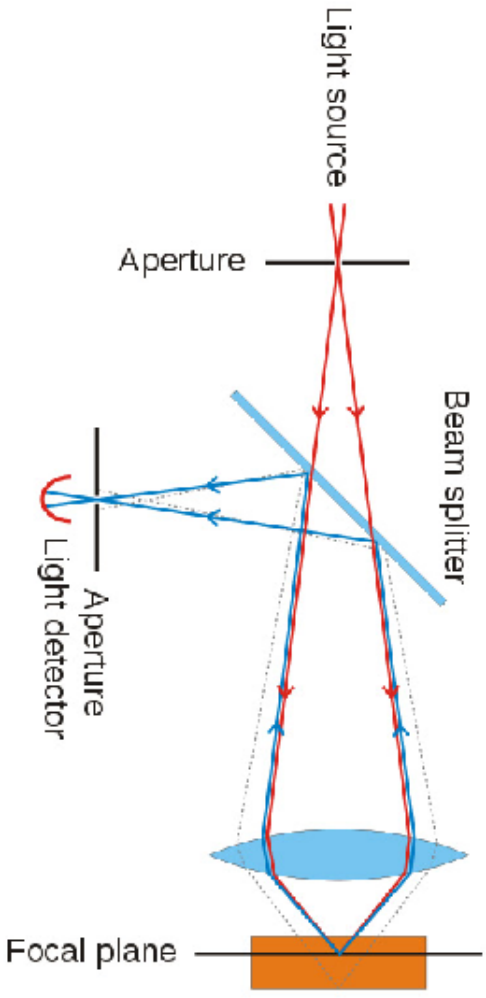


Figure 2.1: Diagram of a single-photon confocal microscope

on the imaged field and vice versa. The higher is the temporal resolution, the lower will be that spatial and the dimensionality of dataset. Furthermore, the resolution is theoretically limited by the diffraction theory. A point-like object is represented by an Airy pattern whose spatially distributed intensity is given by the Point Spread Function (PSF). The properties of PSF either in the image plane or in the axial direction are the most important factors that limit the resolution of a microscope. The generally accepted criterion for the minimum resolvable detail, the Rayleigh criterion, states that two points are resolved when the first minimum of an Airy disk falls on the maximum of the other. Following this criterion, the lateral ($R_{lateral}$) and axial (R_{axial}) resolution of confocal microscopy applications can be defined as

$$R_{lateral} = \frac{0.46\lambda}{NA} \qquad\qquad R_{axial} = \frac{1.4\lambda n}{NA^2} \tag{2.1}$$

where λ represents the wavelength of the emitted light, n the index of refraction of medium in which the lens works and NA the numerical aperture of the instrument. If compared to that in a wide field microscope, $R_{lateral}$ is reduced by about a 30%, as the emitting fluorophores are already located in a small focused volume. But because the axial resolution is generally higher than the lateral, 3D images are usually acquired in a non-uniform grid, with a spacing in z direction greater than on the xy plane. As a consequence objects spatially close in the specimen can appear non-resolved, superimposed in z, as shown in Fig. 2.2,2.3. Moreover, the resolution predicted by the diffraction theory concerns ideal conditions. In the field of live-cell imaging, samples are thick and inhomogeneous. Spatially closed points can be truly resolved depending on the image contrast, which is affected by the number of photons detected and by the corresponding signal-to-noise ratio. Further limits in image contrast and resolution are related to the pixelation. To conclude, the radiation intensity and duration of exposure of live specimen, that if increased could improve the signal-to-noise ratio, are restricted by photobleaching and photodamage. As previously introduced, live-cell imaging is marked by a tricky compromise between antithetic requirements.

2.2.2 Multi Photon Laser Scanning Microscopy

The MPLSM [14] is similar to confocal laser-scanning microscopy (CLSM). In both a laser beam is scanned across a focal plane in a fluorescence-stained specimen, and the light emitted by the excited sample is detected by a photomultiplier tube which

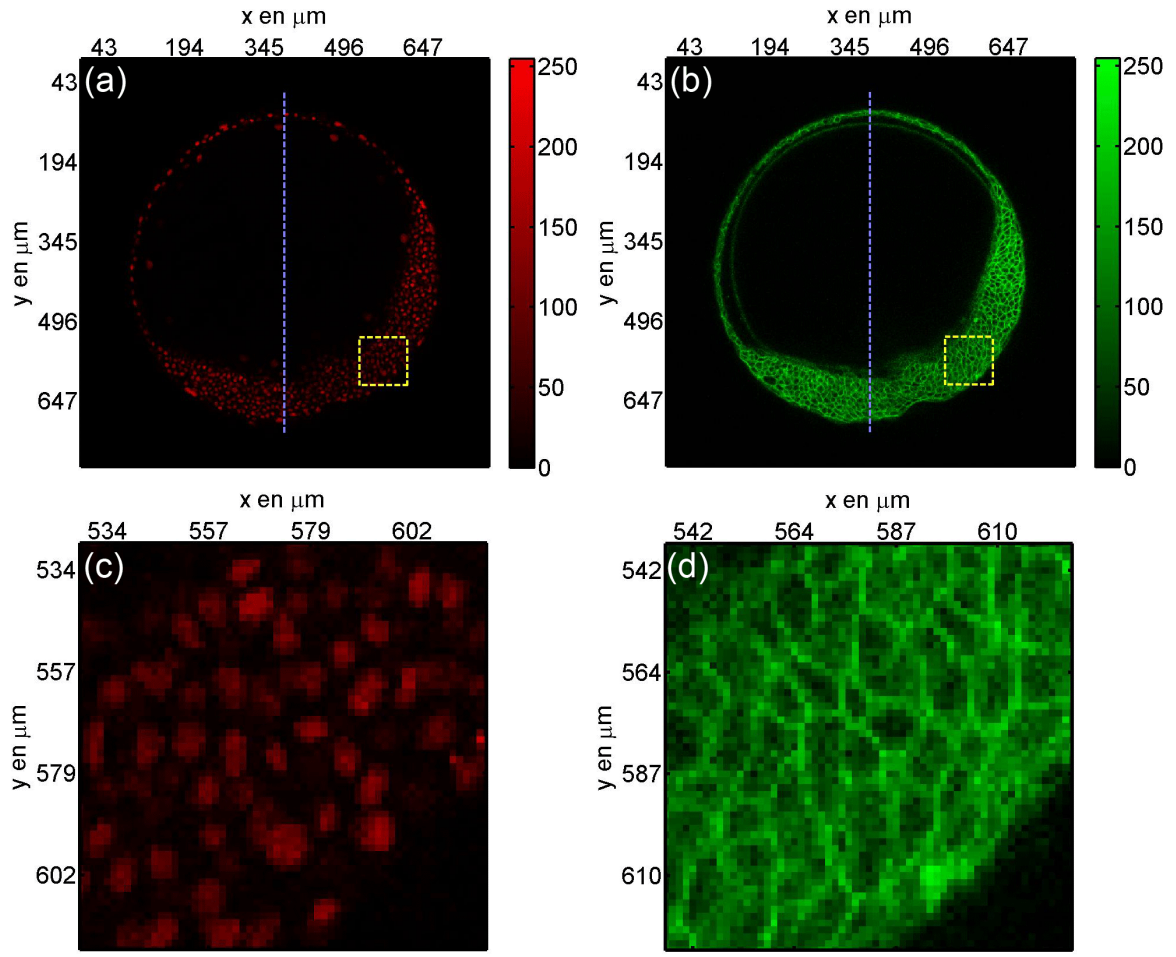


Figure 2.2: Example images of 071221aF experiment in the optical plane (XY). The images were acquired at a depth of $150\mu m$ and 10 hours post fertilisation of the egg. The depth of this image corresponds to the line fuchsia in figures 2.3. The full section of the nuclei image(a) and membrane (b) shows that the embryo is fully capture at this depth. On (c), the detail of the nuclei, which corresponds to the rectangle in yellow dots (a), shows structures at good resolution. (d) presents a detail of cell membranes in the same area.

amplifies the signal and produces digital images (Fig. 2.4). The main difference is that the excitation of fluorophores in the sample is produced, in the MPLSM, by the nearly simultaneous absorption (within $10^{-18}s$) of two (or more) photons, each at a wavelength higher than the wavelength required for single photon excitation. In the most commonly diffused case of a two photon microscope the wavelength is about doubled and the energy halved. As a consequence the fluorescent light emission depends on the square of the excitation intensity and only in the focal volume there is sufficient intensity for the process to occur. Emitted light falls down rapidly outside the focal point. In such a way the out-of-focus fluorescence is intrinsically limited and, unlike in confocal microscopy, it is not necessary to use spatial filtering as a small pinhole. A drawback is that multiphoton excitation requires a very large instantaneous flux of infrared

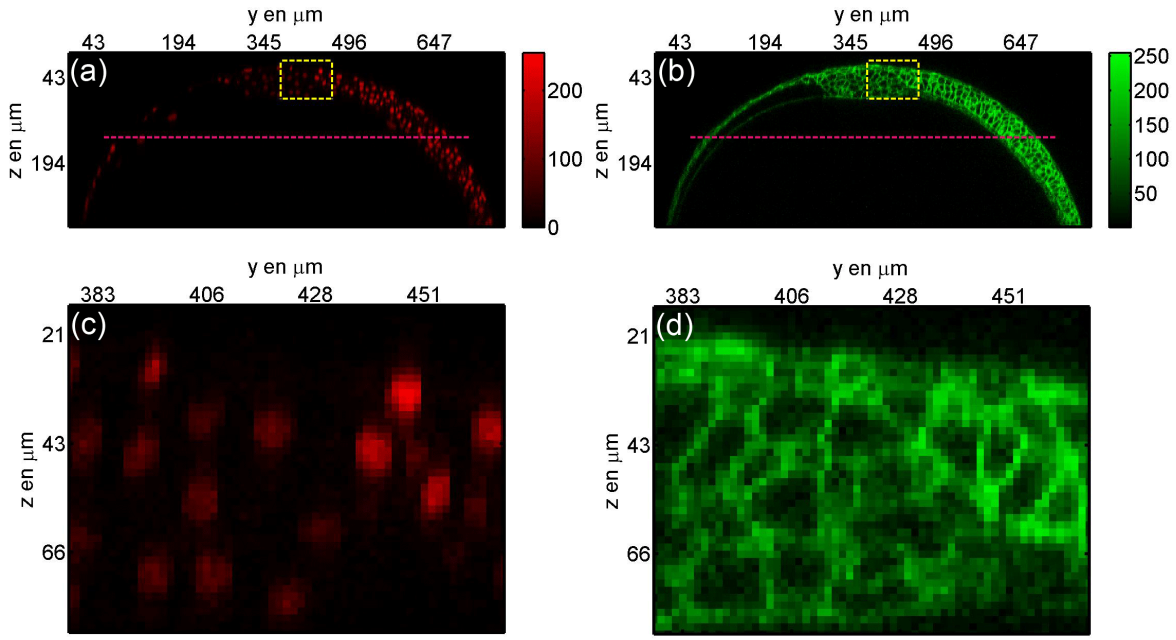


Figure 2.3: Example images of 071221aF experiment in a plane parallel to the optical axis (YZ) (blue dots in figure 2.2 (a) and (b)). The images were acquired between 0 and 300 μm depth at 10 hpf. The image of volume nuclei (a) and membrane (b) shows that it crosses most of the animal tissue. We note that the signal to noise ratio tends to decrease with depth. Thumbnails (c) and (d) corresponding to the rectangles in yellow dot (a) and (b) show a detail of the image of the nuclei and cell membranes along the optical axis. The axial resolution is nearly the double than the lateral and some nuclei appear not resolved along z direction.

photons and thus the use of a laser more expensive and complex than lasers suitable for confocal microscopy (e.g. the titanium-doped sapphire (Ti:Sapphire)). The laser produces pulses extraordinarily brief ($100fs$) at high frequency (usually $80-100MH$) to ensure very high instantaneous energy but with a low average value to avoid sample damage. Because of the longer wavelength of excitation, the theoretical resolution of TPLSM is worse than that of CLSM by approximately a factor of two [15]. But considering that the confocal microscopy does not approaches its theoretical resolution, likewise the MPLSM, the resolution is *de facto* comparable [16]. The two-photon imaging provides indeed an higher signal-to-noise ratio and consequently an higher contrast because of the intrinsic removal of out-of-focus signal. Conversely, the infrared light used in MPLSM penetrates deeper into scattering specimens than commonly used visible light, allowing to image two or three times deeper [17]. Moreover, the infrared light produces only a small photodamage to cells if compared with the phototoxicity at higher frequency. In a comparison with a confocal excitation, photodamage and photobleaching are further reduced as limited to the focal plane. As shown in Fig.

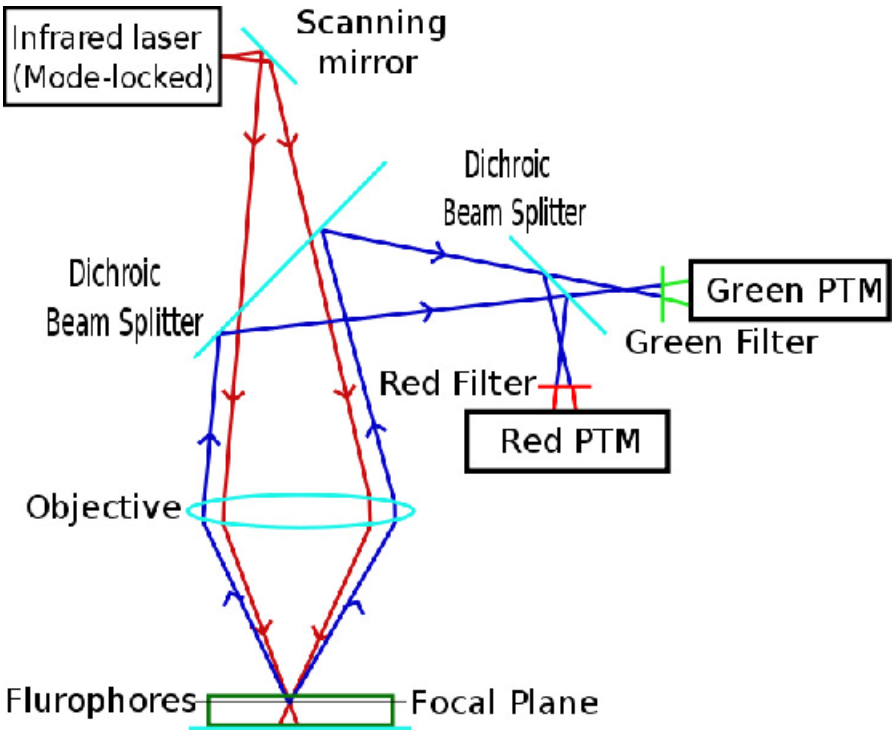


Figure 2.4: Diagram of a two-photon microscope

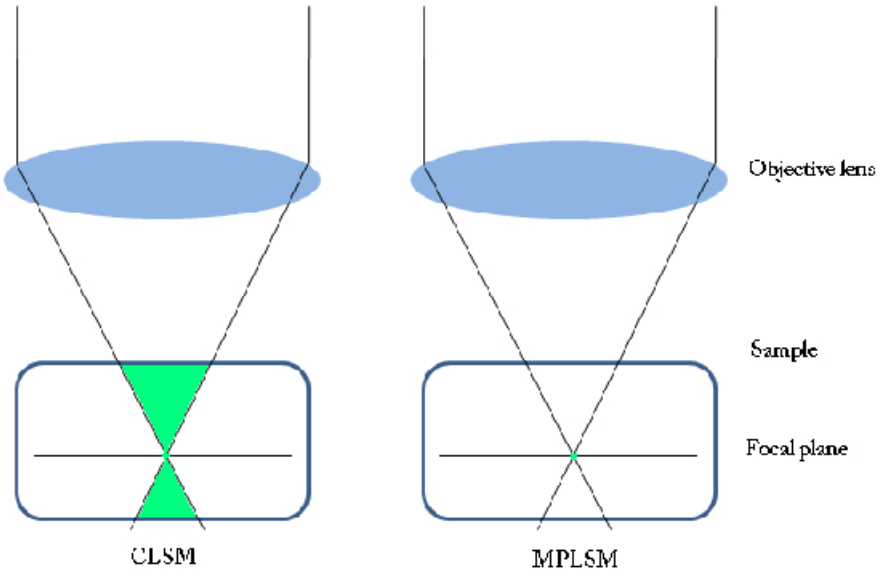


Figure 2.5: Comparison between confocal and multi photon laser scanning microscopy. In CLSM fluorescence, and consequently photodamage and photobleaching, occur throughout the whole thickness of sample along the path of the excitation beam. In multi-photon fluorescence (right) the phenomenon is limited to a small focused volume.

2.5, in the CLSM the specimen is conversely damaged in its whole thickness at every single scan. The reduced phototoxicity and dye bleaching, together with the ability to penetrate deeper into tissues, makes this technique particularly suitable for acquisition of time-lapse sequences of three-dimensional data.

2.3 Embryo staining and mounting

Wild-type zebrafish embryos were injected at the one cell stage with 200pg mCherry/H2B RNA and 200pg eGFP-ras prepared from PCS2+ constructs [5, 6]. Although mCherry, unlike eGFP, bleached significantly through imaging, this color combination was the best compromise allowing proper staining of the cell membranes for further segmentation. Injected embryos were raised at $28.5°C$ for the next 3 hours. Embryos were mounted in a 3cm Petri dish with a glass coverslip bottom, sealing a hole of 0.5mm at the Petri dish center where a Teflon tore (ALPHAnov) with a hole of 780 μm received the dechorionated embryo. The embryo was maintained and properly oriented

Table 2.1: Developmental table (according to [8]) of the imaged embryo ID 070418a

Time step	Image duration	Development time	Drift from the $28.5°C$ table	Equivalent at $28.5°C$	Developmental stage
t0	0h00	4h10	Non detectable	Between 4h00 and 4h20	Late sphere stage
t120	2h13	6h23	Non detectable	6h00 to 6h30	Late shield close to 60% epiboly
t320	5h57	10h07	+ 0h07	10h00	Tail bud
t360	6h39	10h49	+ 0h29	10h20	1 somite
t420	7h46	11h56	+ 0h36	11h20	3-4 somite
t480	8h53	13h03	+ 0h57	12h00	5-6 somite
t540	10h03	14h13	+ 1h13	13h00	7-8 somite

by infiltrating around it 0.5% low melting point agarose (Sigma) in embryo medium [18]. Temperature control in the room resulted in a temperature of about $26°C$ under the objective slightly slowing down development with respect to the standard $28.5°C$ developmental table [8], (Table 2.1). After the imaging procedure, embryo morphology was checked under the dissecting binocular and the animal was raised for at least 24h to assess morphological defects. Embryo survival depends on total imaging duration, average laser power, and image acquisition frequency (or time step Δt). Dataset 070418a and 070420a correspond to a standardized procedure with $50s < \Delta t < 70s$ allowing 75-110 sections per time point for 10 hours with an average laser power of 80 mW delivered to the sample. Lowering the laser power to less than 60 mW and lengthening Δt up to 3.5 min allowed imaging embryos for more than 20 hours, then raising them until adulthood. These conditions were used for making up to 320 sections at 400Hz line scan rate (bidirectional scanning), or 200Hz to improve signal-to-noise ratio, as for the image data set 080322a shown in movie Movie1-3.

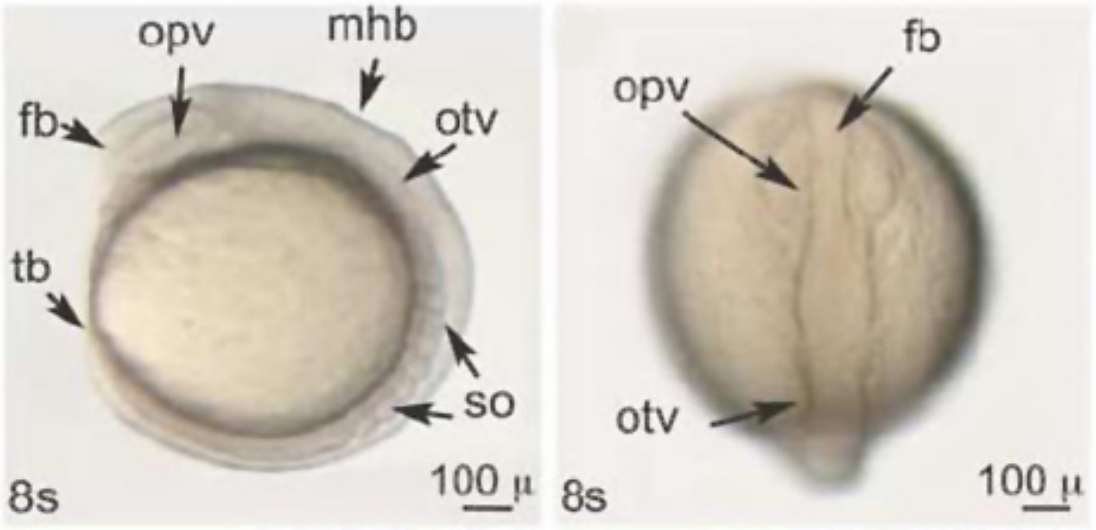


Figure 2.6: Bright field views of live zebrafish embryos. 8-somite stage. (left) lateral view, dorsal to the right. (right) animal pole view, rostral to the top. fb, forebrain; mhb, midbrain-hindbrain boundary; opv, optic vesicle; otv, otic vesicle; so, somites; tb, tail-bud.

2.4 Image acquisition

Two-photon-excited fluorescence (2PEF) microscopy, which allows in vivo imaging of embryos for extended periods of time, has proven to be the most effective technique for deep-tissue fluorescence imaging with sub-cellular resolution [19, 20].

However, achieving the automated tracking of cells throughout the whole living embryo from 4D image data sets involves a difficult compromise between many variables, such as signal-to-noise ratio, spatial and temporal resolution, thickness of the explored volume and cell survival, volume size, image acquisition duration, bleaching and phototoxicity which has not been successfully challenged to date.

Tables 2.1, 2.4 and figure 2.7 gather the characteristics of experiences presenting the best compromise.

Embryo labeling was obtained through RNA injection performed at the one-cell stage to obtain ubiquitous expression of H2B/mCherry fusion protein and farnesylated eGFP, which stained nuclei and membranes respectively [5, 6]. While nuclear staining was instrumental to perform cell tracking, membrane staining was essential to assess cell morphology, behavior and neighborhood, and to reveal morphological landmarks.

Table 2.2:
Experiments details used to reconstruct the zebrafish morphogenesis

Name	Lineage	Orientation	Image size (x,y,z,t),GB	spacing (μm,μm,μm,s)	Start
070128c	Wild	dorsale	1024x1024x100x167, 32GB	0.68x0.68x2.05x260	6 hpf
070418a	Wild	pole-animal	512x512x104x761, 38GB	1.37x1.37x1.37x67	4 hpf
070420a	Wild	pole-animal	512x512x74x702, 25GB	1.37x1.37x1.37x49	4 hpf
070927b	Wild	pole-vegetatif	512x512x131x413, 26GB	1.37x1.37x1.37x87	6 hpf
071126aF	Wild	pole-animal	512x512x234x306, 34GB	1.51x1.51x1.51168	4 hpf
071221aF	Wild	pole-animal	512x512x205x180, 18GB	1.51x1.51x1.51x304	4 hpf
071226a	Wild	pole-animal	512x512x120x231, 13GB	1.37x1.37x1.37x154	4 hpf
080322a	Wild	laterale(90)	512x512x139x420, 28GB	1.37x1.37x1.37x177	4 hpf
081018a	oeptz57-/-	pole-animal	512x512x171x433, 36GB	1.31x1.31x1.31x151	4 hpf
081025a	oeptz57-/-	pole-animal	512x512x179x406, 35GB	1.31x1.31x1.31x158	4 hpf

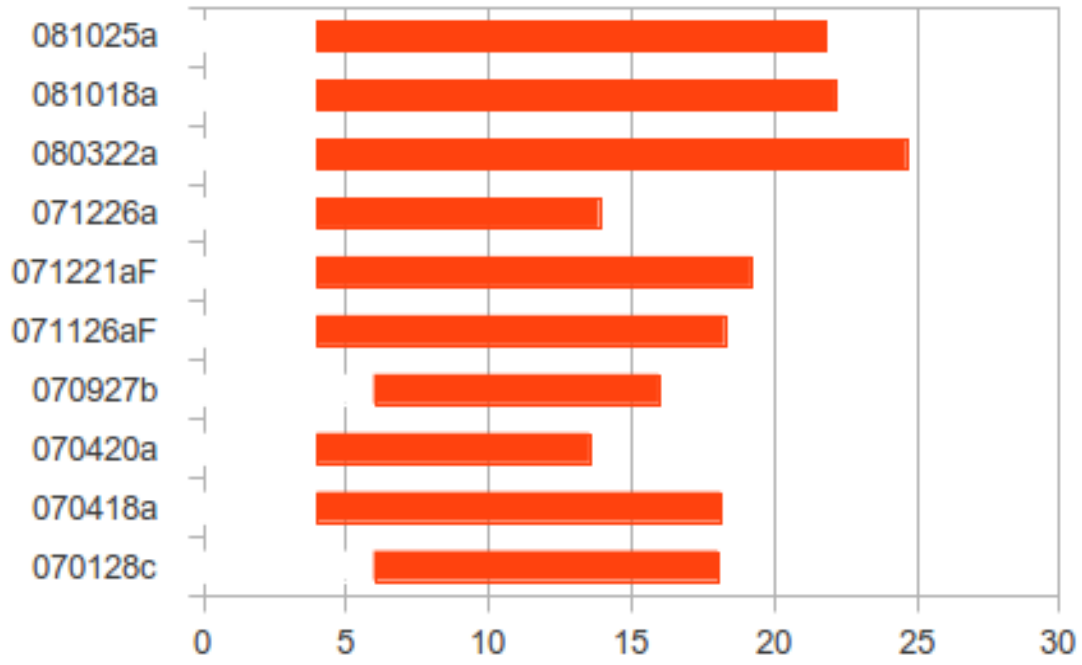


Figure 2.7: Dataset time comparison. x axis: hours post fertilisation.

Embryo imaging started when sufficient signal was detected (4 hours post-fertilisation hpf at $28°C$, i.e., at the sphere stage) and imaging proceeded for the next 10 hours at $26°C$ (Figures 2.6, 2.8, Table 2.1, movie Movie1-1).

Imaging was achieved with a Leica DM6000 upright microscope SP5 MLSM equipped with a 20/0.95NA W dipping lens objective (Olympus). Axial resolution at the sample surface (1.5 μm) was estimated by recording 3D images of 0.1 or 1 μm fluorescent polystyrene beads (Invitrogen) at the surface of an agarose gel. Field size was 700-700 in x, y, 140 μm in z for 070418a and 100 μm in z for 070420a; voxel size was 1,37 x 1.37 x 1.37 μm^3. Simultaneous dual wavelength excitation with pulses at two different wavelengths (980 and 1030nm) with pulsed laser beams (50Mhz, 200fs) was provided by a solid-state Ytterbium femtosecond oscillator (T-pulse 20, Amplitude Systemes) operating at 1030nm. The beam was split by a polarising cube. One of the two beams was launched into a nonlinear photonic crystal fibre (Amplitude Systemes) that shifted the optical spectrum from 1030 to 980nm. Light beams (1030nm and 980nm) were then recombined before being injected into the microscope. Raw data visualisation in Figure 2.8, Movie1-1, Movie1-2, Movie1-3, was done using Amira software (Mercury Computer

Systems).

The duration of the experiments, between 10 and 20h, with an acquisition of one volume per minute does not disturb the development of fish. Too much power from the laser, however, can lead to a necrosis of tissues. Sometimes gastrulation leads to the formation of a more anteroposterior axis that plunges into the yolk during a normal development. Stress exerted by the agarose is certainly the cause but we observed that it does not prevent the implementation of the plan of organization of fish whose structures are normally despite the disruption of the initial morphogenesis.

Figures 2.2 and 2.3 illustrate the type of images obtained at 10 hpf, depth of 150 microns. Imaged structures are have good contrast in the imaging plane (XY) but less along the optical axis (OZ). In this direction, it is more difficult to distinguish cellular structures because of the poor resolution. The theoretical resolution of the microscope is worse along the optical axis which may be sufficient to explain the degradation of the images in this direction. The signal to noise ratio decreases in the deeper regions.

The images show non uniform regional labeling. They have several origins: differences between tissues, the "bleaching" of fluorophores, a lack of labels specificity, imaging depth. Thus, the outer surface of epithelial membranes is not label. The label is different in the hypoblast and the epiblast and there is an accumulation of membrane signal at the border of these two regions. The "bleaching" is first seen at the top of the embryo in the epithelial cells where nuclear signal may completely disappear. Finally, some intracellular structures are sometimes label of nonspecific manner as is the case of the periphery of nuclei in the image of the membranes.

The experimental device used to follow the development of the zebrafish has a cellular scale. Quality images, however, degrades with the depth and duration of the experiment. Below 300 μm and a after 15th hours the images not suited to a automatic reconstruction of cell movements. The acquisition will however cover up to half of the embryo during ten hours of development into a unique experience. It provides for this time quality images suited to the reconstruction and description of the cellular environment.

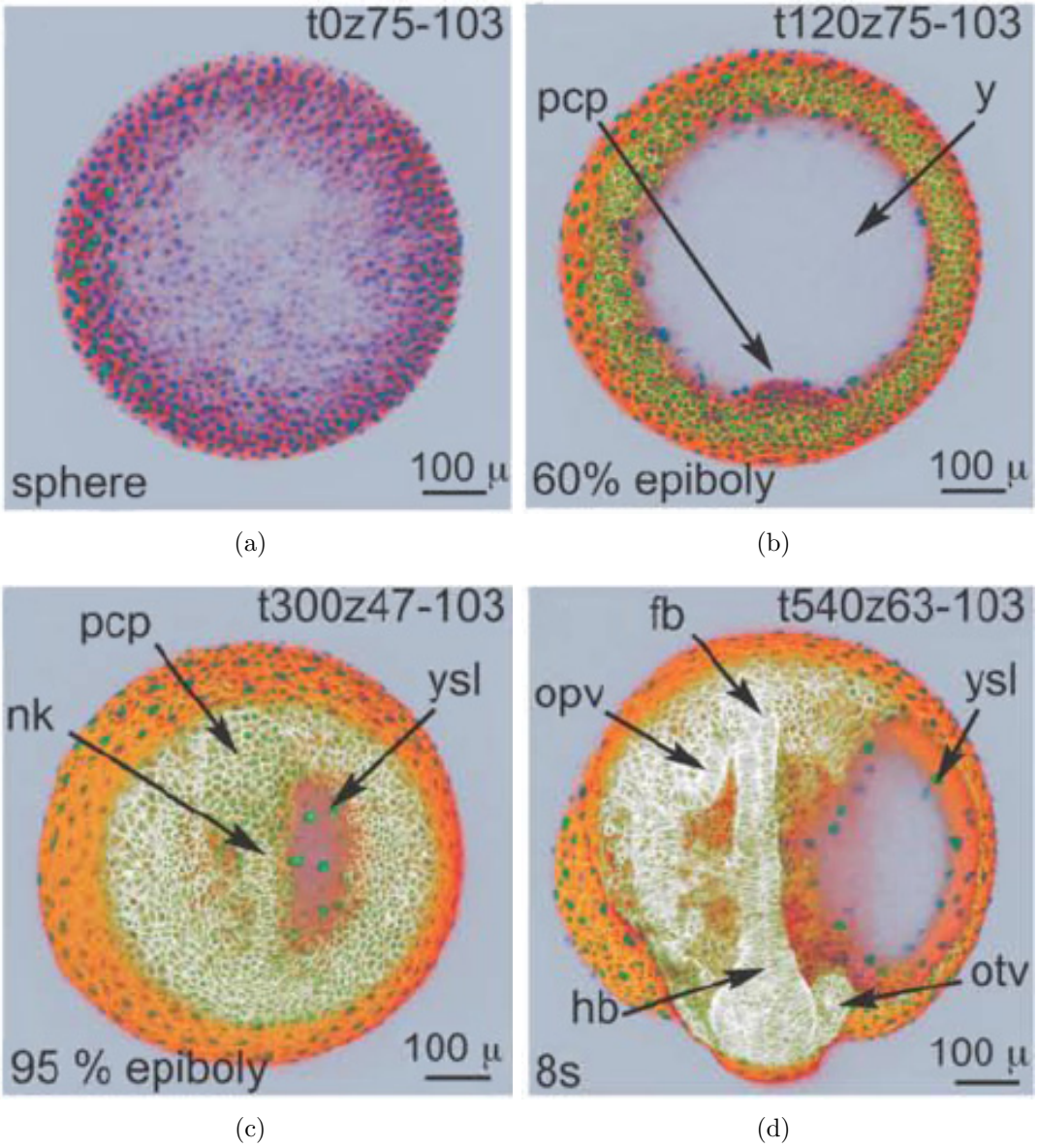


Figure 2.8: Reconstructing zebrafish brain early embryogenesis from time lapse optical sectioning. (a-d) Image data set ID 070418a, raw data 3D visualization, animal pole view, orthoslice removing upper sections, time points and shown (z projected) sections indicated top right, developmental stage indicated bottom left. fb, forebrain; hb, hindbrain; nk, neural keel; opv, optic vesicle; otv, otic vesicle; pcp, prechordal plate; y, yolk; ysl, yolk syncytial layer.

2.5 Mathematical definition

Either 2D or 3D nuclei or membranes images taken by a multi-photon laser microscopy at early stages of zebra-fish embryogenesis are represented by the image intensity functions, which we denote I_n in case of nuclei and I_m in case of membranes. Although nuclei and membranes images are color (either red or green in our case, cf. Figure 2.2), only one of the color channels is nonzero. So both I_n,I_m, can be understood as a scalar functions from an image domain Ω, $\Omega \in \mathbb{N}^2$ or $\Omega \in \mathbb{N}^3$ to $\mathbb{N}$ representing a "grey-level" image intensities. Without lost of generality we assume that $0 < I_n, I_m \leqslant 255$.

The nuclei and membranes images have a specific feature that can be utilized for our goals. Namely, nuclei are given by (highly noisy) humps of the function I_n,and, inner parts of cells bounded by cell membranes are given by (highly noisy) humps of the function $256 - I_m$.

Chapter 3

3D zebrafish embryo image denoising

The biological processes leading to organism formation and development of individuals is a fundamental issue for the biomedical research but is nowadays largely not understood. Achieving of an integrated understanding of such processes needs to analyze the cells individually and in a living embryo. Such goal represents a challenge for imaging techniques and image processing algorithms. In fact, recent advances in imaging strategies open the way to *in toto* 3D+time imaging of live animals with a resolution at cellular level and enough contrast to allow segmentation and tracking of individual cells. However, a noise is intrinsically linked to the scanning technique and image analysis algorithms applied consequently to the time series of 3D zebrafish images need to remove spurious, noisy, structures. The image filtering has to be always a first step in a chain of image processing operations, and it is very important to design appropriate filters and chose their optimal parameters for any particular type of data. The goal of this chapter is to apply the methods of nonlinear diffusion filtering to 3-D confocal images [1] of zebrafish embryogenesis in order to perform the segmentation [21] and the tracking [22] of individual cells. The filtering models are discussed in section I. Section II introduces a modified version of standard edge detector to filter membranes data. In section III we analyze behavior of the methods in processing our 3-D data set. We evaluate quantitatively the filtering results using the mean Hausdorff distance of isosurfaces to a gold standard.

3.1 Nonlinear PDEs Based Models

Let the input processed 3-D image be modelled by a real function $I_0(\mathbf{x})$, $I_0 : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^3$ represents a spatial rectangular domain. Observing that the Gauss function is a fundamental solution of the linear heat equation,

$$I_t = \Delta I = \nabla \cdot (\nabla I) \tag{3.1}$$

it has been possible [23, 24] to replace the classical convolution of an image with the Gaussian kernel of a given variance $v = \sqrt{2\sigma}$ by solving the linear heat equation for a corresponding time $t = \sigma$ and initial condition $I(\mathbf{x}, 0) = I_0(\mathbf{x})$. Applying (3.1) to an image means to diffuse its graylevels in an isotropic way. Despite the fact that (3.1) reduces the noise superimposed to the image, it blurs edges and moves their position.

To overcome these shortcomings, Perona and Malik [25] introduced the first nonlinear diffusion model called *anisotropic diffusion* in the computer vision community

$$I_t = \nabla \cdot (g(|\nabla I|)\nabla I), \tag{3.2}$$

making the diffusion coefficient g dependent on the image features. This model can behave locally as the backward heat equation, depending on the intensity of $|\nabla I|$, which is an ill-posed problem from a mathematical point of view. Therefore Catté, Lions, Morel and Coll [26, 27] proposed to use the convolution of ∇I with the Gaussian kernel to evaluate the diffusion coefficient, keeping all the advantages of the original model and avoiding its drawbacks

$$I_t = \nabla \cdot (g(|\nabla G_\sigma * I|)\nabla I). \tag{3.3}$$

We will refer to (3.3) as *modified Perona Malik* model in the course of this chapter.

Another, geometrical generalization of (3.1) was suggested by Alvarez, Lions and Morel [26]

$$I_t = g(|\nabla G_\sigma * I|)|\nabla I|\nabla \cdot \left(\frac{\nabla I}{|\nabla I|}\right). \tag{3.4}$$

Since the right hand side can be rewritten as $g(|\nabla G_\sigma * I|)I_{\eta\eta}$, where $I_{\eta\eta}$ represents the tangential component of ΔI, it provides a smoothing only in the direction orthogonal to the image gradient and uses g as a weighting term to slow down the diffusion on high image gradients. Since $K = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|}\right)$ is the mean curvature of level sets of I, (3.4) represents a geometrical diffusion of the image isosurfaces driven by their mean curvature, with an image dependent stopping function g. We call this model *slowed mean curvature flow*.

The last equation which we consider in this chapter is so-called *geodesic mean curvature flow*

$$I_t = g(|\nabla G_\sigma * I|)|\nabla I|K + \nabla g(|\nabla G_\sigma * I|) \cdot \nabla I, \tag{3.5}$$

which can be derived from (3.4) by adding the image dependent advective term. The equation was introduced simultaneously in [28, 29, 30] for image segmentation, for the filtering purposes (3.5) was suggested in [31, 32]. The image dependent velocity term

drives the graylevel isosurfaces in the direction of $-\nabla g$, i.e. towards local edges. It can be written in divergence form

$$I_t = |\nabla I| \nabla \cdot \left(g(|\nabla G_\sigma * I|) \frac{\nabla I}{|\nabla I|} \right) \tag{3.6}$$

to see (3.6) as a geometrical generalization of (3.2).

3.2 Image Denoising

We have applied the models introduced in the previous section to an in vivo zebrafish embryo nuclei and membranes 3-D images acquired with a confocal microscope Leica DM6000 upright microscope SP5 MLSM. The data represent stages of development of zebrafish embryo nuclei and membranes that cover a period 4-6 hours, from the sphere to the shield stage. In the entire period of development, the whole embryo is similar to a sphere with a diameter of 700 μ. The data have a physical dimension of 300 x 300 x 30 μ^3 and cover only the top part of the embryo.

The nuclei data are composed of well contrasted objects, with an enough good signal to noise ratio and regions approximately uniform. The membranes images are more complex and difficult to handle. They are hollow and form an "interlacement" continuos in the whole volume. Very often the contrast is very low or the fluorescent signal is completely absent. The membranes thickness is very small, in the best case composed by no more than $3,4$ voxels. Therefore, we adopted a different strategy, depending on the kind of data, to represent the image features.

As introduced in the previous section, in all the methods we use, the image features are given through the function g. In the case of nuclei we represent g as a standard edge indicator, i.e., as a smooth nonincreasing function of the original image gradient,

$$g(\mathbf{x}) = \frac{1}{1 + \frac{(G*\nabla I)^2}{\beta}} \,. \tag{3.7}$$

However, for the membranes g, we found useful to express it as a smooth nonincreasing

function of the image intensity, namely

$$g(\mathbf{x}) = \frac{1}{1 + \frac{(G*I)^2}{\beta}} \,. \tag{3.8}$$

In both (3.7) and (3.8) the Gaussian variance acts like a scale parameter that determines the minimal size of details that can be preserved. The parameter β is instead related to the image contrast and it acts like a scale parameter by which the graylevels of the image features are mapped into the g function.
By using (3.8), we leave the useful signal of membranes unaltered as much as possible. Indeed, despite of the fact that (3.7) strongly reduces the noise, it smooths excessively the membranes. This behaviour is particularly evident where the signal is weak and the thickness of the membranes is thin. An example is shown in Fig. 3.1. Although (3.7) removes slightly better the noise, the membranes appear blurred and part of the information is lost. We verified in [21] that such information about boundaries is important to correctly segment the membranes. Mathematical reliability of such nonlinear models for image processing where the edge indicator may depend on image intensity is given in [33].

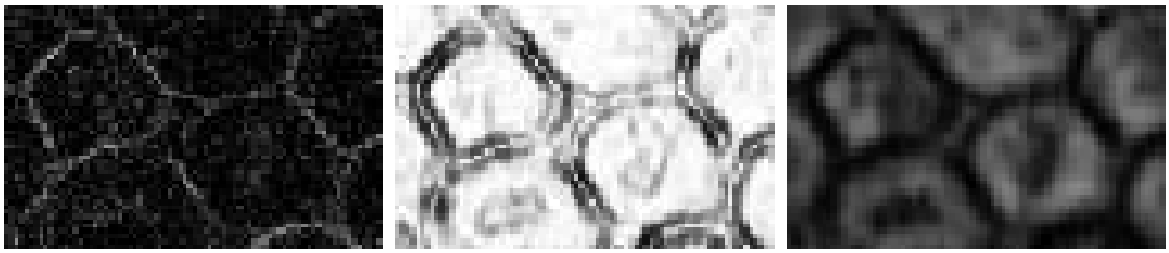
Numerical schemes for solving presented PDEs models can be based either on the explicit, see e.g. [32], or the semi-implicit [27, 34, 35, 36] time discretizations and on the finite difference, finite volume or finite element space discretizations. Since in case of explicit schemes one has to take care about the CFL stability condition, we use semi-implicit schemes, that are unconditionally stable. For the nuclei images, the implementation of such schemes is described in details in [36].

3.3 Filtering Results

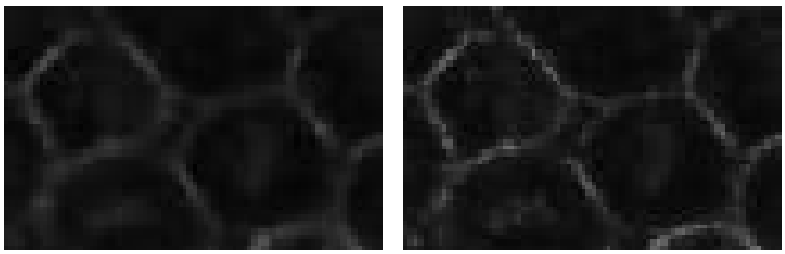
In this section we discuss visually and quantitatively the behaviour of the models on nuclei images, while for membranes we discuss the results by visual inspection. In the Fig. 3.2 we show a detail of an isosurface representation of nuclei, whereas such kind of visualization is not appropriate for membranes, where the membranes are not always closed. A detail of a slice selected in the xy plane is shown in Fig. 3.3. Observing figures, we can see that all the methods reduce the noise superimposed to

the image preserving at the same time its features. We used small values of variance σ in order to enhance the contrast and preserve the small structures of membranes. The slowed mean curvature flow and the modified Perona-Malik models show good behaviour after 10 filtering steps, both for nuclei and membranes images, while the geodesic mean curvature flow requires 15 filtering steps. In all the computations our voxel size $h = 0.01$. Then the time step $\tau = 0.0001$ (to be close to relation $\tau \approx h^2$ which is standard for solving parabolic equations). In Fig. 3.2 we used further parameters as follows: $\beta = 2, \sigma = 5 \cdot 10^{-4}$ for the modified Perona-Malik and smcf models, $\beta = 1, \sigma = 10^{-3}$ for gmcf model. For filtering the membrane images presented in Fig. 3.3 we used $\beta = 10^{-3}, \sigma = 10^{-4}$ for the modified Perona-Malik and smcf models, while $\beta = 10^{-4}, \sigma = 5 \cdot 10^{-4}$ for gmcf model.

Concerning membrane images, it is worth to note that, although the noise intensity



(a) Original data (b) Standard edge detector (c) New edge detector



(d) Data filtered using eq.(3.7) (e) Data filtered using eq.(3.8)

Figure 3.1: The original data and the edge indicators. On the bottom the data filtered using both functions.

is often comparable with the intensity of membranes and the thickness of membranes is very small, the methods are able to distinguish between the noise inside membranes and the membranes itself. Visually the behaviour of all the methods is really satisfactory and comparable. To find a way how to evaluate quantitatively the filtering by different methods for this kind of data will be an objective of our further research.

Concerning nuclei images, by the visual inspection, cf. Fig. 3.2, we can conclude that the methods strongly reduce the noise and smooth small variations in image intensity without changing the shape of nuclei. However, we would like to compare the models also quantitatively. To that goal, we have selected a subvolume of the first unfiltered frame of the embryogenesis time sequence and constructed a gold standard by a manual segmentation. Then we calculated the mean Hausdorff distance[37] between the manually segmented surface of nuclei in the gold standard and isosurfaces of original and filtered data, respectively.

Given two finite point sets, $A = \{a_1, \ldots, a_p\}$ and $B = \{b_1, \ldots, b_q\}$ the *mean Hausdorff distance* is defined as

$$MHD(A, B) = \max\left(mhd(A, B), mhd(B, A)\right),$$

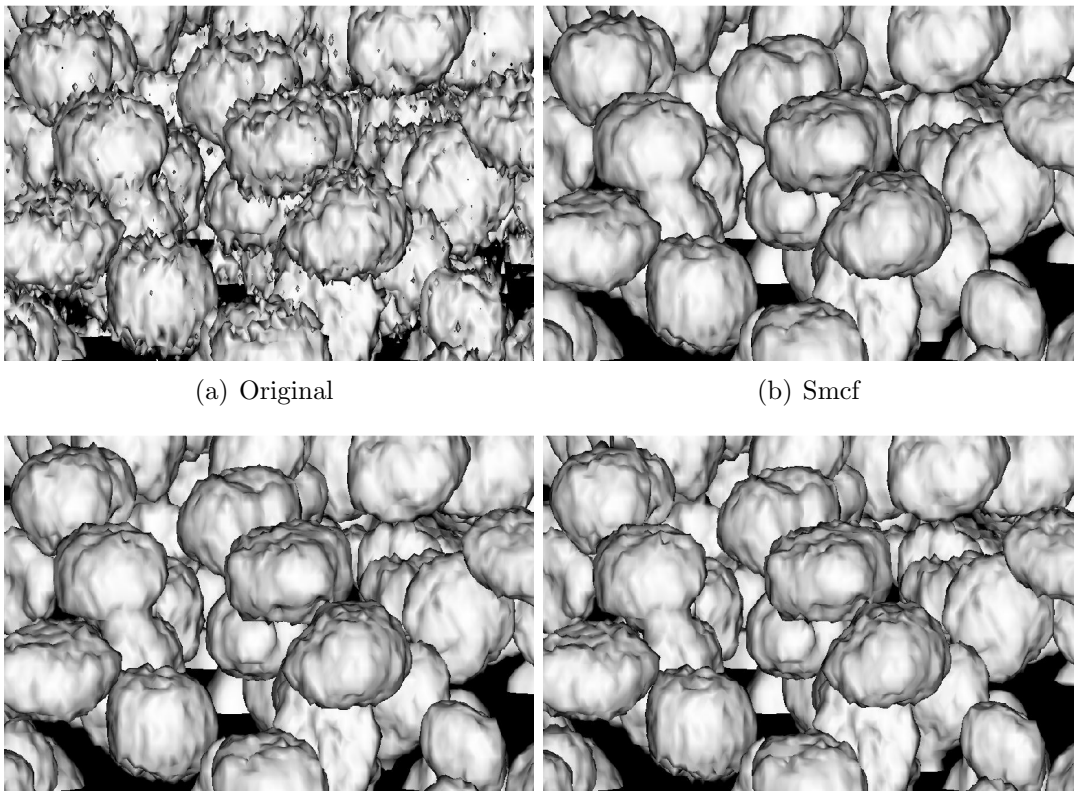
where

$$mhd(A, B) = \frac{1}{p} \sum_{i=1}^{p} \min_{b \in B} \ \|a_i - b\|$$

is called *mean directed Hausdorff distance* and $\|\cdot\|$ is some underlying norm (usually Euclidean) on the points of sets A, B. The $mhd(B, A)$ is defined similarly. The mean Hausdorff distance is widely used to measure the mismatch between two point sets usually to perform an image matching. In our case the sets A and B are given by discrete points that forms the nuclei surface in the gold standard and the image intensity isosurfaces either in original or in filtered volumes, respectively.

An appropriate image filtering should produce not only image smoothing and noise removal but also image enhancement. Then the level sets of the image intensity should accumulate around the boundaries of nuclei. The ideal image smoothing and enhancement would give a profile of image intensity perfectly steep on the nuclei boundaries.

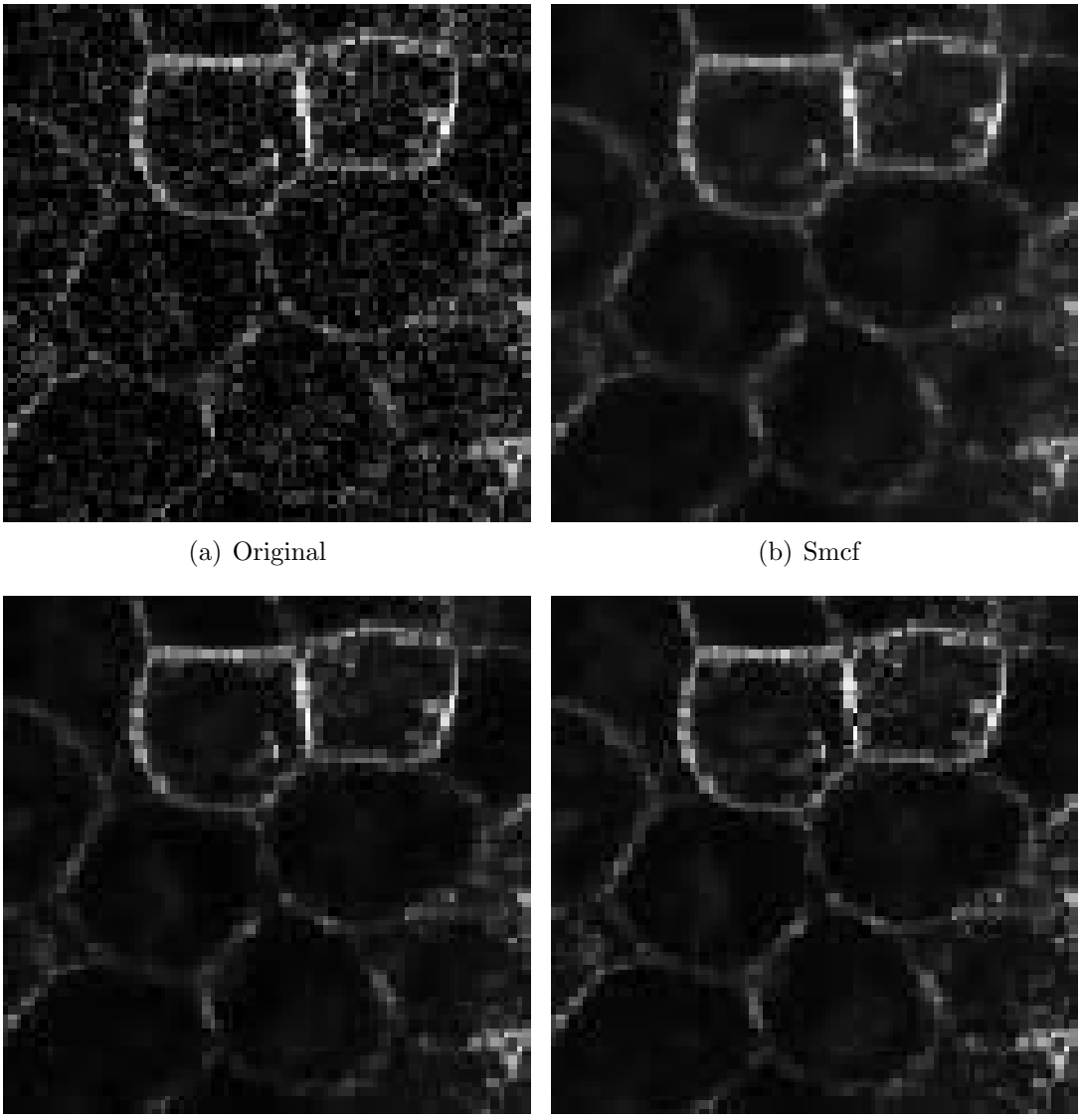
Such profile would correspond to a totally flat graph of the mean Hausdorff distance.



(a) Original (b) Smcf

(c) Mod. P-M (d) Gmcf

Figure 3.2: Isosurface representation of original and filtered nuclei (by isosurface value 28).



(a) Original (b) Smcf

(c) Mod. P-M (d) Gmcf

Figure 3.3: A slice in the xy plane of original and filtered membranes. The slice has been selected in the middle of volumes.

Therefore, by means of the mean Hausdorff distance to gold standard computed for the original and filtered data, it is possible to quantitatively evaluate the capability of smoothing and enhancement of the methods. Particularly, a reduction of the MHD by filtering shows the capability of smoothing, while the flatness of the MHD graph is related to enhancement.

Observing the image histogram and by the visual inspection we estimated the level of intensity 28 as the closest to the real boundaries of nuclei. We selected isosurface levels around this value, from 15 to 45, with step 5, in order to evaluate the mean Hausdorff distances. For every such isosurface either in original or filtered data, we found a nucleus surface in the gold standard and we calculated their mean Hausdorff distance. At the end we averaged the mean Hausdorff distances over all nuclei in the subvolume. To check correct selection of the pairs of nuclei in the gold standard and other data we visualize them overlapped within their bounding boxes, see Fig. 3.4 in case of original data. The values of the mean Hausdorff distances are reported in Table I, and Fig. 3.5 shows the related graphs. We can easily observe that our choice of isosurface value 28 as the closest to the real boundaries of nuclei has been correct. In the original data, the isosurfaces with values between 25 and 30 give the smallest (and approximately the same) mean Hausdorff distance to the gold standard. This fact is expressed in almost flat graph of the mean Hausdorff distance in this interval. It is worth to note that the interval of the flatness for original data is very narrow (in spite of the graphs for modified Perona-Malik, smcf and gmcf filtering results), which means that correct representation of nuclei in the noisy data is very sensitive to the choice of correct isosurface level. Let us note that all the methods show good capability of smoothing, the mean Hausdorff distance with respect to the gold standard is reduced in the whole range of chosen isosurfaces. The graph of the slowed mean curvature flow is almost parallel to that of original data, meaning that it performs a pure edge preserving smoothing. The graph of modified Perona-Malik method is similar, but less convex and slightly more flat for lower isosurfaces. Therefore, this method tends to more accumulate the image graylevels in that range of image intensity. The regions closely outside the nuclei contour moves towards the nuclei boundaries. This kind of behaviour is very strong in the geodesic mean curvature flow, that completely flatten the regions of a low intensity, enhancing thus edge position around nuclei. Further details and deep discussion about the evaluation of such methods using the mean

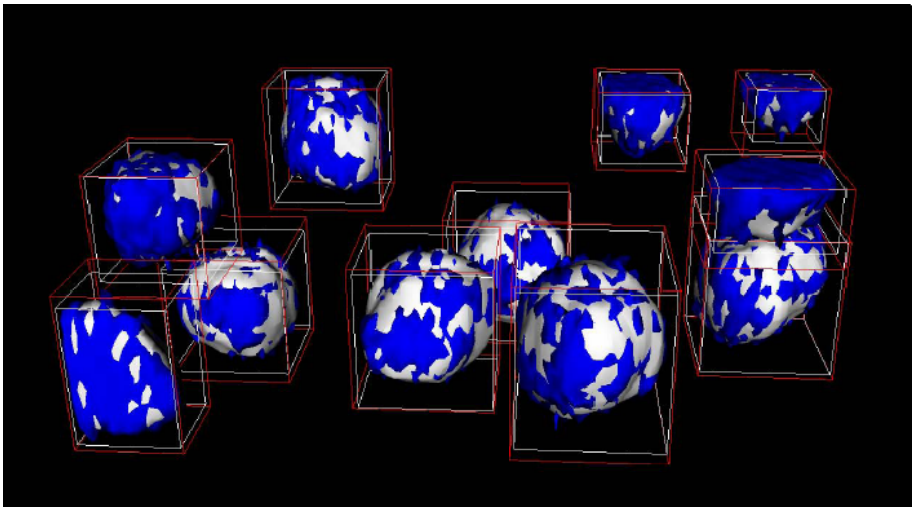


Figure 3.4: The plot of 11 nuclei of the gold standard (white) overlapped with the corresponding component of isosurface 25 in the original data set (blue).

Hausdorff distance can be found in [36].

3.4 Conclusions

In this chapter we have presented PDEs edge preserving denoising methods and we applied them to 3-D confocal images of zebrafish embryo. These images characterized by a low resolution in z direction giving rise to regions partially overlapped. Our goal was to identify the best filtering method in order to facilitate further image processing procedures, as nuclei identification, segmentation and tracking. We studied the behavior of all the methods and evaluated their performances both visually and quantitatively. The quantitative analysis has been carried out first by calculating the mean Hausdorff distance of isosurfaces to a gold standard. We showed that all the analyzed methods are able to reduce the noise, smooth small variations in image intensity of the original data and remove the spurious regions. For the Perona-Malik, slowed and geodesic mean curvature flow models, the filtering reduces the mean Hausdorff distance significantly for a large range of isosurfaces. We have introduced a modified version of classical edge detector to filter membranes data. Our study shows that nonlinear diffusion methods are well suited for processing such type of data.

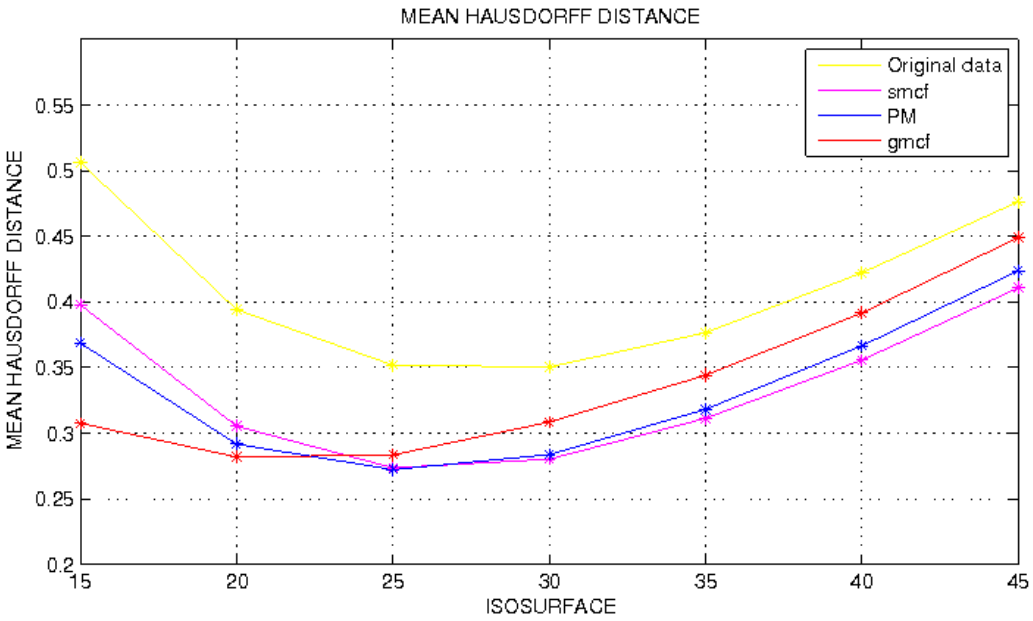


Figure 3.5: The graph of the mean Hausdorff distance for the original and filtered data against the gold standard.

Table 3.1:
Mean Hausdorff Distance

Isosurface	Original	SMCF	PM	GMCF
15	0.506	0.411	0.369	0.308
20	0.394	0.302	0.292	0.282
25	0.352	0.265	0.272	0.283
30	0.350	0.274	0.284	0.308
35	0.377	0.313	0.318	0.344
40	0.422	0.364	0.367	0.392
45	0.476	0.429	0.424	0.449

Chapter 4

Cells detection

We develop a method based on Hough transform for counting and extraction of approximate cell centers in 2D and 3D images of early stages of the zebra-fish embryogenesis. Identifying the cells in a fully automatically algorithm with minimal intervention of an operator is required in this kind of problem, given that the size and cell number is huge, hundreds of cells early in the development become thousands of cells after few hours after the egg fertilization.

The approximate cell centers give us the starting points for multiple algorithms that will allow us to achieve the automatic reconstruction of the embryo development. Finding the cells center is a crucial task in segmentation of nuclei and membranes. Also identifying each cell in the embryo is the first step to track the cells in time. Methods coming from different frameworks could be applied to detect the shape of cells in the images, topological algorithms like watershed-based methods could be used automatically although it has problems separating touching structures, level set methods have being used successfully for this task in [38], difference of Gaussians functions could identify nuclei shapes at different scales. Otsu's method [39] is used to automatically perform histogram shape-based image thresholding, morphologic operators like Chuang et al. [40], elastic deformation [41] or wavelets transformation [42] could be used to find the nuclei shape. All this methods fail when cells nuclei touch each other or signal to noise ratio is weak [43, 44].

In this chapter we will introduce the Hough transform in order to identify regular shapes in the images, the zebrafish nuclei have spherical shape that could be detected by the Hough transform. Later in this chapter experiments on real 3D embryogenesis images are presented and the results are discussed.

4.1 Hough transform to detect image features

The Hough transform is a technique which can be used to isolate features of a particular shape within an image. The classical Hough transform [45] requires that the desired features be specified in some parametric form, it is commonly used for the detection of regular curves such as lines, circles or ellipses.

To illustrate how the Hough transform works this section shows how to find straight lines in a image. The simplest case of Hough transform is the linear transform used

for detecting straight lines. In the image space, the straight line can be described as the set of (x, y) coordinates where $y = mx + b$, m is the slope parameter and b is the intercept parameter. The set of all straight lines in the picture plane constitutes a two-parameter family. In the Hough transform, the idea is to describe the characteristics of the shape in terms of its parameters. If we fix a parameterization for the family, then an arbitrary straight line can be represented by a single point in the parameter space (Hough space). In case of a straight line these parameters are the slope parameter m and the intercept parameter b. However, the slope approaches infinity as the line becomes more vertical. This problem is resolved by describing any straight line with the parametric line equation

$$x \cos\theta + y \sin\theta = \rho \tag{4.1}$$

with $\theta \in [0, \pi]$.

The *normal parameterization* specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin, see figure 4.1. The normal parameters for a line are unique, with this restriction every single line in a plane $x - y$ is mapped to a point in the plane $\theta - \rho$. To find the lines that intersect a pair of points belonging to the set $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ we transform points (x_i, y_i) into sinusoidal curves in the plane $\theta - \rho$ using the equation 4.1.

Plotting the all possible points (θ, ρ) defined by each pair of (x_i, y_i) in image space, they appear as sinusoids functions in the polar Hough parameter space. This point-to-curve transformation is the Hough transformation for straight lines. The colinear points in the plane $x-y$ correspond to sinusoidal curves in the plane $\theta-\rho$ that intersects in one point. Hence the problem of finding colinear points can be converted into the problem of finding concurrent sinusoidal curves.

The Hough transform works by letting each feature point (x, y) vote in $\theta - \rho$ space for each possible line passing through it. These votes are totaled in an accumulator array. Suppose that a particular (θ, ρ) has one vote, this means that there is a feature point through which this line intersects. If a position (θ, ρ) in the accumulator has n votes, this means that n feature points lie on that line.

To extract the corresponding parameter values of the relevants shape in the image from the accumulator we take the coordinates of the maximum values.

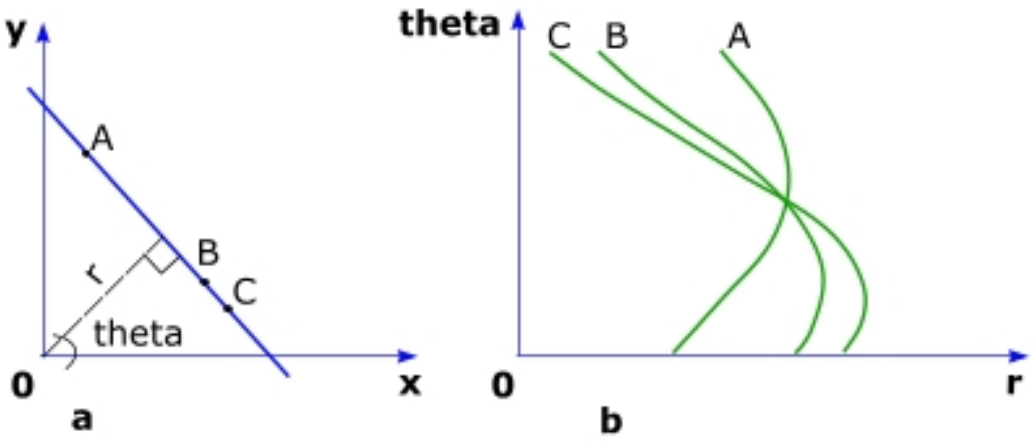


Figure 4.1: Normal parameterization of a line. (a) a straight line in the original coordinates described in terms of the length of a normal from the origin to the line - r and orientation theta; (b) the Hough plane where points A, B, and C are transformed into three sinusoidal curves

4.1.1 3D Hough transform to detect the spherical nuclei shape

The standard Hough transform and the Generalized Hough Transform are well known techniques which can be used to isolate features of a particular shape within an image. The generalized Hough transform use the analytic equation of a shape and its derivative to define a map between points in an image and a reference point in the shape. Figure 4.2 shows the procedure to detect multiple circles in images.

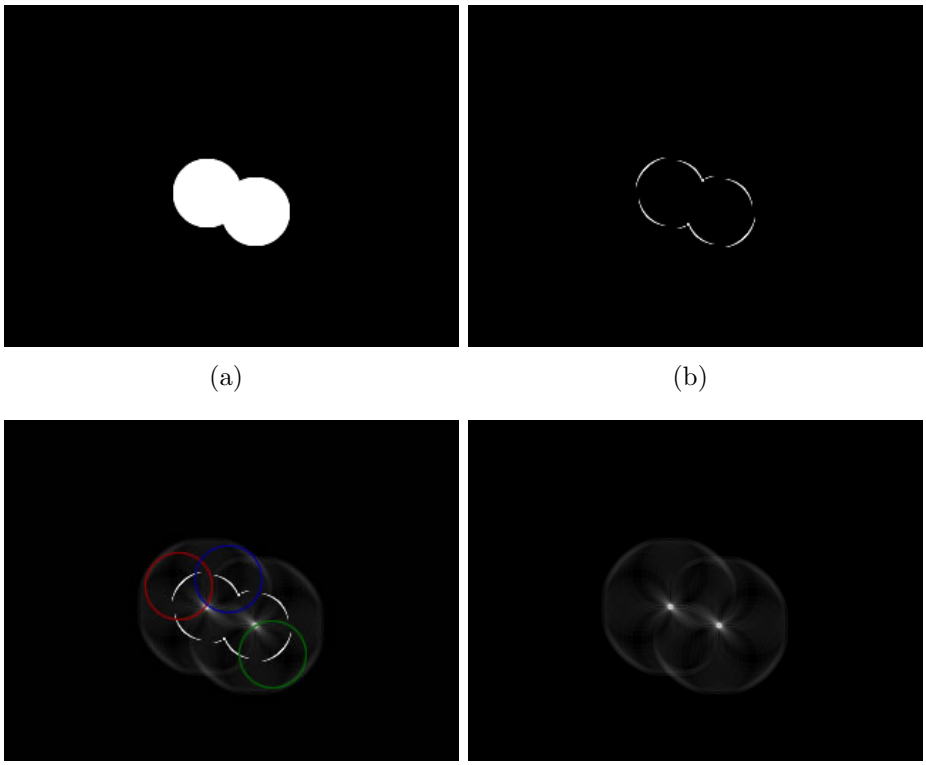
At first sight is easy to see the almost spherical shape of the nuclei of the zebrafish embryo, then the Hough transform is an excellent tool to recognize nuclei. The center of each nuclei is found using the Hough transform with the equation of the sphere (4.2).

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2 \tag{4.2}$$

where (a, b, c) is the center of the sphere with radius r. The dimension of the Hough space is 4 but if we fix the radius, the dimension is reduced to 3.

For each *edge pixel* the locus for the parameters of a sphere is another sphere centered in (x, y, z) with radius r.

The Generalized Hough transform uses the directional information associated with



(a) (b)

(c) (d)

Figure 4.2: Hough transform identifying circles shown in frame a. An edge detection filter is applied to the image (frame b). frame c show votes for random edge points detected, in frame d show the acumulator space after the voting phase. The center of the circles have correspond to the local maximum coordinates.

the edge, if we take into account that the local gradient of the image intensity will necessarily be orthogonal to the edge. Since edge detection generally involves computing the intensity gradient magnitude, the gradient direction is often found as a side effect. This reduces the computation time and has the interesting effect of reducing the number of useless votes, reduces the parameter locus to a point because the center of the sphere is located r units along the direction of the gradient.

4.1.2 Accumulator

We build a procedure that for each edge pixel e with coordinates (x_1, y_1, z_1) in the image space the accumulator is increased in the coordinates:

$$(x_1, y_1, z_1) + r(n_x, n_y, n_z) + \epsilon$$

where (n_x, n_y, n_z) is the normal vector in the point e, r is the radio of the sphere we are looking for, and $|\epsilon| < E$ Each point (a, b, c) in the accumulator represents all the spheres centered in (a, b, c) in the image space. If the accumulator in the point p with coordinates (a', b', c') is equal to k it means that k edge pixels have voted increasing the accumulator.

One interpretation of the accumulator is that each point p in the accumulator represents the probability that in the point with coordinates p in the image space there is the center of a sphere.

The algorithm can be summarized with the following steps:

- An edge detection filter is applied to the image.
- The gradient direction is found using the derivative of Gaussian operator.
- For each edge pixel (x, y, z) belong to the contour of a possible sphere, a votes are set in $r + \epsilon$ pixels in direction of the gradient.
- The local maximum represents the center of the spheres.

There is a correlation between the image space and the accumulator space. The variables in the image space are x, y, z and the variables in the accumulator space are a, b, c that represent the center of a sphere in the image space. In the figure 4.4 we

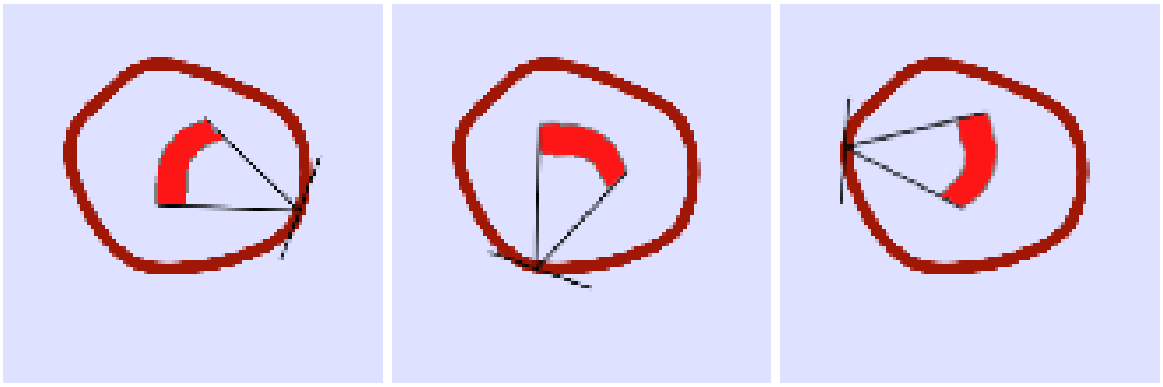


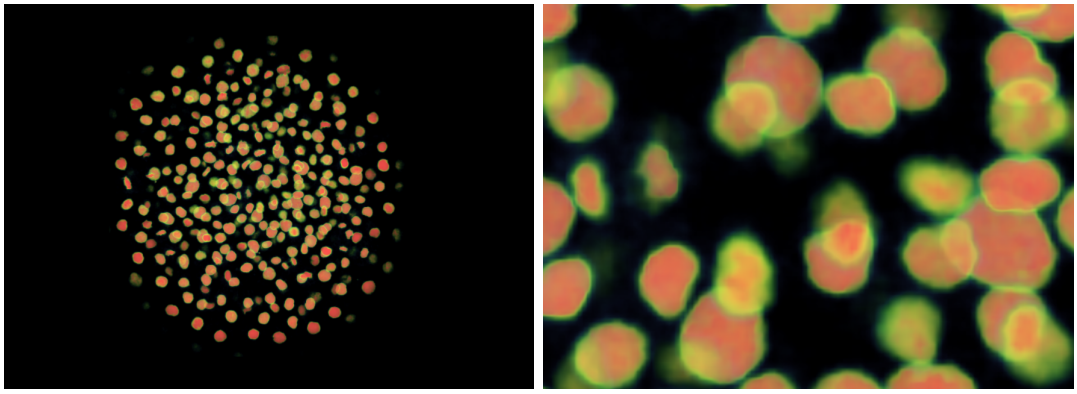
Figure 4.3: The image features vote in the neighbor of the center of the cell (2D example). The red section represents where the votes are spread in the accumulator.

can see the ray casting volume rendering of cells and the ray casting volume of the accumulator. As we can notice the positions of the higher values of the accumulator are in the center of each nuclei. In figure 4.5 there is the simultaneous rendering of the nuclei image and the accumulator. This figure shows how the higher values of the accumulator are concentrated on the center of the cell. The nuclei are in red and the accumulator array is rendered yellow and blue (higher value blue). Therefore we look for the local maximum in the accumulator array to identify the cells coordinates.

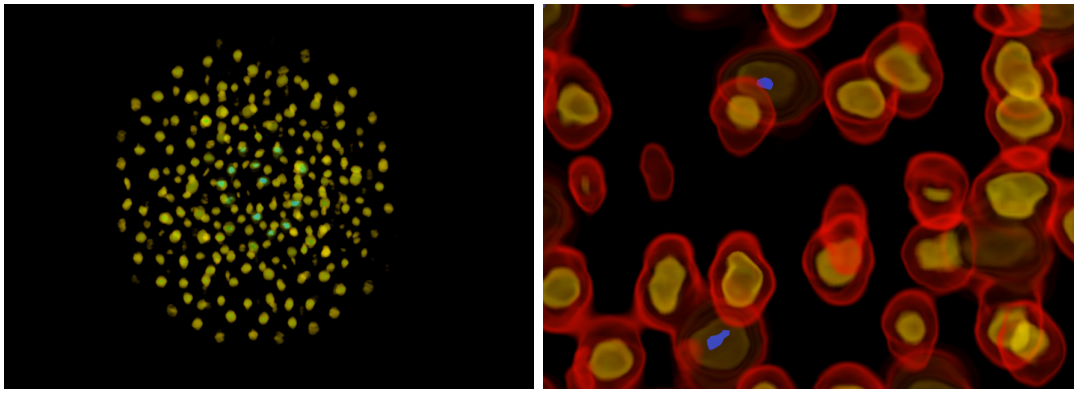
4.2 Hough algorithm

Using the concepts of the previous sections we can sketch a procedure (algorithm 1) based on the Hough transform to find spheres in a 3D image.

The input parameters of this algorithm are: the 3D image volume, the size of the cells we want to find and the variance of the Gaussian kernel. The output of the



(a) Top view of the embryo. Each figure with different zoom.



(b) Top view of the accumulator array. Each figure with different zoom.

Figure 4.4: Top: Ray casting volume of the embryo. Bottom: Ray casting volume of the accumulator.

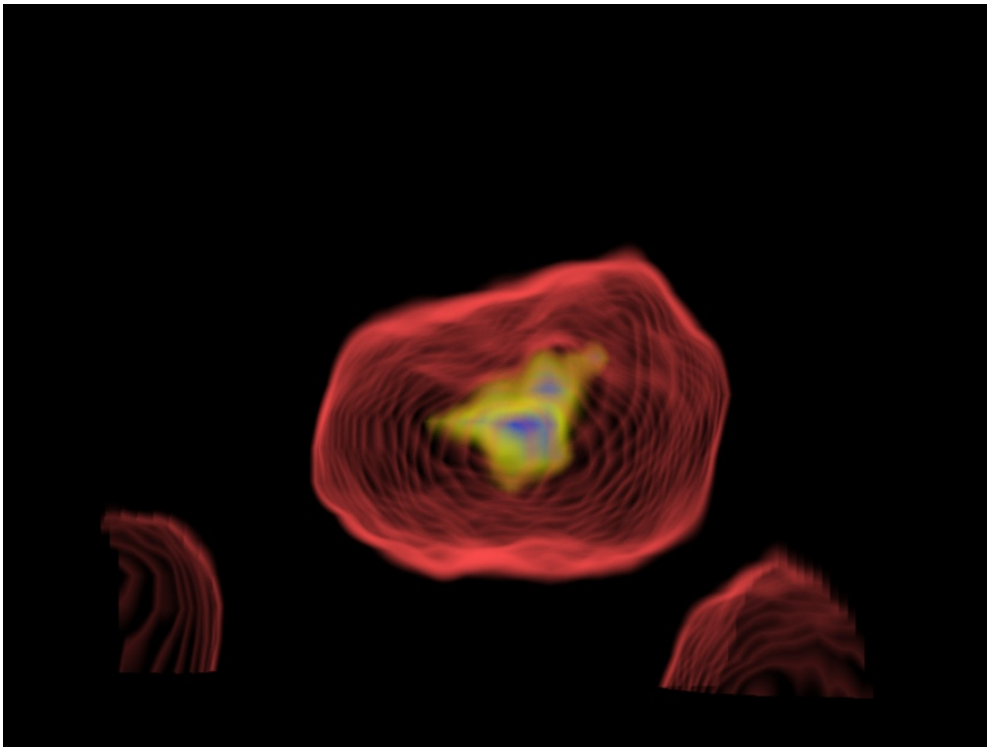


Figure 4.5: Ray casting volume of the embryo and the accumulator. Both channels are rendered with independent opacity and transfer functions. In red we can see the nuclei, in yellow are the lower values of the accumulator, in blue are the higher values of the accumulator.

algorithm is a list of 3D points with the location of the center of the spheres.

input : 3d image, I
input : Set of Radius, $[r_{min}...r_{max}]$
input : Variance, σ^2
output: Set with the centers of the spheres, *sphereSet*

```
Accumulator ← 0;
foreach radius r ∈ [r_min...r_max] do
    foreach edge pixel e with coordinates X∈ I do
        gradient ← Derivative of Gaussian operator(e,x);
        foreach ϵ where |ϵ| < E do
            Accumulator[X + r * gradient + ϵ] + +;
        end
    end
end
sphereSet ← Extract Local Maximum(Accumulator);
```

Algorithm 1: Hough algorithm.

Another important task in this algorithm is the stopping criteria when we look for the local maximum. Several alternatives can be considered. If we do not know how many cells there are in the volume the algorithm can stop when the local maximum is below a threshold or is estable. If we know how many nucleus we are looking for, we can iterate until that number is achieved.

4.3 Results

This section shows the results of the nuclei recognition using the Hough transform to detect spheres in the 3d image data.

Figure 4.6,4.7 shows the ray casting volume rendering of dataset 070420a. A mark with a synthetic white sphere is rendered where a local maximum is found in the accumulator of the Hough space. Lets notice the good location of each mark, all of them are almost in the center of each nucleus and there are not many mistakes (unrecognized nucleus or multiply-marked nuclei). Even the cells which are in the bottom of the volume where only a small portion of the nuclei are visible are recognized. Figure 4.8 shows the number of cells at different stages in the embryo, as not all the embryo is

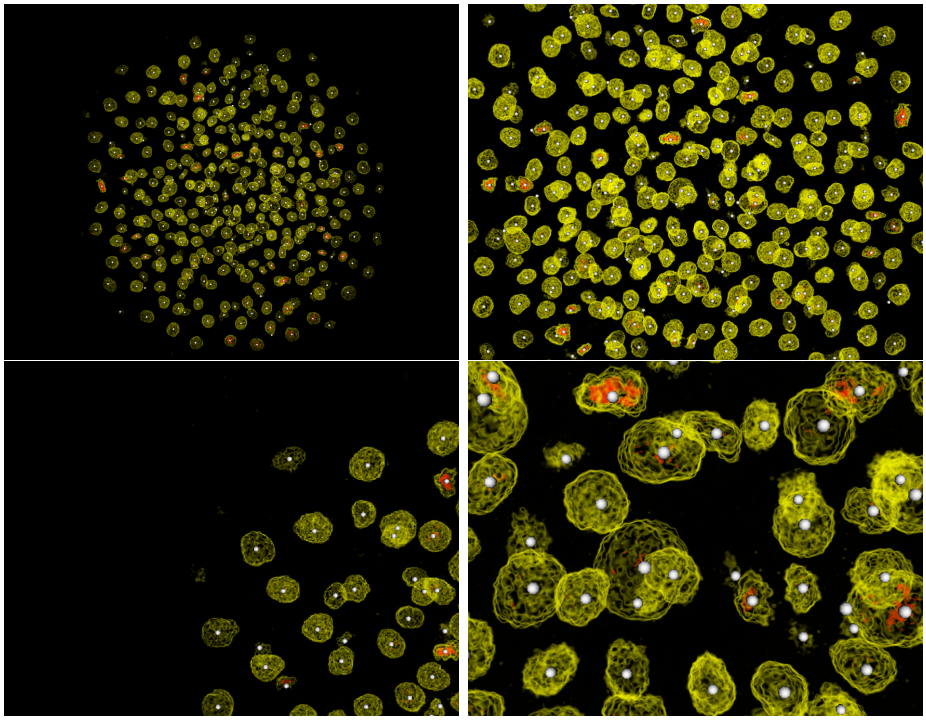


Figure 4.6: Nuclei recognition result using the Hough transform in dataset 070128c. A synthetic white sphere is draw where there is a local maximum in the accumulator. The images are rendered using 3D ray casting at different zoom factor.

capture by the microscope, cells move, enter and exit from the images, therefore the number of detected cells not always grows.

Counting the number of cell in the embryo would give us the proliferation rate or number of cells division (figure 4.9) by time. This could be used to show the synchronization in space-time of cell division of neighbour cells. Also with the cell position and number it is possible to calculate the cell density in the spatio-temporal space (figure 4.10).

In order to validate our object counting algorithm, we performed a manual center detection on a sub-volume of a nuclei image and we tested the correctness of the result obtained by automatic detection. Considering the further applications of the center detection, there are three basic types of errors that can occur:

- False positive detection is the case when the algorithm finds a center that does not correspond to any important object in the image. This happens especially

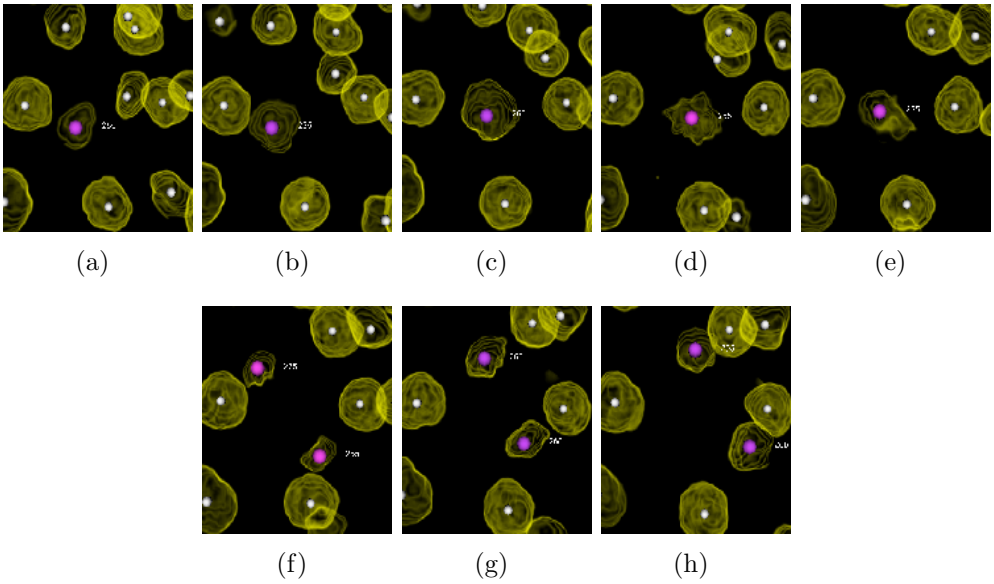


Figure 4.7: Result of cell identification process in dataset 070128c. The white marks are the cells recognized with the Hough transform in consecutive timesteps. The label 255 indicates a single cell trajectory that divides.

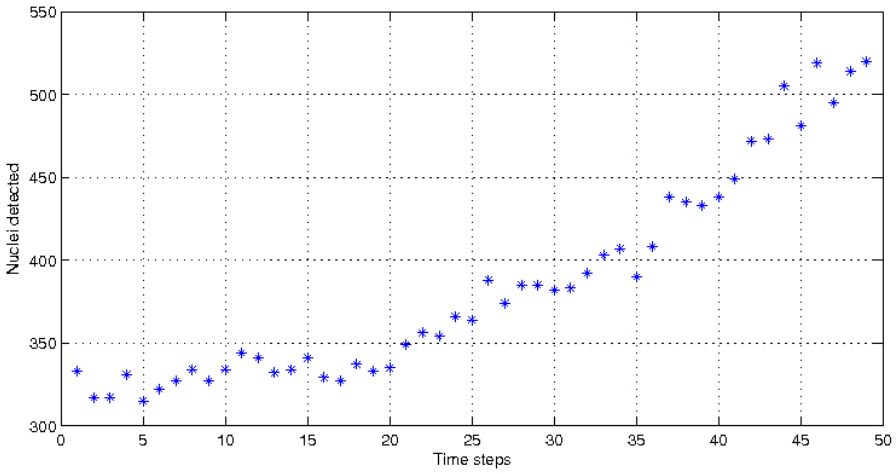


Figure 4.8: Number of nuclei detected in the time lapse series. Dataset 070128c. Not all the embryo is capture by the microscope, cells moves, enter and exit from the images, therefore the number of detected cells not always grows.

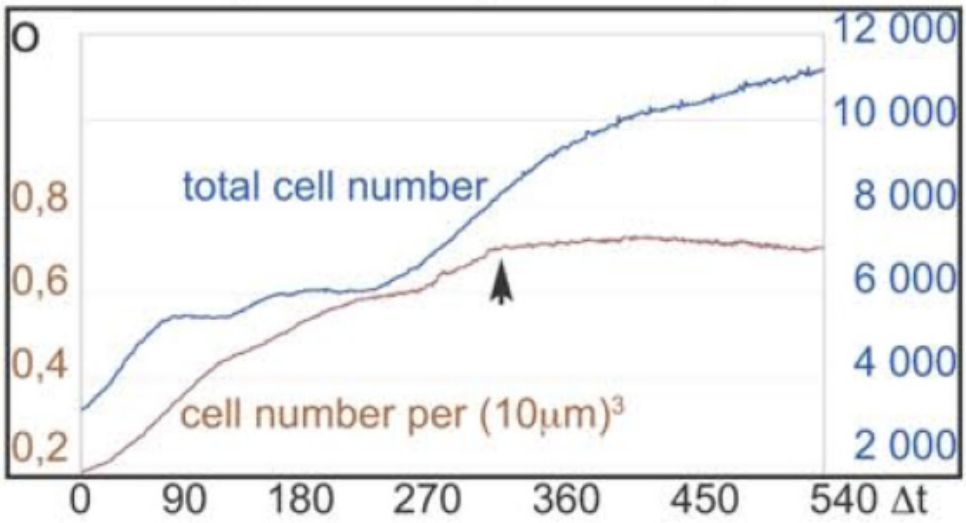
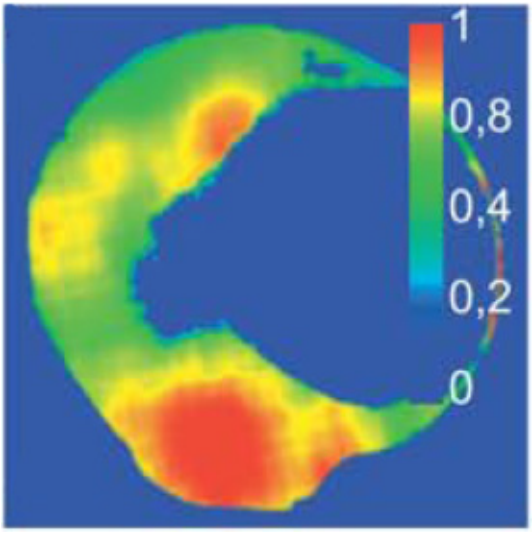


Figure 4.9: Plot of the total cell number (blue curve) and cell density (brown curve) as a function of time (dt=67s), arrow points to the stabilization of cell density by early somitogenesis. Data set 070418a. [46]



(a)

Figure 4.10: (a) data set 070418a, local cell density in the whole volume at t479, z=104,[46]

in large epithelial nuclei that can contain inner structures that cause detection of more than one center in a single nucleus.

- Another false positive case is detection of a center that corresponds to some artifact or noise structure.

- Finally, false negative detection is the case when an object was not detected. This is usual for small nuclei of very low intensity.

In order to measure the quality of our automatic object detection, we evaluated the number of false positive (FP) and false negative (FN) cases and the average distance $\overline{D}$ of an automatically detected center to the corresponding center found manually. For a small subvolume, the algorithm found 70 out of 75 nuclei, with 2 false positives. The false negative detections appear only on the border of the subvolume or deep on the embryo.

Higher rate of false negative cases can occur in the parts with weaker staining. Discovering or correcting false negative detections would be much more difficult than correction of the false positives. Therefore, for data with varying level of staining, the algorithm must be set up in a way that avoids production of false negative detections, even if it often means to accept a certain number of (usually correctable) false positives. In practice it means that the stopping criterion must be adjusted, making the algorithm stop at an earlier when finding local maximum.

After 7 or 8 hours of development, the cell density and cell size, make the individualization of cell very difficult, nuclei seams to stick together and errors like false negative or double detection of the cell increase.

4.4 Conclusions

An essential information concerning the process of embryogenesis is the number and position of cells in each time step of the image sequence. In this chapter we developed an algorithm based on the Hough transform to detect the spherical shape of cell nuclei. We presented how this algorithm could be used to identify the 3D coordinate and count the number of cell at each time step. The cell center detection method that we described provides a good estimation of this number. Counting the number of cells at

each stage of the embryo can provide some important quantitative characteristics of the embryo. We can calculate its global or local density of cells or cell division rate. All this information, is precious for measuring the differences between individuals. Identifying the cell position is a crucial task to perform the segmentation of membranes and nuclei, and to perform the cells tracking.

Chapter 5

Cells segmentation

The 3-D reconstruction of cellular shape is a crucial task for reaching an integrated understanding of biological processes leading to organism formation. Providing automated procedures for reconstructing the shape of all the cells of a living vertebrate embryo is far beyond the current state of art. Achieving such a goal would readily provide measurements for a large number of biological features including cell shape changes and deformation characteristic for cell differentiation and tissue morphogenesis. Cell shape segmentation is also essential to track cell divisions and help reconstructing the cell lineage tree and from that extract the cell proliferation rate in space and time. This kind of data is highly relevant to investigate stem cell populations, early steps of cancerogenesis and drug effects in vivo. Furthermore, the reconstruction of the cellular shape will provide relevant parameters to measure the variability between different individuals of the same species, opening the way for understanding the individual susceptibility to genetic diseases or response to treatments. In this context, our aim is to design an algorithm achieving an automated segmentation of nuclei and membranes from 4-D imaging of live embryos engineered by fluorescent markers. Although interactive methods have better performances (in terms of the percentage of objects correctly segmented), we expected to avoid the need for any manual intervention that becomes unrealistic when manipulating millions of objects. The segmentation technique has to be chosen according to the data features. Typically, 4-D images for living organism provide incomplete information such as objects with missing boundaries and the segmentation technique should deal with that. Many algorithms for the shape reconstruction have been developed by researchers worldwide, and there exist almost as many segmentation methods as there are segmentation problems. The 2D and 3D automatic or semi-automatic nuclei segmentation has been covered in a number of previous works [47, 48, 49, 50, 51, 52, 53, 54, 32]. All the developed algorithms have proved to be very useful for nuclei segmentation, however the reconstruction of the whole cell using membrane protein markers is almost an unexplored area. In a previous work by Sarti et al. [32] confocal microscopy images were processed to extract the shape of nuclei. However, in that case, the analyzed volumes were not acquired from a living organism but from pieces of fixed tissues. On the contrary, the analysis of biological processes during embryogenesis means analyzing the cells within their natural environment, i.e., in a living embryo. In that case, segmentation has to proceed from 4-D data whose quality is much more difficult to handle. Ortiz et al. [55] presented a segmentation

algorithm based on gradient-curvature driven flow, which is suitable for whole cell segmentation. They measured the robustness against noise and resistance to surface discontinuities on synthetic images and demonstrated the suitability of the method on real cell images. However, as they discussed, the resistance to surface discontinuities is strictly dependent on a parameter introduced in curvature term that determines the strength of the regularization. This pose a trade-off choice between surface accuracy and missing boundaries filling that should be solved by the user. Here we present a method to segment a large number of cells from 3-D images characterized by non homogeneous intensity and gradient signal and capable to complete surface discontinuities without any compromise between precision and ability to integrate the uncompleted contours. The segmentation method we propose in this work is a generalized version of the Subjective Surfaces technique [56, 57]: it is distinguishable from the classic formulation by the different weights applied on the two flows constituting the motion equation (curvature and advection). In addition, two different dynamics constitute the same segmentation process: by acting on the matching of level curves, we control the evolutive behavior in order to make it first mostly diffusive then a level set motion. In the biological application we deal with, these strategies are fundamental for reaching satisfactory results, as preliminarily shown in [21]. Here we expose more widely the same base concepts, but including a study on the stability condition, a preliminary algorithm validation and an overview on future developments. The different sections of this chapter follow the steps undertaken to analyze the 3-D confocal images (Fig.5.1). In Section 5.1 we briefly explain the technique for image acquisition. In Section 5.2 we apply a filtering method for image denoising. In Section 5.3 we describe the segmentation algorithm and in Section 5.4 we show some meaningful results. Finally, in Section 5.5 we describe a possible method for the algorithm validation reporting the outcomes of its application on few cells with different shapes.

5.1 Image Acquisition

5.1.1 In Vivo Imaging Technique

In vivo imaging is becoming an increasingly powerful tool for the analysis of morphodynamical patterns in biology. Microscopic imaging, taking advantage of fluorescent

proteins engineering, is able to achieve a resolution at the sub cellular level in a whole living organism, to analyze biological circuits dynamics and quantify molecular components. To obtain accurate measurements of 3-D features at the cellular level in living embryos, it is necessary to use an acquisition technique with micrometrical resolution, able to reconstruct volumetric information and with enough contrast to allow segmentation of individual cells. To fulfill these requirements, the analyzed images have been acquired by confocal microscopy (CLSM) or by multiphoton laser scanning microscopy (MLSM) with the best compromise in terms of spatial and temporal resolution [1].

5.2 Image Denoising

The noise present in the image can disrupt the shape information, therefore denoising is an essential preliminary task in images segmentation. Noise has different sources such as a non homogeneous concentration of fluorescent proteins in the labeled structures or the electronic noise from the instrument.
In order to accurately reconstruct the object shape, the denoising method has to improve the signal-to-noise ratio, faithfully preserving the position of the boundaries that define the shape of the structures. The *geodesic curvature filtering* [32, 36, 58] is able to achieve this task, by evolving the 3-D image I according to the following equation:

$$I_t = g|\nabla I|K + \nabla g \cdot \nabla I, \tag{5.1}$$

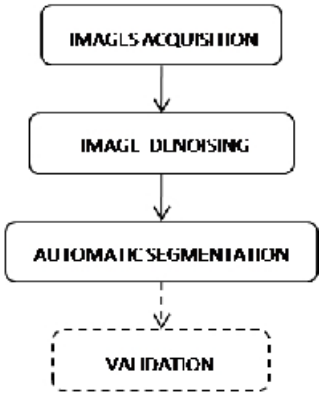
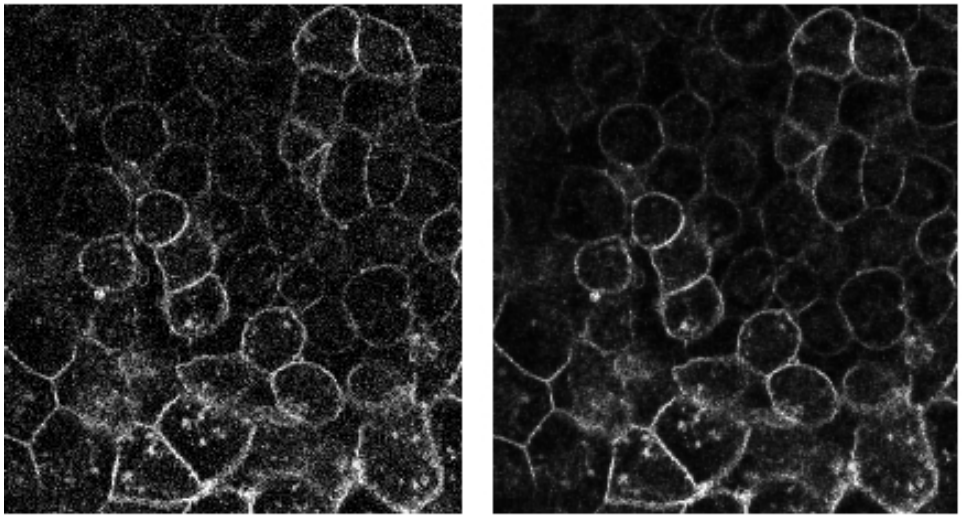


Figure 5.1: Flowchart depicting the sequence of steps we undertook for nuclei and membranes segmentation.

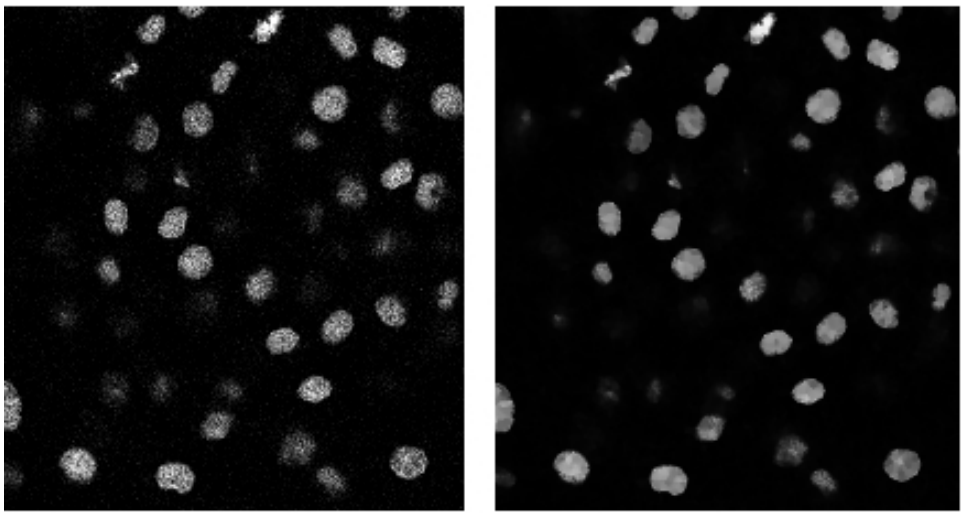
where $K = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|}\right)$ is the mean curvature of level sets of I and $g = g(I)$ is a nonincreasing edge indicator function dependent on image features.

5.3 Algorithm For Cell Segmentation

In this section we introduce our segmentation algorithm through a detailed description of the steps leading to the shape reconstruction of an entire volume of cells. The algorithm allows extracting all the membranes and nuclei shapes in the acquired images by processing the two channels separately. Composed by a chain of image processing operations, the method is completely automated and represents a possible solution to overcome the limits of *apriori* knowledge or manual intervention in segmenting objects of different shape and with missing contours. Our attention is in particular focused on membranes segmentation. As preliminarily mentioned in introduction, the membranes signal is often low or even completely absent, giving rise to images with nonuniform intensity and incomplete objects. Furthermore, as can be observed in figure 5.2a, the useful signal consists in boundaries between adjacent cells and, even when present, is composed by very thin structures. Missing boundaries could be completed by using the Geodesic Active Contour method [28], but the technique greatly depends on the algorithm initialization: at the starting point, the reference level has to be an approximation of the final contour. An interesting solution, that does not require any a priori knowledge about the edges topology, has been introduced in [54] and consists in the use of a Malladi-Sethian approach [59]. Every membrane is segmented using a level-set function initialized in its center and then expanded by a balloon term. The missing boundaries are completed by a manually chosen different weight between the regularization and expansion term. As we are dealing with thousands of cells the user intervention is, in our case, unfeasible. And if the weight term is automatically chosen, the method is often not able to correctly detect the membranes boundaries. An example is shown in figure 5.11. We propose to use a different technique, based on the Subjective Surfaces [56, 57] model, in order to correctly reconstruct the shape of membranes without any manual intervention. The method of Subjective Surfaces has been introduced [56] to segment objects, as the Kanizsa triangle, perceived by human eyes but characterized by a wide absence of information on boundaries. Such peculiarity makes the model especially suitable for this particular application. Besides, we would



(a)



(b)

Figure 5.2: Details of filtering results (on the right) in comparison with the original images (on the left) for membranes (a) and nuclei (b) xy slices. The noise is greatly reduced and the image contrast is enhanced thanks to the intensity level sets accumulation around the boundaries promoted by the geodesic curvature technique.

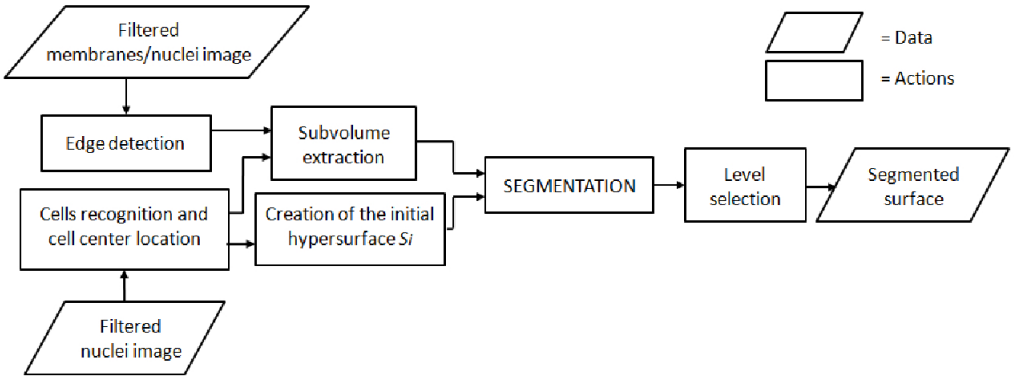


Figure 5.3: Procedural scheme for the segmentation of a single object (i.e. membrane/nucleus).

like also to point out that the Subjective Surfaces can be successfully applied also to nuclei segmentation, automatically solving problems related to nuclei sometimes clustered as the resolution of microscope is not able to distinguish them. In the proposed procedure, each object is processed separately from the others limiting the computation to subvolumes containing only one cell. This structure greatly simplifies the code parallelization, because it allows subdividing the volume in blocks of few cells, sending blocks to different processors for computation and then collecting all the segmented surfaces as a single result. The image processing chain given by the steps followed for the segmentation of a single cell is schematized in Fig. 5.3. The main algorithm feature is the self-action: it has been designed to completely avoid user intervention, because the segmentation of thousands of cells is achievable only if it is fully automated. In order to explain our procedure, we first introduce the classic Subjective Surfaces model. As the method requires the construction of a metric based on the image features, we introduce a classic and a new edge detection function, depending on the kind of managed data (membranes or nuclei), necessary to construct such metric. We then show how an initial reference point, still required for segmentation, is chosen applying the generalized 3-D Hough transform [60] and the method is initialized starting from that point. The section proceeds with the introduction of a modified version of the classic Surfaces method and the explanation of used numerical schemes. It is concluded by showing our results.

Our segmentation algorithm has been implemented using the programming lan-

guage C++ and the ITK [61] VTK [62] libraries for treatment and visualization of medical images.

5.3.1 Subjective Surfaces method

The method of Subjective Surfaces, as introduced in [57] and then improved in [63], consists, in the 3D case, in the volume minimization of a 3-D manifold embedded in a 4-D Riemannian space with a metric constructed on the image itself. Let us consider the volumetric image $\mathcal{I} : (x, y, z) \rightarrow I(x, y, z)$ as a real positive function in some domain $M \subset R^3$ and its low level locale features given by a function $g = g(x, y, z)$. Such function is used to construct a Riemannian metric h in R^4 that will be used as embedding for a 3-D hypersurface evolution:

$$h = \begin{pmatrix} g & 0 & 0 & 0 \\ 0 & g & 0 & 0 \\ 0 & 0 & g & 0 \\ 0 & 0 & 0 & g/a \end{pmatrix} \tag{5.2}$$

Starting from a reference point of view selected in the center of object to be segmented, is then constructed, in the image domain M, an initial function $\Phi = \Phi(x, y, z)$, usually a distance or a peak function. The graph of Φ represents a 3-D manifold $S = (x, y, z, \Phi)$ embedded in (R^4, h) whose volume is represented by

$$V_g = \int_M \frac{g(x, y, z)}{\sqrt{a}} \sqrt{a + \Phi_x^2 + \Phi_y^2 + \Phi_z^2} dxdydz \tag{5.3}$$

The hypersurface S is evolved afterwards to minimize its volume. The motion equation

$$S = H\nu \tag{5.4}$$

represents a mean curvature flow for S, as H is the mean curvature of S in (R^4, h) and ν its inner normal. During the evolution S is attracted towards existing boundaries and smoothed in uniform regions, inside objects. The hypersurface level sets accumulate where exists information, developing discontinuities and completing, by geodesics, regions that are continuation of existing edges. The condition of minimum for V_g cor-

responds to a piece wise constant solution of motion equation (5.4) and easily allows the selection of a level set of S to segment the object. The parameter a introduced in the metric is a stretching factor and represents a weight between two different dynamics, a diffusive or pure level set motion. For a wide treatment about the meaning of this parameter we refer the reader to [63]. A bidimensional example of evolution of hypersurface S to segment a membrane is shown in figure 5.6. Further details are given in next sections.

5.3.2 Low level features extraction: edge detector

The initial task in Subjective Surfaces segmentation is to build a metric on the image low level local features. For such purpose a classic solution is to consider an edge indicator $g = g(x, y, z)$, a smooth nonincreasing function of the image gradient [25]:

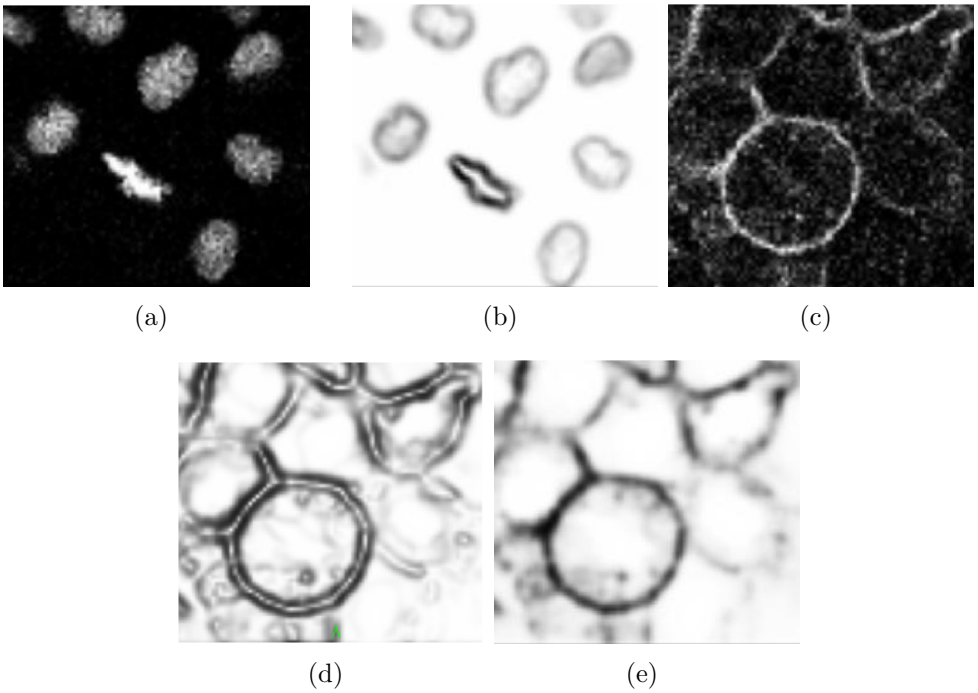
$$g(x, y, z) = \frac{1}{1 + (|\nabla G_\sigma(x, y, z) * I(x, y, z)|/\beta)^n} \tag{5.5}$$

where $G_\sigma(x, y, z)$ is a Gaussian kernel with standard deviation σ, $*$ denotes the convolution and n is typically 1 or 2. The parameter σ determines the minimal size of details that can be preserved, whereas β is related to the image contrast and acts as a scale factor by which the image graylevels are mapped into the g function. The value of g is close to 1 in flat areas ($|\nabla I| \to 0$) and close to 0 in the regions where image gradient is high (i.e. edges). Thus, the minima of g denote the position of the edges and its minus gradient is a force field that can be used to drive the evolution, because it always points in the local edge direction. The analyzed signals (membranes vs nuclei) behave in a completely different way in terms of edge detection: nuclei are solid and well contrasted objects; membranes are hollow, with a thickness of about 3 to 4 voxels and adjacent to each other. In nuclei images, the contours to be segmented are located in the regions where image gradient is higher and the minima of (5.5) denote the position of the edges (Fig. 5.4(b)). On the contrary, the function (5.5) reveals a double contour, on the internal and the external side of the cell (Fig. 5.4(d)), stopping the hypersurface on the internal cell boundary. These specific features require using different functions for the detection of the edges in nuclei and membranes images. In order to locate the minima of g in the middle of membranes thickness, we propose

an alternative edge indicator, using the image itself (not its gradient) as a contours detector. The edge indicator we propose is:

$$g(x,y,z) = \frac{1}{1 + (|G_\sigma(x,y,z) * I(x,y,z)|/\beta)^n} \tag{5.6}$$

As expected, its minima locate the contours in the middle of the membranes thickness (Fig. 5.4(e)). We would like also to point out that we used the expression (5.6) in the denoising of membranes images, because (5.5) is stronger in noise removal but it also excessively blurs image contours, loosing part of the signal if it is weak or thin, whereas (5.6) better preserves the useful information on boundaries [58].



(a) (b) (c)

(d) (e)

Figure 5.4: Details of (a) nuclei and (c) membranes original data and the edge indicators obtained by applying the standard formulation of g, (b) and (d) respectively. (e) is the alternative edge indicator defined in order to detect a single contour in membranes images.

5.3.3 Hypersurface initialization

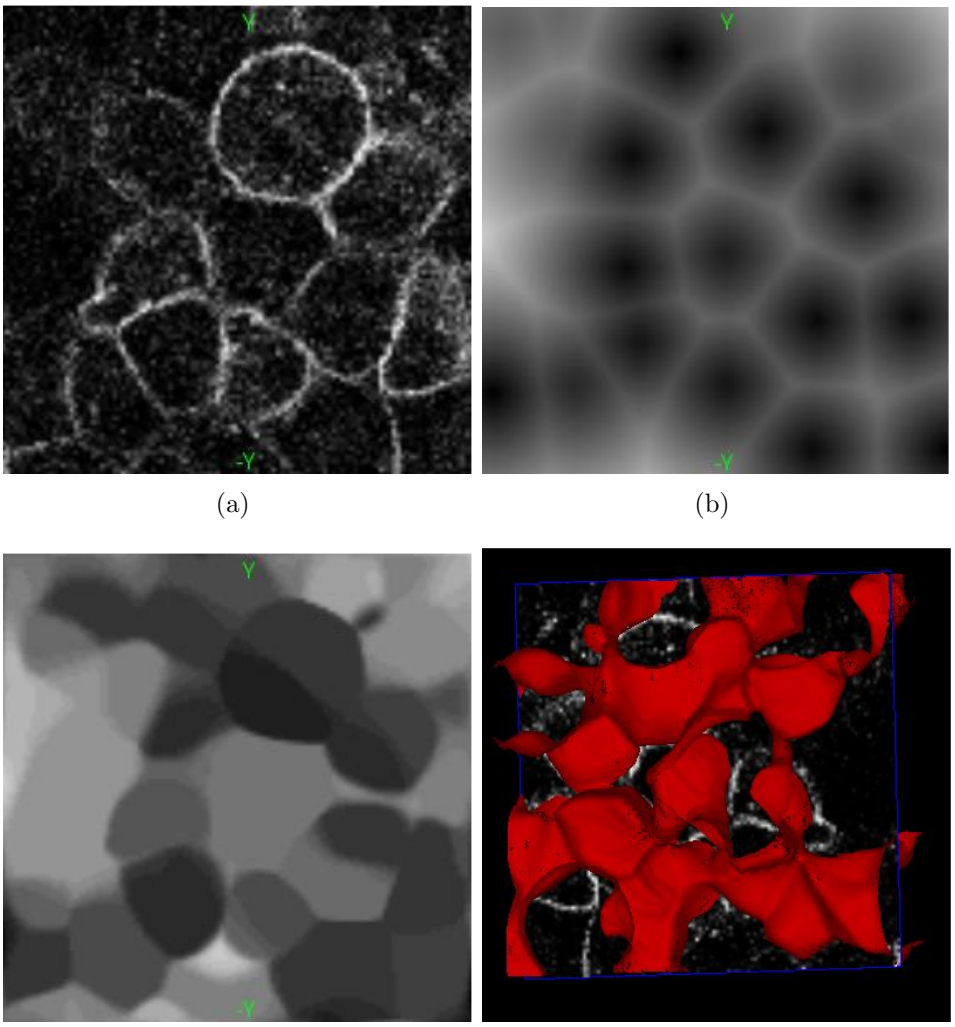
Starting from every reference point detected via the generalized 3-D Hough transform we now construct an hypersurface in (R^4, h) by defining a Φ function in the image domain M. There are some alternative forms for Φ, for example $\Phi = -\alpha\mathcal{D}$ or $\Phi = \alpha/\mathcal{D}$, where $\mathcal{D}$ is the 3-D distance function from the reference point and α is a constant. We used the initial function $\Phi = \alpha/\mathcal{D}$, instead of $\Phi = -\alpha\mathcal{D}$, to have a higher contrast in the processed image. The same expression of S_i can be employed both for nuclei and membranes processing.

We would like to observe that the segmentation of multiple objects within the same volume can be achieved either by the simultaneous segmentation of all the objects, defining only one initial function for the whole image, in a similar way as done in [32] for nuclei segmentation, or by processing sequentially each cell with its own initial function and collecting the intermediate segmentations after their computation. In the early stage of our work, we did some tests applying the first option (Fig. 5.5). This method gave good results only for non-contiguous objects, but failed in case of multiple contact regions, meaning that it was fine for nuclei but not for membranes. Indeed, membranes of closely interacting cells look fused and neighboring cells are merged into a single object by the segmentation process (Fig. 5.5(d)). In addition, the computation of large volumes is costly in time and memory, thus requiring parallelization of the code. For these reasons, we chose the second option both for membranes and nuclei segmentation.

5.3.4 Segmentation: Modified Version of the Subjective Surfaces Technique

Let us still consider equation (5.4) and rewrite it with respect to the Φ function introduced in previous subsection:

$$\Phi = H|\nabla\Phi|. \tag{5.7}$$



(a) (b)

(c) (d)

Figure 5.5: Segmentation results obtained using a single initial function for the whole volume: (a) original image, (b) initial function, (c) processed function and (d) isosurface.

The mean curvature of the hypersurface S, H, is given by:

$$\begin{aligned} H &= \frac{(a^2+\Phi_x^2+\Phi_y^2)\Phi_{zz}+(a^2+\Phi_x^2+\Phi_z^2)\Phi_{yy}+(a^2+\Phi_y^2+\Phi_z^2)\Phi_{xx}}{(a^2+\Phi_x^2+\Phi_y^2+\Phi_z^2)^{3/2}}+ \\ &-2\frac{\Phi_x\Phi_z\Phi_{xz}+\Phi_x\Phi_y\Phi_{xy}+\Phi_y\Phi_z\Phi_{yz}}{(a^2+\Phi_x^2+\Phi_y^2+\Phi_z^2)^{3/2}}+ \\ &+\frac{g_x\Phi_x+g_y\Phi_y+g_z\Phi_z}{(a^2+\Phi_x^2+\Phi_y^2+\Phi_z^2)^{1/2}} \end{aligned} \tag{5.8}$$

Setting

$$K = \frac{(a^2+\Phi_x^2+\Phi_y^2)\Phi_{zz}+(a^2+\Phi_x^2+\Phi_z^2)\Phi_{yy}+(a^2+\Phi_y^2+\Phi_z^2)\Phi_{xx}}{(a^2+\Phi_x^2+\Phi_y^2+\Phi_z^2)^{3/2}}+ \\ -2\frac{\Phi_x\Phi_z\Phi_{xz}+\Phi_x\Phi_y\Phi_{xy}+\Phi_y\Phi_z\Phi_{yz}}{(a^2+\Phi_x^2+\Phi_y^2+\Phi_z^2)^{3/2}} \tag{5.9}$$

and adding two different parameters, λ and ν, we can explicitly write our model equation:

$$\begin{cases} \Phi_t = \lambda g K|\nabla\Phi| + \nu\nabla g\cdot\nabla\Phi & \text{in } M\times]0,T[\\ \Phi(x,y,z,t) = \min(\Phi_0) & \text{in } \partial M\times]0,T[\\ \Phi(x,y,z,0) = \Phi_0 & \text{for } (x,y,z)\in M \end{cases} \tag{5.10}$$

Let us notice that, differently from the classical formulation of the Subjective Surfaces technique, we assigned different weighting factors to the first and second term on the right side of Φ_t expression in (5.10). The term K still represents a mean curvature for S, but with a metric h (cf. eq.(2)) in which the edge detector is equal to 1. Our model equation can then be read as in the following. The first term on the right side of (5.10) represents a mean curvature flow, a parabolic motion that evolves the hypersurface in normal direction with a velocity given by the mean curvature K and weighted by the edge indicator g. The second term is a pure passive advection along the velocity field $-\nabla g$, whose direction and strength depend on position. This term attracts the hypersurface in the direction of the image edges.

Locally, different behaviors can be identified in the image regions according to one of these flows. In the homogeneous regions $g = 1$ and $\nabla g \to 0$, therefore (5.10) reduces to the mean curvature flow: inside the objects the hypersurface levels collapse in a point then disappear. In regions where the edge information exists $g \to 0$ and (5.10) reduces to a simple advection equation: the hypersurface levels are driven towards the edges by the field $-\nabla g$, their accumulation causes the increase of the spatial gradient and S starts to generate discontinuities. In regions with subjective contours (missing boundaries), continuation of existing edge fragments, a is negligible and (5.10) can be approximated by a geodesic flow, allowing the boundary completion with geodesics. The application of these dynamics is clear in Fig. 5.6, showing the effect of boundary completion in a membrane with a missing contour. The use of different weights between the regularization and the advective term ($\nu > \lambda$) facilitates the control of

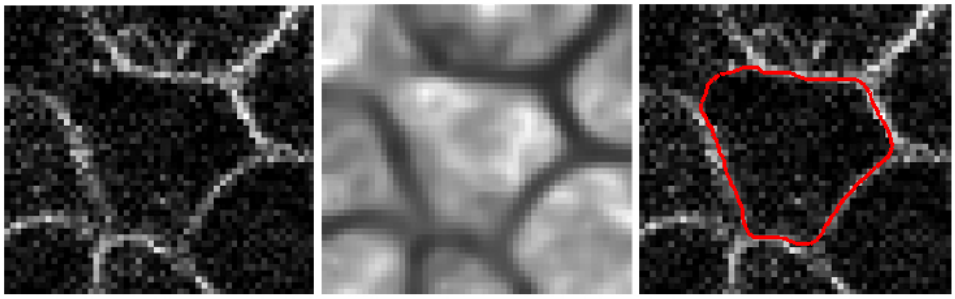
evolutive process. Indeed, the segmentation, together with the missing contours completion, is obtained through the discontinuities developed by the hypersurface on object boundaries, while the hypersurface is simultaneously smoothed and flattened inside the object. An higher weight of advective term ensures a better accumulation of image graylevel around existing contours. The position of the reference point influences the result of segmentation. If it is around the object center, at the end of the evolution the highest hypersurface values correspond to the shape we want to extract. Therefore, selecting a level little lower than the maximum allows extracting the desired contour. The same behavior can be observed if we start from a slightly off center initial condition. On the contrary, if we consider a strongly off center point-of-view other adjacent structures may become predominant, not allowing a correct segmentation. We solved this problem, at least partially, by changing the dynamics of motion equation during the evolution process. Using first an high value of parameter a the process is most diffusive. The hypersurface smooths, moving away from adjacent, external structures and simultaneously flattens inside objects. Then, with a low value of a, the hypersurface evolves driven by a pure level set motion, sharpening its discontinuities. A bidimensional example is shown if Fig.5.7.

5.3.5 Level Selection

In the conclusive step of the algorithm, we automatically picked the level set that describes the desired object. After segmentation, the intensity distribution of the function Φ is typically associated to a bimodal histogram with values range between 0 and 255, because of a linear rescaling. The higher intensity peak (near to 255) corresponds to the segmented object, the lower one to the background. Therefore, the segmented surface could be extracted as the isosurface corresponding to the intermediate value 128. These surfaces are represented and stored through a VTK PolyData format [62]. Some VTK postprocessing filters (vtkImageGaussianSmooth and vtkPolyDataNormals [62]) have been applied to reduce the surface discontinuities.

5.3.6 Numerical Discretization

Concerning the numerical schemes for discretization, the partial derivatives in (5.10) are approximated with finite differences [56, 57, 64]. Time derivatives are discretized



(a)

(b)

Figure 5.6: 2-D example of membrane with missing boundaries segmentation. (a, from left to right): original data, edge detector, segmented contour, in red, superimposed to original data. (b, from left to right): evolution of the initial point-of-view surface and selection of a level set, red line, for segmentation.

with first order forward differences, the parabolic term with central differences and the advective term with upwind schemes, where the direction of the one-sided difference used in a point depends on the direction of the velocity field $v = -\nu\nabla g$ in the same point. Let us consider a uniform grid in space-time (t, x, y, z), then the grid consists of the points $(t_n, x_i, y_j, z_k) = (n\Delta t, i\Delta x, j\Delta y, k\Delta z)$. We denote by Φ^n_{ijk} the value of the function Φ at the grid point (t_n, x_i, y_j, z_k), by g_{ijk} the value of the edge indicator in the grid point (x_i, y_j, z_k) and by v_{ijk} the vector value of the velocity field v in the same spatial grid point. The equation (5.11) is the numerical approximations of (5.10), where D is a finite difference operator on Φ^n_{ijk}, the superscripts $\{2, 0, 1\}$ indicate backward, central, and forward differences and the superscripts $\{x, y, z\}$ indicate the direction of differentiation.

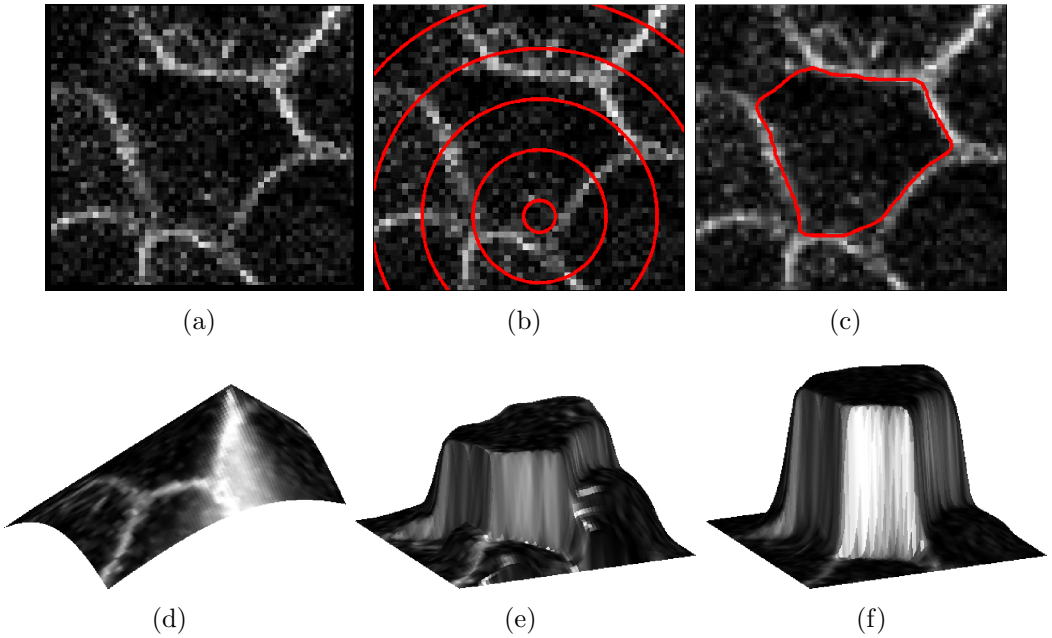


Figure 5.7: 2-D example of membrane with missing boundaries segmentation. The point for the initialization of the Surface is chosen very close to boundaries but the membrane is correctly segmented thank to the use of different values of parameter a during the Surface evolution. (a) Original membrane. (b) Initial distance function Φ, depicted in red, superimposed to original data. (c) Membrane segmentation, in red, superimposed to original data. (d) Original Surface S constructed as graph of Φ. (e) The Surface at the end of the first, diffusive, process ($a = 1$). (f) The Surface at the end of evolution, after a pure level set motion ($a = 10^{-6}$).

$$
\begin{aligned}
\Phi_{ijk}^{n+1} &= \Phi_{ijk}^{n} + \\
&+\Delta t \left\{ \lambda g_{ijk} \left[\frac{(a^2 + D_{ijk}^{0_x^2} + D_{ijk}^{0_y^2}) D_{ijk}^{0_{zz}} + (a^2 + D_{ijk}^{0_x^2} + D_{ijk}^{0_z^2}) D_{ijk}^{0_{yy}} + (a^2 + D_{ijk}^{0_y^2} + D_{ijk}^{0_z^2}) D_{ijk}^{0_{xx}}}{a^2 + D_{ijk}^{0_x^2} + D_{ijk}^{0_y^2} + D_{ijk}^{0_z^2}} + \right.\right. \\
&\left.\left. -2 \frac{D_{ijk}^{0_x} D_{ijk}^{0_z} D_{ijk}^{0_{xz}} + D_{ijk}^{0_x} D_{ijk}^{0_y} D_{ijk}^{0_{xy}} + D_{ijk}^{0_y} D_{ijk}^{0_z} D_{ijk}^{0_{yz}}}{a^2 + D_{ijk}^{0_x^2} + D_{ijk}^{0_y^2} + D_{ijk}^{0_z^2}} \right] - \nu \left\{ \left[\min(v_{ijk}^{0_x}, 0) D_{ijk}^{-x} + \max(v_{ijk}^{0_x}, 0) D_{ijk}^{+x} + \right.\right.\right. \\
&\left.\left.\left. + \min(v_{ijk}^{0_y}, 0) D_{ijk}^{-y} + \max(v_{ijk}^{0_y}, 0) D_{ijk}^{+y} + \min(v_{ijk}^{0_z}, 0) D_{ijk}^{-z} + \max(v_{ijk}^{0_z}, 0) D_{ijk}^{+z} \right] \right\} \right\}
\end{aligned}
\tag{5.11}
$$

5.4 Results and Discussion

We applied our algorithm to membranes and nuclei images. The results reported here are mainly membranes segmentation. Membrane images treatment allowed testing characteristic behaviors of the adopted segmentation technique, such as boundary completion. Furthermore, membranes manipulation is more difficult and reveals specific problems when applied to epithelial cells. Nuclei are more simple to handle and their segmentation does not require particularly sophisticated tools.
The optimal values for the algorithm parameters have been found by visual inspection of the segmented surfaces resulting from the use of different parameters. Through the superimposition of the segmented shape on the original image, we recognized the set of values corresponding to the best fitting.
The parameters set we used for the segmentation of the whole dataset is:

- high value of $a = 1$;
- low value of $a = 10^{-6}$;
- load of the curvature term $\lambda = 0.1$;
- load of the advective term $\nu = 10$;
- time step $\Delta t = 0.065$.

In the first part of computation, the hypersurface evolved with a high value of a to make the flow mostly diffusive and therefore to optimize the completion of the cellular shape, even if it is irregular and indented or if the initialization point is not centered on the object. In the second part, we used a low value of a and the evolutive behavior was mostly a level set motion, which contributed to match the level surfaces to the contour. Since diffusion was faster, the number of iterations with a high value of a was lower than the number of iterations with a low value of a, to make comparable the effects of the two motions on the final contour. We used the same parameters both for nuclei and membranes processing, except for the total number of iterations: nuclei are smaller, thus their segmentation requires less iterations (40000 iterative steps for membranes, 10000 for nuclei).

The discrete time step Δt has been chosen as the maximum value which insures the stability of the advective term in (5.10). The stability condition, which can be deduced from mass balance considerations [65] applied to a single voxel, is given by:

$$\frac{\Delta t}{\Delta x}|v_x| + \frac{\Delta t}{\Delta y}|v_y| + \frac{\Delta t}{\Delta z}|v_z| \leq 1 \tag{5.12}$$

where v_x, v_y, v_z are the velocity components.

In our application we used the following approximation of the components arising in the velocity field:

$$\begin{aligned} g_x &\approx \frac{g_{i+1,j,k} - g_{i-1,j,k}}{2} \\ g_y &\approx \frac{g_{i,j+1,k} - g_{i,j-1,k}}{2} \\ g_z &\approx \frac{g_{i,j,k+1} - g_{i,j,k-1}}{2} \end{aligned} \tag{5.13}$$

Since $0 \leq g \leq 1$, in the worst case $|g_x| = |g_y| = |g_z| = 0.5$. Therefore, substituting this value in (5.13) and setting $\Delta x = \Delta y = \Delta z = 1$, we obtain from (5.12) $\Delta t \leq \frac{2}{3\nu}$. In our case $\nu = 10$, so we set $\Delta t = 0.065$.

A similar analysis is difficult for the curvature term because it depends in nonlinear way on the solution. In the actual implementation we did not use strong curvature weight. In case of strong curvature influence one should approximate the equation through the semi-implicit schemes [34, 66, 67, 68] which are unconditionally stable.

In this Section we show some meaningful results of segmentation on two different cell types distinguishable in the imaged developmental period: *epithelial* cells from the enveloping layer and *inner* cells. Their morphology varies along the cell cycle introducing more morphological categories. The inner cell mass is covered by an epithelial layer (EVL or enveloping cell layer). EVL cells are polarized, i.e. their apical surface and baso-lateral surface have specific properties, polygonal, large, flat and they largely keep their shape when dividing. They sometime have several nuclei, might correlate with some stressed condition linked to manipulation. They also always show intracellular membrane staining, probably corresponding to intra cellular membrane compartments. Inner cells are smaller than EVL cells and not polarized. They fill the space and their nucleus is centered. During division inner cells become spherical and largely loose ad-

hesion to their neighbors. Fig. 5.9 shows the effect of boundary completion on an inner cell: the missing contour, underlined by the red circle, is completed by a straight line. The algorithm showed the same behavior for dividing membranes (Fig. 5.10). When two different nuclei were found inside the same cell and the membrane presented a constriction along the division plane, the algorithm segmented two cells by completing their contours with straight lines. These results demonstrate the suitability of the Subjective Surfaces technique for this scenario, especially if compared with other methods. In Fig. 5.11 we discuss our algorithm against Malladi-Sethian approach [59] in the specific case study of missing membrane boundary. The performances are comparable in the region with well defined contours, whereas the final shape achieved by the classical level set method fails in membrane completion. Before undergoing division, inner cells become spherical, whereas nuclei staining elongates as the chromosomes arrange in the future cell division plane (Fig. 5.12). It should be noted that the nucleus size is underestimated in the last two parts. This is due to the parabolic regularization term in the motion equation (5.10), which prevents the segmented surface to reach the contour if it is concave and with high curvature. . However, the nuclei of not dividing cells were correctly segmented, as confirmed by visual inspection. Fig. 5.13 shows a complete sequence of an inner cell division. In the first stages the cell shape is irregular, because of the adhesion to its neighbors, but becomes spherical before mitosis. In the same way, the nucleus shape changes during cellular division from a spherical or ellipsoidal aspect to a more oblong and flat shape. This morphological features are linked to specific division phases (see figure 5.8):

1. *Prophase* the nucleus starts changing its shape and gaining in intensity, because of chromosomes condensation, and the membrane gradually looses adhesion to the neighbors;

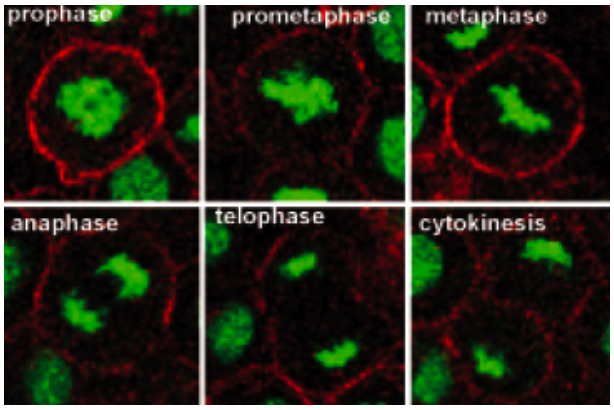
2. *Prometaphase* chromosomes attach to the mitotic spindle;

3. *Methaphase* the chromosomes arrange in the future cell division plane;

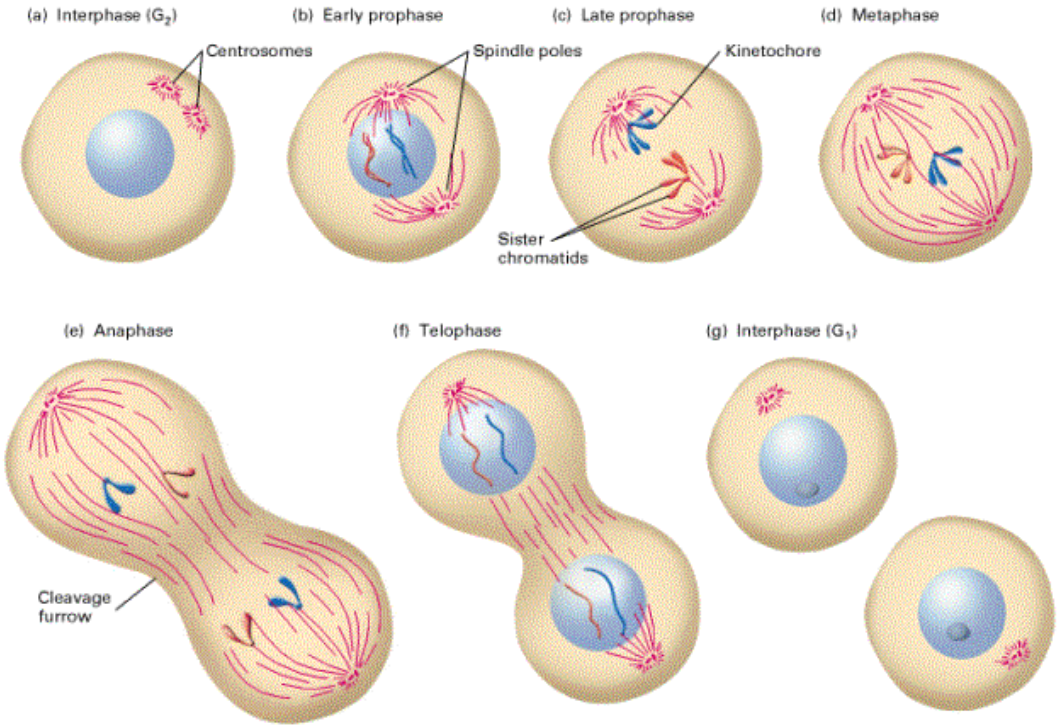
4. *Anaphase* the two sets of chromosomes separate;

5. *Telophase* the membrane shows a constriction along the future cell division plane;

6. *Cytokinesis* the daughter cells separate.



(a)



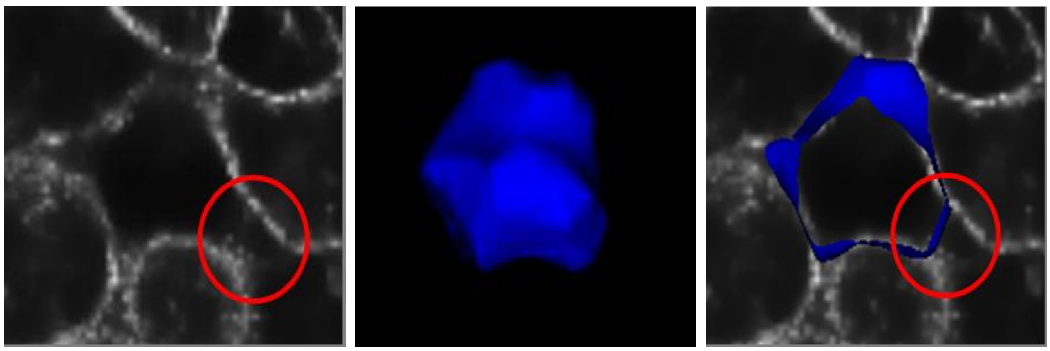
(b)

Figure 5.8: Figure a: The different phases of cell division ovserved by confocal microscopy. Experiment 070128c. Figure b: schema of cell division phases [69]

Eye inspection of the results revealed some problems in the segmentation of EVL membranes. As we described above, these cells surrounding the embryo are very flat. This feature impaired membrane completion by the Subjective Surfaces 3-D technique, because the small extension in depth stopped the evolution process. Furthermore, EVL cells show intense intracellular labeling (as we can see in the central cell of Fig. 5.14(c)), probably corresponding to intracellular membrane compartments (Golgi apparatus or endoplasmic reticulum). When the evolving surface reached this intracellular staining, it was not able to pass on. These considerations led us to think we require a specific method for the segmentation of the epithelial cells. Prior segmentation, they have to be automatically localized within the acquired volumes through a discriminating factor. At the moment, we are developing a simple method for the detection of the epithelial cells based on their position. First, we segmented the surface of the embryo using the Geodesic Active Contours technique [28]. The evolution was not stopped by the external layer of cells, because they have a weak outer contour, so we obtained the profile below (Fig. 5.15(a)) and the epithelial cells remained outside (Fig. 5.15(b)). Since the segmented surface is the zero level of the processed function S_i, the epithelial cells can be detected simply by verifying the value of S_i in the center position: if it is negative the center is outside the surface and the cell can be classified as epithelial. This method is interesting because of its simpleness, but other factors could be used for the detection, such as the polygonal shape or the bigger size of epithelial nuclei. Finally, in Fig. 5.16 we show the segmentation of two subvolumes of nuclei and membranes. Every object is labeled with a different color, whose scalar value corresponds to the cell identity number.

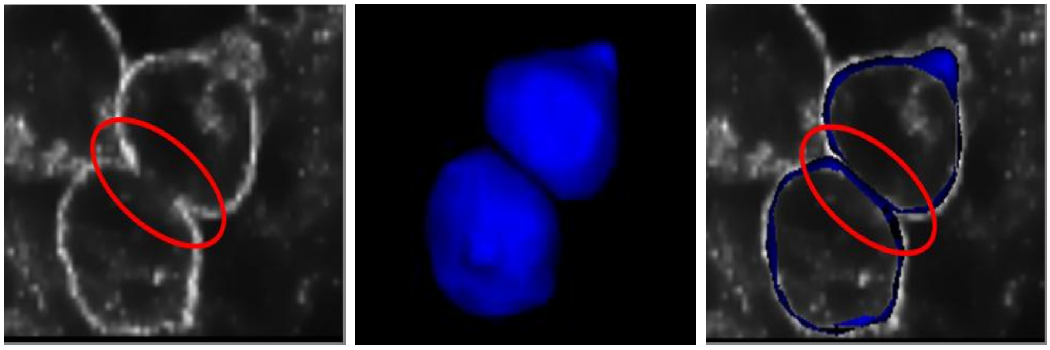
5.5 Validation

In order to visually inspect the results validity, the segmentation algorithm has been first tested using a special framework designed for managing series of 3-D biological images [70]. It provides an helpful support for testing new algorithms or their application on new images and it supplies some visualization tools suitable for the analysis of the processing results. In this framework, the matching between the detected contours and the real edges can be evaluated through the superimposition of the segmented surface and the original data in a tridimensional representation or visualizing a cut of the



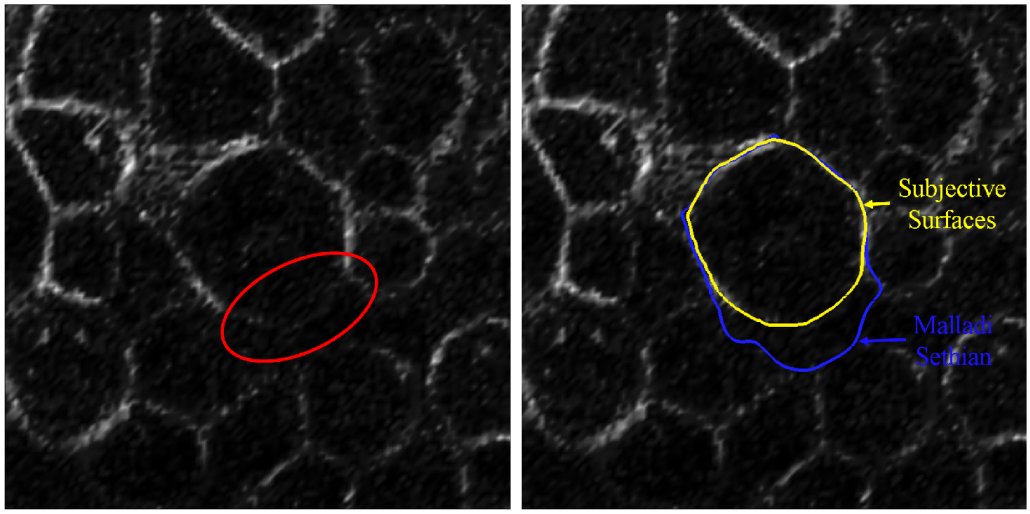
(a) (b) (c)

Figure 5.9: Segmentation of a membrane with an incomplete contour: (a) missing portion underlined by a red circle, (b) segmented surface, (c) cut of the surface superimposed on an image slice.



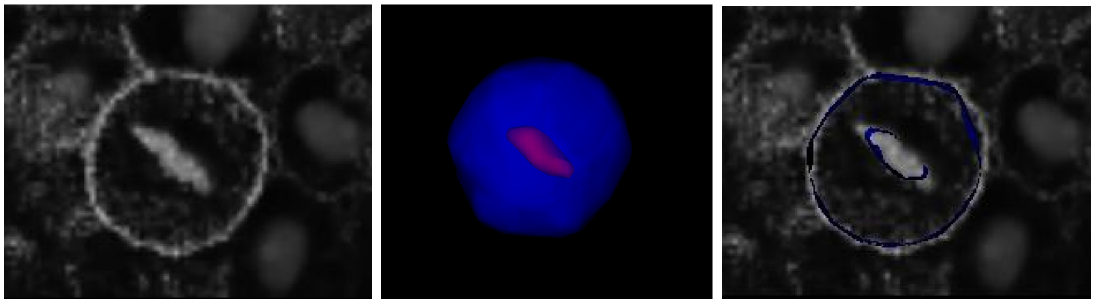
(a) (b) (c)

Figure 5.10: Segmentation of a dividing cell: (a) constriction of the membrane underlined by a red circle, (b) segmented surface, (c) cut of the surface superimposed on an image slice.



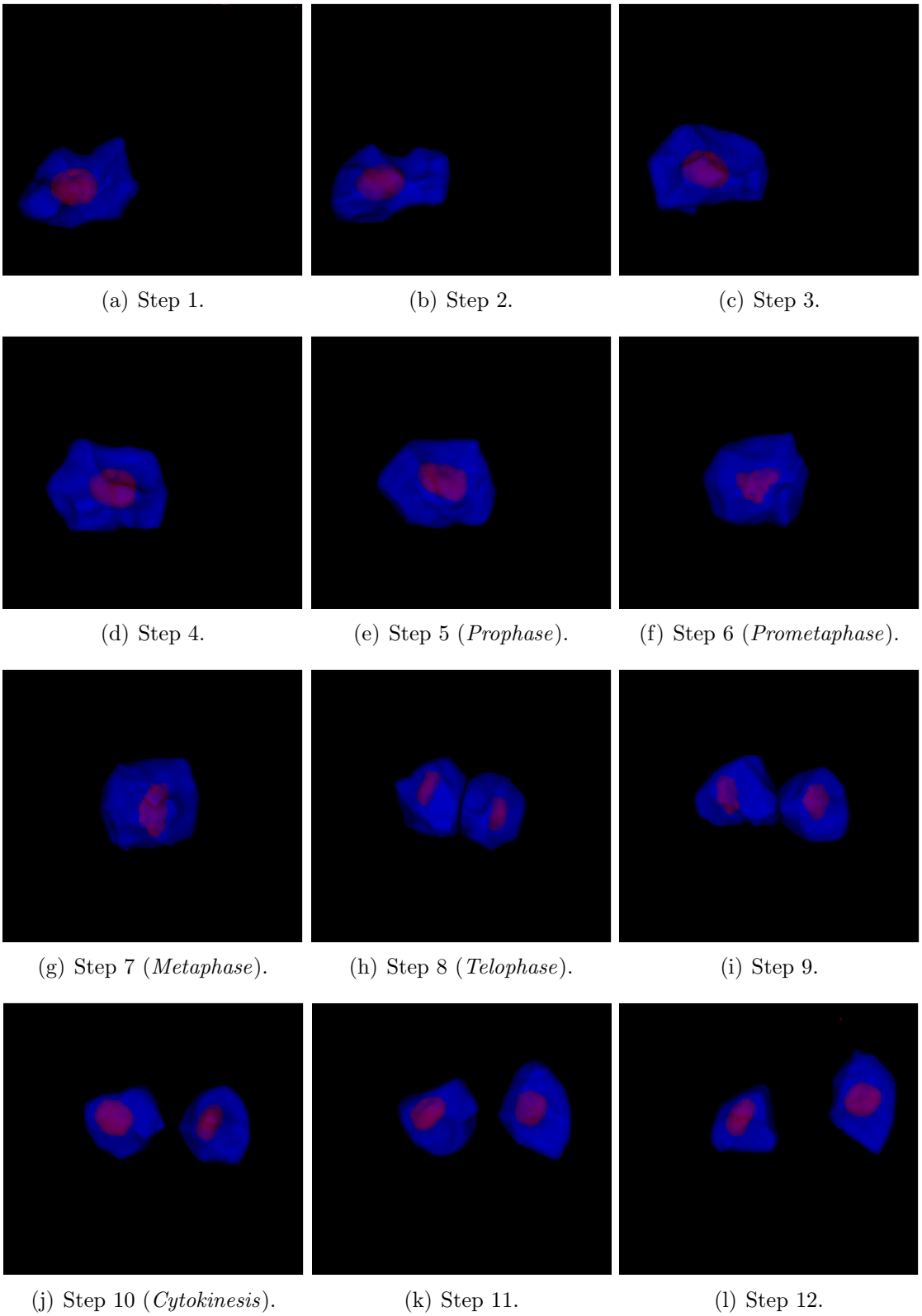
(a) (b)

Figure 5.11: Segmentation of an incomplete membrane by using different segmentation techniques: (a) missing portion underlined by a red circle, (b) segmented contour (yellow line Subjective Surfaces, blue line Malladi-Sethian).



(a) (b) (c)

Figure 5.12: Segmentation of a cell before division: (a) superimposition of membranes and nuclei signals, (b) segmented surfaces, (c) cut of the surfaces superimposed on an image slice.



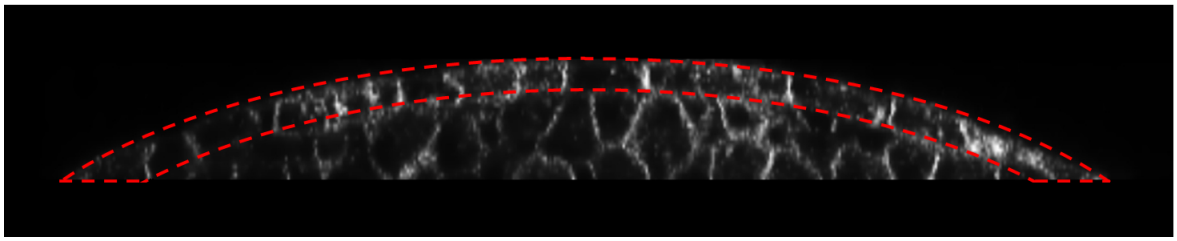
(a) Step 1. (b) Step 2. (c) Step 3.

(d) Step 4. (e) Step 5 (*Prophase*). (f) Step 6 (*Prometaphase*).

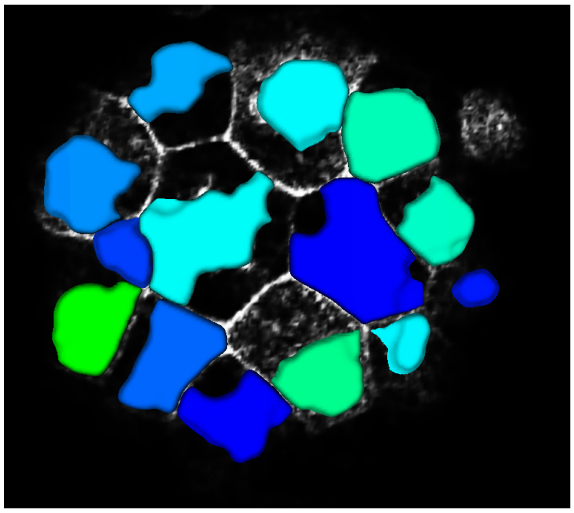
(g) Step 7 (*Metaphase*). (h) Step 8 (*Telophase*). (i) Step 9.

(j) Step 10 (*Cytokinesis*). (k) Step 11. (l) Step 12.

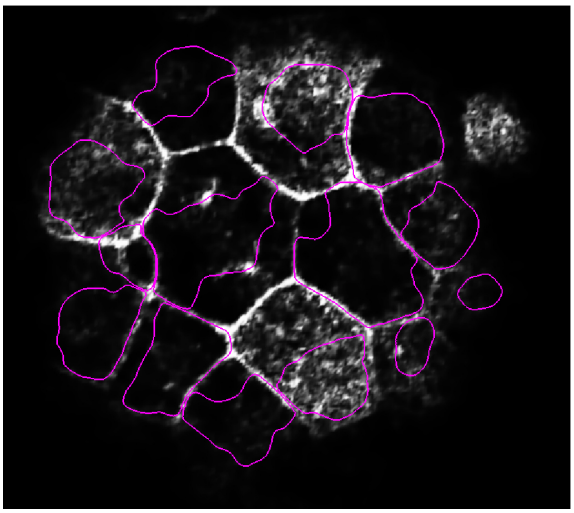
Figure 5.13: Sequence of cell division.



(a)

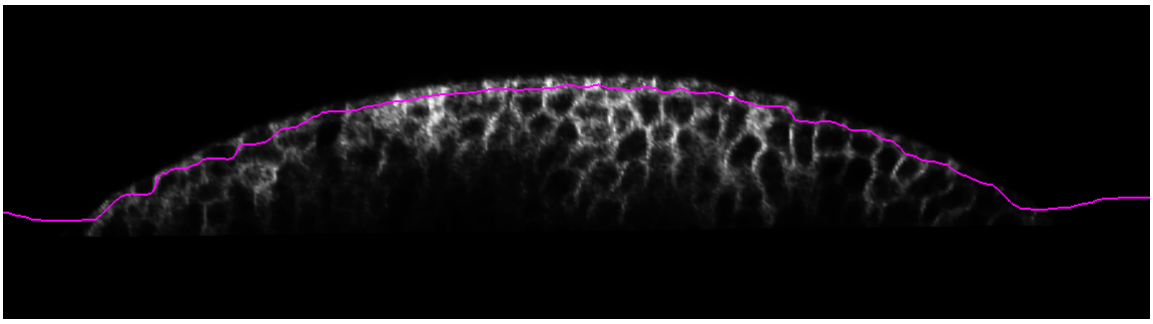


(b)

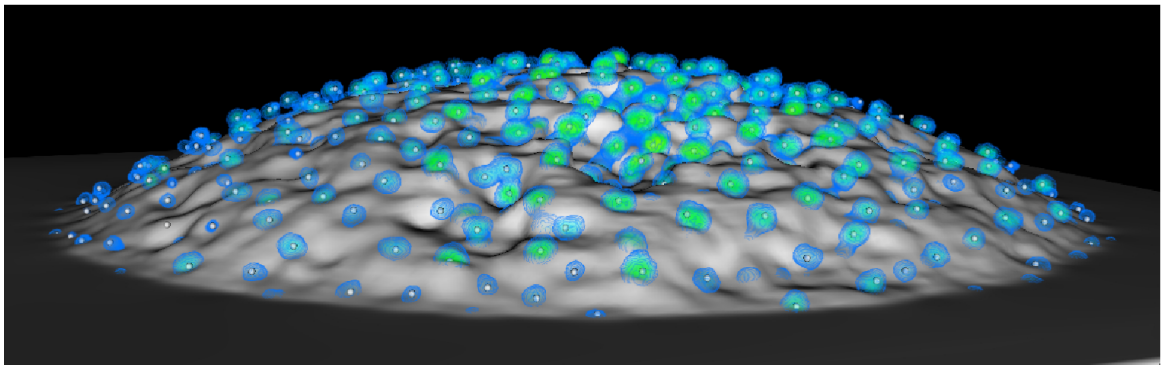


(c)

Figure 5.14: Segmentation of epithelial cells: (a) location of the epithelial cells in the acquired volumes (dashed area), (b) segmented surfaces, (c) slice of the segmented surfaces superimposed on an image slice.

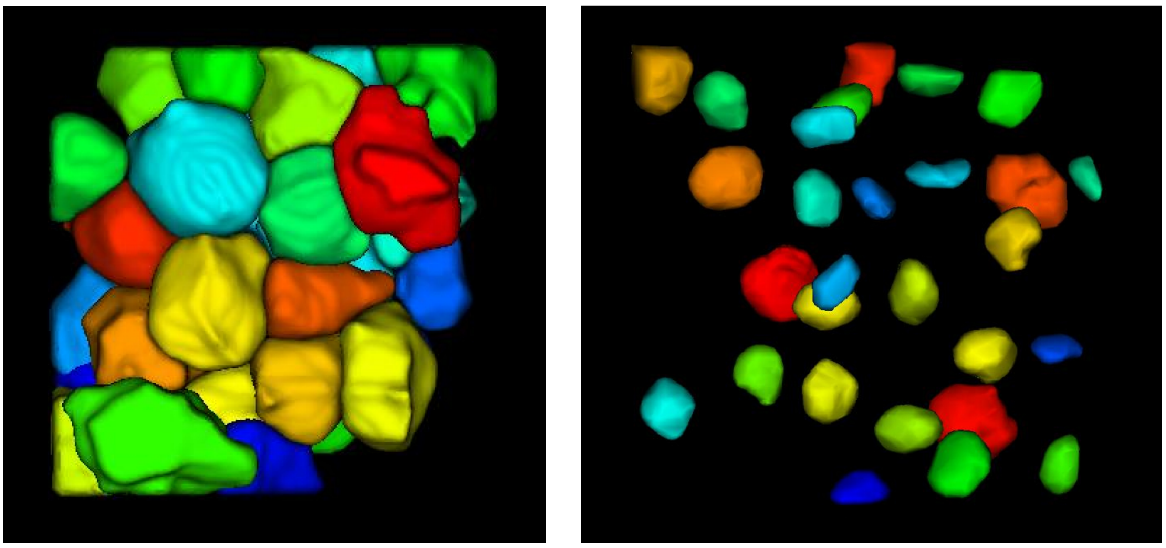


(a)



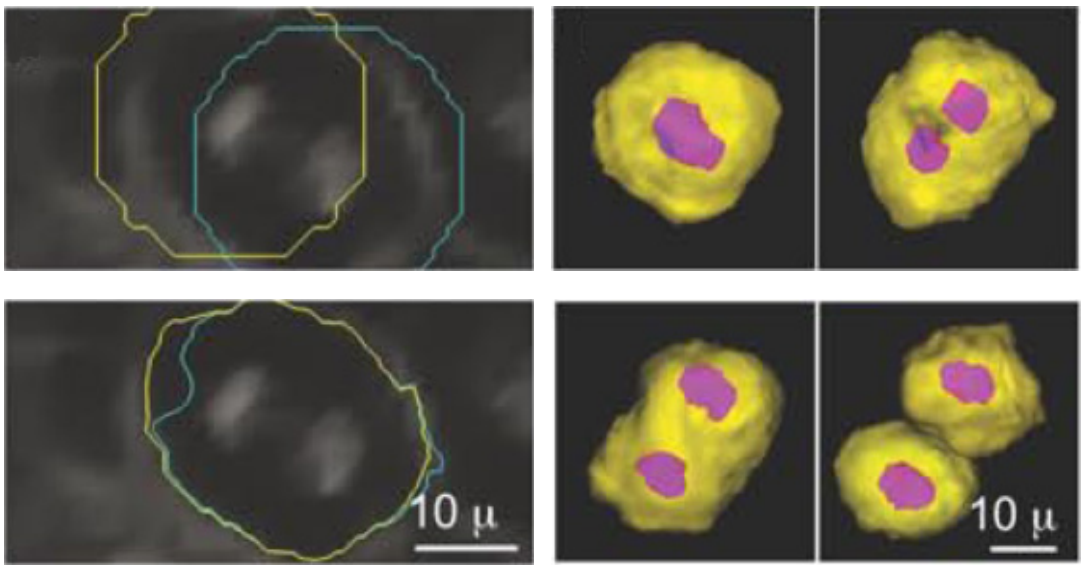
(b)

Figure 5.15: Detection of the epithelial cells on a 4-D MLMS dataset: (a) slice of the segmented surface, (b) superimposition of the segmented surface, nuclei channel in volume rendering representation and detected centers.



(a) (b)

Figure 5.16: Segmentation of an entire subvolume: (a) membranes, (b) nuclei.



(a) (b)

Figure 5.17: Mitosis Detection. (a) Top panel, cell ID 1089/t001, initial segmentation function, two centers. Bottom panel, final segmentation, the two centers are found inside one cell. Filtered data shown in gray scale. (b) Mother cell ID 215/t008, 3D rendering of nucleus and membrane shape throughout mitosis.

surface in xy, xz, yz planes on the image slice (Fig. 5.18). It also allows the selection of a specific cell or of a group of cells simply by specifying their center numbers, so that the user can limit the analysis to the cells he is interested in.

Table 5.1: H.d. (Hausdorff distance) and mean H.d. in reference to 10 different gold standards.

	CELL1		**CELL2**	
	H.d. (μm)	Mean H.d. (μm)	H.d. (μm)	Mean H.d. (μm)
gold std 1	1.4	0.6	2.0	0.3
gold std 2	1.8	0.6	2.3	0.4
gold std 3	1.9	0.7	2.3	0.4
gold std 4	1.2	0.3	2.0	0.3
gold std 5	1.1	0.3	2.0	0.3
gold std 6	1.1	0.3	2.2	0.4
gold std 7	1.0	0.3	2.0	0.3
gold std 8	1.0	0.3	1.9	0.4
gold std 9	1.1	0.3	1.9	0.3
gold std 10	1.1	0.3	2.1	0.3
mean value	1.3	0.4	2.1	0.3
std deviation	0.3	0.2	0.1	0.0

The visual inspection of results allows the detection of glaring mistakes in shape reconstruction, such as surfaces overlapping and incomplete contours. This estimation is not enough to quantify the algorithm precision. To measure the algorithm capability to detect and reproduce the correct edges, we propose to use the *Hausdorff distance*, that calculates the distance between two surfaces, and to compute the distance between the segmented surface and a gold standard obtained by manual segmentation, detecting contours by hand. The reference surface was generated exploiting the functionalities of ITK-SNAP [71], an ad hoc software that collects the contours manually detected in every 2-D slice, returning a binary 3-D image.
Given two finite sets of points $A = \{a_1, ..., a_m\}$ and $B = \{b_1, ..., b_n\}$, the Hausdorff distance is defined as [37]:

$$H(A, B) = max(h(A, B), h(B, A)) \tag{5.14}$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b|| \tag{5.15}$$

in the *classical formulation* and

$$h(A,B) = \frac{1}{m}\sum_{i=1}^{m} \min_{b\in B} \|a_i - b\| \tag{5.16}$$

in the *mean Hausdorff distance*. The term $h(B,A)$ is defined similarly.

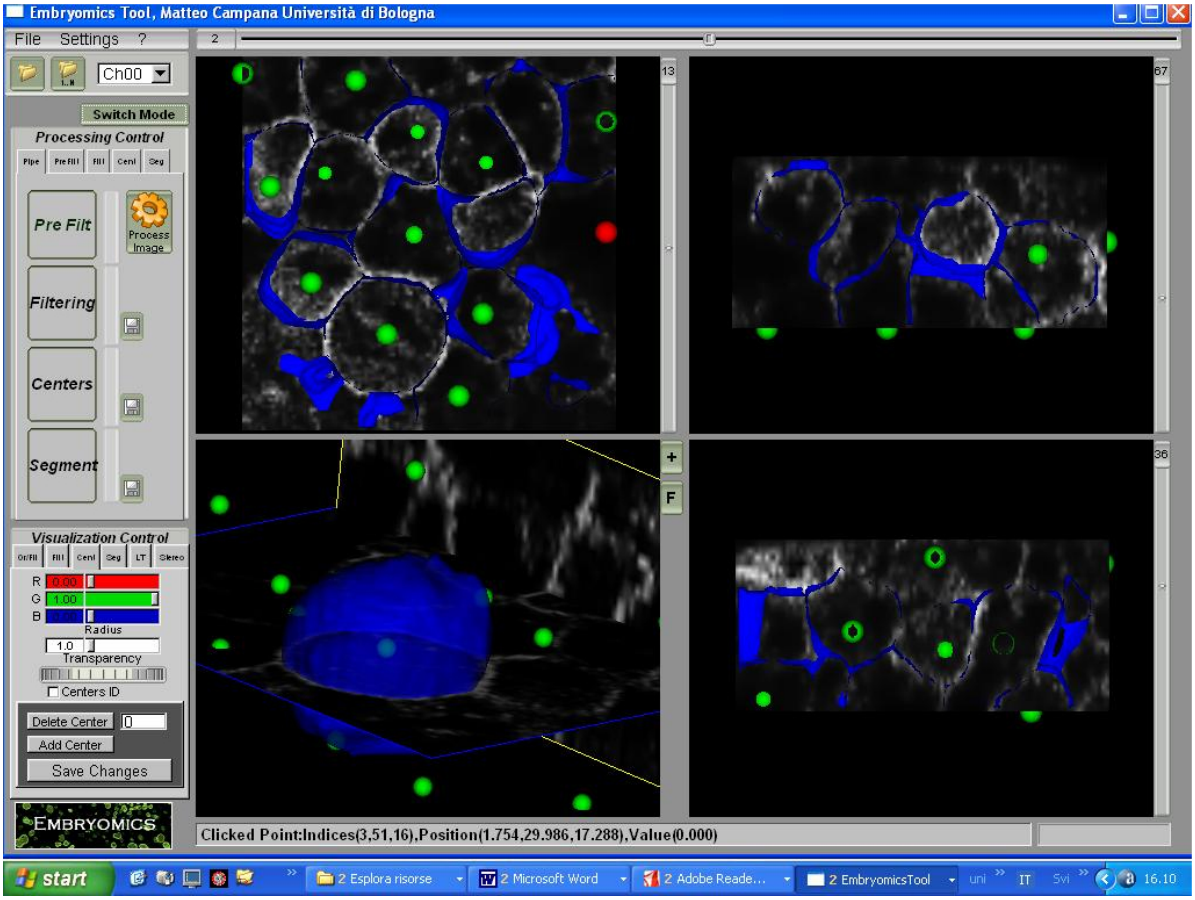


Figure 5.18: Framework interface. The GUI is structured in two parts: the interactive control panel on the left side gives accessibility to the implemented algorithms and to the visualization modalities, whereas the superimposition of the segmented surface and the original data is shown in the central windows (3-D representation in lower left window, 2-D slices in the upper and lower right windows).

The two formulations can be used to quantify different kinds of errors: the maximum segmentation error is associated to the classical expression, whereas the total error (i.e.

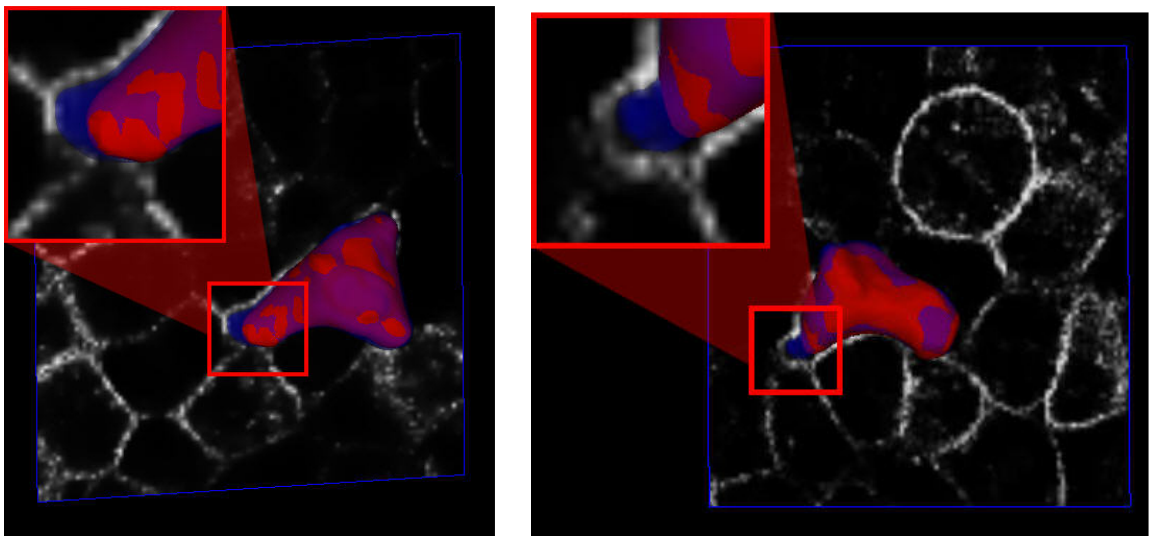


Figure 5.19: Details of segmented surfaces (red) and gold standards (blue) showing the region where the maximum segmentation error is located.

algorithm precision) is associated to the mean Hausdorff distance.

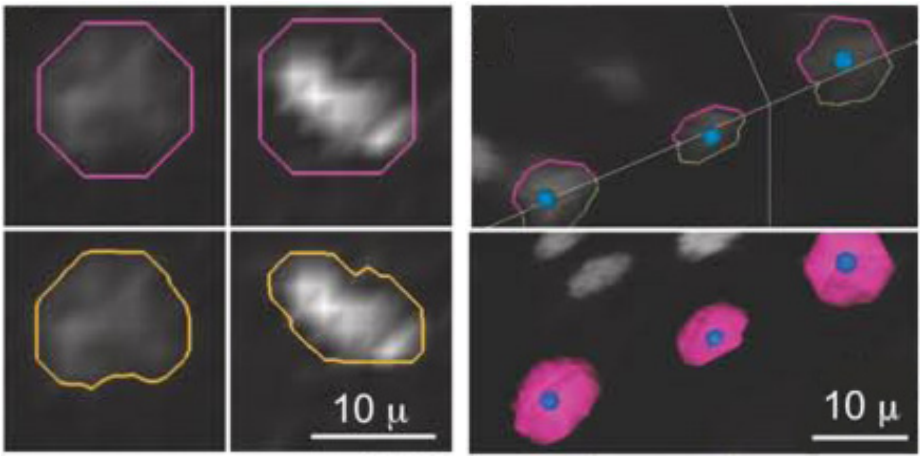
In the foreseen validation method, the calculation of the distance should be repeated considering different gold standards for the same membrane, to execute a statistical analysis (mean value and standard deviation) on the data. This procedure should reduce the influence on results of the user who made the manual segmentation. The proposed procedure has been applied to a few cells. However, the first analysis on two cells with different shapes revealed interesting features (Tab. 5.1). Cell1 was imaged before division and is almost perfectly spherical, whereas cell2 has an irregular shape because of the adhesion to adjacent cells. Examining different conformations, we want to estimate the effect of the "shape factor" on the algorithm precision. The mean Hausdorff distances are comparable for cell1: 0.412 ± 0.160 μm and cell2: 0.347 ± 0.019 μm. On the contrary, the maximum error is different whether the shape is spherical: 1.278 ± 0.311 μm or irregular: 2.088 ± 0.142 μm. These results suggest that the shape influences the maximum error but not the overall precision. This is probably due to the behavior already observed in section 5.4: the inability of the segmented surface to reach the contour if it is concave and with high curvature causes an increase of the maximum error in the irregular shapes (Fig. 5.19). Certainly, these results have to be supported by a larger record of cases. However they indicate that our precision is at

least sufficient to identify mitosis which is a major issue for further reconstructing cell behaviors.

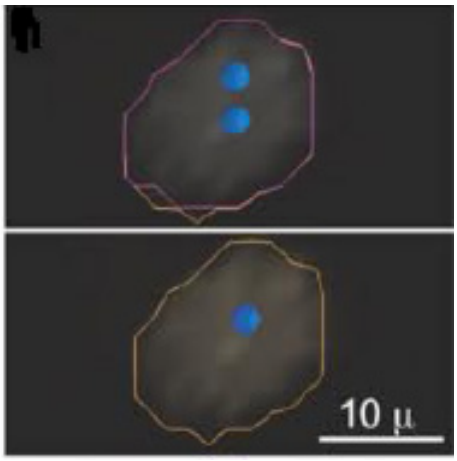
Nuclear segmentation was then used to correct "double centers", nucleus that were double detected from the nuclear center detection step (Figure 5.20). Finally, membrane segmentation was used to detect mitosis (Figure 5.20) given that two nucleus are inside one membrane.

5.6 Conclusions

We designed an algorithm for the automated segmentation of membranes and nuclei based on Subjective Surfaces technique that has good performances on live zebrafish embryos confocal images. Visual inspection of the results has shown the ability of the algorithm to complete the missing contours, especially in membranes images, and to correctly reproduce the objects shape. The local precision seems to decrease for elongated and flat shapes (EVL cells and dividing nuclei). However these observations have to be further confirmed by a quantitative analysis on the segmentation error, according to the procedure proposed in this chapter. It is possible to correct error done in cell detection, this segmentation algorithm could be used to erase double detection of cells. Given that in the mitosis process two nucleus are inside a single membrane it is possible to classify some of the cells stages. The algorithm could be improved by integrating the segmentation of membranes and nuclei, superimposing their edge indicators and defining two different isosurface values for extracting both shapes in the same process. A specific method could be designed for the segmentation of the EVL cells that have to be localized prior segmentation. With this work we have built the basis for future developments toward a deeper understanding of the biological processes involved in the organism formation. In this direction, the next step will be to pass the segmentation results to a specific algorithm for the cell shape analysis, that has to be defined yet, with the final goal to extract information on the cell state and analyze the dynamic of their shape. This should bring us close to an automated segmentation procedure for the whole zebrafish early embryogenesis.

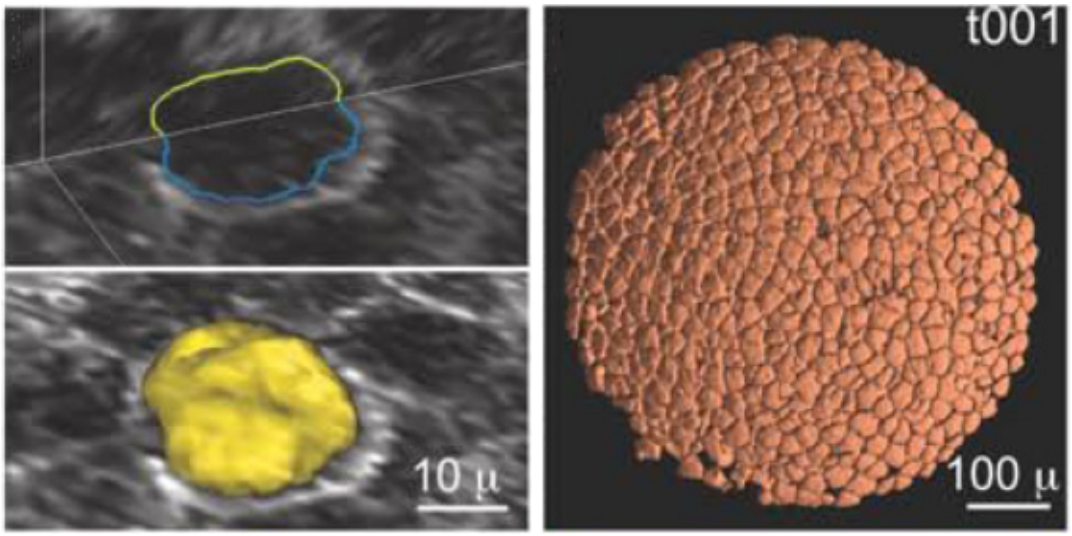


(a) (b)



(c)

Figure 5.20: Nuclei segmentation and double centers correction. Data set ID 070420a. (a,b) Nuclei segmentation. (a) Top panels, 2D cut of initial segmentation function (pink) superimposed with raw data. Bottom panels, final segmentation function (orange). Left panels, cell ID 707/t001. Right panels, cell ID 780/t001, metaphase nucleus. (b) cells ID at t001, 1628, 1638, 1671. Top panel, subjective surface technique applied to nuclei segmentation, orthoslices view. Bottom panel, same as top panel but 3D rendering. Filtered data shown in gray scale. (c) Correction of double centers by comparing nuclei segmentation functions. Cell ID 1089/t001. Top panel, originally detected centers and corresponding segmentations. Bottom panel, correction by averaging the two centers and merging segmentation functions.



(a) (b)

Figure 5.21: Membranes Segmentation. (a) Cell ID 802/t001, superimposed membrane segmentation and raw data. Top panel, two orthoslices. Bottom panel, 3D rendering. (b) Whole volume, 3D rendering of segmented cell shapes.

Chapter 6

Cells movement estimation

Image registration computes the displacements transforming a so-called moving image into a target image. Thus, registration of cell images sequence can be used in cell dynamics analysis to answer the question "when and where do cells move?". This is a first step toward describing the relation between local cell behaviour and macroscopic structures. Numerous studies address this topic in developmental biology [72],[73],[74],[75].

Recent imaging techniques allow observing cell interactions together with embryo body plan formation [1]. Description of this spatio-temporal data with an accurate registration is a preliminary step to investigate the biomechanics of cells and tissues. The zebrafish is a good model for biomechanical study. It is well suited for in vivo imaging and its gastrulation exhibits many typical cell behaviours [8] including collective displacements, shape and adhesion changes.

We aims at characterizing registration algorithm ability to capture cell movements in wide field, micrometer resolution, images sequence of zebrafish. There is a huge amount of literature (see [76] or [77] for general references) about registration methods and their applications in biomedical domain.

Cells images however have their own characteristics; they show ambiguities and their large cell number imply that there is a high number of freedom degrees. The laws to describe the cellular medium at the cell single level are unknown as it is both a discrete and a continuous material. Therefore it can be discussed which deformation model is the most efficient for cell motion estimation using registration: elastic, fluid, optical flow. There are nevertheless studies addressing directly the question of cell registration, [78] evaluating qualitatively cell plants motion estimation with various classical methods. A registration method is applied to 3D confocal cell image registration in [79].

Migration of cells is a crucial issue to describe vertebrate development and explore the relations between cell mechanical activities and the formation of macroscopic structures. Thus motion estimation using registration is a first step towards the understanding of cells and tissues biomechanics. The present chapter aims at characterizing registration algorithm ability to capture cells and tissues displacements. An evaluation

protocol has been set up in order to measure the error produced by a given algorithm. Evaluation relies on a reference, a gold standard of cells trajectories. The difference between this reference and the cell trajectories built by registration is the error measurement. Four non-rigid registration methods used in the biomedical field have been tested: Bspline elastic registration, MIA fluid registration, ITK demons optical flow and ITK levelset optical flow. Two artificial datasets were registered with these methods. We measured errors and their distributions with our gold standard to assess registration quality. Results show that all methods are well suited to achieve nuclei registration with an accumulated error remaining below the voxel size for 80% of the cells. BSpline gave the best registration under five time steps of error accumulation whereas ITK demons was the best for longer accumulation. The same method were used on zebrafish development 3D data.

6.1 Motion estimation using registration methods

The typical problem of motion estimation is to analyze the motion from bidimensional data, where each image (frame), come from a sequence of images.

Each image acquired with an specific frame-rate. In this work we analyze images that come from a confocal microscope, these images can be define as a sequence of multidimensional 3D volumes where 2 channels are acquired.

In the next equations we will consider the bi-dimensional images where $\overline{x} = (x, y)$, the general form of this procedure could be used to process 3D datasets. The initial hypothesis in the classical analysis of the optical flow postulated by Horn and Schunck, is that locally (in space and time) the image intensity is constant [80]. This hypothesis is called *brightness constancy assumption,BCA*. Formally, if $I(\overline{x}, t)$ is the image intensity, we can write

$$I(\overline{x}, t) \approx I(\overline{x} + \delta\overline{x}, t + \delta t), \tag{6.1}$$

where $\overline{x} = (x, y)$ and $\delta\overline{x} = (\delta x, \delta y)^T$ is the displacement of the local region of the image in $(\overline{x}, t)$ after a time δt.

The first order development of the Taylor function of (6.1) is:

$$I(\overline{x} + \delta\overline{x}, t + \delta t) = I(\overline{x}, t) + \nabla I(\overline{x}, t) \cdot \delta\overline{x} + \frac{\partial I(\overline{x}, t)}{\partial t}\delta t \tag{6.2}$$

where ∇ is the gradient operator.

Replacing equation 6.2 in equation 6.1 and dividing by δt, we get

$$\nabla I(\overline{x}, t) \cdot \frac{\delta\overline{x}}{\delta t} + \frac{\partial I(\overline{x}, t)}{\partial t} = 0 \tag{6.3}$$

that, when $\delta t \to 0$

$$\nabla I(\overline{x}, t) \cdot \mathrm{v} + \frac{\partial I(\overline{x}, t)}{\partial t} = 0 \tag{6.4}$$

where $\cdot$ is the dot product and $\mathrm{v} = \frac{\mathrm{d}x}{\delta t}$ represents the trajectory velocity vector field at any point $\overline{x}$, defined in the case of 2D images as:

$$\mathrm{v} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \delta x/\delta t \\ \delta y/\delta t \end{pmatrix} \tag{6.5}$$

The equation 6.4 is valid at any point in the image. It represents the Optical Flow Constraint Equation. It is the fundamental equation of the optical flow problem.

Adopting suffix notation:

$$I_t = \frac{\partial I}{\partial t}$$

,

$$I_x = \frac{\partial I}{\partial x}$$

,

$$I_y = \frac{\partial I}{\partial y}$$

we can rewrite 6.4

$$I_x u + I_y v + I_t = 0 \tag{6.6}$$

represent the classical definition of OFCE due to Horn and Schunck.

6.1.1 Geometric interpretation of OFCE: the aperture problem

In the 2D space, the optical flow equation identifies a line in the velocity plane (u, v) where every point $\overline{x}$ in the image must satisfy that their velocity vector $v(\overline{x})$ must reside in that line (see figure 6.1). This is an ill posed problem where the result is an infinite family of possible solutions $(\mathrm{v}_a, \mathrm{v}_b, \mathrm{v}_c, ...)$. There are more unknown parameters than equations. We have two unknown (u,v) and only a single linear equation. For each point, only $\mathrm{v}_\perp$, the component parallel to the intensity gradient can be calculated. This phenomenon is know as the aperture problem [81], and only the regions of the image where it has some sort of texture (non homogeneous intensity, all the partial derivative non zero) the vector field can be estimated (see figure 6.2).

Knowing that $\mathrm{v}_\perp$ of the vector field v is parallel than ∇I

$$\mathrm{v}_\perp = s\hat{\mathrm{n}} = \frac{-I_t \nabla I}{\|\nabla I\|^2} \tag{6.7}$$

where:

$$s = s(\overline{x}, t) = \frac{-I}{\|\nabla I_t\|}$$

,

$$\hat{\mathrm{n}} = \hat{\mathrm{n}}(\overline{x}, t) = \frac{\nabla I}{\|\nabla I\|}$$

represent the modulus and direction of the normal component of velocity at any point $\overline{x}$.

Then, the *optical flow constraint* just is enough to determine the component of the flow field in the orthogonal direction of the image gradient but not to retrieve the entire vector field. Vemuri et al. [82] overcame this problem by developing a neat and elegant surface evolution approach to achieve the smooth deformation field between two 3D images expressed in a level-sets framework. Registering two consecutive 3D images I_1 and I_2 is equivalent to determine the evolution of the level-sets of I_1 along its normal direction ∇I until it becomes the target image I_2. This evolution can be written as:

$$\begin{aligned} I_t(\overline{x}) &= S\,\|\nabla I_t(\overline{x})\| \\ \text{with } I(\overline{x}, 0) &= I_1(\overline{x}) \end{aligned} \tag{6.8}$$

where S is the velocity term. Choosing the velocity term S equal to $I_2(\overline{x}) - I(\overline{x}, t)$ makes this curve evolution stop when the image I reaches the level-sets of the target image I_2. Equation (6.8) does not give explicitly the transformation vector field between the two images that can be achieved using an analogous surface evolution in

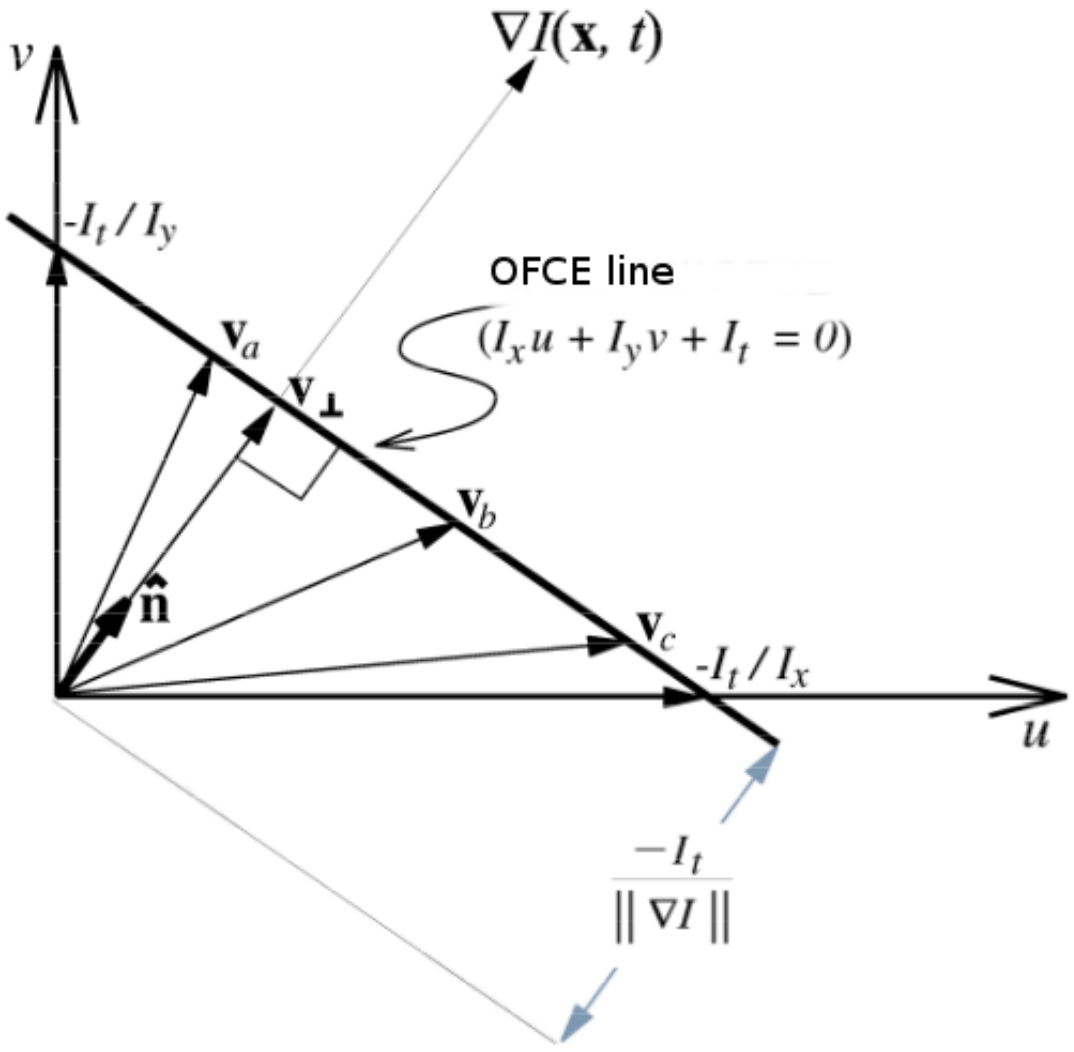


Figure 6.1: Geometric interpretation of OFCE. The OFCE define a line in the velocity plane perpendicular to ∇I. Each velocity vector under the line is a possible solution to the problem. The normal velocity $\mathrm{v}_\perp$ is the vector with minimum modulus.

vector form:

$$\overline{V_t} = (I_2 - I(\overline{V}(\overline{x},t)))\frac{\nabla I(\overline{V}(\overline{x},t))}{\left\|\nabla I(\overline{V}(\overline{x},t))\right|} \qquad (6.9)$$

with $\overline{V}(\overline{x},0) = \overline{0}$

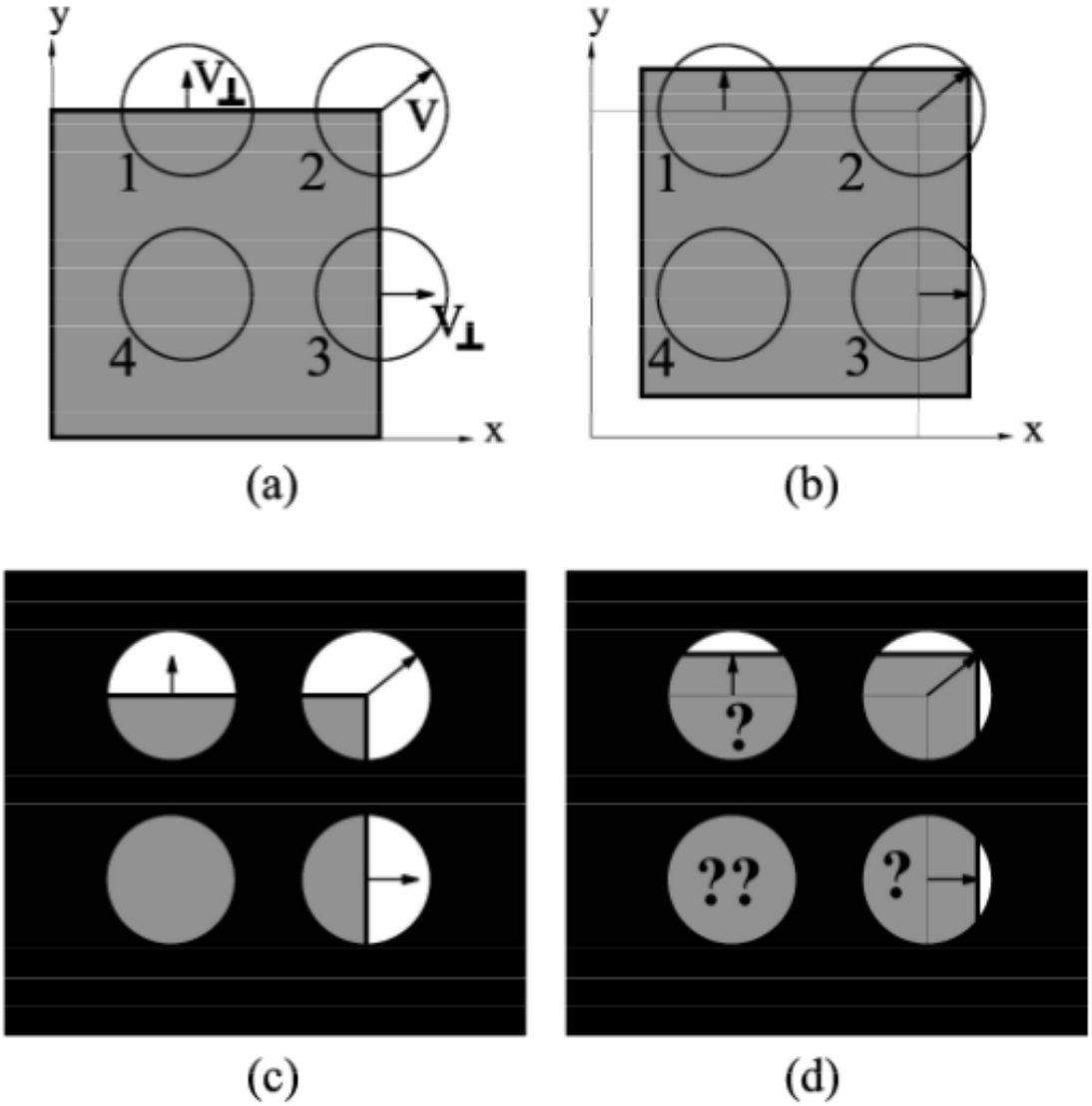


Figure 6.2: Aperture problem: looking at the holes 1 and 3 only is possible to retrieve the motion normal to the square because of the lack of local structure in the image ($I_x = 0$ in 1 and $I_y = 0$ in 3). In the hole 4, $I_x = 0$ and $I_y = 0$ then $I_t = 0$ it is no possible to observe any motion. Hole 2, is the only one where is possible to observe a complete bi-dimensional structure, therefore is possible to retrieve the motion.

where $\overline{V}(\overline{x}) = (x+u, y+v, z+w)$. Since the movement of cells in the embryo depends locally on the neighboring cells, the vector field we are expecting should have a regular (smooth) deformation. In order to achieve a smooth vector field the images are convolved with a Gaussian kernel G_σ before taking its gradient, therefore, expressions (6.9) and (6.8) are modified into:

$$\begin{aligned} I_t(\overline{x},t) &= (I_2 - I(\overline{x},t)) \left\| \nabla G_\sigma * I(\overline{x},t) \right\| \\ \text{with } I(\overline{x},0) &= I_1(\overline{x}) \text{ and} \end{aligned} \tag{6.10}$$

$$\overline{V_t} = (I_2 - I(\overline{V}(\overline{x},t))) \frac{\nabla G_\sigma * I(\overline{V}(\overline{x},t))}{\left| \nabla G_\sigma * I(\overline{V}(\overline{x},t)) \right|} \tag{6.11}$$

with $\overline{V}(\overline{x},0) = \overline{0}$

6.2 Algorithms testing

The aim of the study is evaluating and comparing the abilities of four non-rigid registration algorithms to capture cell movements in zebrafish early development image sequences. All registration algorithms have in common to compute a motion field that deforms a moving image I_{mov} into a reference image I_{ref} . To that aim they are representing I_{mov} in a given transformation model which imposes cinematic constraints for further deformation. Thereafter I_{mov} is deformed by solving a minimization problem on a criterion E that will depend either locally or globally on both I_{mov} and I_{ref}. Thus registration algorithms may differ in the choice of the transformation model imposing cinematic constraints on deformation (elastic like or fluid like for instance). They may also differ in the choice of the optimization strategy. The algorithms chosen here present different implementation of similar principles. Their main characteristics regarding the registration of two consecutive images are described below.

- MIA fluid registration [83]: it is a non-rigid viscous fluid registration. The core algorithm in the fluid registration is based on a linear elastic deformation of the fluid velocity field. This means that viscous constraints are imposed to the velocity field. Concerning the optimization part, the algorithm is a so called

force based model. The forces applied to the pixels of I_{mov} are equivalent to the minimization of a criterion. They are obtained by calculating the gradient of the criterion. The criterion here is a squared difference of I_{mov} and I_{ref} intensity. The algorithm is implemented in a multiscale framework. We used the algorithm implementation available in MIA open source toolbox [84].

- ITK Thirion demons optical flow [85]: it is a non rigid registration method based on optical flow. It is also a force based method. Here the forces are measured thanks to optical flow equation. This one can be interpreted as the conservation of object intensity when the object moves. It has been shown that demons algorithm is an approximation of the fluid registration algorithm with a normalized version of the forces used (meaning that the criterion $E(I_{mov}, I_{ref})$ to optimize is also different)[86]. It is also suggested that the approximation leading to demons implementation may cause problems in terms of topology and stability of the fluid model. In this study we use ITK open source library implementation of Thirion's algorithm [61].

- ITK levelset optical flow [87]: it is a non rigid registration algorithm also based on optical flow forces computed in a levelset framework. Vemuri established existence and uniqueness of the solution for the evolution model in a Sobolev space as opposed to using viscosity methods. ITK's implementation was also used for this algorithm.

- BSpline elastic registration [88]: it is a non rigid parametric motion estimation algorithm. The deformation field is represented using a parametric model based on B-spline spatial basis functions. These functions are used because of their implicit smoothness (minimum curvature properties). Thus B-spline functions impose field smoothness. One can also control the rigidity of the solution with B-spline parameters. The solution to our registration problem is a deformation field that minimizes the criterion $E(I_{mov}, I_{ref})$. This is found by using a multidimensional optimization algorithm acting on the parameter of the spatial model (i.e. spatial basis function position). The optimization algorithm used is gradient descent using an analytical representation of the gradient and following a multiresolution strategy both in the image and transformation spaces.

6.3 Evaluation dataset

The sample volumes 3D image series were obtained through time-lapse biphoton laser scanning microscopy from live embryos engineered to fluorescently label nuclei and membranes. Our experimental data figure 6.3 corresponds to the top part (100μm deep) of the embryo imaged from the animal pole during gastrulation.

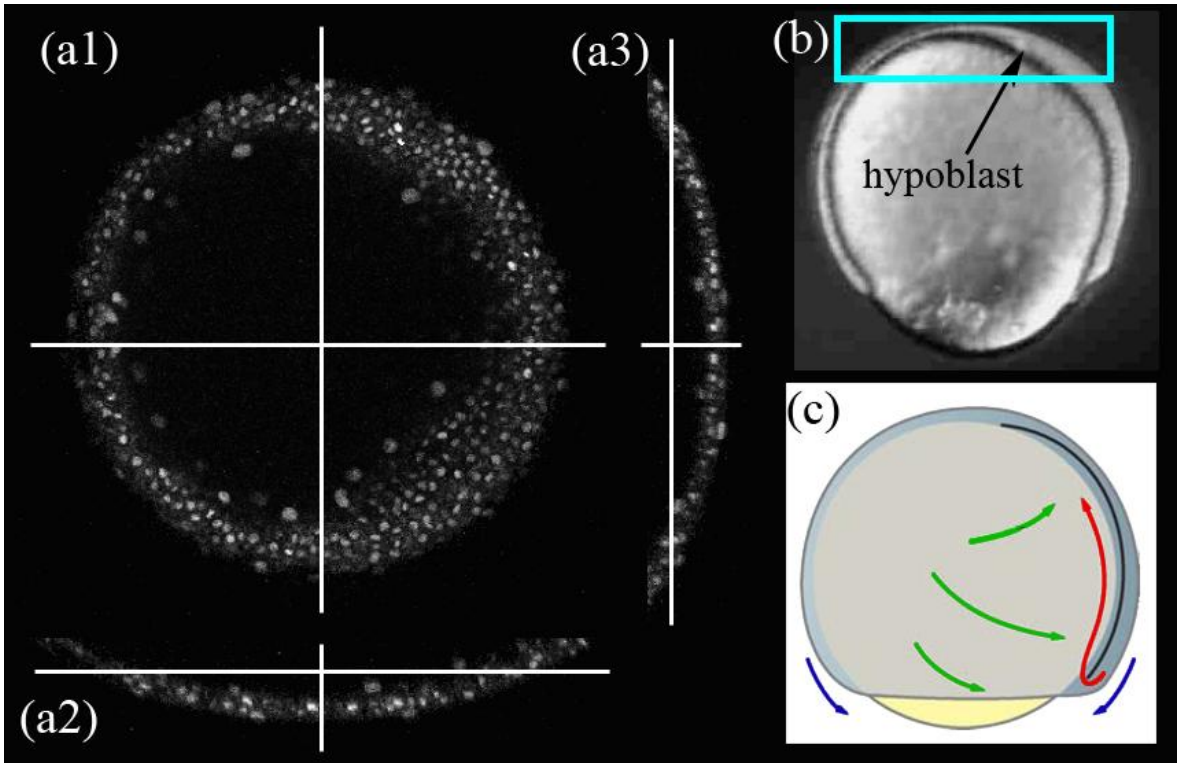


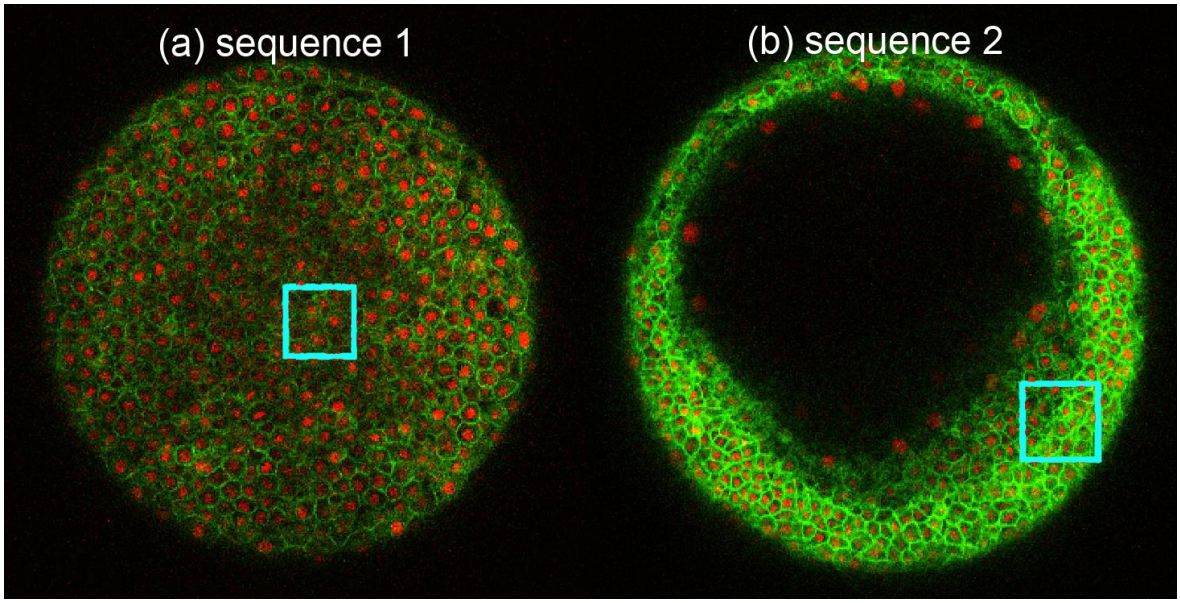
Figure 6.3: The dataset exhibits the top part of a zebrafish embryo imaged from the animal pole. (a1) (a2) and (a3) respectively xy (animal pole view) xz and yz cut of nuclei image at 8hpf (75% epiboly). (b) Lateral view of the embryo at 8 hpf. The arrow points to the hypoblast leading edge on the dorsal side of the embryo. The dataset (blue frame) covers a period starting with epiboly and ending with gastrulation. (c) Schematic representation of gastrulation movements. Tissues undergo during gastrulation a combination of movements to simultaneously cover the yolk (epiboly, blue arrow), migrate towards the animal pole (red) and converge towards the midline (green).

Two test sequences consisting of ten consecutive subvolumes (about 50μm3 each) have been extracted from this dataset. The first one taken in the middle of the embryo

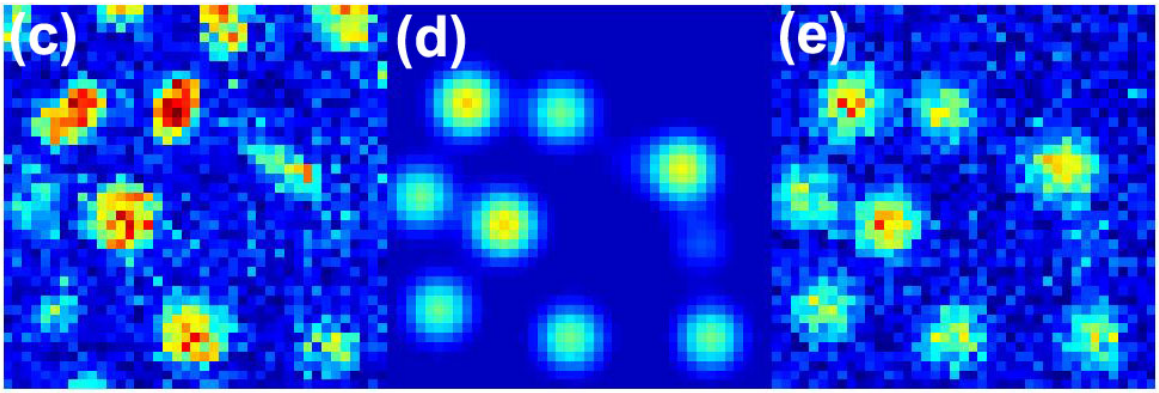
at the end of doming period contains about 30 cells (see figure 6.4,a). Their movements are small, most of them are between 0 and 6μm, some of the cells move more that 10μm. The nuclei mean diameter is about 8μm. The second volume period is matching with the section displayed in figure 6.4,b. It contains cells of the epiblast (outer layer of the embryo) and cells of the hypoblast (inner layer). Both tissues undergo during gastrulation a combination of movements to simultaneously cover the yolk, migrate towards the animal pole and converge towards the midline. The sub- volume covers the boundary between epiblast and hypoblast where a movement discontinuity is expected. The subvolume contains about 70 cells, the nuclei are much smaller and move faster than in the first test sequence. Both test sequences data have a spatial resolution of 1.37μm and a temporal resolution of 49 seconds. The noise level is quite high (see figure 6.4) so that the only detectable subcellular structure are the nuclei. Some nuclei may touch each other (optical artifact) which also makes more difficult the task of registration.

6.4 Registration evaluation protocol

Registration results contain the information to follow image structure. The principle of our registration assessment is to compare movement of structures measured by registration to a gold standard of cell movements, entirely checked manually. Starting from an image sequence of zebrafish, nuclei positions are extracted and followed in time. Nuclei trajectories remaining in our test volumes during all the sequence (ten time steps) have been selected and manually checked to avoid any error. These nuclei trajectories are our gold standard. This reference is available for comparison with any registration result on our test volume. Registration is performed on pairs of consecutive images (I_i,I_{i+1}) and produces a vector field ($F_{i,i+1}$) for each of them. $F_{i,i+1}$ enables to transform (x_i, y_i, z_i) coordinates at time i, into the new positions $(x_{i+1}, y_{i+1}, z_{i+1})$. Therefore, applying $F_{1,2}$ at the initial position in the gold standard, we obtain the new cell positions calculated from registration at time 2. Measuring the difference between gold standard position and the registered position of the nuclei gives the registration method error. The computed trajectories are built by applying the fields iteratively to the positions obtained. The error measured between registered nuclei and gold standard is the accumulation of error from the starting point. This protocol is schematized



(a)



(b)

Figure 6.4: Test volume presentation: (a) the first volume for registration evaluation (blue frame). It lasts 10 time steps during the end of doming period (5 hpf). The second test volume is taken at 8hpf (75% epiboly). (c) A cross section a volume 1 shows the presence of noise and typical nuclei size (8μm). (d) Placing a gaussian blob at the position of every detected and tracked cell allows constructing artificial data with the same reference trajectories as real data. (e) Adding noise mimics real data aspect.

in figure 6.5. A perfect registration method should give an error for each trajectory at each time step of 0μm. We note $Xref_t^i$ the gold standard i (reference) position of cell i at time t and $Xreg_t^i$ the position obtained by registration of cell i at time t. We thus define the absolute distance error to gold standard at time t of cell i:

$$\delta_t^i = ||Xref_t^i - Xreg_t^i|| \tag{6.12}$$

We also define D_t^i the displacement of the cell i from time 0 to t if no registration is done:

$$D_t^i = ||Xref_t^i - Xref_0^i|| \tag{6.13}$$

It is the distance of gold standard trajectories to their starting point. Finally we define the relative error on registration as the ratio of these two distances:

$$\delta_{trel}^i = \frac{\delta_t^i}{D_t^i} \tag{6.14}$$

Apart from plotting the trajectories and the deltas in the real 3D space we measured the δ_t^i distributions at each time step and calculated the distribution mean: $\Delta_t = \frac{1}{N}\Sigma_i\delta_t$ where N is the number of cells).

In addition the δ_t^i distribution allows answering other important questions such as: "Are the errors widely spread around their mean value ?", "What is the percentage of trajectories below pixel size?", etc. Thus confidence interval measures are also used in results analysis. We call $\Delta_{t,k}$ the interval containing the k% smallest δ_t^i . Knowing that δ_t^i is positive and that we are interested in having the smallest possible value for it, the lower bound of the interval is not very important. So we finally define $\Delta_{t,k}$ as the upper bound of the interval containing the k% smallest δ_t^i . This measure provides a meaningful evaluation of registration results. Similarly we call $\Delta_{trel,k}$ the relative counterparts of $\Delta_{t,k}$.

Additionally, some artificial sequences of images have been created from scratch in order to maximize the coherence between reference trajectories and the data to register. Indeed there can be uncertainties in the reference trajectories. Errors remaining even after hand validation, on cell nuclei position for instance, can bias the error

measurement. In our case the nuclei displacements are small (1,2,3 pixel in length). As a consequence a one pixel error on nuclei centre position can significantly modify the observed nuclei displacements and consequently deteriorate artificially registration error evaluation. The two artificial sequences are built in order to match exactly the trajectories of sample test 1 and 2. Starting from ten empty volumes, gaussian blobs are written down at each position of the reference trajectories. Gaussian standard deviation of blob are compatible with cell radius size (4μm) (figure 6.4-d). Afterwards gaussian and poissonian noise are added to mimic real empirical data aspect (see figure 6.4-e). Finally, the artificial sequence built have the same gold standard as real data with the advantage of a perfect correspondence of artificial nuclei centre position with the gold standard.

6.5 Results

In order to optimize the registration process, we investigated the influence of sources and filters. In this section we present and discuss the registrations results obtained

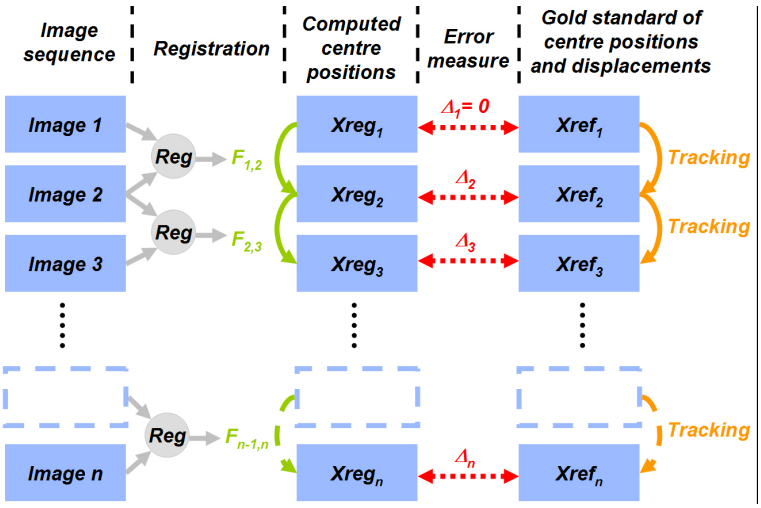


Figure 6.5: This protocol allows to quantify registration result. Given a set of nuclei at t_0 and vector fields $F_{i,i+1}$. We apply F_0^n to the initial set of cells and gives registered nuclei at time n. Error between gold standard and computed trajectories are measured at every time step (Δ_t). The lower the error values are, the better the algorithm performance is.

with the four methods to be compared. Before applying the registration, we apply some noise and filtering algorithm to evaluate their robustness. Then we compare the 4 registration methods results, analyze the error distribution and accumulated error evolution. Finally we discuss the results obtained with real data.

6.5.1 Image preparation

In order to optimize the registration process, we investigated the influence of noise in the images. Images are noisy (see figure 6.4-c,e) and filtering may help diminishing registration error. Two different filters have been applied:

- Gaussian difference smoothing (filter 1). The filter smooth the image and removes small scale image intensity variations.
- Perona-Malik (filter 2). The filter reduce image noise without removing significant parts of the image content.

Registrations have been computed with 4 different data preparation:

- (a) artificial data with noise added and without filtering
- (b) artificial data with noise added with filter 1
- (c) artificial data with noise added with filter 2
- (d) artificial data without noise and filtering

The figure 6.6 shows the measurement of Δ_t. MIA filter exhibits sensitivity to noise. It is possible to observe that the reference (sequence d) is the better registered than sequence a (the reference plus noise). Both filters decrease Δ_t values, filter 1 being better (result almost as good as the reference). With ITK demons and ITK levelset methods there are no difference between the four datasets, given that inside the filter itself the images are smoothed. Thus noise is always filtered out. MIA and BSpline not use any filter inside, use all the information available, even the misleading noise. Consequently, they are more sensitive to noise. In the next experiment gaussian filter is always used for noise removal. The images used in these experiment are the nuclei channel, the membranes images are not used.

6.5.2 Algorithms comparison

Artificial data sequences were analyzed with the 4 methods. We added noise and filtered the artificial data with the gaussian filter. We successively warped the first image with the vector fields obtained after registration. The resulting sequence showed qualitatively the error accumulation after each iteration. We used the warper of MIA toolbox with b-spline of degree 3 interpolator. The resulting sequences conserved the nuclei shape in time. Some degradation is noticeable but becomes important only after 4 or 5 time steps. BSpline performs the best between the algorithms. MIA seems to dissolve the structures, on the contrary ITK demons seems to contract them.

A qualitative evaluation can be observed superimposing the gold standard and the computed nuclei trajectories (fig. 6.7). On vol1 we observe that the 4 algorithms correctly estimate nuclei displacements. The errors seem homogeneous amongst cells; one unique cell is more difficult to follow. MIA registered trajectories are slightly noisier around the reference.

In order to have a more concrete evaluation of algorithm performance, we plotted the 80% confidence curve ($\Delta_{t,80}$) for each algorithm (fig. 6.8). All the algorithms are below or at 1 pixel accuracy even after ten time steps. This means that after ten

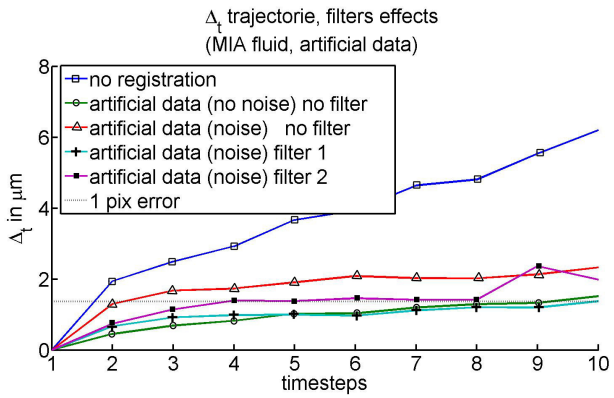


Figure 6.6: Data pre-processing: This figure illustrates the effects of filtering out noise before registration. The mean absolute error of registration (Δ_t) is plotted against time for MIA registration performance on artificial volume 1 (artificial volume sequence matching reference trajectories of volume 1). The smallest the (Δ_t), the best is the registration. Noisy data gives registration error (red) bigger than data without noise (green). Data prepared with filter 1 (gaussian difference convolution, clear blue) gives an error almost as small as data without noise.

time steps of registration 80% of the estimated trajectories have an error of one pixel or less. Meanwhile the D_t is increasing up to 6μm. The curve shapes are similar: a fast increase of $\Delta_{t,80}$ at the first registration step after which the error slowly diverges. Only BSpline is behaving differently. Indeed the slope of the $\Delta_{t,80}$ is constant at all time steps. Registration with the second artificial sequence gives analogous results with $\Delta_{t,80}$ values a bit higher. This result was already anticipated in the qualitative observations. All in all, the four algorithms give good results with error within one pixel accuracy even after ten time steps. It appears that BSpline has better results with small accumulation of error (up to time step 5) and that ITK demons is better after long accumulation (after time step 6). If one is interested in pixel accuracy registration, then the 4 algorithms are equivalent for the nuclei trajectories gold standard that we used. BSpline and ITK demons algorithms have singular behaviour in error accumulation. We now detailed them (fig. 6.9-b,c).

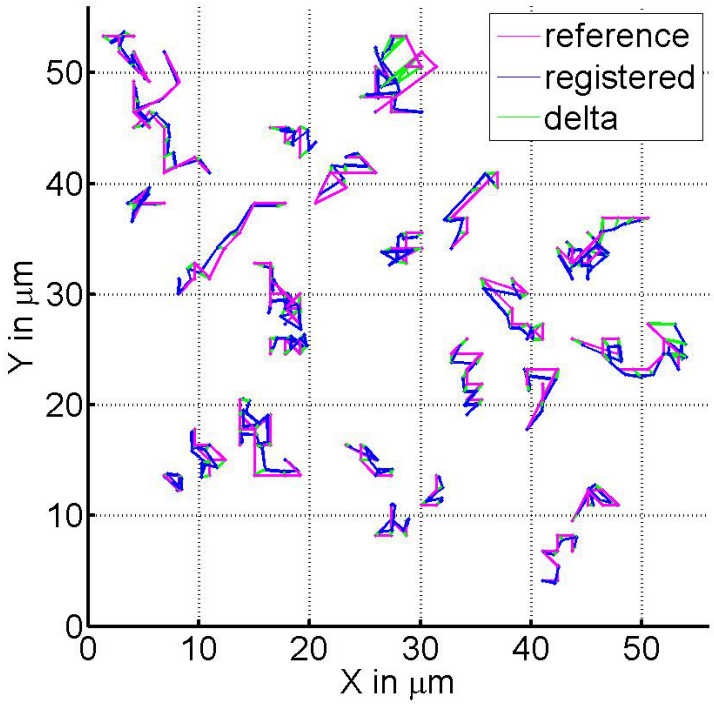


Figure 6.7: Superimposition of nuclei trajectories (gold standard trajectories of volume 1) and the trajectories calculated by registration (ITK demons method). The view is projected in xy plane. Almost all the trajectories are registered with pixel accuracy.

- BSpline: absolute error is growing regularly along the ten time steps. The δ_t^i distribution is very peaked around its mean at the beginning and spread continuously as error is accumulated.

- ITk demons: absolute error has a jump after the first registration and is then constant. The δ_t^i distribution, more spread around its mean than with BSpline at the beginning, is shrinking along time as error accumulate. with ITK demons than with BSpline.

If absolute error is small regarding image sampling it does not say anything about relative error on each cell displacement. Indeed, the relation between relative error and absolute error distribution is not direct. We take again the example of BSpline and ITK demons to illustrate this point (fig 6.9,d,e):

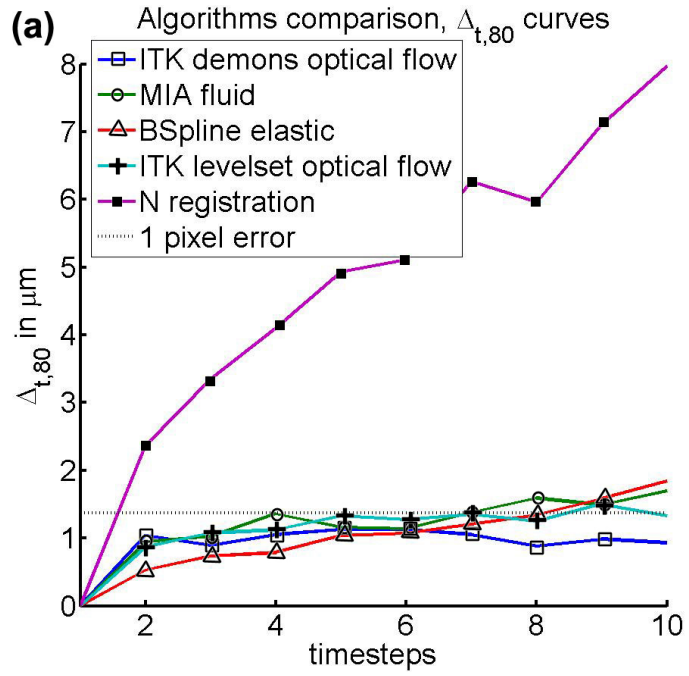
- BSpline: the mean relative error (Δ_{trel}) is constant in time. The distribution, at least for the 90% best cells, is also rather constant. The accuracy for the 90% best registered trajectories is of 35% of relative error on displacement evaluation and can be as small as 5% for the best registered cell. These results are true at every time step of the sequence registration.

- ITk demons: the mean relative error (Δ_{trel}) is following the decrease of the 1 pixel relative error curve (doted line). The support of the $\Delta_{t,90}$ interval is shrinking with time. After ten timestep $\Delta_{trel,80}$ shrinks from 50% to 15%.

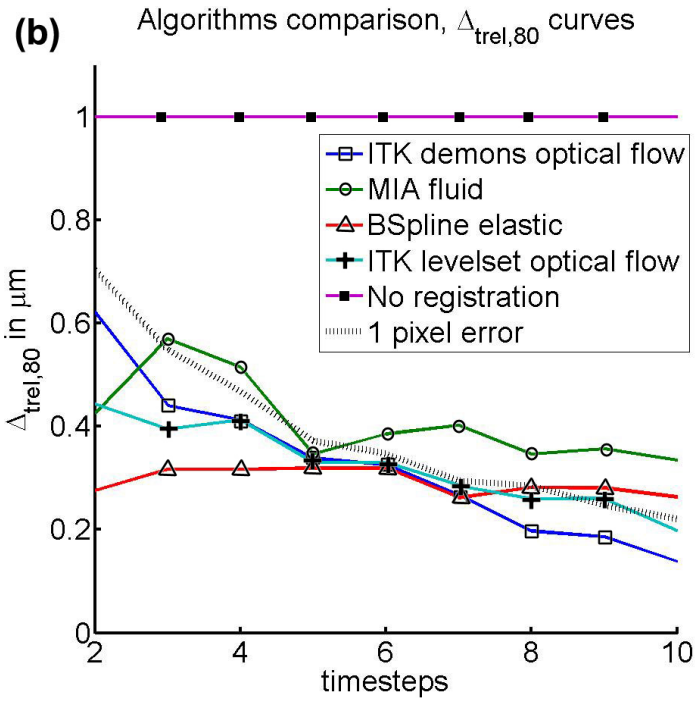
We are probably limited here by the small extent of the nuclei displacements and the fact that algorithm cannot be very good doing subpixel registration. Improving relative accuracy should be possible with higher sampling or in the case of ITK demons by increasing the accumulation of displacement.

6.5.3 Real data registration

Previous results were obtained with artificial data sequences. Here are some comments on real data registration quality. We performed the registration of the 2 real data sequences with each of the algorithms. Qualitative results seemed as satisfying as



(a)



(b)

Figure 6.8: The artificial volume 1 registration is performed for 4 different registration algorithms. The $\Delta_{t,80}$ curves (a) (under which value are 80% of the individual cell registration absolute error) are plotted. All methods have pixel accuracy even when error has been accumulated for ten time steps. Graph (b) shows $\Delta_{trel,80}$ the relative error on registration. As movements are small and given the pixel accuracy of algorithms, relative error remains satisfactory.

with artificial data. Concerning Δ_t measure the scenario is the same for the four algorithms. At the first registration step Δ_t is jumping from 0 to approximately 1.5μm, a value higher than what is obtained with artificial data after 10 time steps of error accumulation. Thereafter, Δ_t increases very slowly. The Δ_t^i distribution is larger than with artificial data. These results are coherent with the hypothesis that our nuclei detection contains some errors on their position. The registration error measured in artificial data is lower than a pixel, as a consequence adding one pixel randomly to registration error may totally change registration error measurement. To test our hypothesis we plotted Δ_t for real data (test volume 1) and corresponding artificial data registration. We also draw the Δ_t curve for artificial data but then using a modified reference by adding noise to the position of cells. The graph can be seen on figure 6.10. The Δt curve of artificial data registration using a noisy reference has a shape similar to the one of real data. This strengthens our hypothesis and points the difficulties of assessing the registration algorithm directly on real data. We are expecting real data registration to be as good as it is on artificial data.

6.6 Conclusions

Four non-rigid registration methods coming from biomedical field have been tested on early zebrafish development sequences. In order to estimate the registration quality we used the tracking of reference structures (nuclei here). We compared trajectories of these structures obtained with and without registration and measured the errors in between. The study of these errors gives a good description of algorithms performance. We are able to measure absolute and relative accuracy of the algorithms along with their confidence intervals. Registration methods were applied to artificial data mimicking zebrafish volume sequences. The 4 registrations methods errors are at the level of the pixel and below even after 10 time steps of errors accumulation. In particular, BSpline methods leads to the smallest error after small accumulation and ITK demons after long accumulation. These results are encouraging for nuclei registration.

These results encourage to apply the method to the set of cells of each time step to facilitate the tracking of each cell (see figure 6.11).

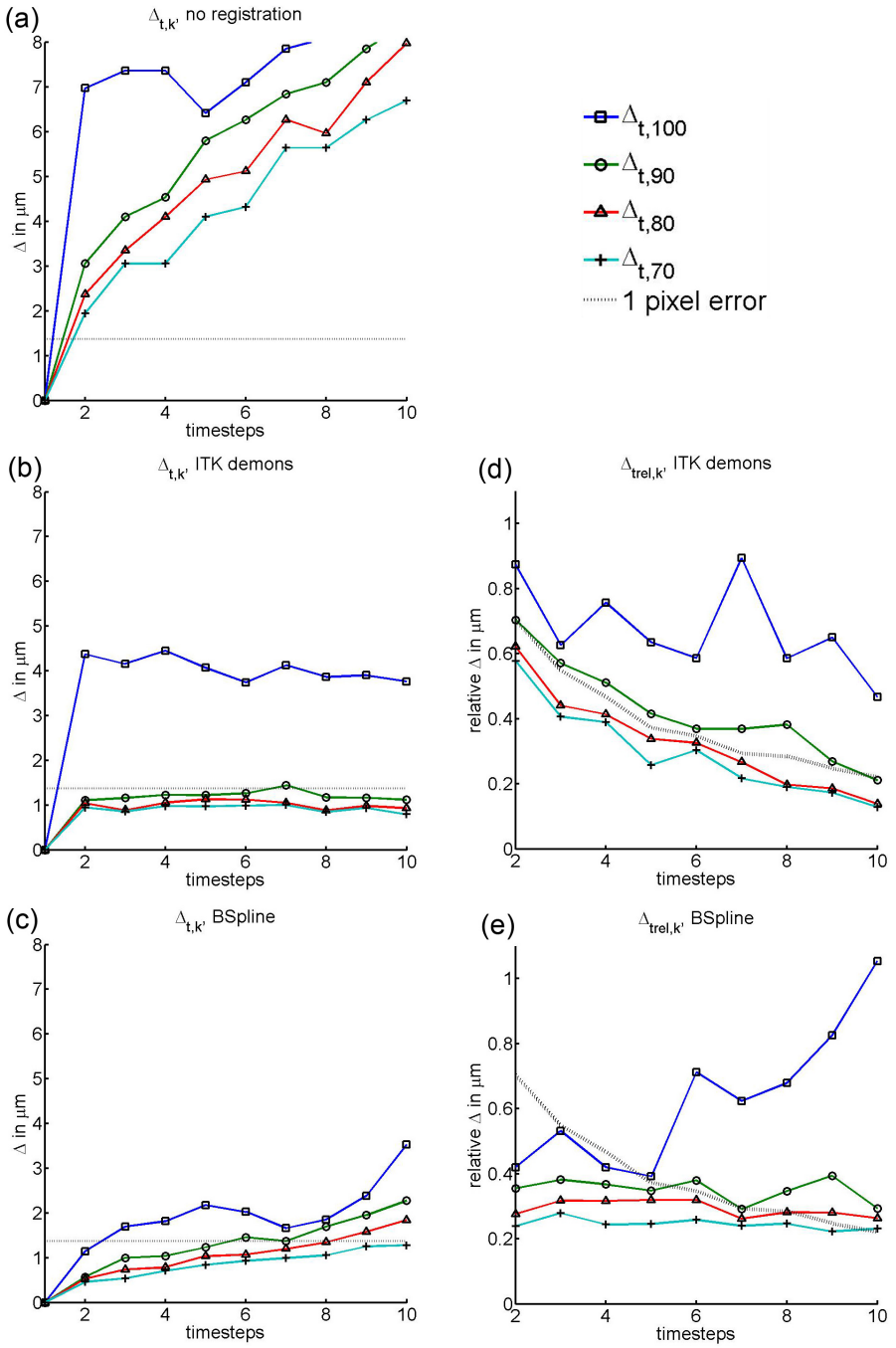


Figure 6.9: Details on BSpline and ITKdemons artificial volume 1 registration performance. (a) Is a reference showing Dt, k. (b), (d) respectively are the absolute and relative error for ITK demons registration. (c), (e) are the analogue for BSpline algorithm. BSpline exhibits excellent results in terms of absolute error support. ITK demons has interesting results in terms of relative error accuracy.

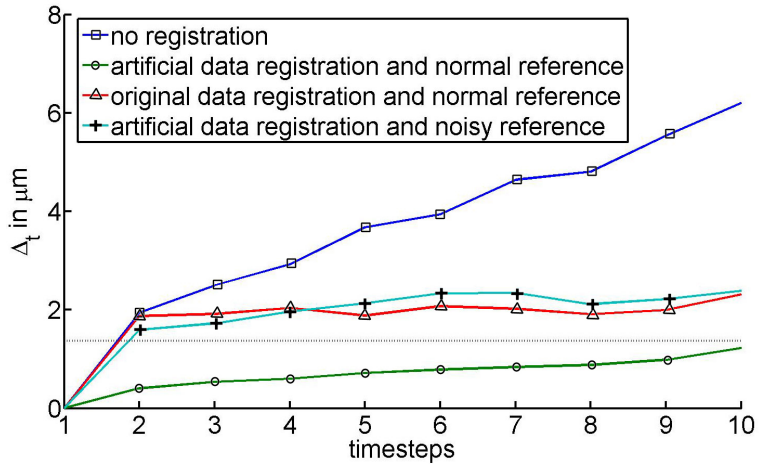
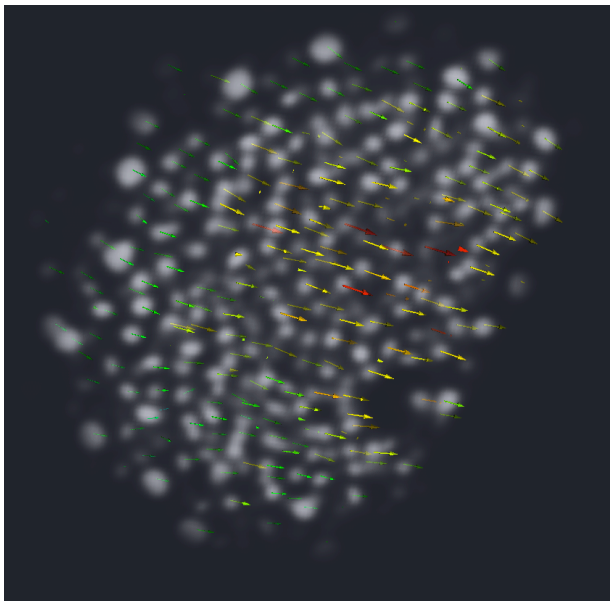
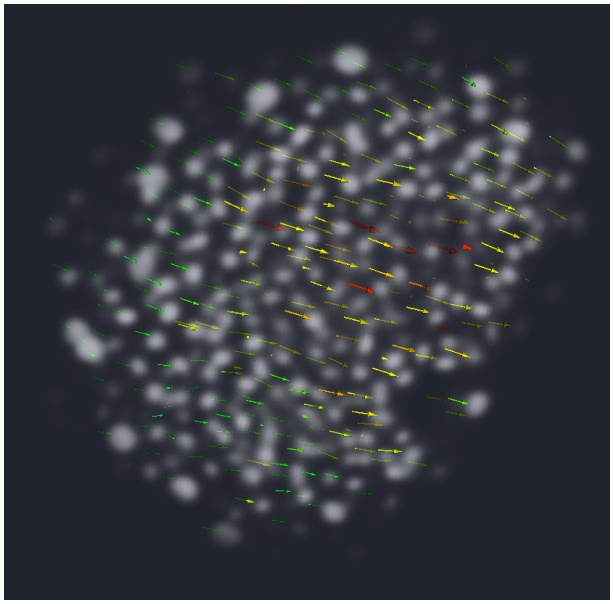


Figure 6.10: This figure illustrates the effect of noise in reference trajectories. Real data and the corresponding artificial data have been registered with BSpline method. With artificial data registration and a gold standard where noise was added on nuclei positions, Δ_t is at the level measured on real data (red).



(a)



(b)

Figure 6.11: Smoothed version of the dataset 070128c, the vector field was acquired using Demons registration method. The vectors with same coordinates of cells are shown, the color of the arrows depends on arrows length. Frame A time step 0, Frame b time step 1, we see that the arrows ends at the position of the cells at next timestep.

Chapter 7

Cells tracking in a live zebrafish embryo

Understanding the cell morphodynamics underlying morphogenetic processes is a fundamental issue for bio-medical research. Such a goal can be achieved through the automated tracking of cell nuclei and cell divisions from 3D+time in vivo imaging [1]. These tasks are the basis for the reconstruction of the cell lineage tree described as the branching process of cell divisions and its deployment in space and time. The complete reconstruction of the cell lineage tree from the egg cell to the adult stage has only been achieved for the worm *Caenorhabditis elegans*. However, in that case, the total cell number in the adult is less than one thousand and the cell lineage is largely invariant. Because of their very large cell number, this challenge has not been taken up so far for vertebrate organisms. Nevertheless, recent advances in imaging strategies open the way to in toto 3D plus time imaging providing data suitable for in vivo cell tracking and cell morphodynamics reconstruction.

The zebrafish (*Danio rerio*) is a vertebrate model that has been chosen for its transparency allowing in vivo inspection at the cellular level deep into the tissues by confocal laser scanning microscopy [2]. The zebrafish exhibits typical vertebrate differentiated cell types and has been largely validated for investigations related to humans including cancerogenesis and a number of genetic diseases [3]. Achieving the automated reconstruction of the zebrafish embryo cell morphodynamics is highly relevant for investigating stem cells populations, early steps of cancerogenesis or drug effects in vivo. Such a goal requires engineering live zebrafish embryos to highlight sub-cellular structures to be imaged by time lapse laser scanning microscopy, designing image processing algorithms and computational methods.

Recent works reveled outstanding advances in image microscopy [89] providing experiments with minimal photobleaching and phototoxicity where the images have super-resolution. On the other hand their work do not provide the details of the algorithms applied.

We propose in this chapter a new tracking algorithm based on a simulated annealing method which take into account the informations provided by cell detection, segmentation and optical flow algorithms and some biological constraints. New results on morphogenetic fields detected in early embryogenesis stage are given and discussed.

To track each cell in the embryo a workflow of image processing algorithms is applied to 3D+t zebrafish confocal images. This algorithm include image denoising, nuclei and membrane segmentation, optical flow methods and function minimization to identify

the same cell at different timesteps.

Image enhancement and denoising is a critical step in order to preserve the shape information in digital images. Furthermore, in laser scanning microscopy imaging, noise is due to different causes, and is difficult to model. In particular, the variation of the fluorescent molecules concentration (synthesized from the injected RNA) superimposes a de facto noise level that is difficult to model. The use of traditional preprocessing algorithms (moving average, median and Gaussian filtering) do reduce the noise superimposed on the image but do not maintain a good definition of the edges or image features. The goal of the filtering should be to reduce the noise from the images without loosing useful details like the edges.

In order to recognize all the cell nuclei in each 3D image with a fully automated procedure we used a 3D version of the Hough transform for the identification of spherical shapes. The Hough transform is an algorithm which can be used to isolate features of a particular shape within an image [45]. It is commonly used for the detection of regular curves such as lines, circles or ellipses. The algorithm uses the duality between points on a curve and parameters of that curve.

Optical flow techniques have been used to estimate pixel correspondence between images obtained at two different times. Therefore, optical flow can give important information about the movement of the objects in the images [80]. This movement is represented by a vector field $V : D_I \to \mathbb{R}^3$. The brightness of the cells in the confocal volume of the zebrafish embryo remain almost constant over time. This is the base for the level set formulation developed by [80] and Vemuri et al. [82].

They developed a neat and elegant surface evolution approach to achieve the smooth deformation field between two 3D images expressed in a level-sets framework. Registering two consecutive 3D images I_1 and I_2 is equivalent to determine the evolution of the level-sets of I_1 along its normal direction ∇I until it becomes the target image I_2.

Since the movement of cells in the embryo depends locally on the neighboring cells, the vector field we are expecting should behave as a smooth deformation. In order to achieve a smooth vector field the images are convolved with a Gaussian kernel G_σ

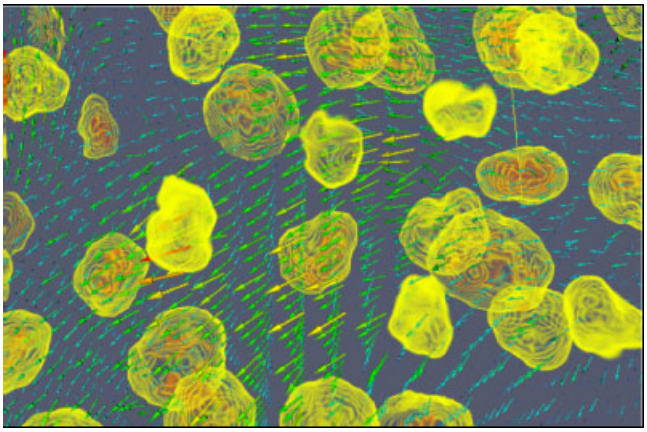


Figure 7.1: 3D optical flow of consecutive frames. The vector field represented with colored arrows is an estimation of the movement of the cells. Parameters: 10 iterations with $\sigma = 4$.

before taking its gradient:

$$\begin{aligned} I_t(\overline{x},t) &= (I_2 - I(\overline{x},t))\,\|\nabla G_\sigma * I(\overline{x},t)\| \\ \text{with } I(\overline{x},0) &= I_1(\overline{x}) \end{aligned} \tag{7.1}$$

The vector formulation of the image deformation could be expressed with the equation:

$$\overline{V_t} = (I_2 - I(\overline{V}(\overline{x},t)))\frac{\nabla G_\sigma * I(\overline{V}(\overline{x},t))}{\left\|\nabla G_\sigma * I(\overline{V}(\overline{x},t))\right\|} \tag{7.2}$$

where $\overline{V} = (u,v,w)$ and $\overline{V}(\overline{x}) = (x+u, y+v, z+w)$ is the displacement vector at X. This curve evolution stop when the image I reaches the level-sets of the target image I_2, $\overline{V}(\overline{x},0) = \overline{0}$ and the gradient is approximated by using the upwind schemes [90].

Segmenting the nuclei and membrane shape allows us to understand the interactions between each cell and their neighbor cells, a further study of their shape could provides us the characteristic parameters of the cell at each stage in the cell cycle. The segmentation method we use in this work is a generalized version of the Subjective Surfaces technique [56, 57],it is distinguishable from the classic formulation by the different weights applied on the two flows constituting the motion equation (curvature

and advection).

The algorithm extract the membranes and nuclei shapes in the acquired images by processing the two channels separately. Composed by a chain of image processing operations, the method is completely automated. Membrane segmentation is used to detect mitosis, given that two nucleus are inside a single membrane.

7.1 Cell tracking

The cells move in the embryo, they migrate, divides and differentiate to form structures. A cell lineage tree can be interpreted as a binary tree where the first cell of the embryo is represented by the root node and the relationship of mother cells and the two daughter cells are represented by edges.

While the embryo grows the cells proliferate and become smaller. The speed of cells could reach 40μm/min while the size of the cell decrease from 30μm radius at 30%-epiboly stage to 20μm or less at 75%-epiboly, usually cells do not overlap their volume in consecutive timesteps. There is a relative smooth movement between cells that belongs to the same layer, while this relative speed increase in neighboring cell belongings different layers.

In biology, mitosis is the process of chromosome segregation and nuclear division that follows replication of the genetic material in eukaryotic cells. This process assures that each daughter nucleus receives a complete copy of the organism's genome. Mitosis is divided into several stages, with the remainder of the cell's growth cycle considered interphase. Properly speaking, a typical cell cycle involves a series of stages: G1, the first growth phase; S, where the genetic material is duplicated; G2, the second growth phase; and M, where the nucleus divides through mitosis. Mitosis is divided into prophase, prometaphase, metaphase, anaphase, and telophase. The Mitosis is characterized by the fact that the nucleus is splitting in two different nuclei and both nuclei are inside a single membrane until the cytokinesis phase ends. This stage could be recognized by segmentation algorithms as seen in chapter 5 and the two newborn cells moves is relative opposite direction until they are dragged by the surrounding cells.

There is also a cell process called apoptosis which is a form of cell death in which a programmed sequence of events leads to the elimination of cells without releasing harmful substances into the surrounding area. Apoptosis plays a crucial role in developing and maintaining the health of the body by eliminating old cells, unnecessary cells, and unhealthy cells. In general, apoptosis confers advantages during an organism's life cycle. For example, the differentiation of fingers and toes in a developing human embryo occurs because cells between the fingers apoptose; the result is that the digits are separate. This process is very difficult to be identified in the images and not a single one was observed in the dataset in witch we worked.

Given that not all the embryo is captured by the image volume, the cells appears and disappears from the image frame. Only cells near the border could disappears and cells that appears are first seen near the border.

The basic principle of single particle tracking is to find for each object in a given time frame its corresponding object in the next frame. The correspondence is based on object features, nearest neighbor information, or other inter-object relationships.

7.2 Functional minimization

In order to efficiently find the unknown lineage tree, we have to track the large number of nuclei over time, finding the optimal joint association between nuclei for every pair of subsequent time frames. Cells moves slowly, therefore the connectivity between neighboring cells is maintain between consecutive frames. We found that elastic graph matching algorithms would solve this problem efficiently.

We consider the for the single cells the following events: *move*, *division*, *disappearance* and *appearance*. While an (apparent) cell disappearance may be caused by apoptosis, it is more typically due to a cell leaving the field of view or misdetection of cell identification; an (apparent) appearance happens when it is detected again at a later time or coming from outside the field of view. In order to rule out implausible events, we only consider at most K nearest neighbors of the cell i within a given distance threshold, this impose an ad hoc maximum speed for cells.

Let I be the whole set of cells at all time steps. The unknown tracking graph f on I, representing the lineage tree, was determined in three phases (see figure 7.2):

1. The first phase use strong local constraints (maximum speed, binary tree structure) for restricting the unknown tracking graph, we link each cell with its nearest neighbor cell at previous time step.

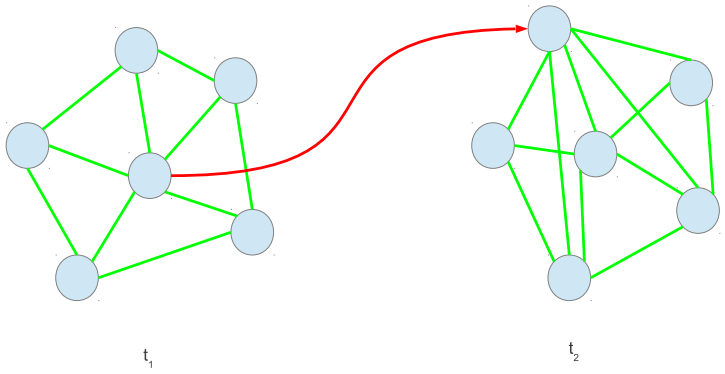
2. Using the lineage tree obtained in phase 1 as initialization, the second phase minimize a heuristic functional, combining information extracted from images (nuclear positions, vector field, mitosis detection using segmentation) with biological regularization constraints (elastic behavior, smooth movement) to measure the correctness of cell matching at consecutive time steps.
   The elastic behavior could be modeled with elastic matching in the first part of equation 7.3 where the cell i and the cell j are near neighboring cells at time step t. With this term we impose an structure we want to conserve at each time step (see figure 2) where the distance between neighbor cells almost do not change.

   Using the vector field obtained by optical flow techniques, it is possible to follow the same cells at consecutive time steps, this is encouraged by the second term of equation 7.3. where V is a function that estimate the movement of each cell from time t to $t-1$ (it means backwards). Minimizing this equation we favor the lineage tree f where the cells are near they are expected to be.

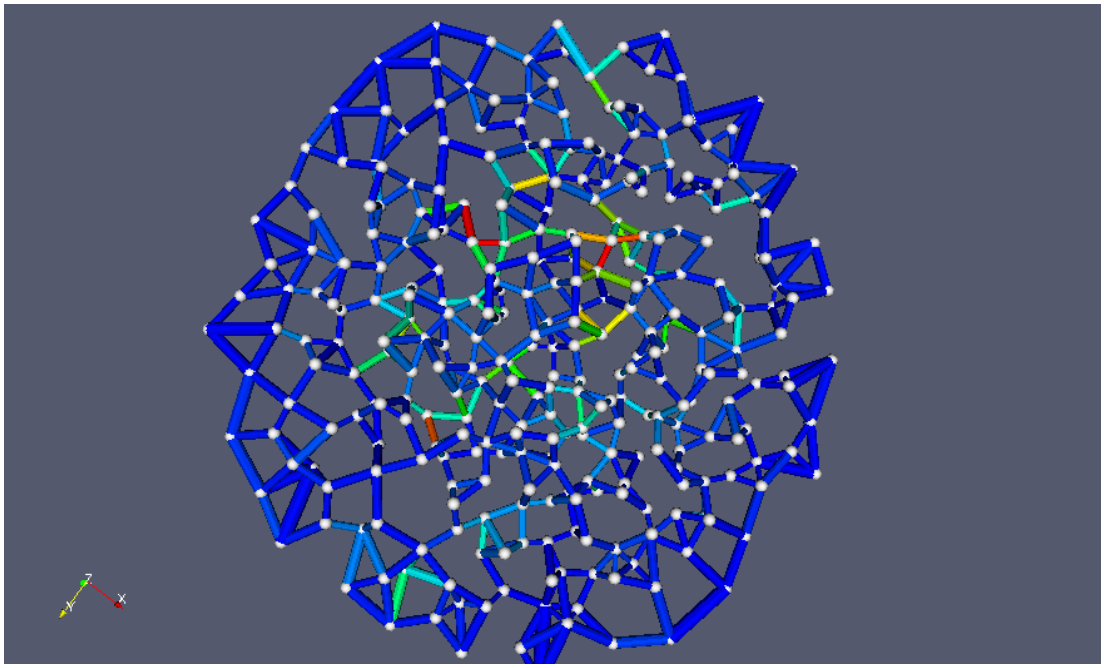
$$E(f) = \sum_{i} \sum_{j\ neighbor\ i} \left\|(X_i - X_j) - (X_{f^{-1}(i)} - X_{f^{-1}(j)})\right\|_2 + \lambda \left\|X_{f^{-1}(i)} - V(X_i)\right\|_2 + D_i + A_i \tag{7.3}$$

   where i and j are neighboring cells at time t, X_i is the cells position in 3D coordinates, D_i and A_i gives penalties if the cell i appears or disappear far from the image boundaries and the function f^{-1} gives the cell at previous time step.

   The minimization of this functional is done by a simulated annealing algorithm [91]. At each iteration of the algorithm 2, we modify the solution f finding a neighbor solution by changing the links between cells. This changing are adding a new link, switching or deleting links. This solution is accepted with a probability that depends on a temperature function. The acceptance of a worst solution is



(a) The underlying structure (green lines) change drastically using elastic matching when there is a tracking error.



(b) Experiment 070128c, The cells are identified by a white sphere, each cell is connected with neighboring cells. The color scale (blue to red) represent the deformation of the structure using the elastic term of equation 7.3, the maximum values (red) match where there are cells division.

done to overcome a local minimum. This temperature function decrease over time, it is higher at the begin of the algorithm and then it cool down. The algorithm keeps in memory the best solution found.

input : Set of cells, I
output: Lineage tree, f

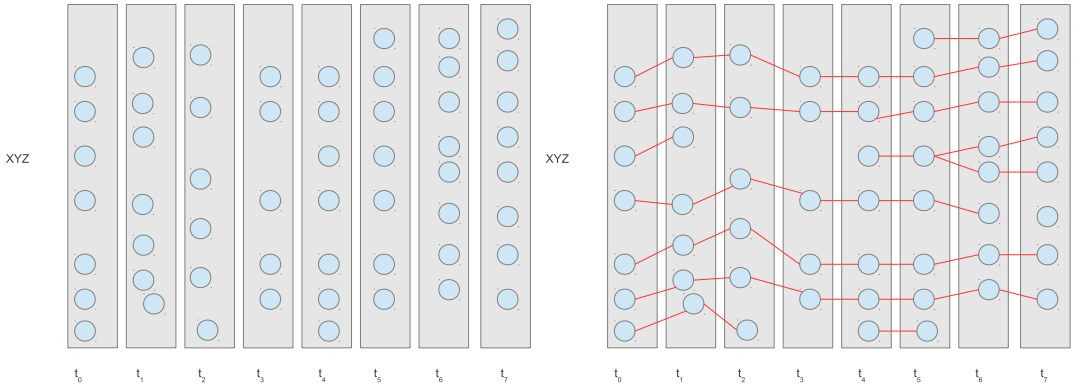
$f \leftarrow Phase1(I)$;
$E \leftarrow E(f)$;
$k \leftarrow 1$;
foreach *timestep* **do**
  while $k < k_{max}$ **do**
    $f_k \leftarrow neighbourSolution(f)$;
    $E_k \leftarrow E(f_k)$
    if $P(E_k - E, temp(k/k_{max})) > random()$ **then**
      $f \leftarrow f_k$;
      $e \leftarrow e_k$;
    end
    $k++$;
  end
end

Algorithm 2: Simulated Annealing

3. The third phase's goal was to minimize false-positive and false-negative errors from the nuclear and mitosis detection steps. This was achieved by looking at the whole biological coherence of the lineage tree identifying:

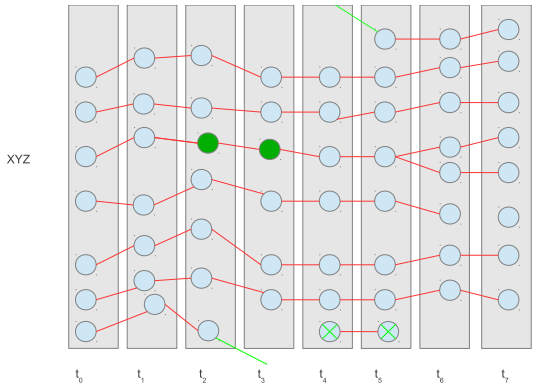
   - Cells divisions not assigned are linked with their mother cell. Tipically new born nuclei moves in relative opposite direction until cytokinesis ends.
   - The discontinuities in cell trajectories are identified as false negative nuclear centres. Usually when there is a false negative cell detection, the trajectory of the cell do not continue for several time steps and later at the expected site there is a false mithosis or a cell with out mother.
   - The cells that lived for only a few times steps are marked as false positive nuclear centers. Cells fate could be after several hours whether a mithosis,

apothosis (never seen in this experiment) or they could leave the image frame. Cell that only live for some minutes are marked as false positives cell detection.



(c) Initial cells set I. Cells at t_2 and t_3 are missing.

(d) Tracking after phase 2, cells division are identified, cells links are assigned



(e) Final result: tracking is corrected based on biological coherence, cells gaps are filled (false negatives errors at t_2 and t_3.) Cells trajectory discontinuities near the boundaries are identified. Cells that live for a short period of time (t_4 and t_5) are marked as false positives.

Figure 7.2: Tracking algorithm example

7.3 Validation

Tracking accuracy was measured by assessing t to $t+1$ link for selections of cells over 10 time steps from time t_{n0} ($t_{n0} \in \{1, 20, 100, 200, 300, 400, 500\}$). Altogether, 15,686 links were manually checked, 1603 at $t_{n0} = 1$ (i.e. from $t0$ to $t10$ and so on for other t_{n0}); 1762 at $t_{n0} = 20$; 1828 at $t_{n0} = 100$; 2948 at $t_{n0} = 200$; 3169 at $t_{n0} = 300$; 2931 at $t_{n0} = 400$; 1445 at $t_{n0} = 500$. In addition to tracking links from t to $t+1$, the checking procedure provided an error measurement on cell center detection. These measurements were used to estimate the tracking error rate. The total error number (n_{error}) was defined as the number of incorrect t to $t+1$ links ($n_{badLinks}$) plus the number of false deaths ($n_{falseDeath}$). The $n_{badLinks}$ were subdivided into two categories: errors due to false positive nuclear centers ($n_{falsePositive}$) and errors due to false tracks: the tracking jumped from one cell trajectory at time t to another cell trajectory at time $t+1$, (n_{change}). The total number of errors was given by $n_{error} = n_{falsePositive} + n_{change} + n_{falseDeath}$. The tracking error rate at a given time point t_{n0} was expressed as a percentage from the ratio of the error number (n_{error}) over the total number of tested links ($n_{allLinks}$).

7.4 Results

Cell tracking in the 4D space of the segmented data was not properly achieved by classical methods using a purely local nearest-neighbor principle. Although we kept temporal resolution as fine as possible (Δt is only 67 seconds in the image data set 070418a), cell displacement during mitosis was a major problem that the tracking algorithm had to solve [22]. The spatio-temporal lineage tree reconstruction was best achieved with a tracking algorithm relying on the minimization of a heuristic functional. Strong constraints were used i) a priori, i.e., before minimization, to restrict the unknown tracking function and ii) a posteriori, i.e. after minimization, to identify biological inconsistencies for correction of false negatives and false positives in the nuclear center and mitosis detection steps. The heuristic functional assigning penalties for high displacements and deformations in cellular feature space was minimized by simulated annealing [91].

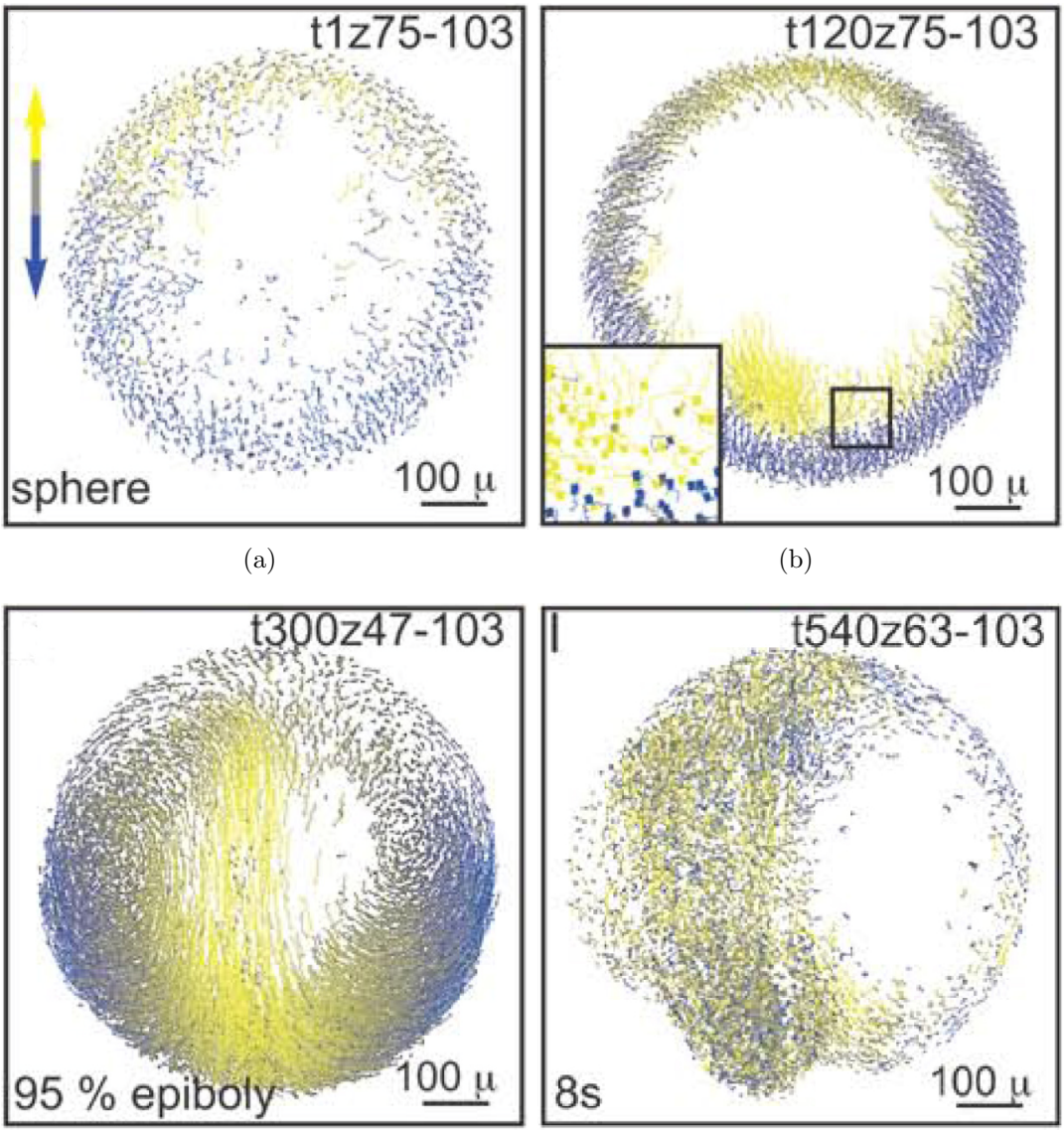
Following the image processing steps of the phenomenological reconstruction, the

virtual embryo consisting of nucleus centers and individual cell trajectories could be explored using tailor made tools reveling the cell's clonal history and allowed to derive cell population dynamics and behaviors (see figure 7.3)

The tracking error rate was measured with the sampling procedure described in Figure 7.6. Altogether, over 15,000 links between the t and t+1 positions of cells have been manually checked at different time points. For the first 300 time steps (5h34 imaging, end of gastrulation) the total error rate is lower than 2%, i.e., more than 98% of the cells were correctly tracked (Figure 7.6d). Tracking accuracy degraded somewhat beyond time step 400 (7h26 imaging, 3-somite stage), but 90% of the links were still correctly found by time step 500 (9h18 minutes imaging, 7-somite stage). False positives in nuclear detection were kept below 0.5% and were not the major source of tracking errors. About half of the tracking errors were identified as false trajectories (the tracking jumped from one cell trajectory at time t to another at time t+1) and the other half was false deaths (the tracking did not find the position of the cell at time t+1, and cell trajectory ended).

The most difficult task in tracking is to recognize the cell division, where a mother cell should link to two daughter cells. Figure 7.4 show the error rate for dataset 070420a. The error rate in mitosis detection is near 10% until the embryo reach 7hs of development, then the cells are smaller and cell density is greater, the task become most difficult. In figure 7.5 a mitosis detected by the tracking algorithm is shown.

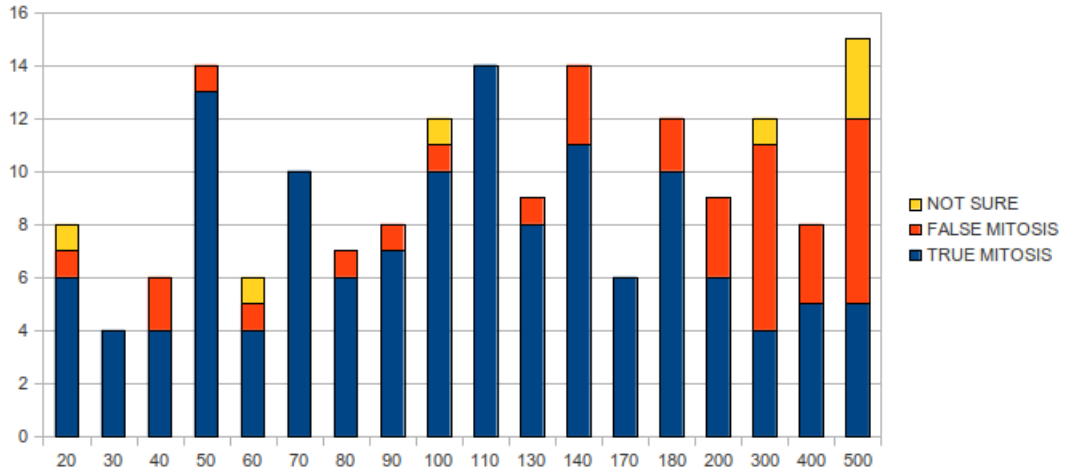
The exploitation of the virtual embryo opens the way to in silico experimental embryology to follow the formation and transformation of 4D morphogenetic fields, backtracking a 3D morphogenetic section towards its earlier 3D fate map. We explored this potential by manually selecting morphogenetic fields in the brain at t=479 (6/7- somite stage), based on morphological landmarks (Figure 7.8d). Backtracking the selected cells, which were categorized in 12 domains, revealed cellular population dynamics consistent with the order and coherence of the brain fate map [92, 93]. The latter - as all fate maps in zebrafish - was however determined no earlier than the shield stage (gastrulation onset), and only for the most superficial cell layer of the 4-5 "deep" cell layers of the embryo, leaving the fate of all deeper cells unknown. Our data reveal that this fate map actually holds for the whole depth of the blastoderm, and that it can be backtracked to the earlier sphere stage, with no major loss in coherence (Figure



(a) (b)

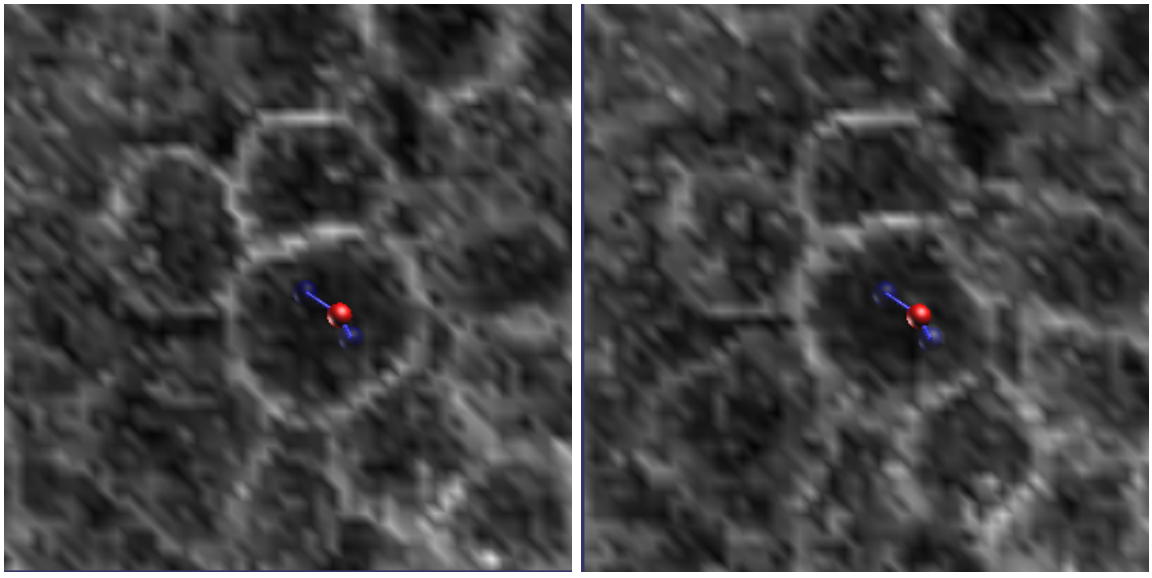
(c) (d)

Figure 7.3: figure (a-d) virtual counterpart of raw data in Fig. 2.8 a-d. (a) Color code for cell displacement from yellow (rostralwards) through grey (null or medio-lateral) to blue (caudalwards) indicating cell displacement orientation. (b) Inset, magnification showing each nucleus center representation as a cube with a vector indicating cell trajectory for the next 12 time steps![note1].



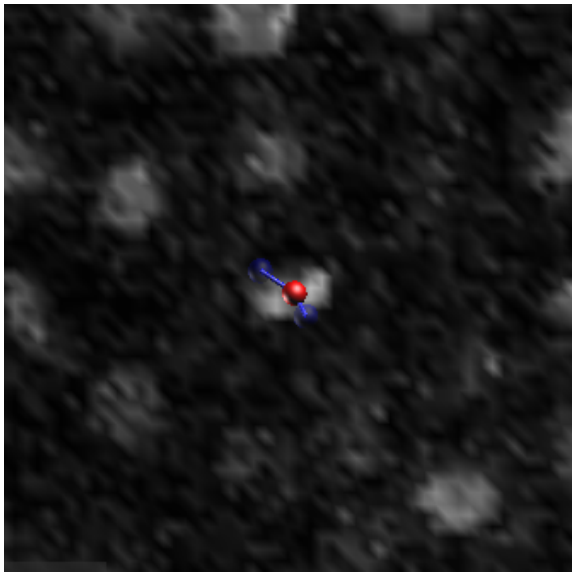
(a)

Figure 7.4: Dataset 070420a. Mitosis detected by the procedure were manually validated at different timesteps. This bar-graph shows the false positive mitosis. This task is so difficult that even by hand not always is possible to detect the mitosis. The x-axis shows the timesteps number, y-axis the number of true mitosis (blue), false positives (red) and known (yellow). Until timestep 300 (7hs of development) the error rate is near 10%. After this point the error rate increase also deeper in z direction, the error rate is bigger.

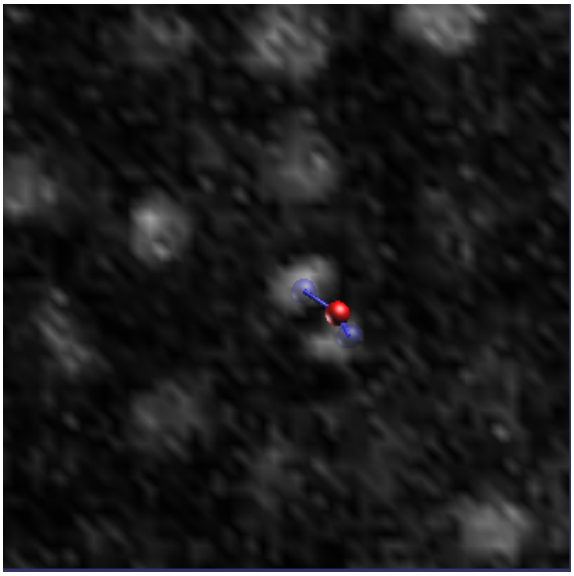


(a)

(b)

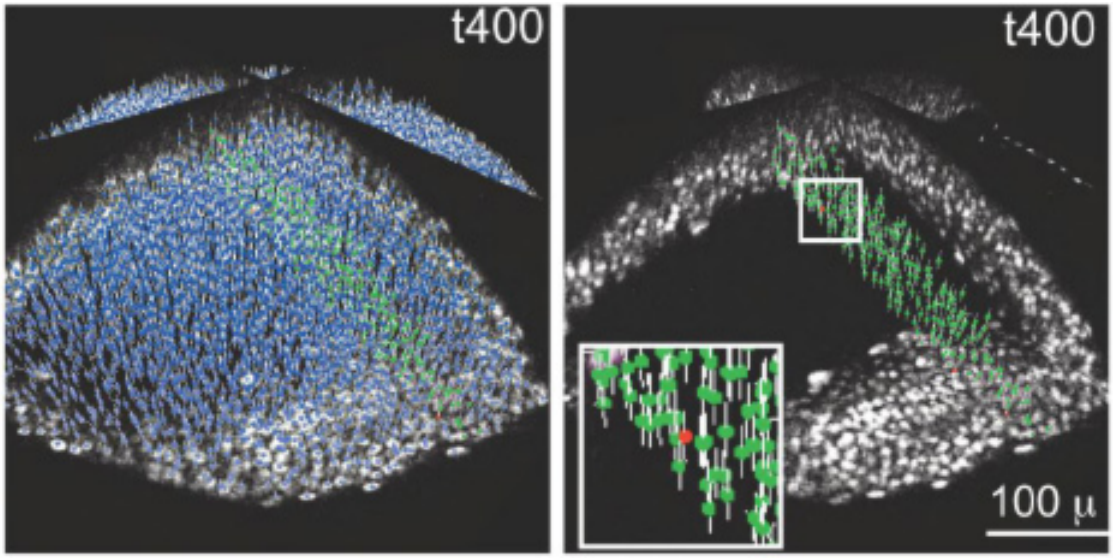


(c)

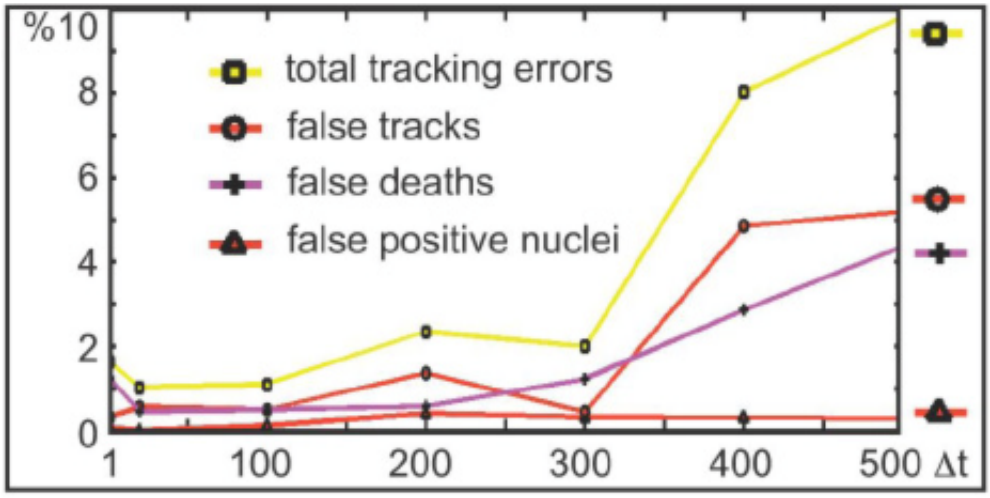


(d)

Figure 7.5: Figures a and b: membrane channel, figures c and d: nucleus channel, figures a,c: time step 100, figures b,d: time step 101. Red dot nuclei at time step 100, blue dot newborn nuclei at timestep 101. Cell division identified using tracking algorithm. While dividing the membrane adopt a spherical shape and nuclei an elliptic shape. Newborn nuclei direction is opposite with similar magnitude.

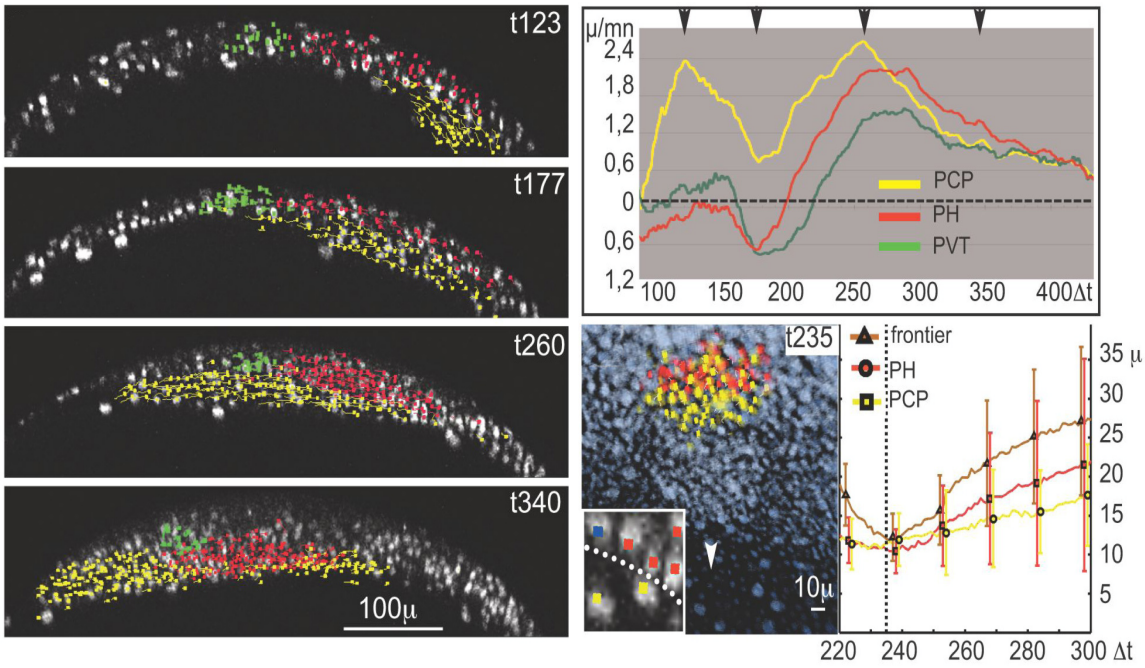


(a)



(b)

Figure 7.6: The lineage tree validation. (a) Nuclear center detection checked for a sub-population of cells here at t400, correct centers indicated in green, false positive in red (right panel, one red nucleus in the inset), vertical white bars through nucleus center used to assess position relative to raw data orthoslice; tracking was manually checked for this population from t400 to t410. (b) Plot of tracking errors.![note1].



(a) (b)

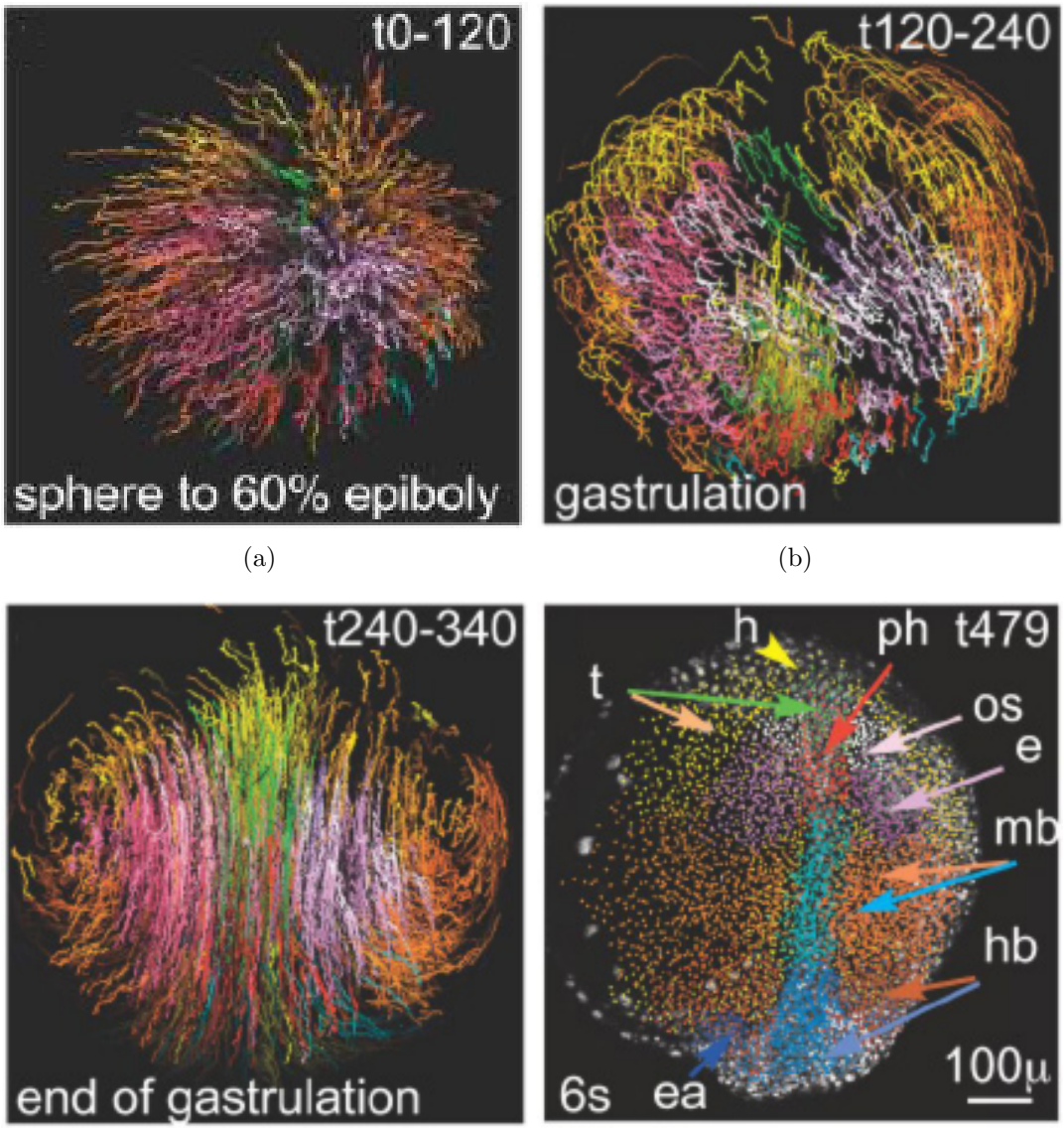
Figure 7.7: Multi scale measurements of cell behaviours from their clonal history. Data set 070418a. (a) Left panels, relative movements of cells selected in the prechordal plate (PCP, yellow), presumptive hypothalamus (PH, red) and presumptive ventral telencephalon (PVT, green), lateral views, rostral to the left, orthoslice of raw data (nuclei) behind selected cells, time step indicated top right of each panel. (b) Top panel, average speed of the selected cells over 10 time steps (in μm/min), positive values indicate anterior movement, arrowheads indicate time points shown in left panels. Bottom left panel, PCP (yellow) and PH (red) cell populations in which pairs of adjacent cells were chosen at t235; view from inside the embryo; white arrowhead points away from the anterior border of PCP migrating from top to bottom of the image; inset shows magnification of adjacent cells across PCP-PH frontier (dotted line); blue, center of non selected cell. Bottom right panel, average distance and variance for the selected cell pairs, tracked between t220 and t300, within PCP (yellow, 27 pairs), within PH (red, 37 pairs), or across the PCP-PH frontier (brown, 38 pairs). Dotted line at t235 when pairs of adjacent cells were chosen, all cell trajectories have been manually checked.![note1].

7.8, movies Movie4-1, Movie4-2 and Movie4-3).

Further analyzing the collective cell behaviors requires integrating at least three levels of organization: single cells, cell neighborhoods and cell populations. We explored such a multiscale approach to investigate cell behavior in the presumptive forebrain at the midline [94, 95, 96, 97]. The speeds of selected cells from the presumptive hypothalamus (PH), presumptive ventral telencephalon (PVT), and prechordal plate (PCP) were averaged over 10 time steps and plotted as a function of time between the onset of gastrulation and early somitogenesis (Figure 7.7). These measurements indicated transitions in cell behaviors correlating with the relative positions of PCP, PH and PVT (Figure 7.7b, top panel, movie Movie4-4). We observed that the speed of PCP decreased as it met PH territory (t123, 60% epiboly). PH, which was moving anti-parallel to PCP, started moving rostrally by 75% epiboly (t200). The three selected cell populations accelerated similarly while keeping different speeds. By the end of gastrulation (t260), PCP slowed down abruptly. For the next hour, PH moved faster than the two others and passed under PVT. These observations, previously inaccessible but readily provided by our in toto reconstruction strategy, definitively validate the subduction scenario proposed earlier [94], hence revealing coordinated behavior between PCP and PH cells. Further measurement of the distance between cells of pairs chosen either within the same field or on both sides of the frontier between the two cell populations, indicate that PCP and PH cells do not remain tightly associated, rather they slide on top of each other (Figure 7.7, bottom panels, Movie4-5). This transient interaction, which is expected to allow proper cell signaling, is part of the biomechanical process driving PH cells rostrally. However, the cell population dynamics revealed by our phenomenological reconstruction indicated that PH displacement might depend on larger scale constraints.

7.4.1 Workflow time-consuming

To achieve the automated reconstruction of cell behaviors in morphogenesis using a workflow of 3D+time image processing of algorithms requires large computation facilities. The algorithm are implemented on C or C++ using MPI to parallelize the algorithms. Each time step of typical image dataset have 512x512x120 pixels and 2 channels (e.g. 071226a dataset). To process the filtering step, typically we use 4 pro-



(a) (b)

(c) (d)

Figure 7.8: Multi scale measurements of cell behaviours from their clonal history. Data set 070418a, animal pole view from the surface. (a-c) Trajectories of cells selected in presumptive organs at t479 and backtracked, temporal window for the displayed trajectories indicated top right, corresponding developmental periods indicated bottom left. (d) 3D rendering at t479, of nuclear centres coloured according to the following code: yellow, hypoblast including PCP and notochord (h); red, presumptive hypothalamus (ph); pink, right eye (e); light pink, right optic stalk (os); purple, left eye; light purple, left optic stalk; green, ventral telencephalon (t); gold, dorsal telencephalon, placodes and neural crest; light blue, ventral midbrain (mb); light brown, dorsal midbrain and dorsal diencephalon; blue, ventral hindbrain (hb); dark brown, dorsal hindbrain; dark blue, ears (ea).![note1].

cessors (communicating by MPI) for every 3D volume, this takes 10 minutes each 3D image. The cell identification step, which takes 20 minutes for one 3D volume. For nuclei segmentation, it takes in average 0.75 second to process each nucleus. In average there are 6000 nuclei in each time step, therefore each volume it takes 1.25 hours while the computation time for the segmentation of each membrane takes 6 seconds (10 hours each time step). To process the Image registration step, tipically we use 4 processors (communicating by MPI), for each pair of 3D consecutive volumes, takes 10 minutes for each time step. The time it takes to track all the cells in the embryo using 8 processors (communicating with MPI) include 45 minutes for restricting the unknown tracking graph, it means to constraint the possible successors for each cell, 4.5 hours to minimise the heuristic functional and 2 hours to identify discontinuities in cell trajectories. The description of the time it takes to process the workflow is in ideal situation, where the processors are always available in the computing center In2p3 [98] and without considering the time it takes to access the datasets and images stored in a Storage Resource Broker (In2p2 [98]). The algorithms were also succesfully implemented to run on the Egee computing grid [99]. In standard load of the computing centre, the overall process of the workflow takes about 48 hours.

7.5 Conclusions

We have presented a workflow of image process algorithms that produce as result the tracking of cells in a live embryo. This tracking allows understand the cell behavior and clonal history which is the basis for understanding morphogenetic processes [100]. This workflow have been applied succesfully to zebrafish embryos from 4 hours post-fertilization through gastrulation and early neurulation stages [8]. The imaged cells, whose number grows from 2000 to more than 15000, were tracked with less than 2% error for at least 6 hours. Cell division were automatically identified with less than 10% of error for 7 hours. Prior to this work, the automated reconstruction of the cell lineage tree attempted for *Caenorhabditis elegans* reached comparable accuracy only for the first 194 cells, despite much more limited cell movements [101, 102]. In the zebrafish, lineage studies previously relied at best on semi-automated cell tracking methods [94]. The virtual embryo allows a systematic exploration of the clonal origin and clonal complexity of organs, as well as the contribution of cell proliferation modes

and cell movements to the formation of local patterns and morphogenetic fields. By providing essential parameters of cell behavior and clonal history, this first assembly of the cell lineage tree underlying zebrafish brain early morphogenesis constitutes the core of the embryome, i.e., the phenomenological and theoretical reconstruction of the entire system's multiscale dynamics.

Note[1]**:** In this experiment the cell centers were obtained using [38], [103].

Conclusions

The zebrafish is a small freshwater fish largely validates as animal model to understand vertebrate development, gene function and healing/repair mechanisms in vertebrates.

We provide here a complete and original 3D+time image analysis framework to achieve the automated reconstruction of cell behaviours in morphogenesis.

The time-lapse series of images has then been analyzed and processed in order to reconstruct the shape and track the position of every cell in the dataset.

Embryo labeling was obtained through RNA injection performed at the one-cell stage to obtain ubiquitous expression of H2B/mCherry fusion protein and farnesylated eGFP, which stained nuclei and membranes respectively. While nuclear staining was instrumental to perform cell tracking, membrane staining was essential to assess cell morphology, behavior and neighborhood, and to reveal morphological landmarks.

We demonstrate tracking performances with single cell accuracy during early embryogenesis of the zebrafish (*danio rerio*) where cell proliferation rate is over 50000 cells per 24 hours. 3D+time image data sets of early embryonic cells, whose number grows from 2000 to more than 15000, were tracked with less than 2% error for at least 6 hours. Prior to this work, the automated reconstruction of the cell lineage tree reached comparable accuracy for the first 194 cells in *caenorhabditis elegans* and other recent trials in drosophila or danio rerio produced only global pictures of cell movements, missing single cell clonal history.

In chapter 1, We have described a series of stages for development of the embryo of the zebrafish. We have defined seven broad periods of embryogenesisthe zygote, cleavage, blastula, gastrula, segmentation, pharyngula, and hatching periods as know in the literature. These divisions highlight the changing spectrum of major developmental processes that occur during the first 3 days of development and we review some of what is known about morphogenesis and other significant events that occur during each of

the periods.

In chapter 2 we have outlined the main features of imaging techniques and the fluorescent proteins that allow us to capture the images. We describe the imaging techniques used for acquiring our experimental dataset. Particular attention is devoted to the main drawbacks of these imaging techniques as photobleaching and photodamage. They influence and the typology of data to be examined, impose physical limits to the dataset dimensionality and resolution, to the image contrast, the signal-to-noise ratio and the specimen survival.

In chapter 3 introduces a set of preprocessing techniques developed with the research consortium based on multiscale analysis and evolutive partial differential equations. Particularly a qualitative and quantitative comparison between the Perona-Malik and the geodesic mean curvature flows is provided. It have been developed a software for the measure of the Hausdorff distance between surfaces allowing to automatically estimate the error between a gold standard and the filtered data.

In chapter 4 we have outlined the importance of the identification of the center of each cell. This procedure was done with the Hough transform. The cell centers found with this procedure are the starting point to nuclei and membrane segmentation and to track the cells.

In chapter 5 we have developed workflow of algorithms for automatically segment the cell shape of nuclei and membranes. This segmentation has then been reconstructed using a modified version of the Subjective Surfaces technique. Such segmentation method has appeared well suitable in particular for membranes segmentation as it is able to complete contours often missing in original membranes data and, likewise other segmentation methods, does require a minimum user intervention.

In chapter 6 we explore four different algorithms that perform the registration of 3D images. We have compared the performance in synthetic data and in real data as well. This registration calculate an estimation of deformation at each point in the embryo. We have used this to predict the movement of each cell.

In chapter 7 we have depicted an algorithm based on the minimization of an a nonlinear heuristic functional. The functional combines the informations extracted from images using the methods already exposed in previous chapters (nuclei positions, vector field, segmentation) with biological regularization constraints (elastic behavior, cell cycle). It then measures the correctness of cell matching at consecutive time steps.

Our results have been manually validated at different time points and reconstruct the formation of prospective brain of zebrafish embryos from 4 hours post-fertilization through gastrulation and 7-8 somites stage.

This reconstruction provides information about millions of cell position, millions of membranes and nuclei segmentation, and thousands of cells tracks cell divisions per embryo.

Our reconstruction opens the way to a systematic and quantitative exploration of the clonal origin and clonal complexity of brain organs, as well as the contribution of cell proliferation modes and cell movements to the formation of local patterns and morphogenetic fields.

Appendix A

Supplementary material

Suplementary material could be download from: **http://goo.gl/OmYYX**

Movie 1-1 data set 070418a, view from the animal pole at 4 hpf at 28.5°C (AP scheme more precisely designated as AP4 to indicate the onset of the imaged developmental period, imaged volume $700x700x140\mu^3$, Dt=67s, 541 time steps), raw data with orthoslice z63, only the sections below the orthoslice are shown (Amira software visualisation).

Movie 1-2 data set 071227b, view from the vegetal pole at the onset of gastrulation (6 hpf at 28.5°C, VP scheme, designated as VP6 to indicate the onset of the imaged developmental period, imaged volume $700x700x178\mu^3$, Dt=87s, 413 time steps), raw data with orthoslice z45, only the sections below the orthoslice are shown (Amira software visualisation).

Movie 1-3 data set 080322a, lateral imaging scheme, starting at 4 hpf (28.5°C, L scheme designated as L4 to indicate the onset of the imaged developmental period). In the L scheme, the position of the dorsal side identified a posteriori is indicated as 0 for a dorsal view, 6 for a ventral view and from 1 to 5 for intermediate positions, here a L4-4 scheme. The whole volume of raw data is shown (Amira visualisation software). For this data set, the light with a wavelength of 980 nm was provided by a mode-locked Ti:Sapphire laser (Mai Tai from Spectra-Physics). The pulse width was less than 100 fs.

Movie 1-4 Virtual embryo 070418a with orthoslice as in Figure 7.3. Colour code from yellow, through grey, to blue indicating cell displacement orientation (projection of the speed vector on the y axis). Yellow indicates migration from caudal to rostral, blue indicates migration from rostral to caudal, grey indicates nul or medio-lateral displacement. Each nucleus centre is represented as a cube with a vector indicating its path for the next 12 time steps. Embryome Visualisation Platform.

Movie 1-5 Same as Movie 1-4 but with the whole volume to show the collective displacement of cells.

Movie 2-1 Data set 070420a at t001. 3D rendering of segmented nucleus and membrane shape for selected cells shown during mitosis. The virtual cells are in the environment of the raw data (green membranes, red nuclei). Transparency was introduced around the segmented cells. Note that the inner virtual cells that are under the epithelium (the enveloping layer, EVL) stand below the junction of EVL cells and rapidly move to align with the EVL junctions after cell division. Embryome Visualisation Platform.

Movie 3-1 Tracking of cell ID 124516142 (data set 070418a). Superimposed raw data and reconstructed data illustrating the augmented phenomenology concept. Here, nuclear center of the tracked cell is superimposed with raw data orthoslices (membrane channel). Embryome Visualisation Platform.

Movie 4-1 Same as Figure 7.8. Cells manually selected in presumptive organs at t479 have been backtracked then allowing forward tracking visualisation, animal pole view. Colour code described in Figure 7.8d. Cell trajectories are kept for the past 100 time steps. Embryome Visualisation Platform.

Movie 4-2 Same colour code as Movie 4-1, only nuclear centers (cubes) of selected cells are shown. Selected cells are backtracked from t480 to t1; at t1, a cell is tracked forward and shown with superimposed raw data orthoslices (nucleus channel) to assess the correctness of the lineage tree assigning its progeny to the optic vesicle (pink). Embryome Visualisation Platform.

Movie 4-3 Same selection of presumptive organs as Movie 4-1 and 4-2; animal pole view from the surface, visualisation of nucleus shape segmentation, cells that are not part of the selections are shown in grey. Embryome Visualisation Platform.

Movie 4-4 Relative movements of cells selected in the prechordal plate (PCP, yellow), presumptive hypothalamus (HP, red) and presumptive ventral telencephalon (PVT, green) as shown in Figure 7.7. Lateral view, rostral to the right. Augmented phenomenology visualisation with nuclear centers, cell trajectories and raw data orthoslices (nucleus channel). Embryome Visualisation Platform.

Movie 4-5 Same as Figure 7.7a. Small populations of PCP (yellow) and PH (red) cells where pairs of adjacent cells were chosen for distance measurements shown in Figure 7.7b. Augmented phenomenology visualization with 3D rendering of raw data (nucleus channel) and nuclear centers throughout gastrulation. Last frames with raw data orthoslice nucleus channel, then membrane channel, to show the position of selected cells relative to morphological landmarks (marked frontier between PCP and PH at t316, end of gastrulation). Embryome Visualisation Platform.

Publications

Publications on international journals

1. C. Zanella, M. Campana, B. Rizzi, C. Melani , P. Bourgine, K. Mikula, N. Peyriéras, A.Sarti: "Cells segmentation from 3-D confocal images of early zebrafish embryogenesis", IEEE Transactions on Image Processing. Volume 19 , Issue 3 (March 2010) Pages: 770-781.

Peer Reviewed Proceeding Articles

1. B. Rizzi, M. Campana, C. Zanella, C. Melani, R. Cunderlik, Z. Kriva, P. Bourgine, K. Mikula, N. Peyriéras and A. Sarti, "3-D Zebrafish Embryo Image Filtering by Nonlinear Partial Differential Equations", Conf.Proc. IEEE Eng. Med. Biol. Soc. 2007.

2. C. Zanella, B. Rizzi, C. Melani, M. Campana, P. Bourgine, K. Mikula, N. Peyriéras, A. Sarti, "Segmentation of Cells from 3D Confocal Images of Live Zebrafish Embryo", Conf.Proc. IEEE Eng. Med. Biol. Soc. 2007.

3. C. Melani, B. Lombardot, M. Campana, B. Rizzi, C. Zanella, P. Bourgine, K. Mikula, N. Peyriéras, A. Sarti, "Cells tracking in a live zebrafish embryo", Conf.Proc. IEEE Eng. Med. Biol. Soc. 2007.

4. B. Lombardot, M. A. Luengo-Oroz, C. Melani, E. Faure, A. Santos, N. Peyriéras, M. Ledesma-Carbayo, P. Bourgine. "Evaluation of four 3D non rigid registration methods applied to early zebrafish development sequences", Conf.Proc MICCAI 2008.

5. C. Castro Gonzalez, M. A. Luengo-Oroz, L. Duloquin, T. Savy, C. Melani, S. Desnoulez, M. J. Ledesma-Carbayo, P. Bourgine, N. Peyrieras, A. Santos. "Towards a Digital Model of Zebrafish Embryogenesis. Integration of Cell Tracking and Gene Expression Quantification". Conf.Proc EMBC'10

6. C. Castro, M. A. Luengo-Oroz, L. Douloquin, T. Savy, C. Melani, S. Desnoulez, P. Bourgine, N. Peyriéras, M.J. Ledesma-Carbayo, A. Santos. "Image Processing Challenges in the Creation of Spatiotemporal Gene Expression Atlases of Developing Embryos". Conf.Proc EMBC'11

Abstracts to International Conferences

1. C. Melani "A Framework for 4-D Biomedical Image Processing Visualization and Analysis ", Pisa, September 04 - 09 September 2006 School on Neuromathematics of Vision.

2. M. Campana, B.Rizzi, C. Melani, P.Bourgine, N. Peyriéras, A.Sarti, "A Framework for 4-D Biomedical Image Processing Visualization and Analysis ", GRAPP 08, 2008, International Conference on Computer Graphics Theory and Applications, Funchal, Portugal.

3. C. Melani, M. Campana, B. Rizzi, E. Faure, B. Lombardot.P. Bourgine, K. Mikula, N. Peyriéras, A. Sarti. "Cell Tracking in a live zebrafish embryo by functional minimization". MLS 2009.

4. C. Melani, A. Sarti, K. Mikula, P. Bourgine, N. Peyriéras. "Salt and Pepper Strategy: a new method to Automatically validate morphogenesys models" MLS 2010.

Abstracts to National Conferences

1. B. Rizzi, C. Zanella, M. Campana, C. Melani, N. Peyriéras, A. Sarti: "Cells shape reconstruction from 3-D+time LSM images of early Zebrafish embryogenesis", Atti del Primo Congresso Nazionale di Bioingegneria, 2008.

2. C. Melani, M. Campana , B. Rizzi, C. Zanella , P. Bourgine , K. Mikula , N. Peyriéras, A. Sarti: "Recognition and tracking of cells in a live zebrafish embryo", Atti del Primo Congresso Nazionale di Bioingegneria, 2008.

Bibliography

[1] S.G. Megason and S.E. Fraser. Digitizing life at the level of the cell: high-performance laser-scanning microscopy and image analysis for in toto imaging of development. *Mech Dev*, 120(11):1407–20, 2003.

[2] M S Cooper, L A Damico, and C A Henry. Confocal microscopic analysis of morphogenetic movements. *Methods Cell Biol*, 59:179–204, 1999.

[3] Kimberly Dooley and Leonard I. Zon. Zebrafish: a model system for the study of human disease. *Genetics of disease*, 10:252–256.

[4] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell, Fourth Edition*. Garland Science, 4 edition, 2002.

[5] K. Moriyoshi, L. J. Richards, C. Akazawa, D. D. O'Leary, and S. Nakanishi. Labeling neural cells using adenoviral gene transfer of membrane-targeted GFP. *Neuron*, 16(2):255–260, Feb 1996.

[6] N.C. Shaner, P.A. Steinbach, and R.Y. Tsien. A guide to choosing fluorescent proteins. *Nat Methods*, 2(12):905–9, 2005.

[7] Judy Sprague, Dave Clements, Tom Conlin, Pat Edwards, Ken Frazer, Kevin Schaper, Erik Segerdell, Peiran Song, Brock Sprunger, and Monte Westerfield. The ZebraFish Information Network (ZFIN): The zebrafish model organism database. *Nucleic Acids Research*, 31(1):241–243, January 2003.

[8] C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling. Stages of embryonic development of the zebrafish. *Am. J. Anat.*, 203(3):253–310, July 1995.

[9] Bioimages database, http://bioimages.vanderbilt.edu, June 2009.

[10] Developmental biology, http://musicalgabbie410.blogspot.com.ar.

[11] H. Morise, O. Shimomura, F. H. Johnson, and J. Winant. Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry*, 13(12):2656–2662, Jun 1974.

[12] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, February 1994.

[13] R. H. Webb. Theoretical basis of confocal microscopy. *Meth. Enzymol.*, 307:3–20, 1999.

[14] S M Potter. Vital imaging: two photons are better than one. *Curr Biol*, 6(12):1595–8, 1996.

[15] M. Gu and C. J. R. Sheppard. Comparison of 3-dimensional imaging properties between 2-photon and single-photon fluorescence microscopy. *J Microsc*, pages 128–137, 1995.

[16] Guy Cox and Colin J. R. Sheppard. Practical limits of resolution in confocal and non-linear microscopy. *Microscopy Research and Technique*, 63(1):18–22, 2004.

[17] S. M. Potter, C. M. Wang, P. A. Garrity, and S. E. Fraser. Intravital imaging of green fluorescent protein using two-photon laser-scanning microscopy. *Gene*, 173(1 Spec No):25–31, 1996.

[18] Monte Westerfield. The zebrafish book: A guide for the laboratory use of zebrafish (brachydanio rerio).

[19] Jayne M. Squirrell, David L. Wokosin, John G. White, and Barry D. Bavister. Long-term two-photon fluorescence imaging of mammalian embryos without compromising viability. *Nature Biotechnology*, 17(8):763–767, August 1999.

[20] F. Helmchen and W. Denk. Deep tissue two-photon microscopy. *Nat. Methods*, 2(12):932–940, Dec 2005.

[21] Cecilia Zanella, Matteo Campana, Barbara Rizzi, Camilo Melani, Gonzalo Sanguinetti, Paul Bourgine, Karol Mikula, Nadine Peyriéras, and Alessandro Sarti. Cells segmentation from 3-d confocal images of early zebrafish embryogenesis. *Trans. Img. Proc.*, 19(3):770–781, March 2010.

[22] C. Melani, M. Campana, B. Lombardot, B. Rizzi, F. Veronesi, C. Zanella, P. Bourgine, K. Mikula, N. Peyrieras, and A. Sarti. Cells tracking in a live zebrafish embryo. *Conf Proc IEEE Eng Med Biol Soc*, 2007:1631–4+, 2007.

[23] Jan J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370–370, August 1984.

[24] Andrew P. Witkin. Scale-Space filtering. In *8th Int. Joint Conf. Artificial Intelligence*, volume 2, pages 1019–1022, Karlsruhe, August 1983.

[25] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[26] Luis Alvarez, Pierre-Louis Lions, and Jean-Michel Morel. Image selective smoothing and edge detection by nonlinear diffusion. ii. *SIAM J. Numer. Anal.*, 29(3):845–866, June 1992.

[27] Jozef Kačur and Karol Mikula. Solution of nonlinear diffusion appearing in image smoothing and edge detection. *Appl. Numer. Math.*, 17(1):47–59, March 1995.

[28] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *Int. J. Comput. Vision*, 22(1):61–79, February 1997.

[29] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: From phase transitions to active vision. *Archive for Rational Mechanics and Analysis*, 134:275–301, 1996.

[30] Ravi Malladi and James A. Sethian. Visualization and mathematics. pages 329–ff., 1997.

[31] Yunmei Chen, Baba C. Vemuri, and Li Wang. Image denoising and segmentation via nonlinear diffusion. *Comput. Math. Appl*, 39:131–149, 2000.

[32] A. Sarti, C. Ortiz, S. Lockett, and R. Malladi. A unified geometric model for 3d confocal image analysis in cytology. pages 69–, 1998.

[33] Jozef Kacur, Jozef Ka Cur, and Karol Mikula. Slow and fast diffusion effects in image processing. *Comput. Visual. Sci*, 3:185–195, 1996.

[34] Karol Mikula and Narisoa Ramarosy. Semi-implicit finite volume scheme for solving nonlinear diffusion equations in image processing.

[35] Z. Kriva and K. Mikula. An adaptive finite volume scheme for solving nonlinear diffusion equations in image processing. 13(1/2):22–35, March 2002.

[36] N. Peyriéras B. Rizzi A. Sarti Z. Kriva, K. Mikula. 3d early embryogenesys image filtering by nonlinear partial differential equations. *Submitted to Medical Image Analysis.*

[37] Jian-Wei Zhang, Guo-Qiang Han, and Yan Wo. Image registration based on generalized and mean Hausdorff distances. *Gastroenterology*, 2005.

[38] N.Peyrieras A.Sarti P.Frolkovic, K.Mikula. A counting number of cells and cell segmentation using advection-diffusion equations. *Kybernetika.*

[39] A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62 –66, jan. 1979.

[40] Tony Chuang, Sharareh Moshir, Yuval Garini, Alice Chuang, Ian Young, Bart Vermolen, Richard van den Doel, Virginie Mougey, Mathilde Perrin, Martina Braun, Paul Kerr, Thierry Fest, Petra Boukamp, and Sabine Mai. The three-dimensional organization of telomeres in the nucleus of mammalian cells. *BMC Biology*, 2(1):12, 2004.

[41] Tianming Liu, Gang Li, Jingxin Nie, Ashley Tarokh, Xiaobo Zhou, Lei Guo, Jarema Malicki, Weiming Xia, and Stephen T. C. Wong. An automated method for cell detection in zebrafish. *Neuroinformatics*, 6(1):5–21, 2008.

[42] Jean-Christophe Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989 – 1996, 2002.

[43] Santos A. Vaquero J.J. Vallcorba I. García Sagredo J.M. del Pozo F. Malpica N., Ortiz de Solórzano C. Applying watershed algorithms to the segmentation of clustered nuclei cytometry.

[44] Carolina Wählby, Patrick Karlsson, Sara Henriksson, Chatarina Larsson, Mats F. Nilsson, and Ewert Bengtsson. Finding cells, finding molecules, finding patterns. In *MDA*, pages 104–114, 2007.

[45] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, January 1972.

[46] Barbara Rizzi. Bioimaging of animal embryogenesis: mathematical methods and computational algorithms. *Thesis*.

[47] B. Herman S. J. Lockett. Automatic detection of clustered, fluorescence-stained nuclei by digital image-based cytometry. *Cytometry*, pages 1–12, 1994.

[48] A I Dow, S A Shafer, J M Kirkwood, R A Mascari, and A S Waggoner.

[49] Norberto Malpica, Carlos O. de Solórzano, Juan J. Vaquero, Andrés Santos, Isabel Vallcorba, José M. García-Sagredo, and Francisco D. Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, 28(4):289–297, 1997.

[50] J. P. Rigaut, J. Vassy, P. Herlin, F. Duigou, E. Masson, D. Briane, J. Foucrier, S. Carvajal-Gonzalez, A. M. Downs, and A. M. Mandard. Three-dimensional DNA image cytometry by confocal scanning laser microscopy in thick tissue blocks. *Cytometry*, 12(6):511–524, 1991.

[51] M. Beil E. Nicopoulou D. Encaoua J. P. Rigaut T. Irinopoulou, J. Vassy. Three-dimensional dna image cytometry by confocal scanning laser microscopy in thick tissue blocks of prostatic lesions. *Cytometry*, pages 99–105, 1997.

[52] K. Rodenacker, M. Aubele, P. Hutzler, and P. S. Adiga. Groping for quantitative digital 3-D image analysis: an approach to quantitative fluorescence in situ hybridization in thick tissue sections of prostate carcinoma. *Anal Cell Pathol*, 15(1):19–29, 1997.

[53] C. T. Thompson D. Pinkel J. W. Gray S. J. Lockett, D. Sudar. Efficient, interactive, three-dimensional segmentation of cell nuclei in thick tissue sections. *Cytometry*, pages 275–286, 1998.

[54] A. Jones D. Pinkel J. W. Gray D. Sudar S. J. Lockett C. Ortiz de Solorzano, E. Garcia Rodriguez. Segmentation of confocal microscope images of cell nuclei in thick tissue section. *J. Microsc*, (193):212–226, 1999.

[55] C.O. De Solorzano, R. Malladi, S.A. Lelièvre, and S.J. Lockett. Segmentation of nuclei and cells using membrane related protein markers. *J Microsc*, 201(Pt 3):404–15, 2001.

[56] A Sarti, R Malladi, and J A Sethian. Subjective surfaces: a method for completing missing boundaries. *Proc Natl Acad Sci U S A*, 97(12):6258–63, 2000.

[57] A. Sarti, R. Malladi, and J. A. Sethian. Subjective surfaces: A geometric model for boundary completion. *Int. J. Comput. Vision*, 46(3):201–221, February 2002.

[58] Barbara Rizzi, Matteo Campana, Cecilia Zanella, Camilo Melani, Robert Cunderlik, Zuzana Krivá, Paul Bourgine, Karol Mikula, Nadine Peyrieras, and Alessandro Sarti.

[59] Ravikanth Malladi, James A. Sethian, and Baba C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(2):158–175, February 1995.

[60] D. H. Ballard. Readings in computer vision: issues, problems, principles, and paradigms. chapter Generalizing the hough transform to detect arbitrary shapes, pages 714–725. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[61] Luis Ibáñez, Will Schroeder, Lydia Ng, Josh Cates, the InsightSoftware Consortium, and Richard Hamming. The itk software guide. 01 2003.

[62] Will Schroeder, Ken Martin, and Bill Lorensen. The visualization toolkit: an object-oriented approach to 3d graphics [visualize data in 3d - medical, engineering or scientific; build your own applications with c++, tcl, java or python; includes source code for vtk (supports unix, windows and mac)]. 01 2006.

[63] A. Sarti and G. Citti. Subjective surfaces and riemannian mean curvature flow of graphs. *Acta Mathematica Universitatis Comenianae. New Series*, 70(1):85–103, 2001.

[64] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, November 1988.

[65] R. J. LeVeque. Finite volume methods for hyperbolic problems. *Cambridge Texts in Applied Mathematics*, 2002.

[66] S. Corsaro, K. Mikula, A. Sarti, and F. Sgallari. Semi-implicit co-volume method in 3d image segmentation. *SIAM JOURNAL ON SCIENTIFIC COMPUTING*, 28(6):2248–2265, 2006.

[67] F. Sgallari K. Mikula, A. Sarti. Co-volume level set method in subjective surface based medical image segmentation. *Handbook of Medical Image Analysis: Segmentation and Registration Models (J. Suri et al., Eds.)*, pages 583–626, 2005.

[68] Karol Mikula, Alessandro Sarti, and Fiorella Sgallari. Co-volume method for riemannian mean curvature flow in subjective surfaces multiscale segmentation.

[69] A level notes,http://alevelnotes.com/eukaryotic-cellular-division-process/147.

[70] C. Melani P. Bourgine N. Peyriéras M. Campana, B. Rizzi and A. Sarti. A framework for 4d-biomedical image processing, visualization and analysis. *Proc. 3rd International Conference on Computer Graphics Theory and Applications, 2008, 403–408.*

[71] Paul A. Yushkevich, Joseph Piven, C Heather Cody Hazlett, C Rachel Gimpel Smith, C Sean Ho, James C. Gee, and Guido Gerig B. User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31:1116–1128, 2006.

[72] Willy Supatto, Delphine Débarre, Bruno Moulia, Eric Brouzés, Jean-Louis Martin, Emmanuel Farge, and Emmanuel Beaurepaire. In vivo modulation of morphogenetic movements in drosophila embryos with femtosecond laser pulses. *Pro-*

ceedings of the National Academy of Sciences of the United States of America, 102(4):1047–1052, January 2005.

[73] L. Solnica-Krezel. Gastrulation in zebrafish – all just about adhesion? *Curr Opin Genet Dev*, 16(4):433–41, 2006.

[74] Ray Keller, Lance A Davidson, and David R Shook. How we are shaped: the biomechanics of gastrulation. *Differentiation*, 71(3):171–205, 2003.

[75] E. Schötz, R.D. Burdine, F. Jülicher, M.S. Steinberg, C. Heisenberg, and R.A. Foty. Quantitative differences in tissue surface tension influence zebrafish germ layer positioning. *HFSP J*, 2(1):42–56, 2008.

[76] W. R. Crum, T. Hartkens, and D. L. Hill. Non-rigid image registration: theory and practice. *The British journal of radiology*, 77 Spec No 2(suppl_2):S140–153, December 2004.

[77] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[78] T. Roberts, S. McKenna, N. Wuyts, T. Valentine, and A. Bengough. Performance of low-level motion estimation methods for confocal microscopy of plant cells in vivo. pages 13–, 2007.

[79] S. Yang, D. Kohler, K. Teller, T. Cremer, P. Le Baccon, E. Heard, R. Eils, and K. Rohr. Nonrigid registration of 3-d multichannel microscopy images of cell nuclei. *IEEE Trans Image Process*, 17(4):493–499, Apr 2008.

[80] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[81] S. Ullman. The interpretation of visual motion. 1979.

[82] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. A level-set based approach to image registration. pages 86–, 2000.

[83] Gert Wollny and Frithjof Kruggel. Computational cost of non-rigid registration algorithms based on fluid dynamics. *IEEE Trans. Med. Imaging*, 21(8):946–952, 2002.

[84] G.Wollny. Mia - a toolchain for medical image analysis, 2002.

[85] Jean-Philippe Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[86] Morten Bronielsen and Claus Gramkow. Fast fluid registration of medical images. pages 267–276, 1996.

[87] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. Image registration via level-set motion: applications to atlas-based segmentation. *Med Image Anal*, 7(1):1–20, March 2003.

[88] Maria J. Ledesma-Carbayo, Andrew J. Derbyshire, Smita Sampath, Andrés Santos, Manuel Desco, and Elliot R. Mcveigh. Unsupervised estimation of myocardial displacement from tagged MR sequences using nonrigid registration. *Magnetic Resonance in Medicine*, 59(1):181–189, 2008.

[89] Philipp J. Keller, Annette D. Schmidt, Joachim Wittbrodt, and Ernst H. Stelzer. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science (New York, N.Y.)*, 322(5904):1065–1069, nov 2008.

[90] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science ... on Applied and Computational Mathematics)*. Cambridge University Press, 2 edition, June 1999.

[91] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[92] R. M. Warga C. B. Kimmel and D. A. Kane. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development*, pages 265–76, 1994.

[93] K. Woo and S.E. Fraser. Order and coherence in the fate map of the zebrafish nervous system. *Development*, 121(8):2595–609, 1995.

[94] S.J. England, G.B. Blanchard, L. Mahadevan, and R.J. Adams. A dynamic fate map of the forebrain shows how vertebrate eyes form and explains two causes of cyclopia. *Development*, 133(23):4613–7, 2006.

[95] Z M Varga, J Wegner, and M Westerfield.

[96] Juliette Mathieu, Anukampa Barth, Frederic M Rosa, Stephen W Wilson, and Nadine Peyriéras.

[97] Stephen W. Wilson and Corinne Houart. Early Steps in the Development of the Forebrain. *Developmental Cell*, 6(2):167–181, February 2004.

[98] In2p3, institut national de physique nucléaire et de physique des particules (in2p3).

[99] http://www.eu-egee.org.

[100] Anne-Cécile Petit, Emilie Legué, and Jean-François Nicolas. Methods in clonal analysis and applications. *Reprod Nutr Dev*, 45(3):321–39.

[101] J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.*, 100(1):64–119, Nov 1983.

[102] Zhirong Bao, John I. Murray, Thomas Boyle, Siew Loon L. Ooi, Matthew J. Sandel, and Robert H. Waterston. Automated cell lineage tracing in caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8):2707–2712, February 2006.

[103] P. Bourgine, M. Campana, R. Cunderlik, O. Drblikova, E. Faure, B. Lombardot, C. Melani M. A. Luengo-Oroz, M. Remesikova, B. Rizzi, T. Savy, C. Zanella, J. Kollar, I. Colin, S. Desnoulez, M. Funabashi, L. Duloquin, S. Randoux, E. Courtade, E. Hirsinger, A. Santos, E. Beaurepaire, P. Herbomel, P. Suret, G. Lutfalla, J. F. Nicolas, R. Doursat, A. Sarti, K. Mikula, and N. Peyriéras. Embryomics: Reconstructing the cell lineage tree as the core of the embryome. *Submitted to Nature*.