RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Antonio Cavallo

Four dimensional particle tracking in biological dynamic processes

Dissertation

HD-KIP-02-19

KIRCHHOFF-INSTITUT FÜR PHYSIK

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diplom-Physik:	Antonio Cavallo
born in:	Roma (Italy)
Oral Examination:	19.06.2002

Four dimensional particle tracking in biological dynamic processes

Referees: Prof. Dr. B. Jähne
Prof. Dr. C. Cremer

Here is presented a general approach to data analysis in multidimensional space using the run length encoding process and volume field method. A study on volume field was carried in order to test the limits on the method and testing against simple Gaussian objects (spots) was made: quantitative measurement was then carried using an angle distribution parameter between the computed results and the expected values. An application of volume field method was tested against results from simulations of chromosome territories using the spherical loop domain (SCD) and the quantitative comparison was made in order to clarify real applications in biologic related field. Two experiment were made, using the run length encoding method. From biological samples subjected to in vivo labeling an then to FISH hybridization, we analyzed the sub-chromosomal movements to see if there were modifications to the underlying structures: we quantify these displacements and compared to other measurements made with other methods. A time recorded HeLA cell sample in vivo labeled was analyzed and his internal marked chromosomes were tracked to analyze displacements related to the activities: here, using the run length encoding, spots were tracked and extracted. A successive Monte Carlo minimization algorithm was used in order to reduce systematic errors.

Hiermit wird ein allgemeiner Annährungsversuch zur Bildanalyse in multidimensionalen Raum vorgelegt, der des run lenght encoding Prozesses und volume field methode sich bedient. Die Forschung über die volume field methode wurde durchgeführt, um die Grenzen der Methode nachzuprüfen und, um die Gaussianobjekte (spots) zu analysieren: quantitative Messungen erfolgten dank dem Vergleich zwischen den Ergebnissen und den vorgesehenen Daten, indem man ein "angle distribution parameter" verwendete. Eine Anwendung von der volume field methode wurde bezüglich der Ergebnisse aus den Simulationen von den chromosomenbereichen getestet, indem man das spherical loop domain (SCD) verwendet hat, und der quantitative Vergleich wurde mit dem Zweck durchgeführt, die tatsächlichen Anwendungen in dem biologischen Bereich zu erklären. Es wurden zwei Experimente mit der Methode run lenght encoding durchgeführt. Aus biologischen Mustern, die in vivo labeling und dann zu FISH hybridization unterzogen wurden, haben wir die subchromosomischen Bewegungen analysiert, um zu beobachten, ob Modifikationen in den unteren Strukturen vorliegen: wir haben diese Verstellungen quantifiziert und wir haben sie mit anderen Messungen verglichen, die mit anderen Methoden durchgeführt wurden. Ein in der Zeit gemessenem Zellmuster HeLA in vivo labeled wurde analysiert und ihre intern markierten chromosomen wurden beobachtet, um die Verstellungen zu analysieren, die mit den Bewegungen gekoppelt sind: hier, indem man die run lenght encoding verwendet, wurden spots beobachtet und herausgezogen. Eine darauf folgende Monte Carlo Algorithmminimisierung wurde dazu verwendet, die systematischen Fehler zu reduzieren.

To the open source movement

Contents

Part I

Biological background

Chapter 1

The cell nucleus organization

1.1 Preamble

The cell nucleus should be a quite chaotic place to live in: there are many "objects", an un-interruptible movement of these and an infinite exchange of many compounds in there as well. On the wall* delimiting the interior of the cell from the "exterior" world there are the gates (protein channels); then we can encounter such an exotic series of objects like Golgi apparati, centriole, lysosome mitochondria chloroplasts and other: all they constitute the living place of the nucleus or the city squares and streets where these "inhabitants" are. But when we go further, we reach the core of this quite strange and active city: the nucleus. That is the place where the action starts, where all the important decisions are made and it is the most conspicuous organelle in the cell (some of 10 % of the total length of a cell): it is a sort of "Wall street" of the cell economy. The cell nucleus is a quite important object from a chemical point of view: it is basically a container for the DNA, separated from the rest of the cell by a double layered membrane. This is a huge improvement from the evolutionary point of view because this separates the two major classes of cells: eucaryotes (with the nuclear membrane) and the procaryote (without), being the former the building blocks of any complex biological system (plants, animals and fungi) and the latter the base for the simpler living systems (bacteria). Our analysis will be based on some studies about the cell nucleus structure, the methods used to analyze it and novel method from digital analysis domain applied to the biology.

*the plasma membrane

1.2 Cell cycle

The cell body is composed of two sharply distinct parts: a cell nucleus and the region surrounding it (cytosol). This separation is marked by the nuclear envelope formed by two concentric membranes punctured by nuclear pores responsible for nucleus-cytosol exchange of chemicals: of course we are talking of eucaryotic cells, because the procaryotic ones lack of this sharp separation, being the DNA directly floating into the cytosol. When the DNA is not splicing (out of the so called M-phase or mitosis as in Figure 1.5), the nucleus appears to be an in-active object (at least it is in-active to the optical observation see Figures 1.6 on page 9 and 1.7 on page 10). The landscape changes suddenly at the beginning of the M-phase, when the already duplicated DNA is being transferred to the child new cells in this sequence (see Figure 1.8 on page 11):

1. The DNA is brought to the highly condensed form (**prophase**)
2. The nuclear envelope is broken (**prometaphase**)
3. The chromosomes are aligned in the cell metaphase plate (**metaphase**)
4. The sister kinetochores are separated and begin the migration to the cell poles (**anaphase**)
5. The nuclear envelopes are regenerated around the two opposite poles of the cell (**telophase**)
6. The cleavage process begins to separate the cytoplasm in two parts to give two sister cells (**cytokinesis**)

This M-phase (or mitosis) is the optically active process as seen in Figure 1.8 on page 11: most of this process is well known and documented[1]. Out the mitosis, there are still three phases called respectively G_1, G_2 (gaps) and S: G_1 and G_2 are time gaps in between the two main M-phase and interphase (Figure 1.5 on page 9). The S phase (or synthesis) is the time for DNA syntetization and with the G_1 and G_2 all they constitute the interphase; this one is the time when the cell nucleus is optically inactive (as in Figure 1.7 on page 10), although is not at all a "stable" status.

1.3 Orders in the interphase

Although there is ample space in the nucleus for the eukaryotic genome, the enormous length of the DNA molecules requires that they be extensively folded, so the *interphase* DNA has to be in a packed status: however this folding structures are very obscures. It is believed that, in this "packaging", there is some room for the histone material also present in the nucleus and accounting for half part of his mass: this histone are present only in eucaryotes cells, confirming that participation. One accredited hypothesis is that chromosomes are tied around an octameric histone (nucleosomal histone or *H2A H2B H3 H4*) core sizing 30 nm diameter (Figure 1.3 on page 8): this is the first stage in the DNA arrangement called nucleosome. The second stage is believed to be formed by "beads on a string" structures (Figure 1.4 on page 8) where the nucleosomes are arranged in a tied structure called 30-nm chromatin fiber: these chromatin

fibers are linked by short chromosome strands (7kbs). This "beads on a string" structure is controversial [44] because only *in vitro* measurement can be done (at least until the scanning force microscopy was introduced); extensions to *the in vivo* should be carefully tested and just the possible presence of one higher structure than nucleosome can be accounted from the experiments: recent informations [50] accredit for a "kinky" fiber sequence instead of a "beads on a string" or solenoid structure, but this is still an open problem. Nowadays the packing problem is a crucial question in life science (for reference see [6]): this packing is related to the general problem of genes activation and duplication.

1.4 Chromosome territories and CT model

Whatever the structure of the chromatin fiber is, there are convincing experimental evidences showing that a structure exists in the arrangement of the chromosome inside the cell nucleus: it has been observed that ithere are well defined territories inside the nucleus, segregated each other[6][48][10]. Experimental results show small movements of these territories[26][4] during an observation period of several hours although this finding can be true only for the few cultured cells so far studied. A relation between human chromosomes 18 an 19 and CT[†] location was recently provided[14]: although both have similar DNA contents (85 and 67 Mb, respectively), the gene-poor chromosome 18 territories were typically found at the nuclear periphery, whereas the gene-rich chromosome 19 territories were located in the nuclear interior. So all we know about CT is that they are quite stable, located in particular and well defined positions: increasing knowledge about their internal structures and their dynamical behavior is very important to understand the mechanisms of the genes activation and duplication.

1.5 Dynamics inside the chromosome territories: the loop models

As we have seen in the previous sections, the chromosome in interphase can show at different orders different structures ("beads on a string", chromatin fibers and/or "kinky" fiber sequences and territories); these are the static part of the cell cycle: being a living object the important things are related to the dynamical behavior. Especially important are the processes of *transition from the condensed phase to the un-condensed one* (where the chromosome forms the CT[49][13]) and the *dynamical property of the CT* itself when it is in the interphase stage. Moreover the unfolding of the chromosome is a process that cannot be visualized (requiring less than few seconds, in the actual simulations): it is possible to analyze this unfolding process through theoretical models and numerical simulations only. The models introduced so far are based on the 30 nm chromatin fiber arranged in a linear sequence:

- Giant Loop / Random Walk (GL/RW)[38]
- Multi Loop Subcompartment (MLS)[29]

[†]Chromosome Territory

- Spherical 1-Mbp Chromatin Domain (SCD)[24]

The first model, or the giant-loop model, is based on the polymer model[‡]: here the chromosomes are approximated by a long chain of "beads" linked by straight lines (Figure 1.1 on the following page). The movements of the "beads" are modeled after a looser random walk for genomic distances [§] below the 2-200Mbp range: when the genomic separation reaches this range a tighter model (Giant Loop) it is introduced linking the "beads" into a greater structure, to account for experimental observations. In the Multi Loop Subcompartment the same polymer model is used to describe a more complex structure as in Figure 1.2 on the next page where a base spring is supposed to join many loops. In the third model, the spherical chromatin domain, the loops from the MLS method are replaced by 1Mbp (500 nm) spheres as found in experiments with marked samples[4][48][5]: this is driven by the limited knowledge concerning the actual folding of the chromatin fiber at the ultrastructural level and to reduce required computing time. Moreover it makes no assumptions on the chromatin topology inside the 1-Mbp chromatin domains itself: in this model the flexible linker connection of adjacent domains is approximated by an entropic spring with the rigidity of a "real" 120kbp linker connection according to the MLS model.

[‡]The $30nm$ chromatin fibers structures

[§]The distance between "beads" computed in Million of base pairs

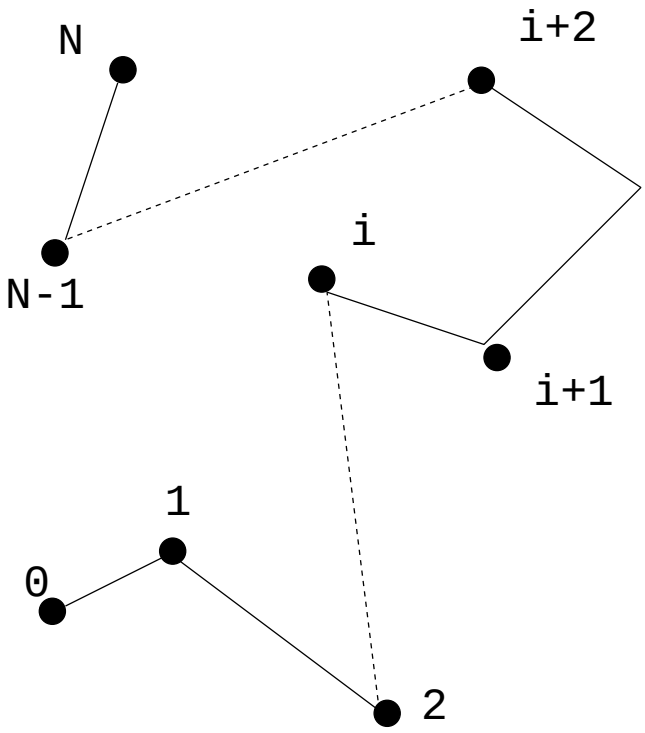


Figure 1.1: Polymer approximation of a chromosome: the lines between "beads" (black spots) are simple approximation to the tortuous path between beads.

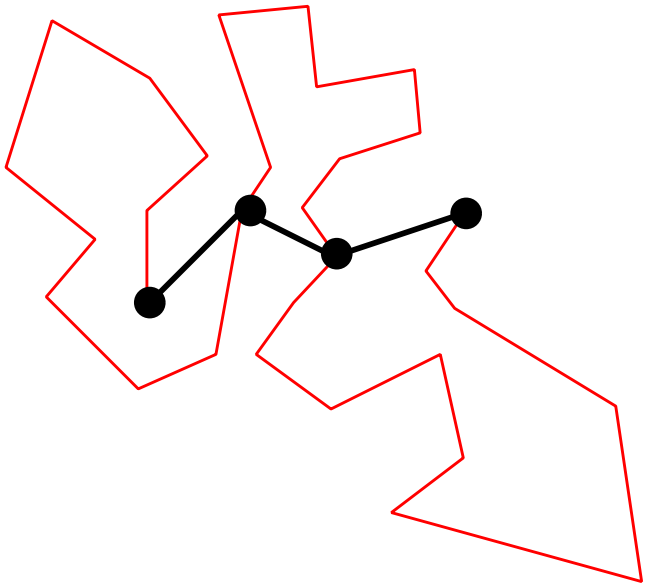


Figure 1.2: Multi loop: in red is represented the chromosome chain and in black the spring like structure joining it

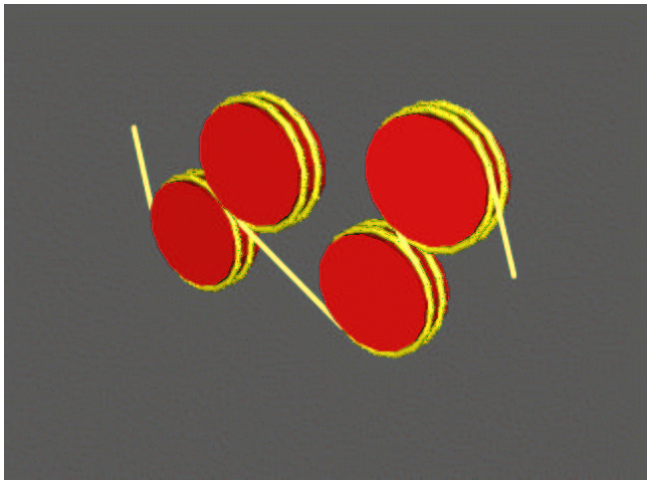


Figure 1.3: The chromosome displacement around the histone core (yellow is the DNA double strand, brown the octameric histone core).

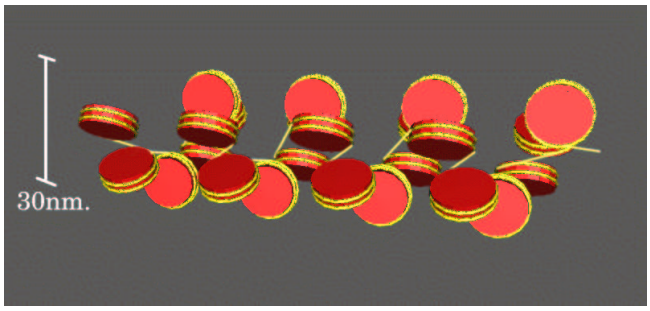


Figure 1.4: The chromosome "beads-on-a-string" (yellow is the DNA double strand, brown the octameric histone core): actually the structure is more compact than this.

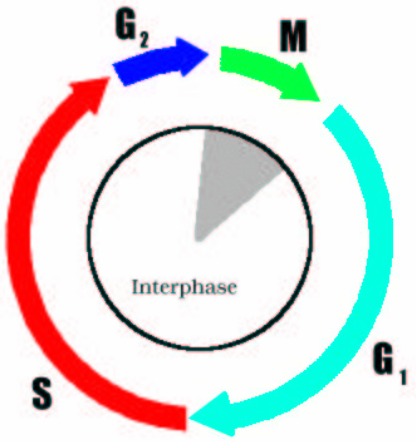


Figure 1.5: The cell living phases. (The time cycle is roughly 24 hours: the M phase accounts for 1/2 hours, S for typical 12 hours and the rest is to G_1 and G_2)

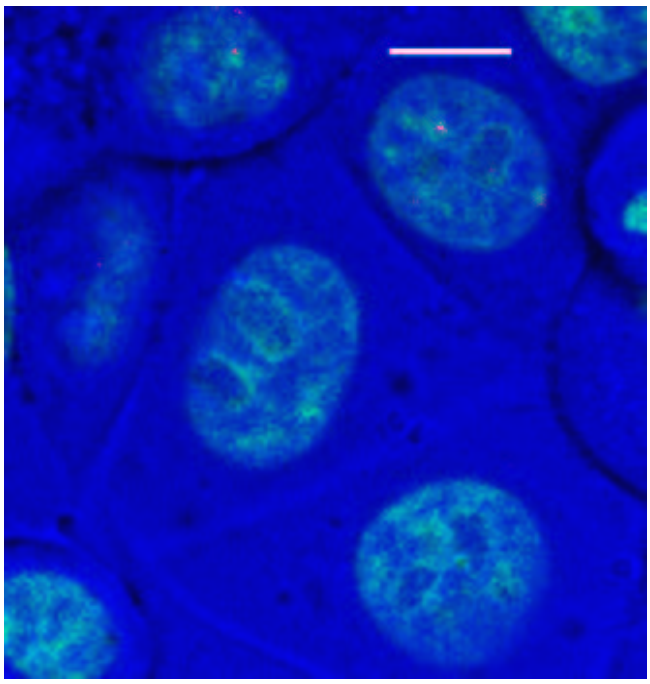


Figure 1.6: Many HeLA cell nuclei (in green is shown the nucleus, in blue the whole cell): on top $10\mu m$ ruler. (Contribute from dipl. phys. L. Schermelleh)

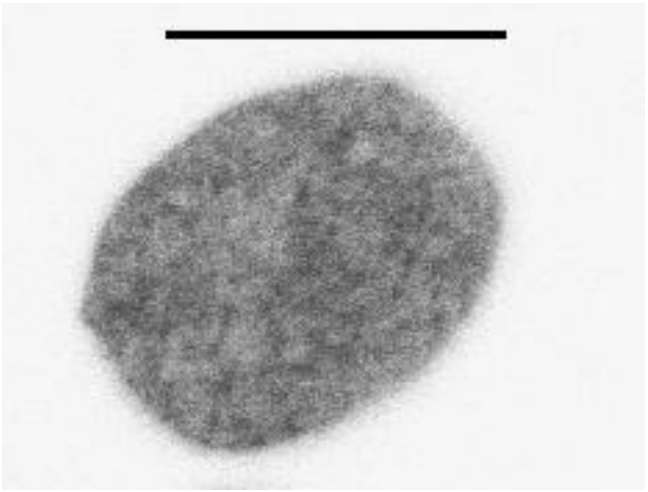
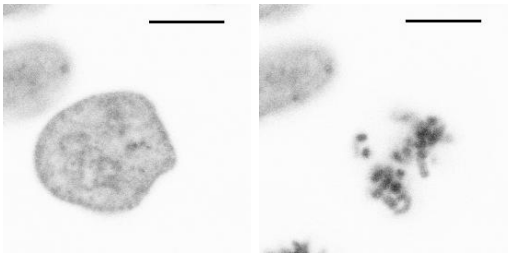
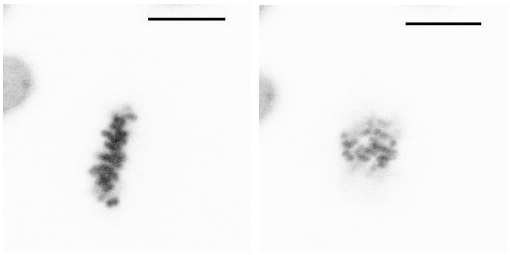


Figure 1.7: An HeLA cell nucleus (magnificated $2.5x$): on top $10\mu m$ ruler. (Contribute from dipl. phys. L. Schermelleh)



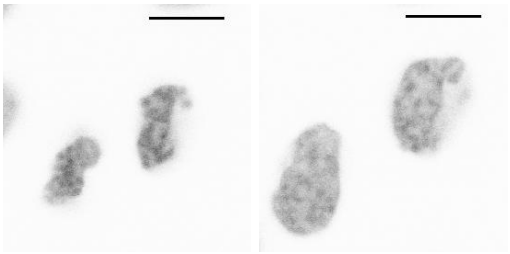
(a) Start of the mitosis: t=0s

(b) Condensation and breakage of the nuclear envelop (prophase-prometaphase): t=900s



(c) Metaphase: t=1800s

(d) Anaphase: t=2700s



(e) Telophase: t=3600s

(f) Cytokinesis: t=4500s

Figure 1.8: Mitosis (on top of each image is a $10\mu m$ ruler: scanning voxel $120x120x500$ nm). (Contribute from dipl. phys. L. Schermelleh)

Chapter 2

Scanning of cell nuclei

2.1 introduction

The cell nucleus is a complex environment: it encloses the DNA the biological recipe for the life, it hosts the production of the RNA and/or production of proteins fundamental to any biological activity. Despite the discoveries about the cell nucleus, the lack of suitable of scanning devices is stronger than ever: the most used tools for inspection so far developed are the confocal scanning[28] device and the electron microscopy* as well Scanning Tunneling Microscopy†. However there are intrinsic limitations to the devices so far introduced: confocal scanning systems are able to do *in vivo* scanning but they lack resolution, electron microscopes require to kill the sample and scanning tunneling devices scan just surfaces. So only an indirect or not so much detailed analysis could be possible at that moment.

2.2 Optical sectioning by confocal scanning system

The confocal scanning microscope is based on a theoretical ground proposed by Minsky [28] in the 1957 but just starting from the 80's were ready commercial systems: the schematic of a confocal scanning microscope is given in the Figure 3.1 on page 17. The schema of the apparatus is tailored to scan along both x and y directions (the scanning plane) as well the z (axial) direction (Figure 3.2 on page 17) : by slicing the sample we obtain a full volume scanning. The z sectioning is made using the effect of the confocal system being its point spread function along the axial direction of a thin lens[18]:

$$I(u) = (\pi N)^2 [\frac{sin(u/4)}{u/4}]^2 \tag{2.1}$$

given $u = \frac{2\pi a^2}{\lambda f^2}\Delta z$, λ wave length of the laser beam, Δz slice thickness in the axial direction, $N = \frac{a^2}{\lambda f}$ Fresnel number and a pinhole aperture: using aperture

*Nobel in 1986, Ernst August Friedrich Ruska

†Nobel in 1986, Gerd Binnig and Heinrich Rohrer

variable pinholes is possible to reduce the field depth (Δz) and analyze the data plane by plane: in a laboratory device like we used in our experiments the typical resolution could exceed the 50 nm along the x and y axes, and on the z axial direction this is nearly 250 nm: this is a practical example to give some figures to the equations so far introduced. However, there are not just the physical limitations to the scanning process but also there are present systematical errors too:

- drift, from the mechanical instability of the moving mechanisms
- color shift, when using different wave lengths there are different focus planes
- optical aberrations, usually due to the optical elements
- drift of the sample itself, when scanning times become long
- photon noise, natural in any optical device

All these are just scratches on the surface of the confocal scanning systems, being that discipline itself a wide field of studies: to our application purposes we will not go into further details, being required just a general understand of the scanning device and of the sources of the problems related to the digital analysis.

2.3 Fluorescent marking techniques

Confocal scanning requires fluorescent marked samples so the device can image them: techniques to mark chromosomes have been developed using many different fluorescent probes making possible to analyze different targets at the same time. Introducing fluorescent markers is a complex task and could be obtained both in living or fixed cells: it depends on the experiment requirement to decide which procedure to apply. In general *in vivo labeling* is the procedure to follow when we manage experiments with living cells: it requires just the incorporating of the fluorescent marker into the cell. On the other side the *FISH* technique (FISH stands for *fluorescence in situ hybridization*) is a strong procedure requiring at least the denaturation of the chromosomes: this makes the procedure unsuited for living analysis.

2.3.1 In vivo labeling

The *in vivo labeling*[16][47] uses the duplication phase in chromosomes: during the S phase (see Figure 1.5 on page 9) there are probabilities that modified nucleoside[‡] could be tied to the DNA strand incorporating fluorescent probes that can be later visualized by laser excitation. This marking requires normally some replication sequences to incorporate the labeling probes, but finally they don't modify too much the underling structure of the chromosomes that continue to be living: so it will be possible to track chromosomes in a living environment.

[‡]We used, for example, the Cy3-dUTP complex, composed by an Halogenated deoxynucleoside dUTP plus the fluorescent probe Cy3 (wavelength 543nm)

2.3.2 FISH technique

The chromosomes painting is not so old[34]: one of the most prominent technique is called FISH or short for *fluorescence in situ hybridization*. Basically FISH is following the protocol here listed[43]:

- Sample preparation (source, fixation and un masking)
- Probe preparation and labeling
- Denaturation of probe and sample
- Hybridization of probe to sample (annealing)
- Post-hybridization washing
- detection

This is a procedure that kills the cell because it is an harsh treatment (it requires denaturation of the sample): so it is unsuitable for in vivo processing (opposite to the *in vivo labeling*). On the other side FISH is site-specific and the fluorescent markers could be incorporated in specific sites into the chromosome giving a very indeep knowledge about the target: progresses in the markers bring an entire collection of library markers specific to many different tasks. Although the usefulness of this technique, a basic important question is raised because the sample is affected from FISH: are the morphological characteristics of the chromosomes preserved in a FISH treatment? This is a fundamental requirement for using such a technique because if FISH changes too much the underling structures, then it will be useless.

Chapter 3

Biological Goals

So far we have introduced a general description of the biological background (for an overview[6]): we would follow the general trends and catch up with some open problems. To our knowledge (quite limited, at the moment) one question is remarkable, deserving further investigation: are the models so far introduced (see Section 1.5 on page 5) related to experimental parameters? We know that macro-parameters * can be extracted from experimental data and then compared to a developed theoretical model: so we would compare the results from speeds detected using the volume field method † to the theoretical computed speeds from a developed model (see 1.5 on page 5). *The first question, aim for our analysis, is to apply a volume field method to correlate experimental data and theoretical previsions.* But in the samples there are important informations hidden in the smallest detectable features‡: so more question are posed every day in works with biological samples but related to a different scale problem. So a second open question stems from the so described FISH processing 2.3.2 on the preceding page: *what degree biological information is preserved when we apply some modifications to our sample? Are the procedures senseless?* Here digital imaging processing is required to give an explicit answer, because any further experiment can be carried only and only if there are evidences that the methods used are reliable. The third question is more biologically relevant: we know that in interphase there is a lot of activity (even if this exhibits a low optical activity), how can we extract informations from this phase? In the interphase period, chromosomes are in the synthesis phase (S in Figure 1.5 on page 9): active foci make small movements related to their activities and their position is not that casual as well (see Section 1.4 on page 5), and tracking these movements is crucial in the understanding of cell nucleus activity. To summarize we have these three questions to cope with:

- Linking between theoretical model of the chromosome, and experimental results
- Analyze the deep correlation between the treatment of the sample and the

*as the $< r^2 >$ distributions of the chromosome density[38]

†see Section 4 on page 19

‡Eg. the transcription sites in a chromosome or nuclear lamina connection points, can not be described with a simple model using "beads", but nevertheless their analysis is quite important in biology today

meaningfulness of the results

- Discover if there are relations between chromosome displacements and their activity in the nucleus

We hope to solve these problems in the current work, as well to identify the problems always present in an ongoing scientific progress.

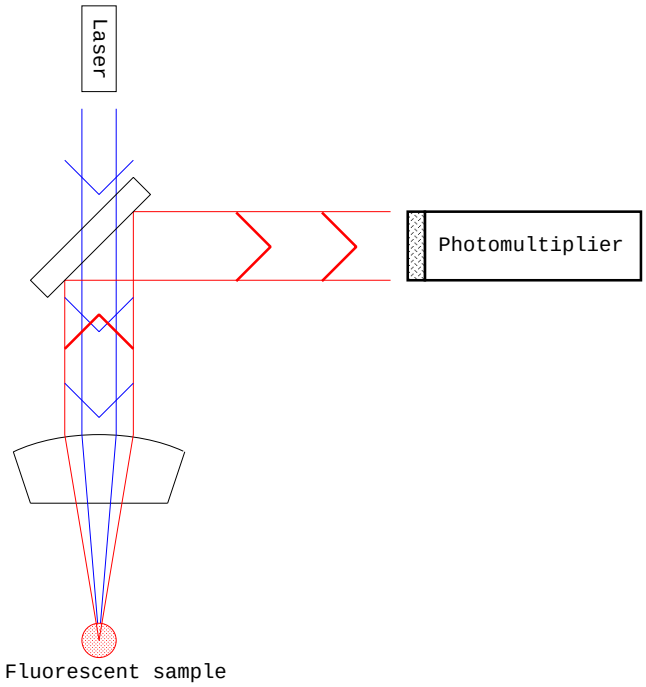


Figure 3.1: Confocal scanning microscope: general schema.

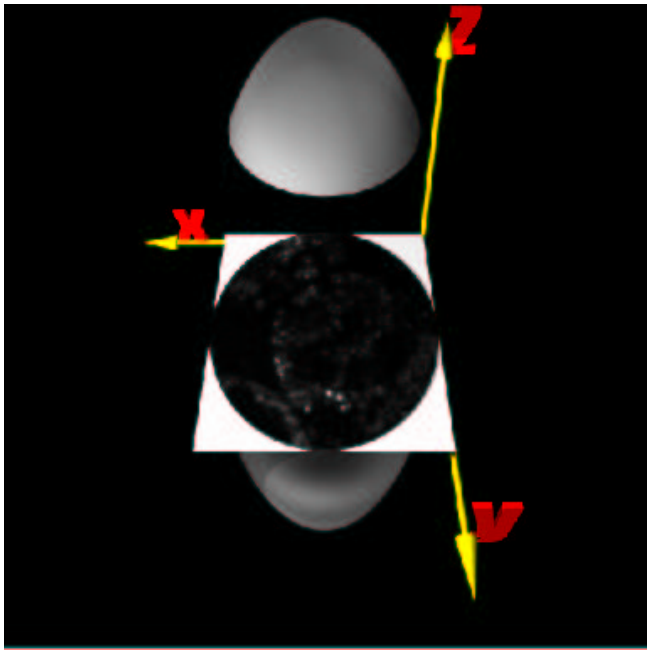


Figure 3.2: Optical sectioning.

Part II

Image processing problem: tracking random motion.

Chapter 4

Introduction

4.1 Goals and methods

So far we have introduced the biological background and the goals of our ongoing analysis: here we would introduce the tools and the strategies we will use. First and foremost important problem is to be able to identify the dimensionality of the problem and we will discuss the connection between measurement regarded as process of dimensional reduction (Section 4.2). Second we will describe what can be expected from a data set and the "parameters" to be extracted (Section 4.3 on the following page): this is the measurement process description. Third we will introduce the "object" description and its propriety: in particular we will link the object property to the "parameters" (Section 4.4 on page 21).

4.2 Dimensional squeezing

In a true/false experiment* it is easy to "display" the outcome; this experiment is a typical 1D problem, but you can image quite a lot of such experiments: any experiment giving a numerical result (parameter) is still one dimensional.

Everything you can give as number is a one dimensional experiment (1D).

When the problem gets more complicate the "space" of the problem raises to the next dimension order: two dimensional problems are quite common today as well as in the past. Representing results of a two dimensional problem is matter of writing "plots": nothing new here.

All you can display within a "plot" is a two dimensional experiment (2D).

What we mean for dimensionality here is not the mathematical dimension of the domain $f(x) : \mathbb{R} -> \mathbb{R}$ but the couple $(x, y) \in \mathbb{R}^2$: introducing the function $f(x)$† *reduce the dimensionality* of one unit on our acception.

Actually a general established trend in biology, the field we are dealing with, is to scan for pictures using the microscope: is not that new that people are looking into an objective to see samples, it was (and it is) full of new discoveries. That path is so successful that more accurate tool are demanded and the first step is, normally, to replace the eyes scientist with a camera for two trivial reasons: it is repeatable and let the people use tools from the digital analysis

*We consider true/false experiment the most well crafted experiment.

†we can call it a "model": we will detail it later.

world. Here we add one dimension more to the problem (the image intensity): so we are dealing with three dimensional problem, and display the result is a matter of showing the nice pictures all we are used to see.

All you can display within a picture is a three dimensional problem (3D).

So imaging the two[‡] dimensional pictures of biological samples into a figure poses no problem: the improvement is to be able to get out of these picture "plot" or two dimensional data (eg. intensity profile) or, better, one dimensional parameters (eg. area surface). The next step is when we are able to scan for volumes along the time, raising the space dimensionality to higher levels: here the resulting space is five dimensions wide[§]. With such a number of dimensions we loose the original ability to explain our results as in one dimensional problem [¶]: we have to manage to reduce, or squeeze, the dimensions to recover our beloved one dimensional problem. We will call the process of dimensional reduction of the original problem down to the smallest dimension possible *dimensional squeezing*[27]: basically this is nothing more than the measurement process described from a different point of view.

4.3 Observable in a multidimensional data set

In a volume data set there are results and it is just matter of getting the desired ones: these could vary from the volume estimation of the cell nucleus to the density displacement of chromosome territories. The multidimensional analysis is driven by the physical or biology related requirements; when we are looking for the spots inside a cell nucleus we would track them, see their speeds and see if there is some trend in their movements: so we focus our attention to analyze the dynamical propriety. Another choice could have been the sizes of these spots and/or intensity bleaching along the time: but characterization of movements or speeds of a point is natural and well understood (actually tracking is a "hot topic" in biology [33][2][21][26][30], still full open to further investigations). *So the first observable to deal with in a multidimensional data set could be the speed of one "spot".* To analyze speeds (or movements) is necessary spot detection, so this is some sort of "object" analysis and the workflow of this experiment class is quite simple [‖]: segment the volume data (for every time step) → track the spot →compute the speed or plot the displacements. There are, also, cross path problems when two distinct trajectories will cross each other: here there is a typical problem of the distinguishable paths (as summarized on Figure 4.1 on page 22).

Another approach to the data analysis is done using the vector field[42][20] (or better volume field) approach: so *the vector field of the speeds for an image* could be chosen as observable. A different approach to the imaging problem is taken here because it does not focus on the "spot" [**] but on the whole data set thought as an intensity field: this approach does not require a spot tracking. With volume field method a speed field is computed for every point in the volume

[‡] We intend the dimensionality here as the dimension of the domain $I(\mathbf{x}, y)$, but we would point the every point in a picture is modeled after the general ensemble (x, y, I)

[§] Actually six if we include multichannel images

[¶] Simply an yes/no answer or a number

[‖] Although problems arise when occlusion problems are present[45] if, for some reason, there are missing spots along the path (not a so un-common case in biological samples)

[**] Or object

space: this speed is pointing along the direction of the brightness constant vector. The tradeoffs to this method stem from the need for a smooth movement of the sampled objects (the method itself poses a limit on the displacement see Section 5.5 on page 25) and from the required absence of any source to the total intensity of the volume (conservation of the intensity, see Section 5.2 on page 23).

4.4 Objects and parameters

In the previous section (4.2 on page 19) we already have outlined a small, although important, difference between two fundamental approaches to the analysis: we pointed that we can extract from volumetric data sets the point paths (particle tracking) or the speed field of intensities (volume/vector field). We would differentiate sharply these two kind of results introducing the concept of "object": this is a trivial concept used, we hope, in an effective way. From a volume scanning appears clear what an object is: sometimes it is a rounded spot, sometimes is a cluster of smaller spots or could be anything resembling a compound structure (see Figure 7.1 on page 34 or Figure 4.2 on the next page). This object detection is trivial for everybody and we are used to do it in a natural way: no matter about it, we show an image (or a volume) and we can point our finger to the "spot" without any further doubt. It is natural, also, to point our finger to this spot and asking something like: how big is it[††]? So in our mind we have a hierarchical scheme:

- The cell nucleus (our multidimensional world)
- The spot (our object)
- The position and/or the size of the spot (our uni-dimensional parameters)

This scheme is about *dimensional squeezing* or reducing our dimensions to the minimum just to obtain the searched parameters: but there is the object in between our world (multidimensional) and our goal (the parameter). Being able to describe in a general way the object(look at Section 6 on page 27) is the same than having all the parameters (or better having a way to evaluate them): the missing link is the spot or the "object" concept in his extended definition. This is a different point of view from the speed field approach because there is no requirement to introduce an "object" concept: basically we are directly analyzing the parameter (speeds) without any knowledge about what an object is.

[††]A parameter, as in 4.2 on page 19

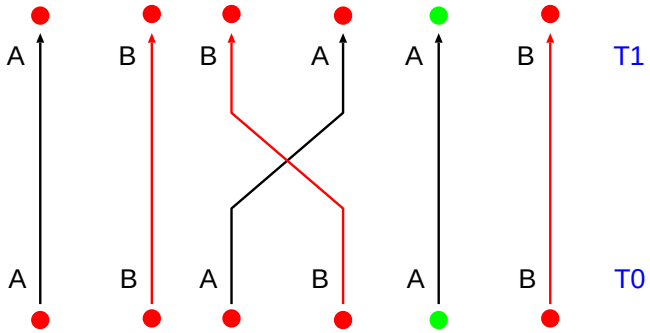


Figure 4.1: Typical cross path problem: the first two pictures show two spots indistinguishable crossing their respective paths, the third one shows two distinguished paths. On side the time frames: if we miss the intermediate step, no information is available of the crossing.

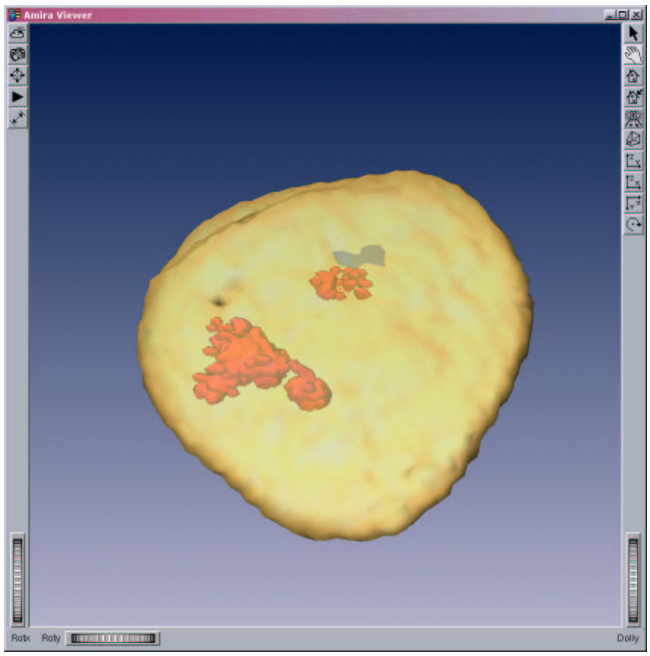


Figure 4.2: A cell nucleus (yellow part) with marked chromosomes (red): the red part could be described as an "object"

Chapter 5

The brightness change constraint equation (B.C.C.E.)

5.1 Introduction

As we pointed in 4.4 on page 21 we are going to introduce the volume field method*: we are assuming here that our scanning volume is a discrete function of the intensity, with no knowledge about his internal composing parts (objects as in the Section 6 on page 27).

5.2 The model of intensity volume

We can model the intensity inside a volume [20],[42], along the time, with a function $g(\overrightarrow{x}, t) : \mathbb{R}^4 \rightarrow \mathbb{R}$. The total derivative is then given by:

$$\frac{dg}{dt} = \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} + \frac{\partial g}{\partial z}\frac{dz}{dt} + \frac{\partial g}{\partial t} \tag{5.1}$$

This is the classical flow conservation equation as seen in [17] extended to a $4D$ space: we can rewrite it using the more compact vector notation, under the assumption of $\frac{dg}{dt} = 0$ (conservation rule).
With:

$$\overrightarrow{f} = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}) \tag{5.2}$$

We have:

$$\nabla g \cdot \overrightarrow{f} + g_t = 0 \tag{5.3}$$

The $\overrightarrow{f}$ in equation 5.3 is called volume flow (it has to be determined) and g_t takes into account for time variations. We can assume as preliminary condition the existence of a vector of constant brightness $\overrightarrow{r}$ (we assume it exists) with such a form:

*We will not differentiate any further vector field from volume filed because they are basically the same method applied to different dimensional problems

$$\begin{aligned} \underline{r} &= [r_1, r_2, r_3, r_4] \\ r_1/r_4 &= f_1 \\ r_2/r_4 &= f_2 \\ r_3/r_4 &= f_3 \end{aligned} \tag{5.4}$$

The equation 5.1 can be rewritten as:

$$[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}, \frac{\partial g}{\partial t}] \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = (\nabla_{\vec{x}\,t} g)^T \underline{r} = 0 \tag{5.5}$$

It's worth noting that $\underline{r}$ should be pointing along the line of constant brightness inside the $\mathbb{R}^4$ space (composed as $3D$ volume and $1D$ time). The search for such $\underline{r}$ vector in the equation 5.5 leads to the following problem:

$$\begin{gathered} \underline{r} = min_{|\underline{r}|=1} \|e\|_2^2 \\ \|e\|_2^2 = \int_{\mathbb{R}^3} w(\vec{x} - \vec{x}')[(\nabla_{\vec{x}\,t} g)^T \underline{r}]^2 \, d\vec{x}' = < [(\nabla_{\vec{x}\,t} g)^T \underline{r}]^2 > \\ \text{where: } < \cdot > = \int_{\mathbb{R}^3} w(\vec{x} - \vec{x}')(\cdot) \, d\vec{x}' \end{gathered} \tag{5.6}$$

5.3 The eigenvalues problem to solve the B.C.C.E.

In the equation 5.6 $w(\vec{x} - \vec{x}')$ is a "windowing" function: this reduces the evaluation to a local neighborhood. This function can be a simple box filter, a binomial filter or a more complex function such a Gaussian shaped filter: choosing the right one is not that crucial but depends on the data. This "windowing" is mainly necessary because we require a constant value of $\underline{r}$ in a small region: if the window is larger then this vector cannot be constant on all the sub-volume (due to the brightness changes). Assuming a constant $\underline{r}$, we can rewrite 5.6 as:

$$\begin{aligned} \|e\|_2^2 &= < \vec{r}^T \nabla_{\vec{x}\,t} g \nabla_{\vec{x}\,t} g^T \vec{r} > \\ &= \vec{r}^T < \nabla_{\vec{x}\,t} g \nabla_{\vec{x}\,t} g^T > \vec{r} \\ &= \vec{r}^T J \vec{r} \end{aligned} \tag{5.7}$$

With that we can reduce the problem 5.6 to an eigenvalue problem ($J\vec{r} = \lambda\vec{r}$) and we are going to minimize 5.7 with constraint $|r| = 1$ using the Lagrange multiplier; the matrix J is defined as:

$$J = \begin{bmatrix} < g_x g_x > & < g_x g_y > & < g_x g_z > & < g_x g_t > \\ < g_y g_x > & < g_y g_y > & < g_y g_z > & < g_y g_t > \\ < g_z g_x > & < g_z g_y > & < g_z g_z > & < g_z g_t > \\ < g_t g_x > & < g_t g_y > & < g_t g_z > & < g_t g_t > \end{bmatrix} \tag{5.8}$$

Where $< g_p g_q > = \int w(\vec{x} - \vec{x'}) g_p g_q$ and the problem to solve is:

$$J\vec{r} = \lambda\vec{r} \tag{5.9}$$

5.4 The discrete approximation

In the relation 5.8 there are continuous equations: converting them to the discrete approximation is necessary; this can be done introducing the linear operators in a discrete $\mathbb{R}^4$ space[19]:

$$\begin{aligned} g_p = \frac{\partial g}{\partial p} &\Longrightarrow \mathcal{D}_p(g) \\ < g_p g_q >= \int w(\overrightarrow{x} - \overrightarrow{x'}) g_p g_q &\Longrightarrow \mathcal{B}(g_p, g_q) \end{aligned} \tag{5.10}$$

Here we have introduced the operators $\mathcal{D}_p$ and $\mathcal{B}$. An explicit representation of such operators in a $\mathbb{R}^1$ subspace is:

- $\mathcal{D}_x = [-1,\ +1]$ (basic two point derivative)
- $\mathcal{D}_x = [-0.0838,\ -0.3323,\ 0.0,\ +0.3323,\ +0.0838]$ (Scharr derivative kernel [39])
- $\mathcal{B}_x = [0.0234,\ 0.2416,\ 0.4700,\ 0.2416,\ 0.0234]$ (Scharr smoothing filter [†])
- $\mathcal{B} = [0.2,\ 0.2,\ 0.2,\ 0.2,\ 0.2]$ (five point box filter)
- $\mathcal{B} = [0.0625,\ 0.25,\ 0.375,\ 0.25,\ 0.0625]$ binomial filter

The building of higher symmetrical operators ($\mathbb{R}^2$, $\mathbb{R}^3$ and so on) is done in the following way[22]:

$$\mathcal{O}p(\mathbb{R}^1) = [a, b] \tag{5.11}$$

$$\mathcal{O}p(\mathbb{R}^2) = [a, b] * \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a & b \\ b & b^2 \end{bmatrix} \tag{5.12}$$

$$\tag{5.13}$$

This "discretization" process imposes constraints to the minimum volume size and, specially, on the number of time steps (assuming there are enough elements in the physical $\mathbb{R}^3$ volume): this means we cannot apply $\mathcal{B} = [1,\ 1,\ 1,\ 1,\ 1]$ if we don't have at least five time steps (and five pixels along x,y and z direction of course).

5.5 An analytical discussion on the speed limit

As in any problem related to digital processing we have to deal with bandwidth limited systems: every step we take is based on the assumption that all the continuum signals are sampled at some rate. There is a useful theorem called *sampling theorem* who relates bandwidth limited signals and sampling necessary to *reconstruct exactly* the signals:

$$\Delta x_{sampling} = \frac{2\pi}{\Delta k_{max}} \tag{5.14}$$

[†]This should be coupled with Scharr derivative kernel

This is true for a function in the Fourier space meeting the band-limited condition: $f(\mathbf{k}) = 0 \ \forall |\mathbf{k}| > \frac{\Delta k_{max}}{2}$. This theorem can used for a linear propagating signal as:

$$f(x,t) = f(x - vt) \tag{5.15}$$

When this signal is separable as $f(x,t) = g(x) * h(vt)$ the Fourier transform

$$F(k_1,k_2) = (\frac{1}{2\pi})^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)e^{-i(k_1x_1+k_2x_2)}dx_1dx_2 \tag{5.16}$$

then,

$$F(k_x,k_t) = (\frac{1}{2\pi})^2 \int_{-\infty}^{\infty} g(x)e^{-i(k_x x)}dx \int_{-\infty}^{\infty} h(vt)e^{-i(k_t t)}dt \tag{5.17}$$

And the limit $\Delta x_{sampling} = \frac{2\pi}{\Delta k_{max}}$ will be:

$$\Delta x_{sampling} = \frac{2\pi}{\sqrt{(\Delta k_t^{max})^2 + (v\Delta k_x^{max})^2}} \tag{5.18}$$

Normally we will not use directly this equation but its reversed form:

$$v = \sqrt{[(\frac{2\pi}{\Delta x_{sampling}})^2 - (\Delta k_x^{max})^2]\frac{1}{(\Delta k_t^{max})^2}} \tag{5.19}$$

Or its reduced form when we have $\Delta k_x^{max} = \Delta k_t^{max}$:

$$v = \sqrt{[(\frac{2\pi}{\Delta x_{sampling}\Delta k^{max}})^2 - 1]} \tag{5.20}$$

This sets up a limit on the speed given the dimension of the moving object (or better his maximum Fourier component) and can be used to give a rough estimation on the maximum theoretical speed for objects with no band limited spectrum as well.

Chapter 6

The R.L.E. approach to the O.O. graphic

6.1 Introduction

As we stated in Section 4 on page 19 and in 4.4 on page 21 a possible path to the data analysis is the ability to describe the spots ("objects"): to do this it is required a coherent approach to the description of the objects. The ability to describe what an object is*, is crucial for extracting the required parameters from the object itself: here we would show a simple approach to description of arbitrary object shapes through the run length encoding mechanism as well the steps to follow to use them in a very effective way. We will introduce some information about the compression algorithms and we will explain how they are suited in describing the object shape.

6.2 The compression problem

The data encoding is a process to transform an original chunk of data (as images or data streams) into another block of data with reduced size and/or with encrypted propriety: so encoding is a broad term and it stands for data compression and/or encryption. In the rest of this chapter we will deal with the data compression problem, because the encryption is too far from our aim. Algorithms for compression are developed because there is a huge demand for bandwidth-limited/memory-constrained applications on the market; every aspect of our life is connected to the compression researches: from Internet (streaming every kind of images, videos and audio contents), the mp3 players to GSM mobiles, everything is in touch with the compression problem. Typical compression schemes had been developed, all fitted to reduce image storage requirement in the pictures (or in data streams); compression can be classified in two main different families[31]:

1. Lossless
2. Lossy

*Its border, its dimension and so on

The former one retains all the "information" of the original data block and let the ability to recover the original set; the latter is a way to retain just the "important" part of the original information, discarding the less "important" data: this will give no ability to recover the original stream of data. In the lossy algorithms (PNG, JPEG when a quality parameter has been settled) the resulting data block is usually smaller than in Lossless counterpart, but it must be pointed that it requires destructive changes driven by a model of the information being compressed: we have to have a clear idea about this model to be able to distinguish between "important" and "unimportant". In the general proceeding of this work we are required to conserve all the important informations as not to make any assumption for the model: so we can exclude any further investigation into the lossless domain compression schemes.

6.3 Basics of the R.L.E.

We already discarded any lossless algorithm from our investigation as encryption schemes as well; we can shortly list some important compression algorithms:

- Run Length Encoding (R.L.E.)
- Lempel-Ziv-Welch (LZW)
- Huffman (CCITT)
- Discrete Cosine Transform (DCS)

All these are well tailored for their application field [†], but one of these is more important for our work: the run length encoding. We can describe how it works following the passages of encoding a 2 bit (monochromatic or black and white) image (see as reference Figure 6.1 on page 30):

1. make a line scanning of that picture.
2. pick the first point of every black line, and record into an array.
3. record the starting position and length of the "chord" into a table.

The basic characteristics of the run length encoding are two:

1. it is lossy (so it retains all of the informations)
2. it is a compression scheme (to represent in a compact form the original data shape)

Although only the shapes are recorded, the data are uniquely identified from this encoding: this let the freedom to recover the initial informations. At first glance this could be assimilated to some sort of generalized region of interest (R.O.I.) as used in digital image processing domain: we will generalize this to higher dimensional spaces in the next section.

[†] so they choose a tradeoff between performance and compression ratio

6.4 Using run length encoding with multicolor pictures and higher dimensional data sets

We can extend the process of run length encoding shown in the previous section not only to the two level pictures but even to the n-bits gray scale images; it is just necessary to extract from the gray scaled image a two level picture using a threshold: this can be repeated for the three colors of a RGB picture as well extending the same procedure to multicolor pictures (Figure 6.1 on the next page). In that scheme we assumed the pictures are two dimensional but this is not a limit: natural extension can be realized, taking the X axis of the picture (or higher dimensional "picture") along the fast moving coordinate. So run length encoding respects two fundamental requirement of our research:

1. records the shape of any object (generality)
2. independence from the space dimension

Finally we hope to have found a suitable way to describe an object: in the next section we will show the intimate connections between an object description and measurement of its parameters.

6.5 The transition to the object oriented data handling

As we have seen so far, we are able to reduce (squeeze) our original picture (or volume) in a smaller series of "chords" (the reed lines in 6.1(c) on the following page): this requires just storing the starting points and the lengths of these "chords". The next step is grouping these "chords" into coherent groups forming all the objects present into the original data set: this is a way to record informations about the objects. If you recall the steps involved in creating the run length encoding version of the original data set (6.3 on the page before), you can see we omitted one point: the segmentation part (actually the passage between Figure 6.1(a) and Figure 6.1(b) on the following page). This step is important: if we are able to use some unique parameter[‡] we can then associate one object just to a sequential unique index and to a threshold value. Here is the sequence involved in handling in an object oriented way our data analysis (see as reference Figures 6.1 on the next page and 7.5 on page 37):

1. Use a threshold value to create a black and white picture
2. Create a run length encoded representation of the data
3. Group all the chords in coherent groups
4. Assign to every group an unique index (the object index)

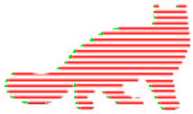
To extract the required object (and all the related informations about it) are just necessary two parameters: the index number and the thresholding value. So from our two [§] dimensional picture we are able to extract one object using just two (discrete, and limited in range) variables: this is just an example of dimensional squeezing so far introduced.

[‡] like thresholding parameter value

[§] or three as pointed in Section 4.2 on page 19

(a) Original picture

(b) Thresholded image



(c) Run length encoding

(d) Encoded shape superimposed to the original image

Figure 6.1: Run length encoding

Chapter 7

The segmentation and visualization

7.1 Introduction

In the following sections we will illustrate a segmentation algorithm, the random Poisson noise as found in experimental section and the graphics tools: all these constitute the machinery "under the hood".

7.2 The global threshold

Segmentation algorithms have as final goal object detection but what we mean for an object is a relative question: we can see from images acquire with the same scanning device (a confocal scanning microscope) different "spots". For example, in Figure 7.1(a) on page 34 it is clear we identify "spots" as the small brighter points lying inside the cell nucleus (the cell nucleus border is not visible) *; in Figure 7.1(c) on page 34 we can see the cell nucleus bodies of three cells: here our "spots" are the nuclei theirself; finally in the third Figure 7.1(b) on page 34 we have the simple form of spot and we have no trouble here because we know we are dealing with bead like objects. So "object" (or "spot" from now) is a concept related both to the scale of the problem and to the kind of object to study: this is not a trivial question because many different approach should be used to detect what we mean for spot. To make more complex the general landscape of "spot" recording, there are *domain related* requirement: in the analysis of spot movements, it is necessary just to track the centroids, but for other purposes (as "spot growing analysis") could be better to record the spot shapes for further statistical analysis †. Without prior knowledge of the domain problem it is hard to decide the full path to the data analysis: looking to the Figures in 7.1 on page 34 we have to manage very different objects. Although many different segmentation algorithms exist (for a comparative overview [9]) we chose to stay more general as possible using a global thresh-

*These spots are not rounded objects and, generally, they could have different shapes.

†This will be one of the roles of run length encoding mechanism as seen in Section 6 on page 27

olding method[37][32]. In this simple segmentation approach, we generate an histogram plot of the intensities of our data set; this histogram will show (see Figure 7.2 on page 34) two region: one bigger containing the background noise and the smaller one containing the distribution from the objects ("spots"). This could lead to errors if this threshold happens to be not unique in different regions of the volume (as in case of large variations in intensity due to the not uniform illumination/response of the system and/or really different classes of spot intensities[8][36], see Figure 7.4 on page 36): so it is mandatory requiring an user feedback to adapt manually the thresholding value (when there is not a clear model for the analyzed object spot[4][5], of course). This user-guided thresholding or "hybrid"[‡] is quite effective in resolving the problems related to the correct segmentation: in a comparative study[9] on phantom tests results that the difference between algorithm quality is in the percent range anyway[§].

7.2.1 Thresholding with run length encoding

Although the segmentation with a global threshold is usually computed with connectivity rules[22][32][37], we want to show that with run length encoded is possible to reach the same goal: it is not an high performance procedure but with the current computing resources it is nearly real time process. The steps involved are (see Figure 7.5 on page 37):

1. create a chorded image (as discussed in 6 on page 27)
2. assign to every chord an unique initial identity number (id)
3. join chords when they are touching and reassign them the same id

This is just here to demonstrate how far can be pushed the run length encoding in domains not immediately related each other.

7.3 The noise problem

As we noted in 7.2 the segmentation problem is strictly related to the noise analysis: without prior knowledge about the noise is not possible to analyze the dataset, because we are required to separate sharply between this and signals. Many physics related events have noise that is related to the underlying propriety of the scanning system and/or of the scanned target; it is strictly tied to the experiment and can not be set a general rule to model it: electronic f-noise, Poisson noise, shot noise and many other are just examples of the many different forms. Noise reduction (a primary task in any analysis) could be obtained in many different ways and/or at different levels: shot noise could be eliminated easily using simple point analysis or thermal noise as found in CCD systems can be removed simply cooling the device itself. In confocal scanning system, as we are dealing with, we can analyze the source of the noise by intensity histogram inspection that shows immediately the source type for our noise: it is a Poisson noise. as in $e^{-\lambda}\lambda^k/k!$ (being λ the expected mean value). In Figure 7.4 on page 35 is shown a confocal scanning section of an HeLA cell and the intensity histogram plot; superimposed there is a fit with the Poisson distribution

[‡] In literature it is called supervised

[§] And here the precision of the method is connected to a life/death problem

($POISSON(k, \lambda, A) = Ae^{-\lambda}\frac{\lambda^k}{k!}$) with three different parameters: segmenting the image data could be done just setting a threshold value to eliminate the Poisson noise.

7.4 The visualization of 4D datasets.

As we pointed in 7.2 on page 31 we are going to introduce user interaction in the process of segmentation and spot tracking: this requires a flexible program to visualize the data. Basically this program is acting as an interface between the data segmentation program and the results, as shown in Figure 7.6 on page 37; the main procedure followed is:

1. segment the data with a fixed starting threshold
2. extract the unique index number [¶], barycenters and sizes of the spot
3. display a data slice at the time step (user selectable)
4. superimpose both the image and the barycenters with sizes
5. under user supervision change the threshold value or select a spot circle
6. repeat for all the spot

For every spot marked there will result an unique index and a thresholding value as selected from the user: this two parameters will suffice to extract all the spot informations as described in Section 6.5 on page 29. Two important facts have to be pointed:

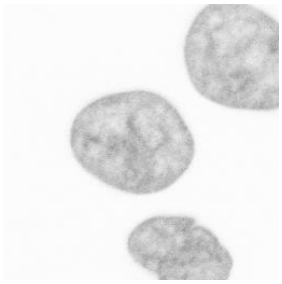
- user click on the border of the spot but the data are from a repeatable segmentation algorithm
- the user knowledge of the problem is preserved

The first point is the scientific minimum requirement and the second one is a trial to not trust blindly on a computer driven program that can fail for condition far from ideal (as we have in biological samples): this is more important when the data requires "one-shoot" analysis.

[¶] As in Section 6.5 on page 29

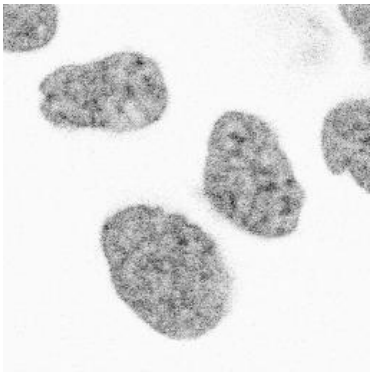
(a) HeLA cell single nucleus (Cy3 channel)

(b) 500 nm test beads (Cy3 channel)

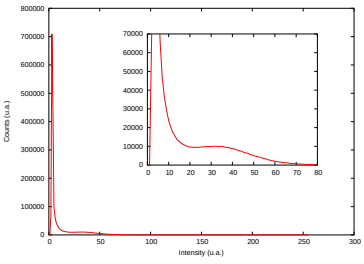


(c) HeLa cell nuclei (GFP channel)

Figure 7.1: Different "objects"

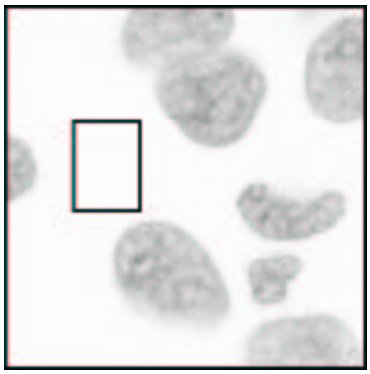


(a) HeLa cell nuclei (GFP channel)

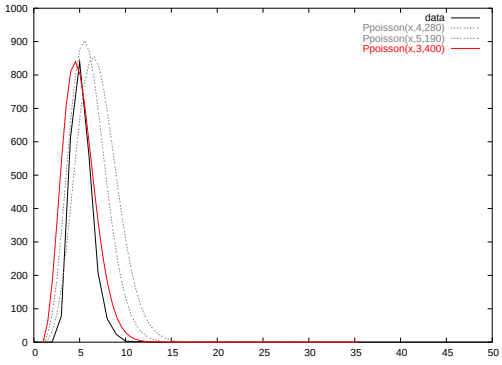


(b) Hystogram Intensity distribution

Figure 7.2: Different "objects" histogram distributions

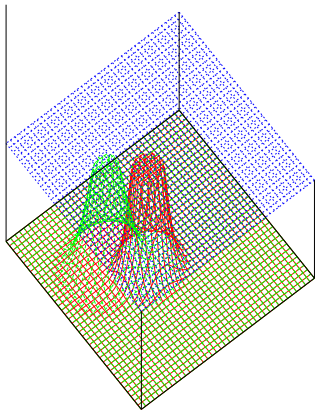


(a) HeLa cell nuclei (GFP channel)

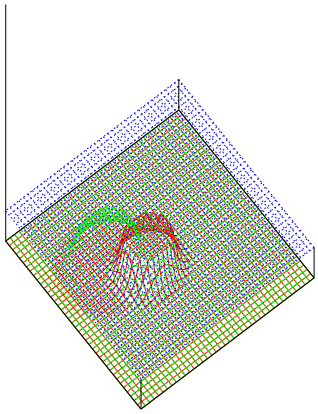


(b) Histogram distribution of noise (poisson fit: 3,4 and 5 λ)

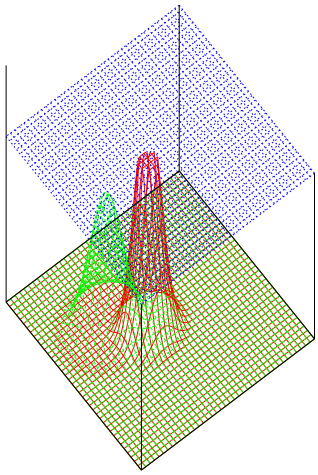
Figure 7.3: Poisson noise in confocal scanning system



(a) Two spots correctly detected

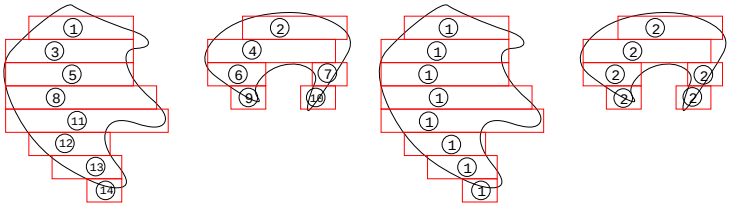


(b) Threshold level set too low, here the spot detected is only one big spot.



(c) Threshold level set too high, here there is only one.

Figure 7.4: The global threshold



(a) Every chord (represented by a red block) has its own index

(b) The chord have been reassigned, when touching, to the same index: finally we have an object no. 1 and an object no. 2

Figure 7.5: The global threshold using the run length encoding

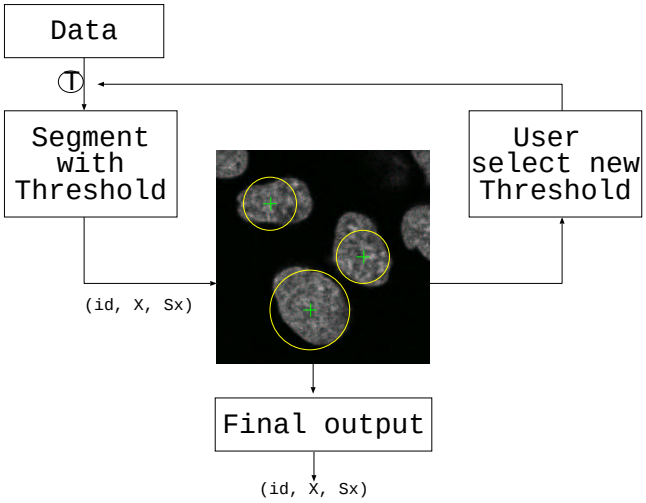


Figure 7.6: Work flow for the four dimensional program

Part III

Results

Chapter 8

Vector field limits in sintetic samples and simulations

8.1 Goal

In this section we would demonstrate some application of the Volume Field method on some simple testing objects: we would also show the limits of the method, and creating the bases for real applications. As shown in section 5.5 on page 25 there is an upper limit to the detectable object in movement: in our on going analysis we used as testing object a Gaussian spot object because its Fourier transform is easy to compute by hand (basically it is again a Gaussian spot in the frequency space):

$$f(x) = Ge^{-\frac{x^2}{2\sigma^2}} \Longrightarrow G'e^{-\frac{k^2}{2\frac{1}{\sigma^2}}} \tag{8.1}$$

Then from Equation 5.20 on page 26 we derived a limit when we assumed as upper limit for our signal the standard deviation $2\Delta k^{max} = 1/\sigma$ * being σ the distribution of our spot along the x direction: then for a $\sigma = 3$ we obtain $v = \sqrt{(\sigma\pi)^2 - 1} \sim 2.9$, that is a rough estimation for testing purposes. To numerically estimate the divergence between the computed speed and the "real" speed of the moving objects we introduced an aperture angle as in test Figure 9.9 on page 54: we ploted that angles distribution.

8.2 Materials and methods

8.2.1 One Gaussian spot

To test how stable is the vector field against different speeds, we generated a Gaussian spot:

$$Ae^{-\frac{x^2}{2\sigma^2}} \tag{8.2}$$

With parameters: $A = 127$ and $\sigma = 3$, all given in pixels. The spot was generated in a volume of $128x20x20$ voxels in size, being the X and Y axes shown in Figure 8.1 and the Z axis (20 pixel actually) the deep axis not shown (all the pictures report just a section of the full data). The spot was moved along the main axis [X] at different speeds: in the first column of Figure 8.1, it has been displayed one section in the middle of the volume (Z slice no. 10) for three different speeds (1, 4 and 6 pixels/frame respectively). The vector field was computed for the three dimensional volumes (plus time, of course) and the vector field was then extracted a the same central Z level: the results are shown on the right column in Figure 8.1.

8.2.2 Two Gaussian spots

We tested the behavior of the algorithm against two colliding spots: in this case we checked the worst case in which the two spots are collapsing into one bigger. Again we generated two spots with the same parameters as in previous section ($A = 127$ and $\sigma = 3$ and in a space of $128x20x20$ voxels): obviously we started in a condition where the two spots are far apart, so no crossing effects were present until the collision happened. With these conditions we were running

*In the sampling theorem we assumed that the signal should be band-limited, we relax a bit this condition.

the simulations at three different speeds (as in previous section) of 1, 4 and 6 pixels/frame respectively (left column in Figure 8.2) The volume field method was computed first with a single processor machine then with a self assembled cluster of in-expensive machines: the result is reported in Figure 8.4 on page 45

8.3 Results

8.3.1 One spot

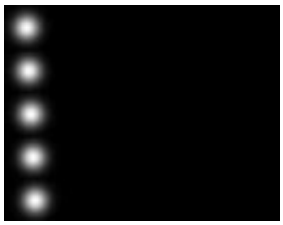
As we stated in 8.1 on the preceding page there is a limit to the sampling of any signal: we stressed this limit in the trial to see if is possible to analyze a spot moving. We can see from the pictures in 8.1 on the next page that the vector field method started to break at speed nearly of 4 pixels per frame; however there were different reasons to this, clearly not related to the sampling theorem. Our estimation of 2.9 pixel per frame in the displacement was an order of magnitude limit, but was an interesting parameter estimation based on a minimum knowledge of the problem (it stems from just the sampling theorem and it is generally true for any dynamical process). We computed the angle distribution between the computed vector field and the real simulated speeds: the plots are in Figure 8.3 on page 44 with fitted Gaussian distributions. From these results was clear that breaking behavior was random, and it summed up to the real motion of a detectable object: this is more important because when we analyze real random spots.

8.3.2 Two spots

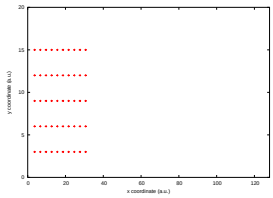
In the pictures in 8.2 on page 43 we did not to set a limit on a measurable parameter as in previous section, but we showed the behavior of the algorithm under a limit case when two "spots" were colliding: this was, at least, a quite un-physical case because two bodies normally could not join at the same time in the same place. However this was interesting case study when we were dealing with objects with no well defined body borders.

8.4 Discussion

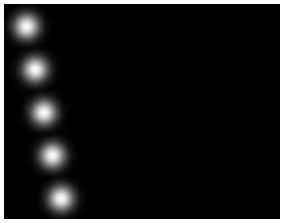
In a trial to estimate the vector field limit, we tested against some simple test objects; the results we have here shown that an upper limit to the vector field could be set for one spot and two spots as well: we derived these results just as an extension to the sampling theorem, and in real simple test cases, just to clearly show the underlying problems before we apply this method to more complex cases.



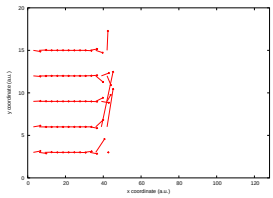
(a) 1 pixel / frame



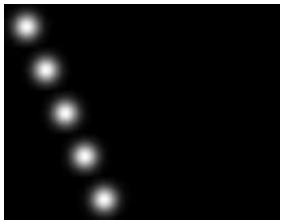
(b) Vector field



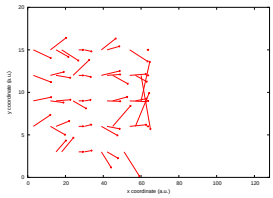
(c) 4 pixel / frame



(d) Vector field



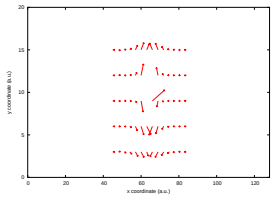
(e) 6 pixel / frame



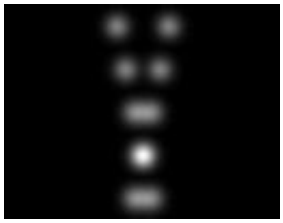
(f) Vector field

Figure 8.1: Vector field of one Gaussian spot at different speeds

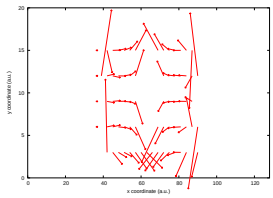
(a) 1 pixel / frame



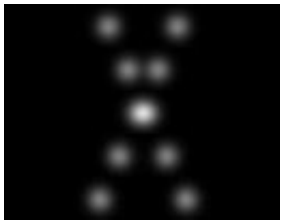
(b) Vector field



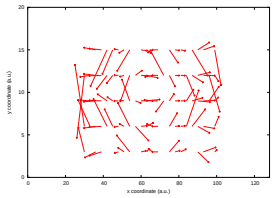
(c) 4 pixel / frame



(d) Vector field

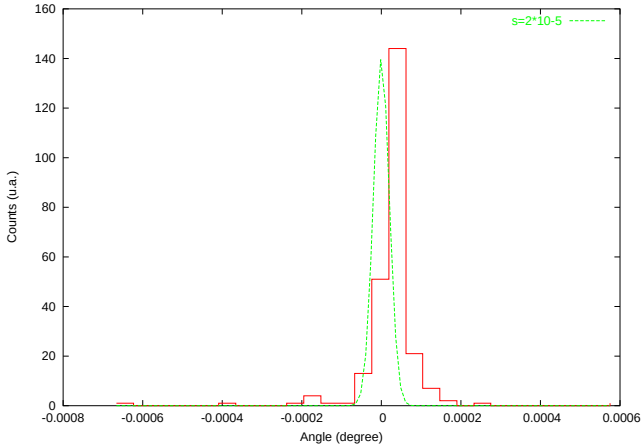


(e) 6 pixel / frame

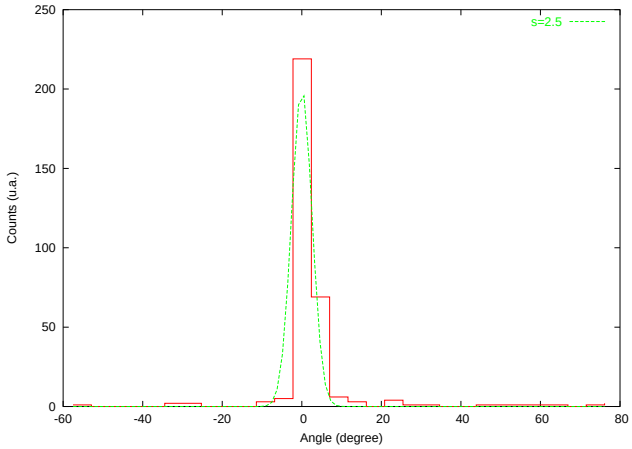


(f) Vector field

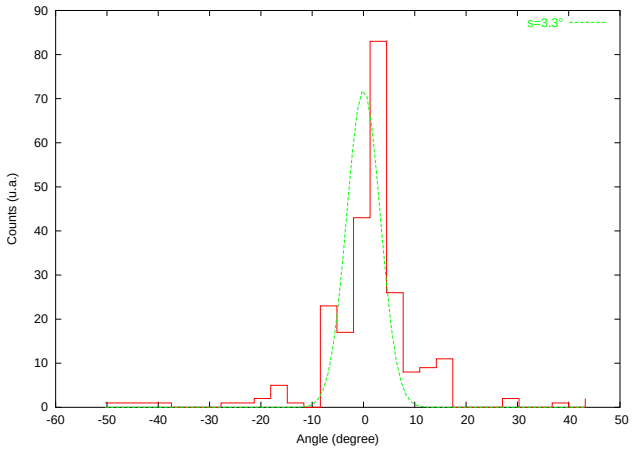
Figure 8.2: Vector field of two Gaussian spots colliding at different speeds



(a) Angle distribution at 1 pixel / frame



(b) Angle distribution at 4 pixel / frame



(c) Angle distribution at 6 pixel / frame

Figure 8.3: Vector field of two Gaussian spots colliding at different speeds: comparison distributions between computed and simulated results.

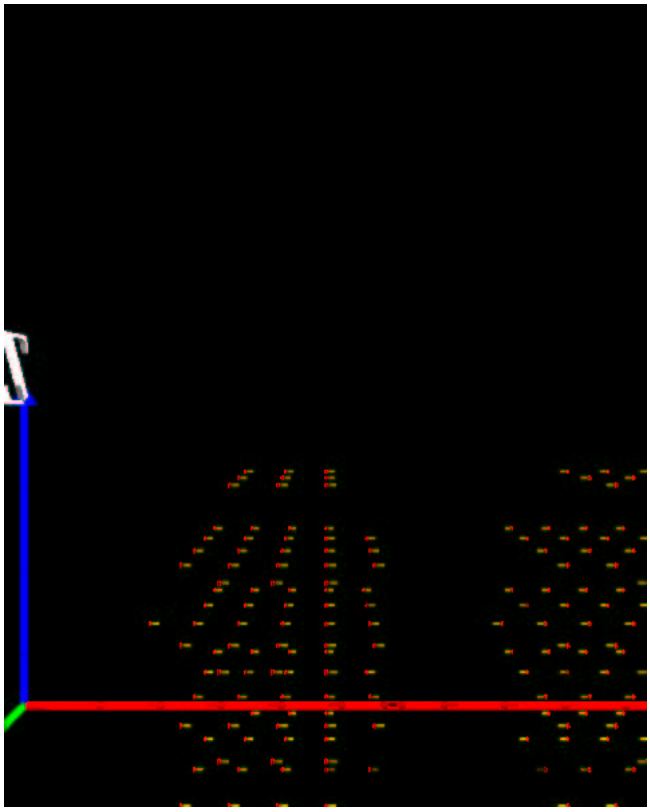


Figure 8.4: Parallel simulation of two Gaussian spots colliding

Chapter 9

Vector field application to simulated chromosome territories

9.1 Goal

As we have seen in 1.5 on page 5 there are theoretical prediction about the phase transition of chromosomes (*interphase* $\Rightarrow$ *metaphase*): this given model has been highly regarded as solution to the unfolding problem. Here is presented a method to link theoretical results from computer simulated chromosomes territories and vector field analysis; it is not intended as complete solution, lacking of any experimental evidence (the required time scanning periods down to the second are far below any conceivable experiment): it is just a trial to open a path for further analysis as the scanning methods and computing resources will be available in the near future.

9.2 Materials and methods

A chromosome marked region was simulated (Contribute from Dr. G. Kreth)using the Spherical 1-Mbp Chromatin Domain (SCD) as in 1.5 on page 5: a random Brownian motion has been used to describe the movement of the chained spheres. These simulated spheres were used to create data volumes with unit sizes, then they were "dumped" in a voxel world of $64x64x64$ units: these volumes represented the cell nucleus during his evolution from the metaphase to the interphase. The time step between two sequential frames was 300 ns and the total sequence was nearly two seconds long. We followed the development of an initial structure (simulating the chromosome in his condensed form) to an unfolded structure representing the final shape of the chromosome in the nucleus. The data were analyzed with a vector field approach and the computed speeds (from the simulated data) were compared with the detected ones: then an histogram containing the angle between these was created and a fit was made to test the results.

9.2.1 One spot

From the total dataset of the simulated region we extracted just one spot (corresponding to one single part of the chromosome being marked), to compare against the simulations made on the section 8.2.1 on page 40: in that case we already had an upper estimate on the maximal displacement of a Gaussian spot (actually in the simulations there were constant density balls). The extracted spot was just a test of a single chromosome unfolding: in the Figure 9.1 on page 50 is shown the 3d plot of the trajectory of the spot in the space and in Figure 9.2 on page 50 is shown instead the "round problem". In effect when we were dealing with simulations we assumed the world being continuous: this was not true if we had systems limited from physical rules* or, as in our case, by the sampling when we dumped the data into a "voxelized" volume. At least this "voxelization" resulted in a different path from the original one (red and green path in Figure 9.2 on page 50): this introduced un-expected changing in the speed that reduced the upper limit as we stated in 8.2.1 on page 40.

*For example, in scanning procedures with a camera there is the upper limit to the resolution given by the pixel size of the CCD scanning device, from the noise and from any other physical effect present.

9.2.2 Multiple spots

Here we used all the simulated data set: this was the most complex case we were dealing with, because both overlapping and speed problems were present (see Figure 9.5 on page 52). Here we collected the informations about the angle between the simulated data speeds and the detected vector field speeds: again was plotted the angle aperture distribution. To extend the analysis was introduced a correlation between speeds at different time steps to try to account for a "long range" correlation (see Figure 9.8 on page 54): so speed detected was compared against the speed of the simulation computed as difference between not only the next step but also between the second, the third and so on. In that case if we had any trend in movement should had been possible to detect it from peaks into the aperture histograms.

9.3 Results

9.3.1 One spot

A plot with the computed vector field speed and the simulated speed was made in Figure 9.4 on page 51: there was a good agreement between them and the analytical proof was the histogram in Figure 9.6 on page 52. It was computed a fit to this histogram and the aperture angle of 17^o was extracted: it was quite larger than we had in Section 8.1 on page 40 (3^o). The difference between the expected value and the obtained one could be accounted by the path rounding problem: anyway a good agreement was found between the real simulated speeds and the vector filed computed ones, making possible (with some care) such a kind of one spot analysis when faster scanning devices will be available.

9.3.2 Multiple spots

We had greater problems with multiple marked sites: we stressed here the same problems we had with one but multiplied by the many object moving in the simulated volume. By looking to the Figure 9.8 on page 54 it was immediately clear that the situation was less than good: the same agreement between the simulated spot speeds and the vector filed detected ones would not be possible. Anyway the same histogram containing the angle spread distribution was computed and the result is shown in Figure 9.6 on page 52: this was not showing any kind of relation, being the distribution running uniformly on all the range of possible angles. To test against the long range correlation, the same histogram was computed using different steps as shown in Figure 9.8 on page 54: again no correlation was found.

9.4 Discussion

We should conclude from that analysis that was not possible to overcome the limits of applicability of the vector field methods in case we were dealing with more than one spot. For the single spot limit, instead, there is some possibility to get results but with a wide error (for the angle direction this was inside 17^o):

so we had to conclude that, based on our experience, applicability to the biology related problems is quite limited. We can summarize the results in such a way:

1. volume field methods could be applied to digitally generated data
2. simulations for multi body objects are not suited to be analyzed with volume field methods

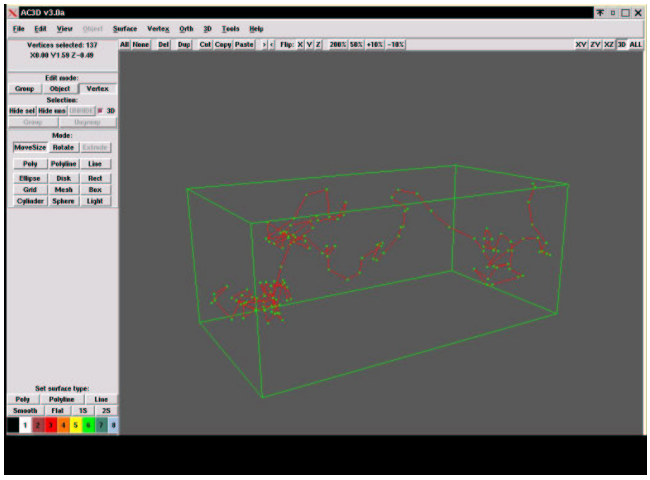


Figure 9.1: A three dimensional representation of one Brownian spot

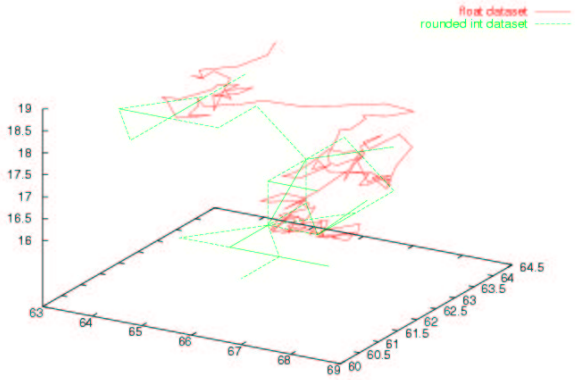


Figure 9.2: The rounding problem with one spot in Brownian movement

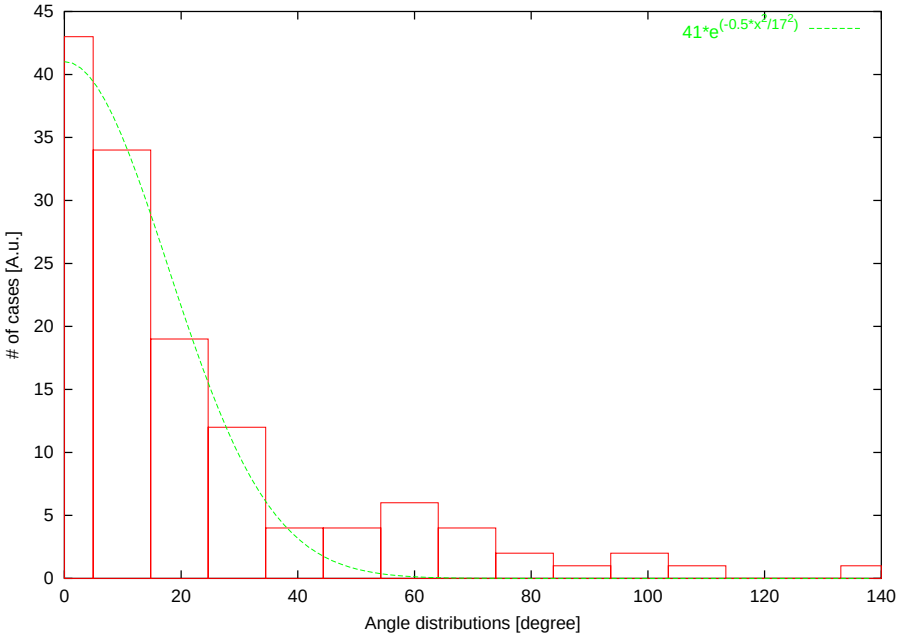


Figure 9.3: Angle dispersion between the real and the computed speeds

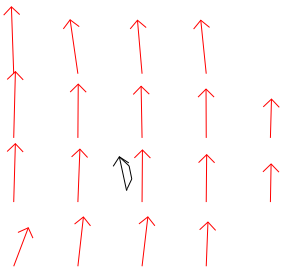


Figure 9.4: Visual comparison between vector field (red) and the real spot movements (black)

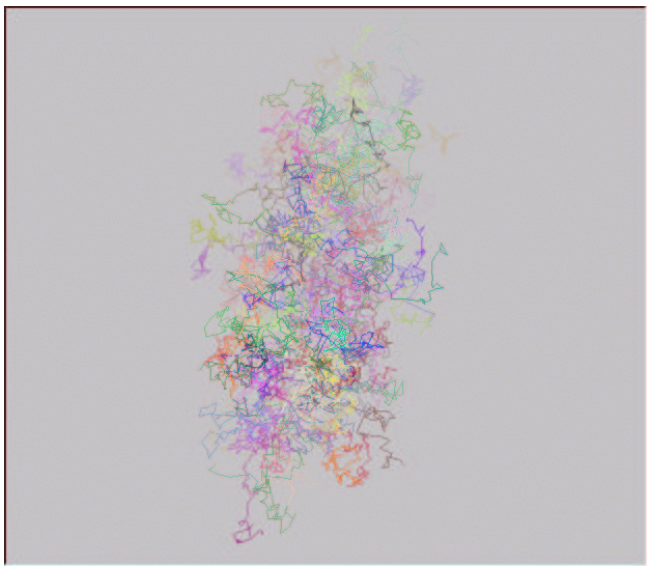


Figure 9.5: A three dimensional representation of multiple Brownian spots

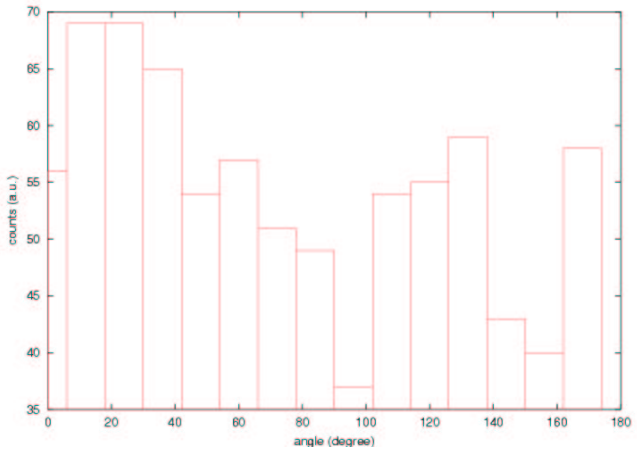


Figure 9.6: Angle dispersion between the real and the computed speeds

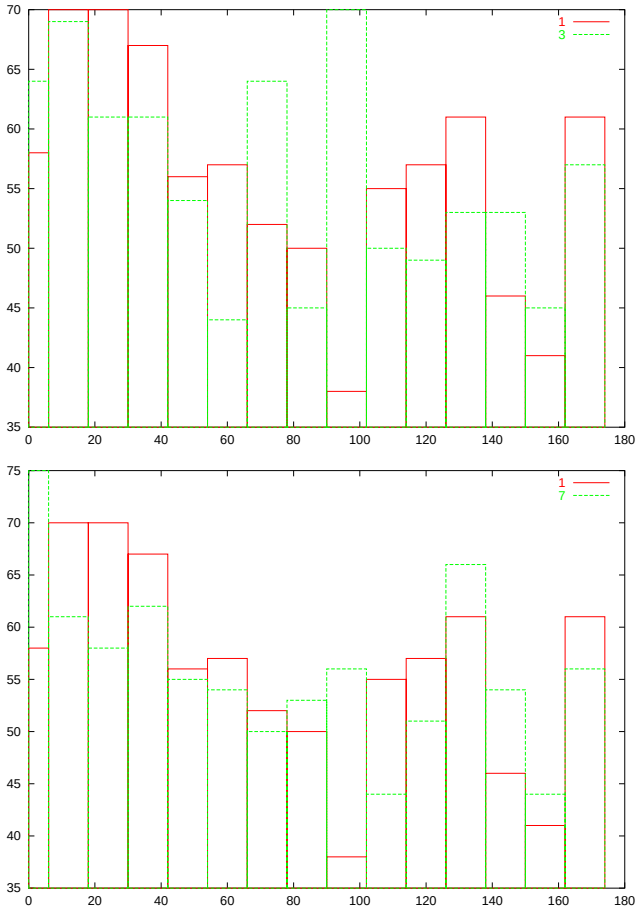


Figure 9.7: Angle dispersion comparison between long range correlations: the topmost figure shows the angle aperture between the detected vector field speed and the speed from simulation computed as difference between the first and the fourth step (topmost figure) and the difference between the first and the eighth step (bottom figure).

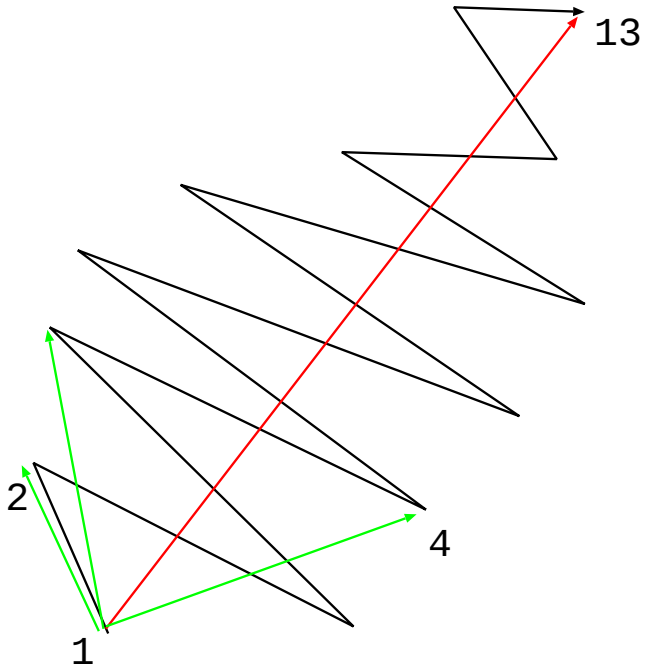


Figure 9.8: The angle between the 1th and the 2th is generally greater than the 1th and the 13th in presence of a directional movement: taking the speed as average movement between two steps far apart will show such a directionality.

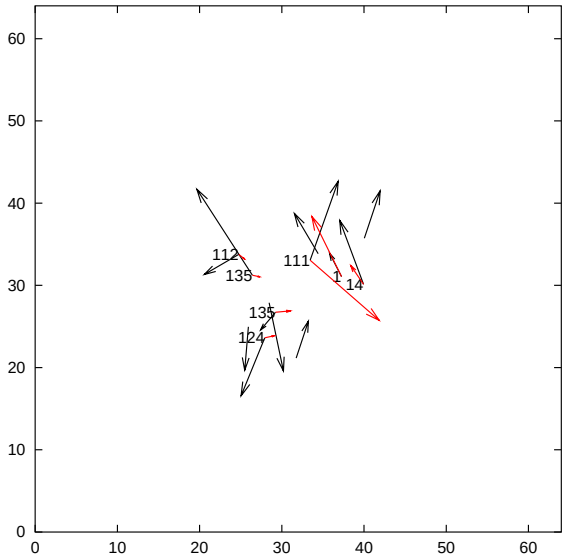


Figure 9.9: Speed comparison between computed vector field and (black) and simulated data (red): central plane in a 64x64x64 voxels world.

Chapter 10

The first biological experiments

10.1 HeLa cell (short) history

Henrietta Lacks was a black woman died in Baltimore in 1951 of cervical cancer at Johns Hopkins Hospital [23], [46]: her cells (initially taken from a cervical biopsy) were discovered having special reproductive property from Dr. George Gey (actually they can survive and reproduce with the available technology of the 50'th) and now they are one of the most important cell line in the biological world.

10.2 Materials and methods

10.2.1 The biological specimens

HeLa cells and SH-EP N14 neuroblastoma cells were in vivo replication labeled with Cy3-dUTP and then followed through several cell cycles to obtain segregated chromosome territories [48], [40]. Chromosome territories were represented by sub-chromosomal domains with a characteristic shape of bright fluorescence spots. The respective cells were first analyzed in vivo using a live cell observation chamber, then the cells were fixed with 4% paraformaldehyde, and finally a FISH procedure has been applied [12], [41]. In each stage (living, fixed, and hybridized) a confocal image stack of the same nucleus was taken (see below and Figure 10.2 on page 61).

10.2.2 Scanning procedure

The registration of the 3D-images of the fluorescent intensity distribution of the "spots" in the cell nuclei in the living, the fixed and the mock-FISH state was performed using a confocal scanning microscope Zeiss (equipped with 63 x 1.4 N.A. objective lens). For each cell nucleus, about 20 optical sections were registered resulting in a 3D volume scan with a voxel size of x=y=50nm in the object plane and z=250nm in the direction of the optical axis. The data were saved from the system on files (512 x 512 pixels wide in 8 bit gray-scale). For each cell nucleus three different volume files* were obtained: one for the nucleus in living state; one for the fixed state and one for the FISH state.

10.2.3 Correction of the natural shift and rotations of the target object

Imaging of a volume was done by an optical sectioning (as described above): during the registration the cell (in case of living ones) and the whole system can shift and rotate from the previous acquisition position. Although this movement is small, the positions will be strongly biased. To solve this problem, a realignment program based on cross-correlation algorithm was developed: firstly a reference section ("slice") inside the volume was fixed: then a subsequent shift/rotation of the rest of the slices was performed. Finally a correlation † was computed for each shift (or rotation). When an optimal maximum for the correlation parameter was reached, the slices were considered aligned.

*$512 \cdot 512 \cdot 20\ voxel = 25.6\mu m \cdot 25.6\mu m \cdot 5\mu m$

†$Correlation(k, k') = \sum_{x,y} I_k(x, y) * I_{k'}(x, y)$

10.2.4 The segmentation algorithm and detection procedure

In the analysis of the nuclear 3D-images like those presented here, correct segmentation poses a fundamental problem, due to the high level of noise and the different shapes of the fluorescent intranuclear objects ("spots")[5]. The most simple segmentation algorithm (as in Section 7.2 on page 31) is based on a specially implemented $N-$dimensional global threshold: all the points are considered inside a "spot" when their intensity is above some fixed threshold (using the same value for the entire image volume).

10.2.5 The computer visualization system

The basic work-flow of the program [35],[3] used to detect the particle position from the data volumes is shown in Figure 10.1 on page 60. The visualization part is shown in Figure 10.3 on page 61: the three volumes are superimposed with three different colors (red for the living state; green for the fixed, and blue for the FISH state). The user can use the handlers on the border of the main window to change the slice displayed inside the volumes; he can modify the contrast (gamma curve) to enhance the visualization; he can set the thresholding value to pass to the segmentation routine: all is done in an interactive way to facilitate navigation through the volumes. The segmentation algorithm(running as a separate process inside the main program[‡]) generates, for each spot detected, the position vector of the barycentre (fluorescence intensity gravity centers) and the standard deviation of distribution intensity. From these data a scaled representation is displayed on the screen taking into account the current slice (Figure 10.4 on page 62): the user can click in the nearest position, and a cross will be displayed at the center of the spot indicating the calculated coordinates of the position of the spot barycentre. In any moment, the thresholding value for the segmentation algorithm can be changed to locate with better precision the barycentre of the spot.

10.3 Results

10.3.1 The computed distances

The output of the program was a data table of the interactively marked spots to feed the computing programs (mostly written in python with numerical python extensions [25], [15]). With these points, the 3D distances between "spots" in a nucleus were determined. The same "spots" was selected in the image volumes of the same nuclei in the living, the fixed and the FISH state. This was done interactively using the criterium of distribution similarity. Only clearly identified spots were used for further evaluation (see Figure 10.6 on page 64).

10.3.2 Changes in spot positioning

The lines $a_1 - a_2$ and $a_2 - a_3$ in Figure 10.6 on page 64 refer to the displacement of the position of one "spot" from his original position a_1 in the living cell state

[‡]This allows to change the thresholding algorithm in a easy way and permits to choose the best one depending on the data set.

(state 1) to the new position a_2 when the cell is "fixed" (state 2) and then "FISH" hybridized a_3 (state 3), after the alignment of the cellular images. In Figure 10.6 on page 64 this displacement is shown as distances $d_{l-f} = \overrightarrow{a_1\, a_2}$, $d_{f-h} = \overrightarrow{a_2\, a_3}$ and $d_{l-h} = \overrightarrow{a_1\, a_3}$: all the distances were computed in 3D. The total number of the nuclei that have been analyzed was 8; nuclei with incomplete data sets were excluded; also the data sets which could not be processed, e.g. due to high noise level, were excluded resulting in a total of 5 nuclei out of 8. A total of about 300 "spots" in 5 nuclei (with a minimum of 42 to a maximum of 114 "spots" pro nucleus) was marked resulting in number of about 140 distances for all the nuclei: it is worth note that these distance differences are comparable to the diagonal of the voxel (circa $260nm$) meaning that are close to single voxel movement. From the data points, Table 10.1 was calculated: the mean values for distances with standard deviation[§], are listed for each nucleus and for each status.

		[l-f] (nm)	[l-h] (nm)	[h-f] (nm)
Nucleus 2	average	500±180	530±230	390±230
Nucleus 3	average	360± 170	620±360	470±340
Nucleus 5	average	560±350	390±120	650±390
Nucleus 6	average	330±190	410±250	400±190
Nucleus 8	average	400±270	440±250	430±180

Table 10.1: Table of 3D distances of the marked spot from the position "in vivo" [l] to fixed [f] and hybridized states [h]: the distances $[l-f] = \|a_1 - a_2\|$, $[f-h] = \|a_2 - a_3\|$ and $[l-h] = \|a_1 - a_3\|$ were computed.

10.3.3 Change in nuclear organization

The second measurement was for intranuclear long distance measurement: we looked for changes in the 3D distances between "zones" (eg. marked as c1,2,3-b1,2,3 and so on in Figure 10.6 on page 64) in the three different states (living, fixed and FISH). For this purpose the same "spot" data as above were used. In one nucleus in the living state [l] we computed the 3D distances between couple of spots at opposite sides ($Distance[l] = \|\overrightarrow{b_1} - \overrightarrow{c_1}\|$ in Figure 10.6 on page 64): then we computed the same distance in the fixed [f] ($Distance[f] = \|\overrightarrow{b_2} - \overrightarrow{c_2}\|$) and in the FISH [h] state ($Distance[h] = \|\overrightarrow{b_3} - \overrightarrow{c_3}\|$) using the same spots. From these distances we computed three values: $\Delta_{lf} = Distance[l] - Distance[f], \Delta_{fh} = Distance[f] - Distance[h], \Delta_{lh} = Distance[l] - Distance[h]$. To increase the number of data in the set we used permutations between groups, using not only distances between A-B but also between A-C, A-D and so on. A plot of $\Delta_{lf,fh,lh}$ against $Distance[l]$ was made (Figure 10.5 on page 63): then this procedure was repeated for the same nucleus in fixed [f] and the FISH state [h]. A linear fit (with the least-squares method) was done to estimate the slope in these data sets ¶ Table 10.2 on the next page contains the results of the fitted data.

[§] $\sigma = \sqrt{\frac{N}{N-1}(\frac{\sum x^2}{N} - \overline{x}^2)}$

¶ We assume there is a relation $\Delta_{l-f} = A \cdot Distance[l], \Delta_{f-h} = B \cdot Distance[l], \Delta_{l-h} = C \cdot Distance[l]$

Nucleus 2	
A	-0.0593 ± 0.0016
B	0.0208 ± 0.0015
C	-0.0373 ± 0.0021
Nucleus 3	
A	-0.0441 ± 0.0016
B	-0.0007 ± 0.0015
C	-0.0445 ± 0.0020
Nucleus 5	
A	-0.0456 ± 0.0048
B	0.0465 ± 0.0029
C	0.0009 ± 0.0040
Nucleus 6	
A	-0.0321 ± 0.0014
B	0.0325 ± 0.0011
C	-0.0004 ± 0.0015
Nucleus 8	
A	0.189 ± 0.021
B	-0.0254 ± 0.0016
C	-0.182 ± 0.021

Table 10.2: Slopes for 3D large distance displacements of "zones" inside the same nucleus in the same state (Figure 10.6 on page 64 and Chapter 10.3.3 on the page before): A is the slope for Δ_{lf} vs $Distance[l]$; B slope for Δ_{fh} vs $Distance[l]$; C is for Δ_{lh} vs $Distance[l]$. We assume a linear relation between abscissa and ordinate: $\Delta_l = A \cdot Distance[l-f]$, $\Delta_f = B \cdot Distance[f-h]$, $\Delta_h = C \cdot Distance[l-h]$.

10.4 Discussion

Here we presented a quantitative analysis of the variability in 3D-distances between the same labeled chromatin sites [spots] in the living [l] ("in vivo") ; in the fixed [f] ; and in the mock-FISH [h] state. Although the dependence of the distance changes between these 3 states, as estimated by the slope (Table 10.2), is relatively small, at any large range distance, the calculated difference in distance changes between the [l-f] state, the [f-h] state and the [l-h] state are usually in the order of several hundred of nanometers and sometimes larger (Figure 10.5 on page 63). A final conclusion could be summarized only with help from other experimental evidences[11]: the reliable conclusions about the relative positions of CTs down to the level of 1Mb-chromatine domain can be reliabily obtained by 3D-FISH. Attempts to study gene positions visualized by 3D-FISH at nm levels run a great risk that artefacts of altered chromatin configurations are measured, wich provide littel, if any, information on chromatin architecture at the nanometer scale in the nucleus of a living cell.

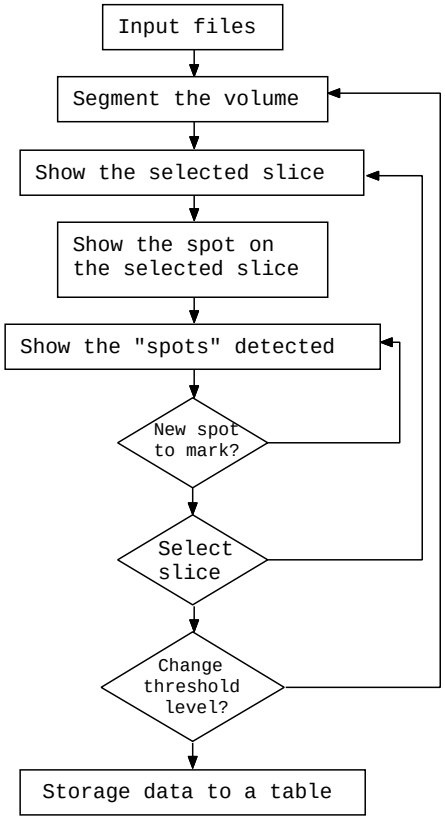


Figure 10.1: Basic work-flow of segmentation program

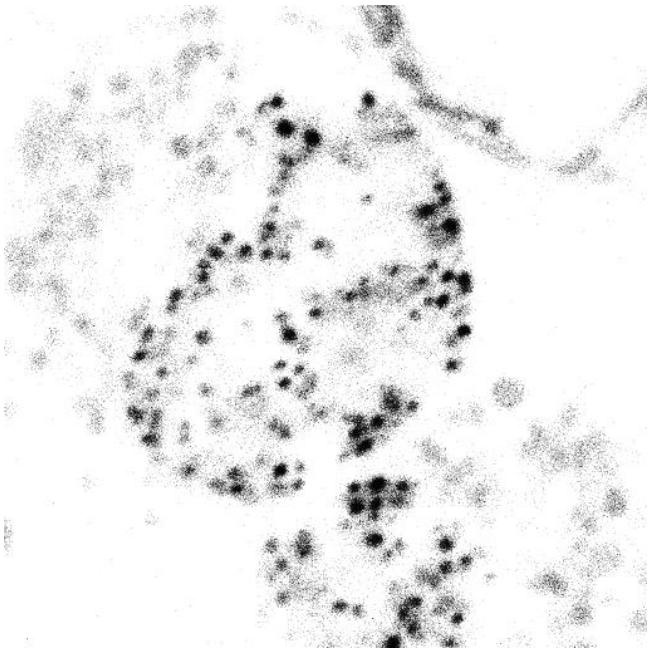


Figure 10.2: HeLa cell nucleus in living state (data preprocessed)

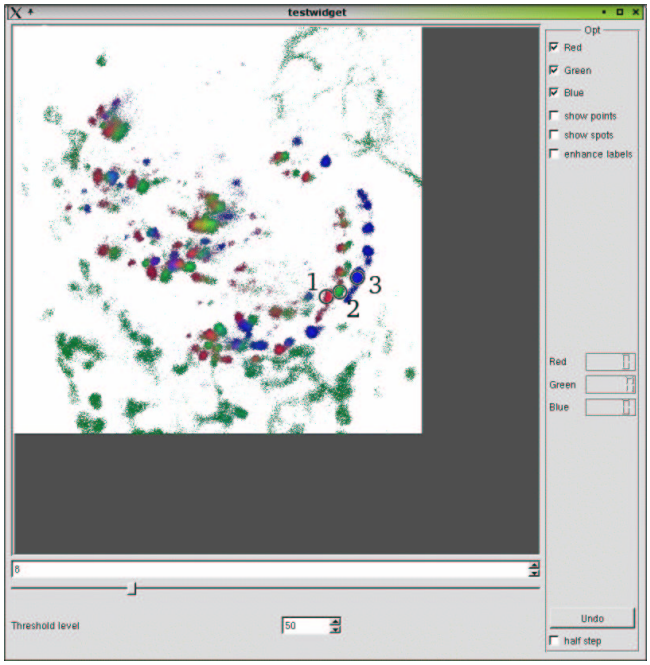


Figure 10.3: Main program display

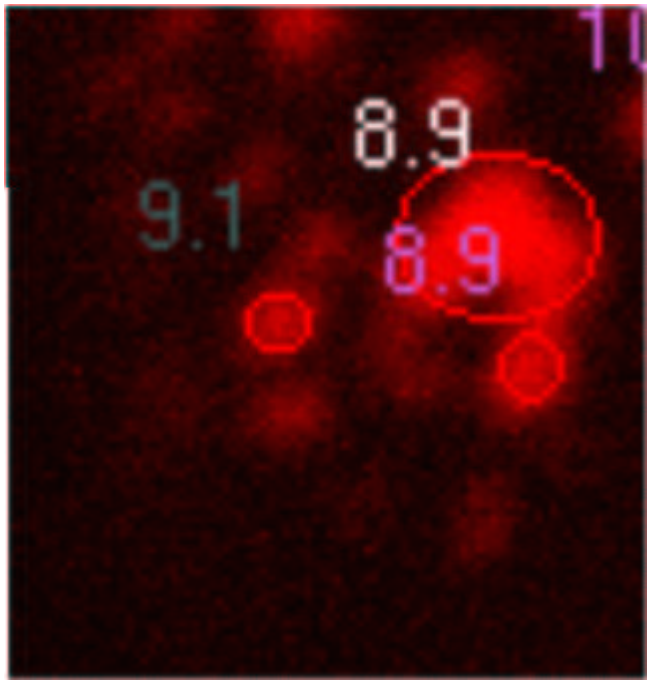


Figure 10.4: Particular of the displayed detected spot

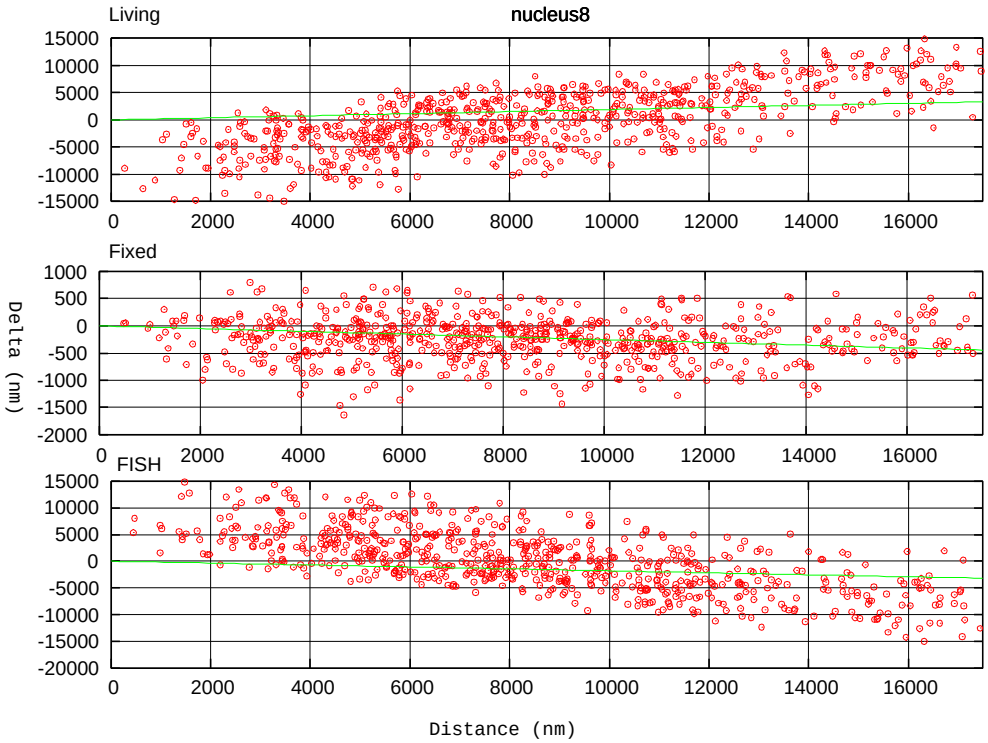


Figure 10.5: Plot of large range displacements of "zones" inside the same nucleus in the same state (Figure 10.6 on the following page and Chapter 10.3.3 on page 58): on the abscissa axis there is $Distance[l]$ against (from the top plot to the bottom one) $\Delta_{lf}, \Delta_{fh}, \Delta_{lh}$. On green there is the fitted line.

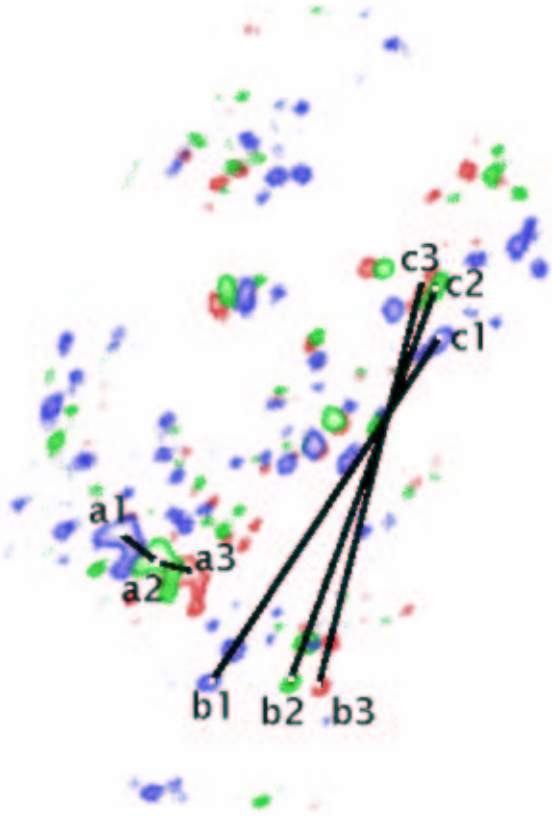


Figure 10.6: Schema for distances measurement

Chapter 11

Following chromosome foci in five dimensions

11.1 Goals

11.2 Materials and methods

11.2.1 The biological specimens

As in the first part of Section 10.2.1 on page 56 HeLa cells were in vivo replication labeled with Cy3-dUTP and with GFP*, then followed through several cell cycles to obtain segregated chromosome territories: the samples so far produced showed two colors, a green one (GFP at 488 nm) and a blue one (Cy3 at 543 nm).

11.2.2 Scanning procedure

Using a confocal scanning microscope Zeiss (equipped with 63 x 1.4 N.A. objective lens) slice scanning images were recorded with a voxel dimension of $180x180x400nm^3$: the scanning region was $256x256x30$ voxels wide and the scanning process was repeated fourteen times ($\delta t = 6min$). During the recording procedure two different channels were acquired: the green one produced by GFP marked histone ($\lambda = 488nm$) and the red one produced by Cy3-dUTP probe ($\lambda = 543nm$) linked to the chromosome sites. The data were saved in two separated 8 bit wide file series: the tiff file format was used requiring further conversion to a different file format. The complete data set was composed by three dimensions for the volume set, one for time and two color channels and the total dimensionality of the problem was so five†.

11.2.3 Processing of the data

We followed the work flow listed on Figure 11.7 on page 74 to process the data; in Figure 11.4 on page 71 is show a sliced view of the time sequence to be processed (first and second columns) and the resulted matched images (third column).

First channel (GFP) processing

First from the GFP channel (in the first column in Figure 11.4 on page 71) we segmented the data with a simple global threshold method (as seen in Section 7.2 on page 31): with the general approach shown in 6 on page 27 we extracted for every time step a list of indexes and thresholds enough to fully describe the central nucleus along all the time sequence. Using the run length encoded data we extracted from this data set only the nucleus to be matched (see Figure 6.1 on page 30 and Section 6 on page 27): we matched the sequence using a correlation algorithm[37] applied to the whole time series. The match has been carried against the previous time step (so there were correlated the step at time t and at time t-1); this has been chosen because the nucleus changed its own shape and rotate too much: using as reference volume the initial one could had been lead to greater errors. We choose to correct for shift along the x and y

*green fluorescent protein

†From now we will use the terms first and second channel to indicate the GFP and the Cy3 channels

axes and for rotations along the z only: this has been chosen because a final matching algorithm was used not requiring precise matching at this stage. The displacement vectors, the rotation vectors as computed from correlation and the run length encoded shape of the nucleus were recorded and used for further analysis against the second channel (Cy3)[‡]. The run length encoded step in the procedure so far used was the same used in Section 10 on page 55.

Second channel (Cy3) processing

We collected the shift and the rotation vectors as computed in the previous section: these were used to correct the data in the second channel. We used the run length encoded informations to extract the part of the whole scanning volume inside the nucleus as detected from the GFP channel (the yellow closed line in the third column of Figure 11.4 on page 71 where is shown the two dimensional projection) At this point we finally had the Cy3 channel data extracted from the whole scanning volume and correlated so removing the motion contribution from the cell movements (see in Figure 11.5 on page 72 for a result of this process). We segmented these data with different thresholding values as in the previous section, we collected for every threshold the index table of the spot detected and we feeded the visualization program.

11.2.4 The computer visualization system

The data so far collected, correlated and segmented were sent to the visualization program shown in Figure 11.5 on page 72: here the user was able to do "hybrid" tracking. This "hybrid" tracking was done following the sequence:

1. The user looked through the different slices in the same volume finding a suitable "spot"
2. He visually followed this spot along all the time steps and all the slices to see if it was always present (no occlusions)
3. He choose a thresholding value that matched the chosen spot, changing it if was necessary
4. "Clicked" the ellipse that described the segmented so a group of parameters were saved on disk

This procedure was repeated for all the spots and the final data recorded for every "clicked" spot were in that order:

- the time step
- the thresholding value
- the index for the spot

According to the Section 6.5 on page 29 these three values are enough to uniquely identify the spot and recover its data: the final spot barycenters were then plotted in 11.6 on page 73.

[‡]The recorded informations were just the segmentation threshold and the index as listed in Section 6 on page 27, because it was enough

11.2.5 Path contraction

Here the data were analyzed to extract informations about the movement of the Cy3 marked spot inside the cell nucleus: so we extracted the data belonging to the single spot and then we repeated for all the spots the measurements. Before analyzing the data a minimization algorithm was applied in order to correct for the nucleus drifts and increasing the quality of the results. As shown in Figure 11.2 on page 70 the spot movement was composed of two parts: a "systematic" one, due to the nucleus drift, and the "real" relative movement. We assumed that with a rigid movement (rotation and shift) we matched the two nuclei, so given the operator $T(\vec{\Delta}, \vec{\Theta})$:

$$\begin{aligned} \vec{x}' &= T(\vec{\Delta}, \vec{\Theta})\vec{x} \\ \vec{\Delta} &= (d_x, d_y d_z) \\ \vec{\Theta} &= (\phi, \theta, \psi) \end{aligned} \tag{11.1}$$

Where $\vec{\Delta}, \vec{\Theta}$ were respectively the offset vector and the rotation vector (see Figure 11.3 on page 71 for reference); a suitable form for the operator $T(\vec{\Delta}, \vec{\Theta})$ could be:

$$\begin{bmatrix} cos(\phi) & -sin(\phi) & 0 & d_x \\ sin(\phi) & cos(\phi) & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11.2}$$

With the constrains:

$$\begin{aligned} \phi = \phi, \quad \theta = 0, \quad \psi = 0 \\ d_x = d_x, \quad d_y = d_y, \quad d_z = d_z \end{aligned} \tag{11.3}$$

And the vector given by:

$$\vec{x} = (x_0, x_1, x_2, 1) \tag{11.4}$$

A more complex operator had been used to model the systematic effects using more degrees of freedom (θ, ψ not zeroed): the form was a trivial extension to the Equation 11.2. Point by point we computed the distances $d_i = |\vec{y_i}' - \vec{x_i}'|$ between the $\vec{x_i} = T(\vec{\Delta}, \vec{\Theta})\vec{x}$ transformed point and the $\vec{y_i}'$ then we minimized the quantity:

$$\begin{aligned} D &= min_{(\vec{\Delta}, \vec{\Theta})} ||e||_2^2 \\ ||e||_2^2 &= \sum_i |\vec{y_i}' - \vec{x_i}'|^2 \end{aligned} \tag{11.5}$$

A monte Carlo based minimization algorithm was then used to minimize the equation 11.5 that gave an $T(\vec{\Delta}, \vec{\Theta})$ operator describing the rigid roto-translation of the whole nucleus that minimized the spot internal movements.

11.3 Results

11.3.1 Distances evaluation

For every time step we extracted the positions of all the spots (actually five) and we computed inter-spot distances using permutations: so the first distance

Average displacements (in nm)
Without contraction
800 ± 400
540 ± 160
720 ± 270
730 ± 280
720 ± 350
After contraction
700 ± 300
160 ± 140
430 ± 180
410 ± 200
850 ± 460

Table 11.1: Average displacements

between the first spot and the second one was computed, then between the first and the third one and so on for the remaining spots. A final table was created out of these distances (Table 11.2 on the next page) and a plot of evolution of these distances was plotted against the time on Figure 11.1 on the following page.

11.3.2 Displacements evaluation

Here we analyzed the spot movements relative to the cell nucleus: we applied to the data extracted (the "original" ones) the minimization algorithm as detailed in 11.2.5 on the page before resulting in a "contracted" path. Then a final plot was made (Figure 11.6 on page 73) reporting the original path and the "contracted" one: a table of the resulting displacements was computed in 11.1 to quantitatively analyze these movements.

11.4 Discussion

From the data acquired and the result obtained we had to note two things:

- the inter spot distances remained almost stable
- the displacement are greater for spots in the external positions

The first point just confirmed that chromosome territories are quite stable: there are experimental evidences and a general trend to justify this statement (Section 10 on page 55). The second point was more obscure and deserves more investigations: was the displacement higher on the border a real effect or a systematic effect of the minimization algorithm or of the model underlying it? We just recorded this strange result deserving further investigation when more data will be available with greater quality.

Variation in distances	
Couple	Average distance(nm)
0 - 1	$9210 \pm 370(4)$
0 - 2	$3270 \pm 190(6)$
0 - 3	$11020 \pm 370(3)$
0 - 4	$17610 \pm 430(2)$
1 - 2	$6160 \pm 190(3)$
1 - 3	$1910 \pm 130(7)$
1 - 4	$8500 \pm 370(4)$
2 - 3	$7910 \pm 220(3)$
2 - 4	$14550 \pm 370(2)$
3 - 4	$6680 \pm 330(5)$

Table 11.2: Average distances

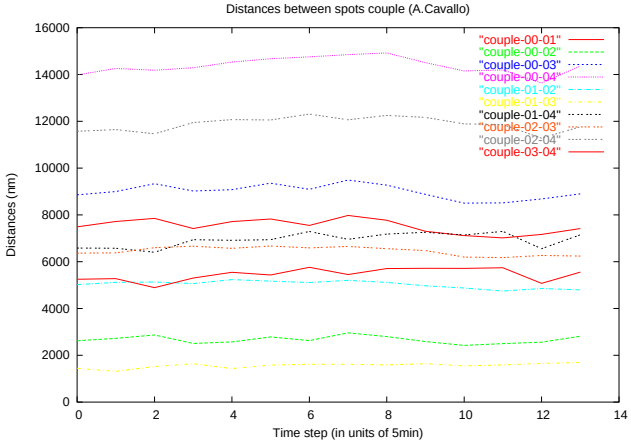


Figure 11.1: Distances between spot couples

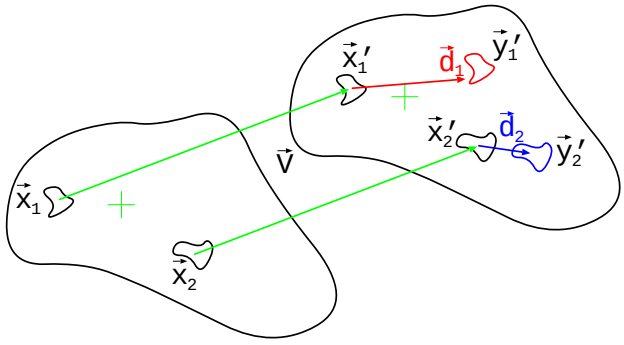


Figure 11.2: Displacement at two different time step of one cell nucleus: the vector $\mathbf{d_1}$ is the real movement of the chromosome marked spot, and the vector $\mathbf{x_1'} - \mathbf{x_1}$ is the "systematic part".

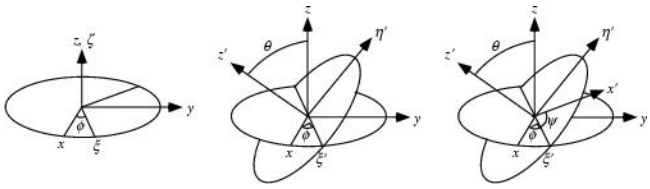
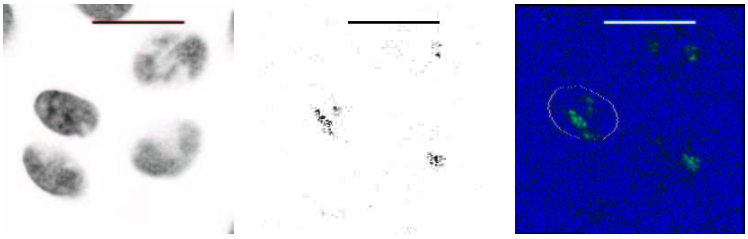
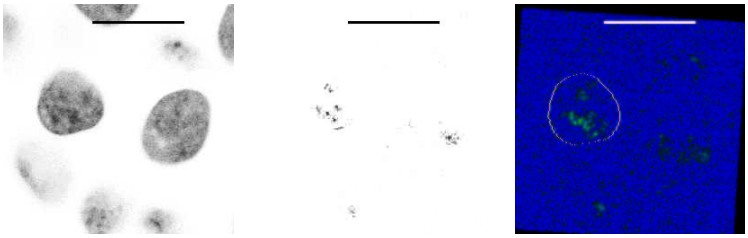


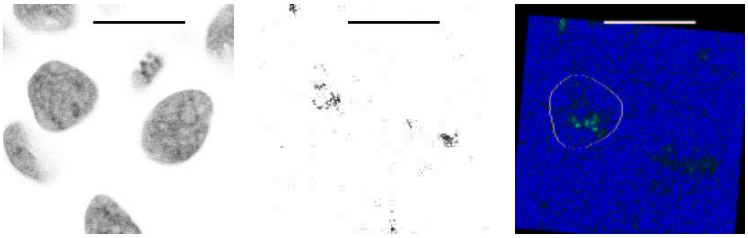
Figure 11.3: The Euler rotation angles



(a) Time=0 (b) Time=5 (c) Time=10



(d) Time=0 (e) Time=5 (f) Time=10



(g) Time=0 (h) Time=5 (i) Time=10

Figure 11.4: Five dimensional data set: first row GFP data (intensities reverted), second row Cy3 channel and third row both channel superimposed and correlated (yellow the GFP channel border of the segmented cell nucleus). Voxel space $256x256x30$ every voxel is $180x180x400nm^3$: on top of every image there is a ruler $20\mu m$ wide, the slice for all is the sixth (out of 30).

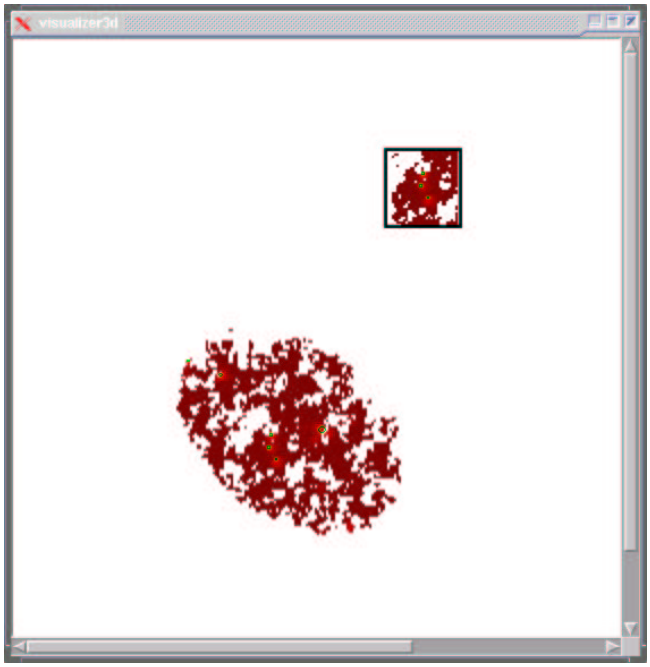


Figure 11.5: A view on the displaying system showing the data from the Cy3 channel segmented and extracted using the shape computed from the first channel (red): in green are reported the segmented ellipses (here are represented the sections). An user can just click on the green spots and select them: with buttons (not shown) he can move the slice plan and/or the time step exploring the full 4d world. In the inset there is an enlarged zone showing the detail of the ellipses: note that these ellipses are just section of the spot ellipsoids.

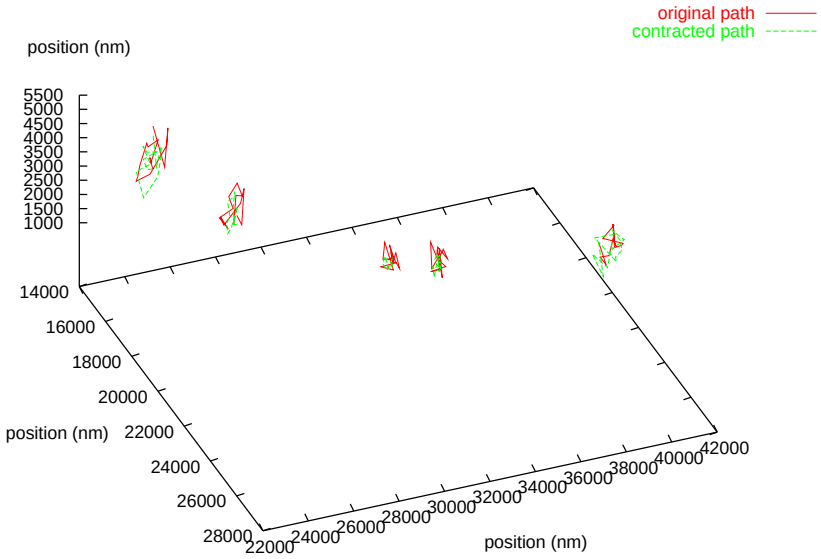


Figure 11.6: The plot of three dimensional displacement of marked spots: in green there is the original path and in red the contracted one

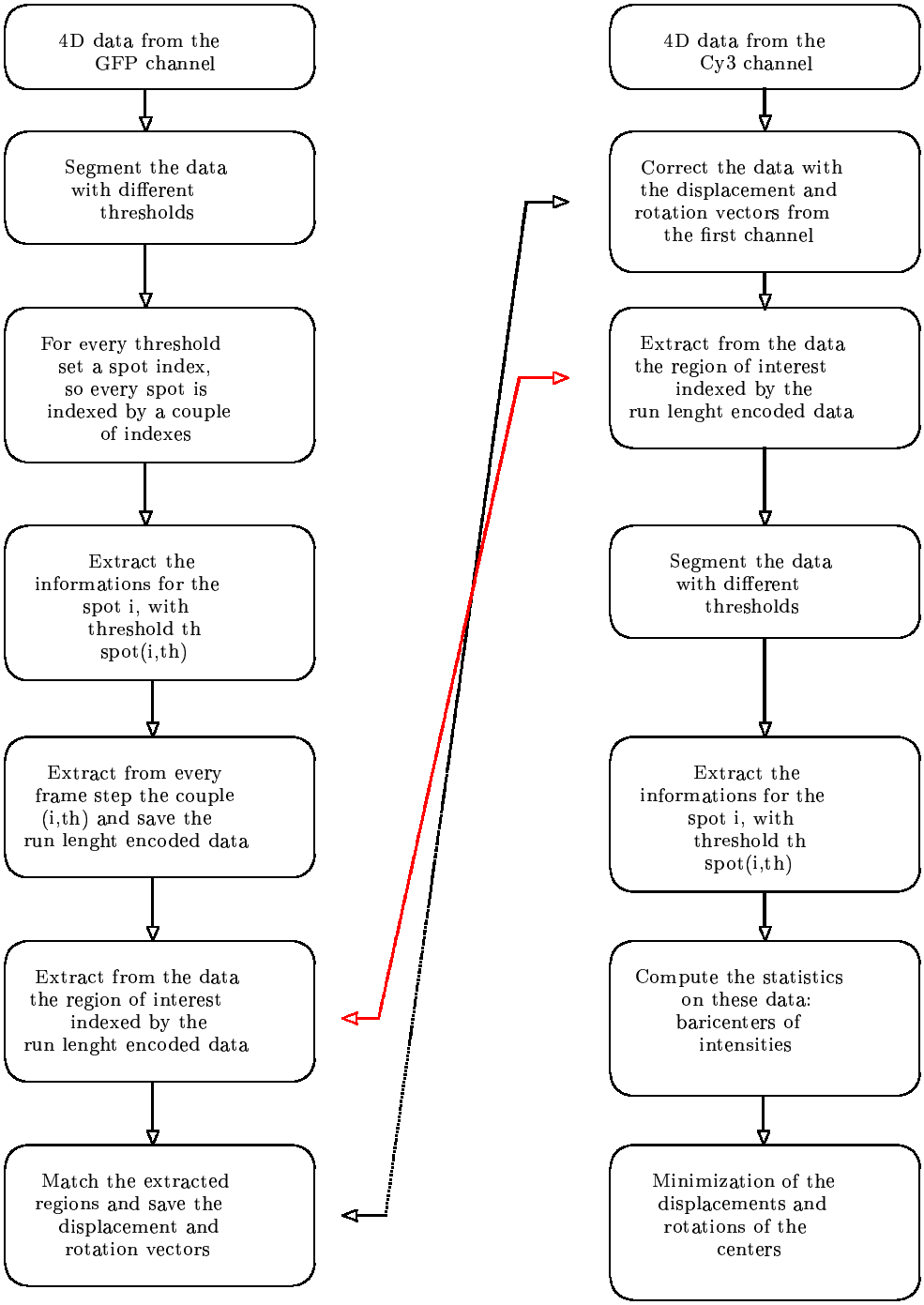


Figure 11.7: Work flow for five dimensional particle tracking

Chapter 12

Conclusions

12.1 The problems and the answers

The biology has become the most prominent discipline in the natural science field: the improvement in methods (optical microscopy, nuclear magnetic resonance and electron scanning devices) made possible to analyze the indeep internals of the cell nucleus and of the chromosome (the structure of whom being a "recent" discovery). So is not at all strange to see every month new works made on the cell topic that is fundamental to the life and, in turn, to all of us. Here we gave our modest contribution of biology related work: we have tried to use digital image processing to help the biologist in their experimental work. First we have tried to apply the volume/vector field method to biological samples in a way that make possible to compare simulations of the chromosomes to the real experimental data: we have been un-lucky because this work resulted in a too hard to manage problem. It is not un-usual in scientific investigations to follow such hard to solve problems: if there could be someone who can say this in advance, well we were not following an original or new path. On the other side we followed a particle tracking approach based on a segmentation-tracking method: we were able to give some answer to some biology related problems[11] and to found a method to analyze dynamical movements[7] in a very effective way. From a digital image processing side we exploited an algorithm (run length encoding) to generate a compact shape representation of multidimensional objects: we used this in conjunction with object oriented style data analysis.

12.2 Future development path

Based on our (little) experience acquired we can plot here what could be done to improve the particle tracking strategy in the near future. A first key point is in the general availability of user-supervised programs to explore four dimensional data sets: we are not talking about visualization system (now generally available), but very specialized programs on the tracking problem who can act as blackboards for scientist. We have used an "hybrid" approach to the particle tracking because the always changing environment made impossible to plan any strategy: during this development we felt the necessity to join the data with some computed parameter (spot detected barycenters in our case). This flexibility is more important in a biology environment where the requirement is to have the 100% probability of data analysis because the experiments are so hard to repeat. This is the general trend as we have seen in many real life cases.

Bibliography

[1] B. Albert, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular biology of the cell.* Garland Publishing, Inc, 1994.

[2] C. A. Bertrand. Color by number. *Science*, 293:1335–1336, 2001.

[3] The blitz++ matrix library. http://www.oonumerics.org/blitz.

[4] H. Bornfleth, P. Edelmann, D. Zink, T. Cremer, and C. Cremer. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophys. J.*, 77:2871–2886, 1999.

[5] H. Bornfleth, K. Sätzler, R. Eils, and C. Cremer. High-precision distance measurements and volume-conserving segmentation of objects near and below the resolution limit in three-dimensional confocal fluorescence microscopy. *J. Microsc.*, 189:118–136, 1998.

[6] T. Cremer C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2:292–300, April 2001.

[7] A. Cavallo, L. Schermeleh, R. Heintzmann, T. Cremer, and C. Cremer. A hybrid approach to the 3d high precision position measurement and particle tracking in human cell nuclei. Accepted for publication in Real-Time Imaging's Special Issue on Bioinformatics.

[8] M. Cercignani, M. Inglese, M. Siger-Zajdel, and M. Filippi. Segmenting brain white matter, gray matter and cerebro-spinal fluid using difusion tensor-mri derived index. *Magnetic resonance iamging*, 19:1167–1172, 2001.

[9] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, and J. J. et al. Heine. Mri segmentation: methods and applications. *Magnetic resonance imaging*, 13:343–368, 1995.

[10] T. Cremer, C. Cremer, H. Baumann, C. K. Luedtke, K. Sperling, V. Teubner, and C. Zorn. Rabl's model of the interphase chromosome arrangement tested in chinese hamster cells by premature chromosome condensation and laser uv microbeam experiments. *Human genetics*, 60:46–56, 1982.

[11] T. Cremer, C. Cremer, A. Cavallo, L. Schermelleh, and I. Solovei. Conservation of dna distribution during fluorescence in situ hybridization on three-dimensionally preserved nuclei (3d fish). Accepted for publication in Exp. Cell. Res.

[12] T. Cremer, A. Kurz, R. Zirbel, S. Dietzel, B. Rinke, E. Schröck, M. R. Speicher, U. Mathieu, A. Jauch, P. Emmerich, H. Scherthan, T. Ried, C. Cremer, and P. Lichter. Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Symp. Quant Biol*, 58:777–792, 1993.

[13] T. Cremer, A. Kurz, R. M. Zirbel, S. Dietzel, B. Rinke, E. Schröck, and al. Role of the chromosome territories in the functional compartimentalization of the cell nucleus. *Cold Spring Harb. Symp. Quant. Bio.*, 58:777–792, 1993.

[14] J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. Bickmore. Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of cell biology*, 145:1119–1131, 1999.

[15] The Python Numerical extension. http://numpy.sourceforge.net.

[16] J. Ferreira, C. Paolella, G. Ramos, and A. I. Lamond. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *The journal of cell biology*, 139:1597–1610, 1997.

[17] P. Geisler, B. Jähne, and H. Haußecker. *Handbook of Computer Vision and Application.* Academic Press, 1999.

[18] M. Gu. *Principles of three dimensional imaging in confocal microscopes.* 1996.

[19] Hanno. *Handbook of Computer Vision and Application*, volume 2, chapter 13, pages 309–397. Academic Press, 1999.

[20] Horst Haußecker and Hagen Spies. *Handbook of Computer Vision and Application*, volume 2, chapter 13, pages 309–397. Academic Press, 1999.

[21] P. Heun, T. Laroche, K. Shimada, P. Furrer, and S. Gasser. Chromosome dynamics in the yeast interphase nucleus. *Science*, 294:2181–2186, 2001.

[22] B. Jähne. *Digital Image Processing.* Springer, 1997.

[23] Sandhya Jha. Henrietta lacks - an unwitting part of african-american history at hopkins. http://www.jhu.edu/ñewslett/02-21-97/Features/Beyond_her_years.html.

[24] G. Kreth, P. Edelmann, C. Münkel, J. Langowski, and C. Cremer. Chromatine structure and chromosome aberrations: modeling of damage induced by isotropic and localized irradiation. *Mutat. Res.*, 404:77–78, 1998.

[25] The Python language. http://www.python.org.

[26] M. M. E. Manders, H. Kimura, and P.R. Cook. Direct imaging of dna in living cells reveals the dynamics of chromosome formation. *Journal of cell biology*, 144:813–821, 1999.

[27] J. Meyering. gnu tr command. man page.

[28] M. Minsky. Microscopy apparatus. US Patent 3013467, 1957.

[29] C. Münkel and J. Langowski. Chromosome structure predicted by a polimer model. *Physical Rev E*, 57:5888–5896, 1998.

[30] W. A. Mohler. Visual reality: using computer recontruction and animation to magnify the microscopist's perception. *Molecular Biology of the cell*, 10:3061–3065, 1999.

[31] D. J. Murray and W. Van Ryper. *Encyclopedia of graphics file formats.* 1994.

[32] H. Myler and A. R. Weeks. *Computer imaging recipes in C.* 1995.

[33] S. Paddock. Optical sectioning - slices of life. *Science*, 295:1319–1321, 2002.

[34] D. Pinkel, T. Straume, and J.W. Gray. Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci.*, 83:2934–2938, 1986.

[35] The qt toolkit. http://www.troll.no.

[36] M. Rovaris, M. Filippi, G. Calori, M. Rodegher, A. Campi, B. Colombo, and G. Comi. Intra-observer reproducibility in mesuring new putative mr markers demyelination and axonal loss in multiple sclerosis: a comparison with conventional t2-weighted images. *J. Neurol.*, 244:266–270, 1997.

[37] John C. Russ. *The Image Processing Handbook.* CRC Press, 1994.

[38] R. K. Sachs, G. van den Engh, B. Trask, H. Yokota, and J. E. Hearst. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci.*, 92:2710–2714, 1995.

[39] Hanno Scharr. *Optimal Operators in Digital Image Processing.* PhD thesis, University of Heidelberg, May 2000.

[40] L. Schermelleh, I. Solovei, D. Zink, and T. Cremer. Two-color fluorescence labelling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.*, 9:77–80, 2001.

[41] I. Solovei, D. Kienle, G. Little, R. Eils, L. Savelyeva, M. Schwab, W. Jäger, C. Cremer, and T. Cremer. Topology of double minutes (dmins) and homogeneously staining regions (hsrs) in nuclei of human neuroblastoma cell lines. *Genes Chromosomes and cancer*, 29:297–308, 2000.

[42] Hagen Spies. *Bildfolgenanalyse mit Tiefenbildern.* PhD thesis, University of Heidelberg, August 2001.

[43] Springer-Verlag, editor. *Fluorescence in situ hybridization - Laboratory manuals.* 2002.

[44] Kensal van Holde and J. Zlatanova. Chromatin higher order structure: chasing a mirage? *The journal og biological chemistry*, 270:8373–8376, 1995.

[45] J. Verestóy and D. Chetverikov. Feature point tracking algorithms. http://visual.ipan.sztaki.hu/psmweb/index.html.

[46] Lisa H. Weasel. Race and gender through the microscope: A feminist perspective on henrietta lacks and the hela cell line (abstract from a conference). http://depts.washington.edu/hssexec/annual/abstractsp9.html.

[47] D. Zink, H. Bornfleth, A. Visser, C. Cremer, and T. Cremer. Organization of early and late replicating dna in human chromosome territories. *Exp. Cell. Res.*, 247:176–188, 1999.

[48] D. Zink, T. Cremer, R. Saffrich, R. Fischer, M.F. Trendelenburg, W. Ansorge, and E.H. Stelzer. Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.*, 102:241–251, 1998.

[49] R. M. Zirbel, U. R. Mathieu, A. Kurz, T. Cremer, and P. Lichter. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Research*, 1:93–106, 1993.

[50] J. Zlatanova, H. S. Leuba, and Kensal van Holde. chromatin fiber structure: morphology molecular determinants, structural transitions. *Biophysical journal*, 74:2554–2566, 1998.