Quantitative Analysis of Intracellular Processes in Living Cells

Vladimír Ulman

Abstract

In modern cell biology visual description of cellular processes as they are seen in microscope is not sufficient. The evolution of technology enables us to support our claims or suggestions in a more precise way as the resolution and speed of optical systems increase. Quantitative characterization of processes may reveal functional dependence among objects under study.

In this review, image preprocessing, extraction of objects as well as labeling is expected to be performed prior to quantitative analysis. Establishing of correspondence between objects in images captured at subsequent time points is expected to be precomputed as well.

We will summarize several successful approaches to live cell analysis and divide them according to their input parameters demand. In particular, we will divide them according to the direct use of a priori knowledge of the correspondence: either the correspondence is directly involved in the computation of quantitative characteristics or we make use of correspondence just for creating time dependence graphs of analyzed parameters.

As for the first case, methods analyzing course of time-lapse objects’ positions will be presented. For instance mean velocity, acceleration, mean squared distance and diffusion constant, from which we may conclude the object motion behavior, will be described. For the latter case, we will show methods that characterize either the structure of a particular object or spatial object distribution at particular time point. We will mention the object volume, main axes and circularity as representatives of best-known characteristics. Also the spatial distribution of particles in a given object can be quantitatively characterized. The most important methods will be accompanied by practical examples.

1. Introduction

1.1 Computer Aided Live Cell Studies

Computer aided live cell studies are truly interdisciplinary scientific field where one can meet mathematicians, physicists, biologists and computer scientists working together. The biological problem is, often with the help of physicists, turned into a mathematical problem. The mathematicians together with computer scientists develop a tool that is processed by a computer yielding semi-automatical or full-automatical computer system. Such systems handle microscope self-sufficient, process acquired image data and perform data analysis ending up with tables or graphs with results. Note the classical interconnection between physics and mathematics since the biological problem typically inherits the physical matter of live cell.

The advance in development of new optical instruments allows biologists to observe more of biological material at higher resolutions as well as in higher dimensions (Eils et al., 2003). Today’s technologies are capable of not only flat grayscale images (three dimensions) but they can handle time-lapse observation of specimen in volume and in color channels (at least five dimensions). The amount of acquired data then be-
comes overwhelming for human being resulting in computer aided processing of data. On the other hand, the higher the dimensionality of analyzed data is the better seeking functional dependence among biological structures can be performed in the sense that we can describe the searched dependency more accurately. It follows the basic observation that whenever we need to better specify a situation given by some state vector we simply add another element into this state vector. This embeds original state space as a subspace in a new more dimensional state space.

Not only the amount of data complicates the situation but also, as we get deeper into intracellular structures, high precision quantitative analysis must be performed in order to get confident and compatible results across laboratories world-wide. The human observer is typically not able to produce unbiased analysis not mentioning the amount of time required for that. A computer program then acts as a fast independent arbitrer in this respect.

1.2 Review Outline
In this short review we will present few basic techniques that enable us to document basic observations in uniform quantitative way in opossite to qualitative non-objective observation of each scientists. We will present the most frequent mathematical descriptors of physical properties of intracellular objects under study. The aim is to provide a set of mathematical tools allowing for documentation of respective physical property in a uniform manner. The list of techniques is not ment to be exhaustive. Instead, we will just summarize the most frequent and most reliable techniques we use in our laboratory.

The rest of this text is according to the following layout. In the next section we will describe the input data model and establish notation. The third and fourth sections will describe quantitative analysis tools from mathematician's and biologist's point of view. Some of them will be accompanied with an example. The final section is conclusion.

2. Input Data Model
2.1 Live Cell Studies
Acquiring and analyzing time-lapse series of images of biological specimen (one or more whole cells or their interior) is called live cell study (Gerlich et al., 2002). Each time step might be also represented by a series of images. One can acquire images at the same planar position (fixed $x$ and $y$ coordinates) while moving the specimen along the optical axis and scanning 2D image at each $z$ coordinate. A stack of images representing volume of specimen is created in this way.

One can also acquire images of the same specimen region (2D area or 3D volume) at different color channels. The purpose of different color channels is to distinguish between intracellular objects since it is possible, in special color channel, to display only particular object types (e.g. only nucleus or only cytoplasm or only gene locus) among all objects comprising the whole cell. Such use of several color channels is typical in fluorescence microscopy (Kozubek et al., 2004).

Thus, the image can be represented as discrete function with several parameters

$$I(x,y,z,t,c) \in G$$

where $x$, $y$ and $z$ are the coordinates within the specimen, $t$ is the time step at which the acquisition took place and $c$ is the color channel which one may think of as an object selector. $G$ represents the values of pixel intensities, typically $G \in [0,255]$. 

18.11.2005, UlmanV.rtf
2.2 Image Preprocessing

The data acquired from optical system are useless without preprocessing steps. We won't get into any details in this respect. Refer to (Šonka et al., 1995) for more detailed description. Instead, we will just expect the preprocessing is done somehow and beside the raw image data (the discrete function $I$) we get another representations.

First of all, all objects of interest are extracted and labeled at every time step and in every color channel. The function $L$,

$$L(l,t,c) = COORD$$

where $l$ is an arbitrary label, returns a set $COORD$ which contains all coordinate vectors $(x,y,z)$ for which $I(x,y,z,t,c)$ holds a voxel (volume element) of the given object that we've labeled $l$. The parameter $t$ is the time step and $c$ is the color channel again.

Second, the correspondence among labels over the time has to be established. Two approaches can be employed. Either the condition on all labels used within the entire data set must be held or a simple mapping function must be determined. We will make use of the first case, the condition. The definition of the condition is: the label of selected object in given color channel must be unique and preserved during the time sequence, if the object is not present in given channel at particular time step then the $L$ function returns empty set. In this way, we can be sure that we are always processing voxels of the selected object in the time-lapse image sequence whenever we use the very same label.

2.3 One Possible Division of Quantitative Property Descriptors

Object properties, from now on called object statistics or object features, can be divided according to several perspectives. We will divide them with respect to time-lapse studies and consider whether the given statistics inherently requires the correspondence information, retrieved from time-lapse sequence, for its computation or single time step is enough to retrieve this statistics. However, we will make use of the correspondence information even in the second case, in which single time step is enough, just for the ability to document changes in the given statistics of some object during the time sequence.

In the following text we will assume time-lapse 3D images acquired in one color channel. Any other assumption will be explicitly mentioned.

3. Single Time Step Quantitative PropertyDescriptors

3.1 One Color Channel Descriptors

We will first begin with the single time step statistics in spite that these are the second case according to our division mentioned in previous subsection. These statistics quantitatively document basic observations like the area or volume of object under study, its circularity or elongation, etc. For the purpose of this section we will choose some arbitrary $t$ that is valid within the time sequence.

First of all, the simplest statistics is the volume estimation $V(l)$ of object $l$ at single time $t$. It is simply defined as the sum of all voxels comprising given object:

$$V(l) = \sum_{v \in L(l,t,c)} 1$$
Second basic feature is the surface area $S(l)$ of object $l$ at single time $t$. Unlike volume, the computation of this statistics involves two problems: first, we must determine boundary voxels and, second, the number of boundary voxels does not approximate the expected volume quite exactly. The boundary is typically defined as the set of voxels $B'(l,t,c) \subseteq L(l,t,c)$ for which it holds that for every of them there exists at least one voxel from its neighborhood that does not belong to the given object $l$. The neighborhood of voxel is usually defined as those voxels whose coordinate vectors differ at exactly one element (coordinate) by exactly one. According to this definition the neighborhood of voxel contains six adjacent voxels. The surface area $S(l)$ of an object $l$ is:

$$S(l) = \sum_{v \in B'(l,t,c)} 1$$

Unfortunately, this approach is biased by error. The error is 7% in average when measuring the surface area of a sphere with different diameters (Hrstíř, 2004).

The units of volume and surface area, according to the given definitions, are voxels. This does not say anything about the real volume and surface area without knowledge of resolution of the image. The resolution is given for each coordinate axis and defines the number of voxels in a line along that axis so that the line is one unit metric distance long. In optical microscopy the resolution is usually given in voxels per micrometer.

The roundness statistics $Ro(l)$ informs how much the object with label $l$ resembles a sphere of similar size. It is defined as the ratio and it has no unit:

$$Ro(l) = \frac{S^3(l)}{V^2(l)}$$

The sphere minimizes the ratio. For any other objects’ ratios it holds that the higher ratio is the less round the object is.

This statistics can be used for detecting cell division in special cases. If the cell can be considered as a rather planar object, a circle in 2D, then 3D image can be projected into single 2D image using maximum projection along $z$ axis:

$$I_{2D}(x,y,t,c) = \max_{z}(I(x,y,z,t,c))$$

An example of $I_{2D}(x,y,t,c)$ of artificial cell before and during the division is in Figure 1.

Very useful quantities are the statistical moments. The general expression is:

$$M_{i,j,k}(l) = \sum_{(x,y,z) \in L(l,t,c)} x^i y^j z^k$$

Useful statistics can be retrieved by proper combination of moments with tuned parameters $i$, $j$, $k$. For example the geometric center $C(l) = (\bar{x}, \bar{y}, \bar{z})$ of object labeled $l$ can be computed:

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{M_{1,0,0}(l)}{M_{0,1,0}(l)}, \frac{M_{0,1,0}(l)}{M_{0,0,1}(l)}, \frac{M_{0,0,1}(l)}{M_{0,0,0}(l)}\right)$$

Statistical central moments are another useful objects’ properties. We denote them as $\overline{M}_{i,j,k}(l)$:

$$\overline{M}_{i,j,k}(l) = \sum_{(x,y,z) \in L(l,t,c)} (x - \bar{x})^i (y - \bar{y})^j (z - \bar{z})^k$$
Division of an artificial cell. The left cell A) is before division and has the roundness property equal to $4\pi$. The right cell B) is, say, in the middle of its division and the roundness property it poses is 127% of the roundness of left cell. The change in roundness of the same object implies change in its shape.

Difference between geometrical and intensity center of cell. There are always two black crosses in pictures A) to C), the upper left one represents geometrical center and does not change in pictures. The position of this center is identical to the geometrical center one would get in picture D) in which the binary mask of this cell is depicted. The lower right cross represents intensity center of cell. It is assumed that the brighter the spot is the denser the cell at that particular spot can be expected. Note that intensity center got slightly further in B) and C) in comparison to A) but performed no movement between B) and C). Also note the brightness differences between B) and C). The distance between both centers only indicates the loss of homogeneity.

Beside of geometric center of an object we can define an intensity center of an object, sometimes referred to as center of mass. The only difference is that intensity center will be sensitive to mass disproportion within the object under analysis. We will denote the intensity center of object labeled $l$ with $C'(l) = (x', y', z')$ and compute it:

$$M'_{i,j,k}(l) = \sum_{(x,y,z) \in L(l,i,j,k)} x'y'z'I(x,y,z,i,j,k)$$

$$(x', y', z') = \left(\frac{M'_{1,0,0}(l)}{M'_{0,0,0}(l)}, \frac{M'_{0,1,0}(l)}{M'_{0,0,0}(l)}, \frac{M'_{0,0,1}(l)}{M'_{0,0,0}(l)}\right)$$

See Figure 2 where the difference between $C(l)$ and $C'(l)$ is shown.
The radius $R_a(l)$ of an object $l$ is the distance from the center of an object to its boundary. We usually distinguish between minimum, maximum and average radius. In the following text we will always refer to maximum radius measured from geometrical center of object. Now that we can compute central moments, we can compute main axes of given object. Main axes of 3D object are three orthogonal axes establishing new coordinate system with its center identical to the center of given object. The direction of new $x$ axis should align with the main direction of object’s matter.

The most basic method of computing main axes of an object is via the eigen values and eigen vectors of covariance matrix $Cov(l)$ (Smith, 2002):

$$Cov(l) = \begin{pmatrix}
M_{2,0,0}(l) & M_{1,1,0}(l) & M_{1,0,1}(l) \\
M_{1,1,0}(l) & M_{0,2,0}(l) & M_{0,1,1}(l) \\
M_{1,0,1}(l) & M_{0,1,1}(l) & M_{0,0,2}(l)
\end{pmatrix}$$

The $x$ axis is the axis given by the eigen vector which corresponds to the largest eigen value.

### 3.2 Descriptors Using More Than One Color Channel

The last statistics we will present in this section is the characterization of distribution of objects from one channel within a large object in another color channel, in particular the distribution of histons within the relatively flat cell nucleus (Cremer et al., 2001). The histons were acquired in channel $h$ and the cell nucleus was acquired in channel $n$. Both channels were projected into 2D. The nucleus (in color channel $n$) serves as a mask for selecting only those histons (in color channel $h$) that lie within the nucleus since there are often histons or just noise outside the nucleus that would harm the distribution statistics. The nucleus was then approximated by system of 25 adjacent non-overlapping stripes (shells). Each stripe was generated by concentric ellipse with its center identical to the nucleus center and with its main axis parallel to main axis of flatten nucleus (main axes were computed from $I_{2D}(x,y,t,n)$), see Figure 4.

Because of the projection of images into 2D the $z$ coordinate becomes useless and we can replace the coordinate set $L(l,t,n)$ by the set $L_{2D}(l,t,n)$ in which we remove the $z$ coordinate from all vectors. We will denote the set of coordinate vectors of all voxels from nucleus mask within each stripe as $Stripe(l,t,n) \subseteq L_{2D}(l,t,n)$ where $i \in \{1,25\}$ and also the average intensity of histons within nucleus mask as $avg(l,t,h)$:

$$avg(l,t,h) = \sum_{(x,y) \in L_{2D}(l,t,n)} I_{2D}(x,y,t,h) / \sum_{(x,y) \in L_{2D}(l,t,n)} 1$$

The average intensity difference of histons within every stripe was calculated and denoted as $avg_{i}(l,t,h)$:

$$avg_{i}(l,t,h) = \sum_{(x,y) \in Stripe_{i}(l,t,n)} (I_{2D}(x,y,t,h) - avg(l,t,h)) / \sum_{(x,y) \in Stripe_{i}(l,t,n)} 1$$

The values were normalized (divided by constant) and the resulting graph documenting the distribution of histons shown in Figure 4 is displayed in Figure 5.
Figure 3
Main axes determination. Main axes are well determined in elongated object in picture A). The ambiguity of determination is shown in pictures B) and C) where both objects hold the same roundness property.

Figure 4
Sample multi-channel data. Pictures A) and B) display the color channels $h$ and $c$, histons and cell nucleus, respectively. Picture C) demonstrates the arrangement of stripes (only 5 stripes are visualized), the further the stripe from the center is the darker color it has.

Figure 5
Sample distribution statistics averaged over four time steps. The horizontal axis contains the relative distance from the geometrical center of nucleus towards the maximum radius. The vertical axis contains mean of normalized statistics. Small vertical bars inside the graph represent the variance of given $\text{avg}_{(l,t,h)}$.
4. Quantitative Property Descriptors Requiring Time-lapse Information

4.1 Movement Parameters Estimation

Let us assume that the time steps are equally spread within the period of acquisition, i.e. the delay $\Delta t$ between two consecutive time steps is always the same. For the sake of trajectory analysis we can represent each labeled object $l$ with its geometrical center $C(l,t)$ in which the second parameter was added since in this section the time step parameter $t$ is not fixed anymore.

The most obvious documentation of object movement is simply visualization of its trajectory. That can be conducted in a computer generated 3D graph containing a continuous line connecting subsequent positions of the center. The color of the line may represent the time, for instance the brightest color represents the starting time and the color of a particular line point fades as we are moving towards the time step at the end of the time sequence. However, this is not quite exactly quantitative characterization in comparision to what we are typically expecting from quantitative characterization.

The average speed $v(l,t)$ at certain time is a quantitative characterization that can be simply determined as:

$$v(l,t) = \frac{|C(l,t + \Delta t) - C(l,t)|}{\Delta t}$$

The expression $|u - v|$ is the Euclidean distance between two vectors $u$ and $v$. In the limit, where $\Delta t$ proceeds close to zero, the average speed becomes the immediate speed. Generally speaking, shorter delays between consecutive time steps can do no harm in terms of data analysis vice versa it can become more precise. Nevertheless, sometimes shorter delays may harm the analysis since the specimen is more frequently exposed to the microscope light which might affect the living intracellular organs (objects, according to the terminology used in this document).

Acceleration $a(l,t)$ of selected object can be estimated too:

$$a(l,t) = \frac{(v(l,t) - v(l,t + \Delta t))}{\Delta t} = \frac{|C(l,t + 2\Delta t) + C(l,t)|}{\Delta t^2}$$

The acceleration or just the change in average velocity can be useful when observing some long-term interaction of two intracellular objects. The scientist doesn’t have to see entire image sequence right from the beginning. Instead he/she can let the computer find the time step from which the interaction actually began. It is presumed that the beginning of interaction is indicated by change in movement of objects under interest.

4.2 Movement Characterization

(Molenaar et al., 2003) published the quantitative characterization of dynamics of telomeres in living mammalian cells. They observed in their experiments that telomeres undergo their motion at three speed rates and they managed to document that in quantitative fashion.

They used the mean squared displacement (MSD) and the diffusion constant (Vasquez et al., 2001). The MSD is defined as average $|C(l,t) - C(l,t + j\Delta t)|^2$ aggregated over many objects $l$ for fixed initial time $t$ and some valid $j$ (the value of $t + j\Delta t$ must not exceed the time-lapse interval). The MSD plotted against the $j\Delta t$ is shown in Figure 6A. Also note that $j\Delta t$ is displayed along the horizontal axis and designated as $\Delta t$ in Figure 6A.

18.11.2005, UlmanV.rtf
Illustration of MSD and diffusion constant. The MSD is plotted in picture A). It shows average displacement of several telomeres over the time from given position at fixed initial time. The telomeres were observed to change their positions at three different speed rates which resulted in their clustering. Each of three curves in picture A) was therefore computed separately for its cluster of telomeres. Note that all three lines are almost straight. The diffusion constant for the same data is presented in picture B). The figure is reprinted from (Molenaar et al., 2003).

The diffusion constant is simply the MSD divided by $\Delta t$ describing the slope of MSD curve in this way. It is also proportional to the rate at which the object changes its position over time. The observation of three speed levels in that particular experiment is well documented in Figure 6B.

Another situation occurs when objects tend to stay at certain small area, objects waggle, then it may be of interest to compute the average displacement of every object or the variance of every object’s positions during the observation period. This quantity has the ability to document the density of local environment around the selected object. Obviously, the smaller the average displacement is the denser the environment or the more stable the cell is. All forces acting on the object are relatively balanced.

5. Conclusion

We have outlined a few simple quantitative characteristics of common features demanded in live cell study. These include the characteristics that can be computed purely from single time step images allowing for the documentation of time development of certain properties of intracellular objects or even of cells themselves. We have also presented movement characterizing parameters. The information about movement itself is, in fact, the only information we gain from time-lapse observation of live cells in addition to the development of single time step descriptors. Thus, this type of descriptors makes direct use of advantages of time-lapse microscopy.

The advantages of computer processed analysis are obvious. The computer program is typically faster than human being and is objective provided that programs behave deterministically. Unfortunately, the analysis is at the end of chain of processes. Far more complex and thus less confident is the preceding element in this chain called image preprocessing that often prevents us from fully automatic specimen processing. Image preprocessing is often more time demanding too. Moreover, we were focused on simple methods that are able to provide useful information while their outcomes can be still considered reliable. The computational demand of feature descriptors was also taken into account.

Last but not least, one should keep in mind that the list of methods for quantitative analysis will probably be never final. The advances in technology enable us to see more and to learn more about intracellular processes. Technology advances as well as new types of experiments may require the development of new
methodology or new descriptors as new knowledge is gained and new features emerge or instruments become able to measure them.

Acknowledgement

This work has been supported by the Ministry of Education of the Czech Republic (Grant No.

Bibliography


