Faculty of Informatics
Masaryk University

Correspondence Establishing in Image Data of Live Cell Studies

Ph.D. Thesis Proposal

Vladimír Ulman

Supervisor: Michal Kozubek

Signature :

Date   :   Brno, September 2005

Contents

Chapter 1

Introduction

1.1 Live cell studies

1.1.1 Time-lapse observation and its analysis

Acquiring and analyzing time-lapse series of images of biological specimen is called live cell study. The goal is to study the evolution of selected structures *during the time*. The specimen under observation is typically a cell including its intracellular matter or just a part of a cell, e.g. cytoplasm, organells, chromosomes or nucleus. An optical microscope connected to digital camera is used for observation and acquisition of image data. The microscope equipment is connected to and driven by a computer. The acquired data can be computer processed and analyzed resulting in fully automated system. In live cell studies snapshots of specimen are recorded periodically [44, 25]. It is not usually called video recording since the delay between two consecutive time steps can be in tens of minutes whereas in video it is typically in tenths of seconds. Time-lapse observation of living specimen is often called *in vivo* observation.

Image data analysis is, perhaps, the most difficult step from the computer scientists point of view. Image data preprocessing may be required depending on the type of biological investigation and selected methodology. This includes object extraction and correspondence establishment in time-lapse studies. The extraction finds and labels all selected objects from acquired image data with unique label. The correspondence binds all labels of the very same object in the experiment. The correspondence establishment is an inherent property of live cell studies. Last, but not least, the evaluation and visualization of computed data, e.g. list of position coordinates of selected object, is important too [21].

1.1.2 The acquisition and parameters of input data

Since the intracellular structures are rather transparent, a special technique is employed [42]. Basically, a special DNA sequence is inserted into the DNA of live cell allowing for production of fluorescent proteins. These attach to the intracellular structures acting as fluorescence markers. The cell is then exposed to the light of certain frequency band.

Owing to the fluorescence, only labeled structures are visualized due to emitted light which is recorded into a CCD camera. Only a certain type of objects, chromosomes for instance, are displayed in acquired image in this way. It is also possible to have more markers inserted into specimen. The microscope then has to be equiped with changeable excitation and emission filters allowing for image acquisition of selected types of objects. A pseudocoloring is often used for displaying purposes with some of three color channels containing image data of respective marker.

The quality of data also depends on the setup of optical instruments. The FISH technology poses several drawbacks. Owing to the photobleaching (the markers fade proportionally to the time being exposed to light) and phototoxicity (the specimen is harmed by the incoming light), the trend is to expose specimen to the light as little as possible. Unfortunately, this is in contradiction to the need of CCD camera which suffers from several kinds of noise [42]. See [23] for discussion of signal to noise analysis too. But it is not the noise only that complicates automatic acquisition. Optical aberration corrections [43] or robust autofocusing [26] have to be solved too when developing a fully automated system. Moreover, the proper selection of fluorescent markers is important to prevent the cross-talks of excitation and emission frequency bands of used markers.

The dimensionality of data also depends on the mode of optical microscope. The wide-field mode produces "flat" (2D) image data. The acquired image is grayscale rectangle. The intensity value of every pixel is aggregated from the specimen along the optical axis at particular lateral coordinate. Whereas in the confocal mode, only a thin plate of specimen is acquired [63]. It is possible to change the focused distance and obtain a stack of 2D images, each displaying different plate from specimen. The volumetric (3D) image is assessed in this way as a set of 2D images, each with different z coordinate.

1.1.3 The difficulties of input data processing

The typical content of acquired data can be divided into two categories. We deal either with high-intensity well-separable objects, e.g. genes or nuclei, or low-intensity poor-quality objects hampered by noise, e.g. histons or telomeres. The first category is usually well segmentable. Moreover, small objects can be represented by points well which tempts to make use of point-based registration methods.

The second category poses an issue. While both categories are affected by noise, this category, in addition, contains objects with blending unsharp edges or tight clusters of objects. Such images are difficult to segment even manually and, consequently, it is difficult to retrieve the representing point set. Because of that, point-based methods cannot be used reliably. Nevertheless, there exist another approaches which may be more successful. However, these still hardly determine correspondence at the objects level within the cluster.

The occlusion of objects hardly occurs in 3D images. We also don't have to cope with transparent objects which get in front of the observed ones. Occasionally, some objects disappear from image for just one particular time step within the entire observation time period. Objects can appear suddenly in the time sequence. They can also disappear

entirely from the time-lapse observation. Furthermore, objects can split or join which makes the correspondence establishment yet more complex.

Movement of all objects from one time step to another doesn't have to be necessarily rigid, i.e. all objects do not have to move uniformly according to the same (even complex) transformation.

1.2 The aim of the thesis

1.2.1 Segmentation using temporal information

The classical image preprocessing sequence, i.e. noise suppression by filtering and segmentation by thresholding [44], introduces errors in segmentation or even fails completely. Over-segmentation or no segmentation of some objects as well as extraction of two — for human observer obvious — objects as one is the most typical erroneous behavior of this approach. Therefore, I would like to avoid the "classical" image preprocessing [55, 69]. Even in spite of the fact that data is corrupted by noise, I would still like to work directly with input data. I would also like to develop a method suitable for the second category — the category of worse segmentable objects whose images show low signal-to-noise ratio.

The segmentation will be conducted anyway. It is unavoidable whenever object labeling is in question. In my approach, I would like to make use of information gained from the temporal observation [37]. Whenever the delay between two consecutive time steps is reasonably short, the situation in the image is relatively the same. The information, gained from the repetition or similarity, shall support the segmentation procedure when deciding whether given voxel belongs to either object or background. The segmentation is expected to be realized using background subtraction technique. In this approach a temporal information is collected to setup statistical background model [53, 22]. The advantage of this approach is its tolerance to background value changes due to noise.

A similar approach is taken by methods used in video surveillance applications. Tracking of moving objects in poor quality and low-rate sampled video is typical task in this field. The nature of video surveillance data resembles biological data presented in the Laboratory of Optical Microscopy which provides me with testing data. The resemblance is mainly due to the non-static nature of background. Also, an object is mainly recognized due to its movement and different texture [40, 12]. Whereas in classical segmentation, segmentation from single frame, the object is identified as cluster of voxels having roughly the same intensity different from its surrounding or as cluster bounded by detected edge. Nevertheless, the edge is detected as rapid change of intensity that brings us back to intensity related segmentation from single time step data.

1.2.2 Correspondence establishment

The segmented data is processed using temporal approach inspired by video surveillance applications [9, 59]. The method will be modified or enhanced for the need of biological data. The idea is to first detect type of motion. Second, according to the type of

motion we get the motion model and we determine its parameters resulting in tracking of given segmented object. The object tracking, from which it is possible to establish correspondence, is claimed to be able of real-time processing [19]. The correspondence is then determined easily when the object is tracked successfully.

The graph representation will be constructed in the final step in which the correspondence assessment should be completed. Each segmented object within the entire time-lapse data will comprise a graph vertex. Note that the same object will be presented several times in the graph. A feature vector will be attached to each vertex. Vector elements should hold at least the spatial coordinates of the object center of mass, time step at which this object was segmented and some other supplementary information, for example the volume of object, main axis direction etc. Initially, all vertices will be connected with each other. An edge should indicate the correspondence between connected objects (vertices). Edges not satisfying any from the set of criteria will be removed except those supported from tracking phase. An example of criterion is the distance between the positions at n-th and $(n + 1)$-th time step should not be significantly different from the distance between positions at $(n + 1)$-th and $(n + 2)$-th. Sort of similar criteria will be used as the one in the example. This technique resembles the scene labeling according to [69].

1.3   Key approaches in the field

The point-based registration is the most often encountered approach in literature for this kind of biological data. It utilizes rigid transformation and is reported to work satisfactory [50, 44]. The uniform movement of vast majority of points provides enough information to successful registration. In fact, the registration itself usually works reliably whenever reasonably large sets of points are available. A small number of points are allowed to join, divide, remove or emerge when registering two consecutive sets. The most demanding constraint is the feasibility of reliable segmentation and point set retrieval which is not always possible to achieve. Sometimes, just a certain feature of every object can be used to represent it, e.g. mass center of larger object.

Voxel-based registration methods provide an approach when reliable point sets retrieval is not available. Instead, these methods generally work directly on the raw or filtered image data which represents an advantage over the point-based methods. Nevertheless, object labeling has to be done somehow anyway for the purpose of correspondence finding. The registration is determined to be the best alignment by means of searching and evaluating the space of all reasonable alignments of both images to be registered. The evaluation should indicate the optimality of alignment. A corelation-like evaluation methods are usually used [73, 14]. The dimensionality of searched space is given by the number of expected transformation parameters. However, for complex transformation (4 or more parameters) the search becomes too demanding without any optimizing reductions of alignment space. The same holds whenever the image data is large, for instance more than $1000 \times 1000 \times 40$ pixels. Translation and/or scale transformation can be efficiently

computed in the Fourier domain [73, 14].

Optical flow technique is another relatively often used solution. It is not a typical registration method since the outcome of optical flow are not any parameters of expected transformation. Instead, the output of computation is a vector field, called flow field, computed for every time step of the time-lapse observation. A velocity vector is associated to each voxel in the flow field. The velocity vector should estimate the direction and the distance of the given voxel in the next time step image data. Since no camera movement is expected in optical microscopy, the optical field of background voxels is expected to be close to null vector. The movement of certain object is detected in the flow field provided the object has some non-constant texture and the constant brightness assumption is met [33]. From the flow field we can track every moving object and estimate the correspondence in this way. The non-moving objects are simply to track because they remained at their positions. Note again the presumption of available object labeling.

Since the optical flow technique, in fact, allows for tracking of objects, as opposite to registration, a separate movement of every object is detectable and represents no problem. It can also perform on raw or modified data as well as voxel-based registration methods. Moreover, implementation using temporal coherence exists [15] which may improve the estimation of coarse flow field. On the other hand, the parameter tuning for reliable flow field computation is rather complex. This technique is not often reported for this kind of image data. The issue of live cell study is in the amount and textural structure of moving objects resulting in so called coarse flow field.

Most of those methods are developed for 2D images. Extensions of these to 3D or special purpose methods for 3D also exist but their implementations are rather rare in comparison to the number of methods for 2D.

1.4 Advantages of the outcome of the work

The main goal of this work is to achieve a solution that, embedded into a computer program, would be easy to handle while robust enough. The biological data presents a great deal of diversity from the computer scientist point of view. Thus, virtually every reasonable method must have parameters that must be tuned for particular type of image data, for particular biological experiment. The tuning of proposed method should be easy and the method itself should work fast.

Expected operating personnel interaction prior to computer processed analysis should only consist of the selection of small background area for the purpose of background model parameters determination. I believe that there is always sufficiently large area within the entire time-lapse data containing solely background. Moreover, no special knowledge is required for selection of background from image which is advantageous.

The method is also expected to be rather general. The design of method permits its application on the image data from the first category, according to section 1.1.3, while the method being originally developed for the second category. In this way, it should be usable for vast of biological data produced by Laboratory of Optical Microscopy.

Chapter 2

State of the art

2.1 Preface

I will briefly discuss some thoughts and ideas that lead me to the selection of studied areas in the field and consequently to the selection of sections that shall follow.

The ability to extract and label objects in the data is crucial in order to be able even to use the term "correspondence". Otherwise, there are simply no items among which the correspondence can be determined. In both registration and tracking methods, which leads to correspondence, objects must be described somehow, some object properties must be retrieved. This information is then helpful when estimating the correspondence itself. Thus, object data must be specified, segmented, in order to be able to compute the property values.

Human observer usually can find the objects of interest in the image data quite easily. Nevertheless, this is often highly non-trivial task for the computer. Even sophisticated segmentation methods fail whenever the input data change slightly. More and more segmentation methods emerge as new data kinds are available. Sadly, this is also true for the situation in live cell studies where, for example, a change of fluorescence marker may change the visibility of marked object resulting in different kind of image data from the computer scientist's point of view.

A pitfall of most segmentation methods is that they use single image for processing that image. See [55, 69, 39] and [64] for review of classical segmentation and image preprocessing algorithms at all. Only the information present in the processed image is considered. The nature of testing data doesn't seem to provide enough information in spatial domain that can be used for reliable segmentation. More information is possible and yet easy to gain by making use of spatio-temporal domain, i.e. by employing the temporal information. However, I believe that the temporal information can be useful only when the delay between consecutive acquisition time steps is reasonably short. This should ensure the continuity of temporal information. In other words, the change in spatial arrangement should not be too dramatic to make the temporal aggregation completely useless.

The correspondence can be established in several different ways with manual deter-

mination of relation between labels among many other approaches. I will definitely focus on the automatic ones. However, the point I would like to emphasize at the moment is that all such techniques must take into account the data from at least two different time steps, typically two consecutive time steps.

Exactly two time steps are used by basic registration methods and optical flow. Object tracking based on the movement prediction, on the other hand, uses quite a few consecutive time steps. Enhancements of basic registration and optical flow methods may, of course, make use of more time steps too. I would like to make use of more consecutive time steps because of the same reasons described in the previous text.

I will focus on the following topics in the following sections: temporal based segmentation methods, "classical" registration techniques (they are still used successfully) and temporal based methods that allow for correspondence retrieval. The boundaries of following chapters are not strict in the sense that the ideas, and sometimes even entire methods, may be shared by more sections. This may, for example, result in a segmentation which also performs object tracking. This is sometimes compared to the chicken and egg dilemma in the literature.

2.2 Segmentation from motion

If the time-lapse series of images is not considered from the point of view of information theory, as it was in the previous section, but from the field of video processing, then the sequence of images may represent a record of a particular object movement. Object properties like volume or roundness can be extracted from just the information presented in image data from single time step (a single frame) — spatial information. I believe that the only properties we may get from the spatio-temporal information, in addition to those from spatial domain, are those characterizing the selected object's movement itself. Of course, the same holds for everything derived from the movement characterization — properties inherently requiring the knowledge of motion parameters. This should justify the heading of this section.

Note that all this techniques describe, or just detect, moving objects. Objects preserving the same coordinate during the time-lapse observation must be detected, if required by the biological experiment, using some other technique. Aggregation of image data over the time coordinate might be advantageous since it might improve the signal to noise ratio.

The review paper [72] focuses on the segmentation methods which should extract semantically meaningful moving objects. Several categorizations of methods are possible. In the paper the division into two more general groups was outlined: motion-based and spatio-temporal methods. Motion-based methods typically proceed in the following three steps. A movement stable features (corners, lines, regions etc.) or individual pixels pose data elements and has to be determined first. Secondly, a motion model and estimation of motion is suggested. In the third step the segmentation is performed by clustering those data elements that act similarly according to motion model. The last two steps are

usually repeated iteratively to improve the result. In the case of 2D images, the most often used motion representation is the optical flow or just the detection of change of given data element over the time. See [12] for an overview of motion models suitable for video surveillance applications. Spectral domain may be used for representation too [70].

Spatio-temporal methods try to overcome the over-segmentation problem in image segmentation and overcome the sensitivity to noise and inaccuracy of motion-based segmentation. These methods perform the motion-based segmentation. Then the "classical" segmentation is conducted to rectify and improve the segmentation results. In spite of the fact that this approach could be more powerful, according to [72], it is not often used. A combination of spatio-temporal segmentation and tracking of segmented regions improved by graph labeling was reported [24].

A combination of both approaches is demonstrated in [30] where color images of traffic scenes were examined. At the very first, pixels were grouped in position-color space and each group was segmented. Segmentation of following frames was done similarly under the supervision of the results from previous frame. A temporal coherence was maintained in this way.

2.2.1 Motion detection

A simple background subtraction from reference background image may be used to detect regions where motion occurs [3, 19]. This mostly works for situations where the background is truly static, i.e. a wall in a room. A differencing between two consecutive frames can be used too [72, 18].

A more sophisticated approach, based on the same principle, is estimating the background model to serve as a reference image. Followed by subtraction of examined image poses quite a favorite technique used for representation of motion in motion-based segmentation. Single Gaussian or mixture of Gaussians is typically used to predict value of particular pixel, each pixels has its own probability function [22, 61, 60, 18].

Normal flow field can be used to detect independent motion even when the camera undergoes arbitrary motion [53]. In their implementation, each pixel was associated with spatio-temporal gradient vector in the normal flow field. Optical flow field was computed for the reference scene that contained no objects that would be detected later. A residual flow, that is the pixel-wise vector difference between the normal flow vector of tested image and optical flow vector of reference image projected onto the normal vector flow of tested image, was used to detect motion. The temporal coherence was also maintained in their implementation. An advantage over intensity differencing is that it is less susceptible to noise and errors in reference flow field estimation.

2.3 Registration techniques

Image registration, in general, is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors

[73, 10]. In the case of live cell study, an optical microscopy respectively, the definition can be narrowed to matching same scene taken at different times only.

The outcome of registration are the parameters of expected geometric transformation (relationship) paring as much corresponding pixels as possible [14]. The ability of transformation to conform the real relationship of registered data is measured by the number of degrees of freedom. Generally, the higher the number is the more complex relationships it can describe and also the more time demanding the computation is. The same holds for dimensionality of examined data, the more data to be aligned the more time is required.

A correspondence can be simply retrieved from aligned image data whenever all objects undergo a global motion. Given object should be at the same place (coordinates) in the next time-step image data after the alignment as it was in the previous time-step image data. Objects are, therefore, expected to move uniformly resulting in some global motion. A confidence is often maintained by ensuring that the feature vector of object and its aligned counterpart object are almost the same.

There exist several approaches to problem of registration [73, 10, 11, 20, 14, 65]. Two basic groups are the pixel-based and voxel-based registration techniques acting on spatial domain of image data. There exists a Fourier-based approaches to registration which I haven't studied and, hence, I'm omitting them from this review.

Recently, a data distributed parallel approach for image registration was reported. Ino et al. [35] reported registration of two 3D images where non-rigid transformation was searched to be speeded up from 15h on single CPU system to less than 10min on cluster of 127 CPUs. Furthermore, even a toolkit utilizing data parallelism for GRID-enabled image analysis applications was developed [29].

2.3.1 Transformation models

Transformation model should *only* reflect the motion of the entire specimen [10], for example when the entire cell shifts or rotates. The registration should, using the chosen transformation, just rectify the image data to be compatible with the previous image data. In live cell studies, in optical microscopy respectively, the settings of optical instrument doesn't change for each particular pseudo-color channel throughout the entire acquisition. Hence, the global transformation accounting for translation and rotation is appropriate.

Assume 3D image data. The reference image data point is at the coordinate $X = (x, y, z)^T$ while registered image data point is at *registered image's* coordinate X'. The transformation given by rotation matrix R and translation vector T relates the coordinate systems of both images:

$$X = R \cdot X' + T.$$

2.3.2 Point-based registration

Point-based registration methods rely on extraction of transformation invariant features. Both image data, that are subject to registration, are represented as two sets of points. The goal is to establish all transformation parameters so that maximum number of points

from reference set fit its counterpart in the registered set [73]. Two points are said to create a pair in such case. The idea behind is that global movement of points is assumed and, hence, most of points fit closely to its counterpart except those performing their own independent motion. Besides the noise, hampering feature extraction, and not quite exactly global movement of points, it is more usual to minimize the sum of distances between points that are supposed to comprise a pair instead of counting pairs.

Basically, the best alignment is mostly determined using clustering technique. All possible subsets of input sets of feature points are examined. The parameters of expected transformation are determined from the subsets which requires the size of subsets to be equal to the number of degrees of freedom to determine all parameters. Parameters form a vector that is stored in the parameter space. Sometimes the parameters are evaluated and those, which evaluation does not exceed given threshold, are discarded from further processing. For instance, circularity is computed from the estimated parameters of affine transformation in [11] and those having less than 0.9 are rejected. After all possible subsets are processed, the biggest cluster of parameter vectors is detected and the most appropriate alignment is determined from the center of that cluster.

A similar approach was taken by [13] when registering 2D images. A pair from point sets is selected and the matching pairs support is computed. This support claims the number of pairs that are possible to establish under the same parameters of given transformation when the selected pair remains fixed. The support is again computed in a clustering fashion as outlined above. The proper alignment is finally determined from pairs, that were established when maximum number was realized, by least squares technique. The extension of algorithm for 3D images was reported in [20] and made fasten in [50]. Note that evaluation is made again according to the number of matched pairs instead of evaluation the quality of transformation parameters which is further supported by the number of similar parameters.

A least-squares fitting determines the transformation parameters from already established pairs. Fortunately, the closed-form solution of least-squares fitting is established even for different representations of transformation [31, 32, 4] since establishing of pairs themselves represents already quite a computational cost. A comparison of those closed-form solutions is given in [46].

An iterative approach was taken by Besl when introducing an iterative closest point algorithm [8]. For each point in reference set find the closest point in the registered set and establish a pairing in this way. Determine the transformation parameters according to [31]. Apply the transformation on all points from reference set and leave there new values (transformed ones). Iteratively repeat the process by finding the closest point from reference set to every point in newly defined registered set. The iteration stops whenever the difference between values of last two least-squares sum is below given threshold.

To conclude the point-based registration I would like to note that the advantage of this methodology is that correspondence is established simultaneously with transformation parameters. In fact, when movement characterization of each object is the goal of experiment, a transformation parameters are byproduct of the process. The correspondence is central objective that enables computation of the movement characterization.

That might also be, perhaps, why the point-based methods are more often used in comparison to other approaches — to support this claim, count the respective references in a recent survey, for example in [73].

2.3.3 Voxel-based registration

Raw or preprocessed image data is aligned by voxel-based registration methods. Data are aligned regardless the actual objects' distances meaning that no image understanding or object extraction is performed. Correspondence is eventually established on the aligned data set.

The most appropriate alignment is determined by searching the space of translation parameters. A sub-space is scanned and all reasonable alignments are evaluated. The robustness depends mainly on the chosen evaluation method. However, inappropriate optimization may mislead the alignment search since the aim of optimization is to reduce the amount of tested alignments until the best one is found. The searching speed depends on the optimization techniques used, amount of processed data and selection of evaluation algorithm.

Both most frequently used evaluation methods and optimization techniques suitable for voxel-based registration are listed in [11]. The registration time consumption greatly depends on the number of tested alignments which can be reduced [65]. The most important factor is the smoothness of evaluation of neighboring alignments.

The correlation-like evaluations are frequently noted in the literature. The most often used, or its variants, are the sum of sign change, the sum of absolute valued differences and normalized correlation coefficient [11]. Quite a similar one is the correlation ratio [58, 57] which estimates, based on the total variance theorem and orthogonality principle, the arbitrary functional dependence of registered data on reference data.

A robust evaluation method based on information theory is the mutual information [14, 73] for registration purposes originally proposed by Viola [68]. Since the mutual information estimates entropies that, in turn, require probability density estimations, an approach using Parzen-window estimator [67] or estimation of densities from histograms [71] can be employed.

Another approach was recently reported [49, 21]. The image data can be imagined as density function, image points are clustered according to density based classification methods, clusters are stored into a confinement (density) tree representation of the image [48]. The shape and topology represented by such graph structure is then matched onto another image graph representation resulting in registration of those two images. Thus, the method allows for non-rigid movement of objects within images.

The Hu's legendary article [34] introduced moment based theory into image shape recognition. The ability of moments to evaluate geometrical transformations, sufficient for needs of optical microscopy [52], from shape characterizations was also reported [34, 47]. A typical approach for transformations, composed from translation, rotation and scaling, is to determine center of mass and tensors [39] (Chapter 7) for both consecutive object's image data. A singular value decomposition and singular value and vector ex-

traction (mathematical background can be found in [27]) is conducted from both moment characterizations. The transformation parameters can then be determined.

A pitfall of such approach is that it is applicable to just one object's image data. It can register only between corresponding objects allowing for local rectification of registered data. Without a priori correspondence knowledge the method capabilities are limited to image data displaying at most one object. Nevertheless, this is not of importance since there are experiments imaging a single nucleus in the whole time-lapse series [56].

2.4   Tracking techniques

Tracking of an object needs to determine all locations of this particular object within a time-lapse observation. I will refer to a *frame* as to a particular time step image data. Either all objects are segmented in every frame and the tracked one is matched by some of its properties or a technique of movement prediction and correction is employed.

A correspondence can be established using tracking of selected objects. In fact, the essence of tracking involves the frame-to-frame correspondence establishment of tracked object. Furthermore, a complete correspondence information within an image data sequence can be retrieved by tracking all moving (and also stationary) objects independently [37, 36, 66, 3, 60, 61]. An advantage over image registration methods is that no transformation is used. This approach is therefore capable of handling an independent local movement as well as global movement of all objects in the data.

In the following sections, I will focus on two techniques which allows for realization of tracking multiple objects. Moreover, both techniques can work directly on raw data which is in accordance with ideas presented in introduction.

2.4.1   Optical flow

The aim of optical flow technique is to estimate the motion field. The motion field consists of velocity vectors characterizing, from the observer position, the true movement of respective pixel or larger element when some granularity or pyramidal approach is used. The output of optical flow algorithm is a flow field. A flow field is the same as motion field if and only if a special constrains are held [69].

The tracking, the correspondence consequently, can be easily realized as long as all objects of interest can be extracted in every frame. The objects relation between consecutive frames can be determined from the flow field because the velocity vector from given frame estimate the position of attached pixel in the next frame.

The idea of optical flow was first introduced in early 80' [33]. Although many specializations and variant approaches exist, still the basic idea is always preserved. The optical flow computation is based on the expectation that the observed brightness of any point of the object is constant over time. Thus, for 2D images, we can relate pixel intensities by

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

where I is the intensity function of pixel at spatio-temporal position (x, y, t), t is the time step. The $(\delta x, \delta y)$ is spatial displacement occurred between two given frames, δt is the delay between frames. By letting

$$u = \frac{\delta x}{\delta t}, v = \frac{\delta y}{\delta t}$$

we can estimate the velocity (u, v). So called *optical flow constraint* equation

$$\nabla I \cdot (x, y) + I_t = 0,$$

where ∇I is spatial gradient vector, $\cdot$ is scalar product and I_t is temporal derivation, can be derived from that. Nevertheless, optical flow constraint is just one equation with two unknowns. Flow vector cannon be therefore determined by using only this equation at just one pixel. That is often termed as aperture problem in literature.

Horn and Schunck, in their paper [33], required the flow field to be smooth and developed an iterative scheme that allowed to compute the (u, v) for every pixel. That remained a classical approach even until these days. Nevertheless, a temporal coherence constraint demanding the flow smoothness between several consecutive frames was developed over the Horn and Schunck scheme [15]. However, such criteria prohibits the sudden change of velocity which might occur in live cell studies — observed object performs kind of shaky movement or, more often, several independently moving objects are located close to each other.

A more adaptable, while locally smoothness preserving, variant may be an accumulator scheme explained in [51]. Since at least two optical flow constraint equations are required in order to estimate the velocity, all two-pixels combinations from a small neighborhood are examined and each velocity is stored. Resulting velocity is estimated as the center of the largest cluster of velocities. Basically the same idea is employed in [17] where instead of accumulator a multi-modal probability density is employed [16].

Major drawback of methods above is that δt should be small such that a spatial pixel displacement between consecutive frames is less than one pixel. It is caused by the Taylor development, that is used in derivation of optical flow constraint equation, where higher order terms are neglected. Sometimes, a multiscale approach is taken where the solution from low-resolution level is used in higher-resolution level. However, this and the selection of second constraint pose the most difficulties with computing a reliable optical flow. Moreover, the extension to 3D images is not straightforward.

More variants exist for computation of optical flow. An acknowledged review [7] presents an overview of possible approaches while [6] compares the performance of few techniques. According to these, previous methods are termed as *gradient-based* methods. Also, they are the most used ones.

A *correlation-based* approach belongs among the most used too. Recently a block matching scheme for estimation of volume movement was reported [45]. The 3D image is partitioned into regular-sized blocks. Each block in given frame is searched in its neighborhood in the following frame using some voxel-based registration technique. A similar approach was taken by Adorni et al. [2]. They labeled each block with one of 16

labels and searched for the same label in the next frame. The density of flow field can be regulated by the size of block. The simple extension to 3D images is balanced by the accuracy.

Methods working either indirectly or directly in the Fourier domain of spatio-temporal data (like above the image function I) are so called *frequency-based* methods. Indirect branch is mostly presented by Gabor filters [1]. A local region is filtered by set of differently tuned Gabor filters. The motion parameters are estimated from the parameters of filter with the biggest response. Direct approach for velocity estimation stems from the eigen value analysis of inertia tensor of spectral energy of spatio-temporal image data [38].

2.4.2 Tracking in video surveillance applications

Techniques presented in this section originally belong to the field of human tracking and motion recognition that is mostly used for video surveillance. The video data shares some similar characterizations as live cell studies data acquired from optical microscope. Moreover, human movement is highly non-rigid and complex so that methods designed for tracking of humans and recognition of their motion should be, theoretically, able to handle the movement and behavior of cell or intracellular structures. See introduction for discussion of this matter.

Recent papers [61, 62] make extensive use of extended Kalman filter [5] to estimate the position of selected object in the next frame. It is also possible to be following multiple objects within the time-lapse series image data [60, 61]. All procedures suggested in given papers are also capable of segmentation of moving objects. The segmentation is always a variant of background subtraction technique (see section 2.2.1).

Note that the correspondence is easily determined when a data set from tracking phase is available.

2.4.3 Moment-based recognition

Recognition of complex non-rigid motions by template matching of spatio-temporal information, compiled into a feature vector of image moments statistics, is a relatively recently developed technique that is capable of real-time processing of low-bandwidth low-rate grayscale video sequences [19].

First impression, that a relatively little information is actually needed to recognize a human motion, was published by Johansson in 1973 [40]. An audience proved to recognize human actions of an actor in a dark room very quickly while only a few lights attached to actor's head, arms and legs were visible. An automatic system for recognition of repetitive human motions (such as gait) was reported in [54]. The recognition was based on matching the spatio-temporal low-level image features against a few templates of repetitive human actions. This is termed as *view-based* approach in opposite to *model-based* approach where structure is detected (such as arms and legs) and compared to a specific motion model.

Further enhancements allow for recognition of arbitrary movement [9, 19]. Video sequences of template human action are converted into MHIs (motion history images).

Each MHI is converted into a feature vector in which elements are Hu's image moments [34] computed from MHI. Such feature vector describes spatio-temporal characteristics of given motion. The input time-lapse image data is processed as described and using Mahalanobis distance compared to templates of know actions [19].

Several statistical classifiers for template matching of MHI-generated vectors are tested in [59]. Also, a hierarchical MHIs can be computed to estimate motion flow for 2D images [18].

For the correspondence establishment, the basic idea is to extract all objects under interest from the first frame. Classify a type of motion performed by given object in a reasonably large local surrounding. The recognized type of motion serves as guideline to determination of objects correspondence. This technique should be mainly helpful in resolving special situations occurring in live cell studies, for example a division of object should be apprehended by template matching well and, hence, the correspondence should be established correctly.

2.5   Scene labeling

Scene labeling belongs to the field of image understanding. Originally, the goal is to label all segmented objects in a way such that the semantical meaning of labeling achieves an appropriate image interpretation.

A graph representation is constructed. All segmented objects are associated with single vertex in a region adjacency graph of that scene (image). Vertices also include object's features (e.g. volume of given object) while edges contain features regarding two adjacent objects (e.g. a distance between centers of mass of two objects). A labeling should be consistent which means it should be valid according to the set of many rules.

Since the size of input data to scene labeling algorithm is expected to remain rather small and, hence, the computation demand reasonable low, I haven't paid lot of attention to this field. According to [69], I just cite a basic literature to relaxation labeling: [41, 28]. A relaxation labeling allows only one label to be assigned to each object as opposite to probabilistic labeling where multiple labels are allowed and are probabilistically weighted.

Bibliography

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *journal of the optical society of America A*, 2(2):284–299, 1985.

[2] G. Adorni, S. Cagnoni, and M. Mordonini. Cellular automata-based optical flow computation for "just-in-time" applications. *Proceedings. International Conference on Image Analysis and Processing*, 1999.

[3] M. B. H. Ali, M. Ebbecke, and A. Dengel. Real time object detection, tracking and classification in monocular image sequences of road traffic scenes. In *ICIP97*, 1997.

[4] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, 1987.

[5] A. Azarbayejani and A. P. Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(6):562–575, 1995.

[6] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Int. J. Comput. Vision*, 12(1):43–77, 1994.

[7] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Comput. Surv.*, 27(3):433–466, 1995.

[8] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992.

[9] A. Bobick and J. Davis. An appearance-based representation of action. In *ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition (ICPR '96) Volume I*, page 307. IEEE Computer Society, 1996.

[10] L. G. Brown. A survey of image registration techniques. Technical report, Columbia University, January 1992.

[11] M. Čapek. *Registrace snímků z konfokálního mikroskopu*. Disertační práce, Czech Technical University in Prague, 1999.

[12] C. Cédras and M. A. Shah. Motion based recognition: A survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[13] S. H. Chang, F. H. Cheng, W. H. Hsu, and G. Z. Wu. Fast algorithm for point pattern-matching: Invariant to translations, rotations and scale changes. *PR*, 30(2):311–320, February 1997.

[14] Q. S. Chen. *Image Registration and its Applications in Medical Imaging*. PhD thesis, Faculty of Applied Sciences, Vrije Universiteit Brussel, 1993.

[15] T. M. Chin, W. C. Karl, and A. S. Willsky. Probabilistic and sequential computation of optical flow usingtemporal coherence. *IEEE Transactions on Image Processing*, 3(6), 1994.

[16] W. Clocksin. A new method for estimating optical flow. Technical report, Computer Laboratory, University of Cambridge, 1997.

[17] W. Clocksin. A new method for computing optical flow. In *BMVC00*, 2000.

[18] J. W. Davis. Hierarchical motion history images for recognizing human motion. In *IEEE Workshop on Detection and Recognition of Events in Video*, pages 39–46, 2001.

[19] J. W. Davis and A. F. Bobick. The representation and recognition of human movement using temporal templates. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 928. IEEE Computer Society, 1997.

[20] V. Dvořák. Párování bodových vzorů v 3D prostoru. Master's thesis, Faculty of Informatics, Masaryk University, 2005.

[21] R. Eils and Ch. Athale. Computational imaging in cell biology. *The Journal of Cell Biology*, 161:447–481, 2003.

[22] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. *6th European Conference on Computer Vision*, 2000.

[23] Y. Garini, A. Gil, I. Bar-Am, D. Cabib, and N. Katzir. Signal to noise analysis of multiple color fluorescence imaging microscopy. *Cytometry*, pages 214–226, 1999.

[24] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *PR*, 33(4):725–740, April 2000.

[25] D. Gerlich, J. Mattes, and R. Eils. Quantitative motion analysis and visualization of cellular structures. *Methods*, 29(1):3–13, 2003.

[26] J.M. Geusebroek, F. Cornelissen, A. W.M. Smeulders, and H. Geerts. Robust auto-focusing in microscopy. *Cytometry*, pages 1–9, 1999.

[27] G.H. Golub and Ch.F. van Loan. *Matrix computations*. John Hopkins University Press, 1993.

[28] E. R. Hancock and J. Kittler. Discrete relaxation. *Pattern Recogn.*, 23(7):711–733, 1990.

[29] S. Hastings, T. Kurc, S. Langella, U. Catalyurek, T. Pan, and J. Saltz. Image processing or the grid: A toolkit or building grid-enabled image processing applications. In *CCGRID '03: Proceedings of the 3st International Symposium on Cluster Computing and the Grid*, page 36. IEEE Computer Society, 2003.

[30] B. Heisele. Motion-based object detection and tracking in color image sequences. In *Fourth Asian Conference on Computer Vision*, pages 1028–1033, 2000.

[31] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629 – 642, 1987.

[32] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *journal of the optical society of America A*, 5, 1988.

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

[34] M. K. Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, pages 179–187, 1962.

[35] F. Ino, K. Ooyama, and K. Hagihara. A data distributed parallel algorithm for nonrigid image registration. *Parallel Comput.*, 31(1):19–43, 2005.

[36] M. Irani and P. Anandan. A unified approach to moving object detection in 2d and 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(6):577–589, 1998.

[37] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV '92: Proceedings of the Second European Conference on Computer Vision*, pages 282–287. Springer-Verlag, 1992.

[38] B. Jähne. Motion determination in space-time images. In *ECCV '90: Proceedings of the First European Conference on Computer Vision*, pages 161–173. Springer-Verlag, 1990.

[39] B. Jähne. *digital image processing*. Springer-Verlag, 1993.

[40] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[41] J. Kittler and J. Illingworth. Relaxation labelling algorithms-a review. *Image Vision Comput.*, 3(4):206–216, 1986.

[42] M. Kozubek. *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances*, chapter FISH imaging, pages 389–429. Wiley-Liss, Inc., 2002.

[43] M. Kozubek and Pe. Matula. An efficient algorithm for measurement and correction of chromatic aberrations in fluorescence microscopy. *Journal of Microscopy*, pages 206–217, 2000.

[44] M. Kozubek, Pe. Matula, Pa. Matula, and S. Kozubek. Automated acquisition and processing of multidimensional image data in confocal in vivo microscopy. *Microscopy Research and Technique*, 64:164–175, 2004.

[45] W. De Leeuw and R. van Liere. Bm3d: Motion estimation in time dependent volume data. In *Proceedings IEEE Visualization 2002*, pages 427–434, 2002.

[46] A. Lorusso, D. Eggert, and R. Fisher. A comparison of four algorithms for estimating 3-d rigid transformations. In *In British Machine Vision Conference*, pages 237–246, 1995.

[47] S. Maitra. Moment invariants. *Proc. of the IEEE*, 67:697–699, 1979.

[48] J. Mattes and J. Demongeot. Tree representation and implicit tree matching for a coarse to fine image matching algorithm. In *MICCAI '99: Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 646–655. Springer-Verlag, 1999.

[49] J. Mattes, J. Fieres, J. Beaudouin, D. Gerlich, J. Ellenberg, and R. Eils. New tools for visualization and quantification in dynamic processes: Application to the nuclear envelope dynamics during mitosis. In *MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1323–1325. Springer-Verlag, 2001.

[50] Pe. Matula, Pa. Matula, and M. Kozubek. Fast point-based 3D alignment of cells in *in vivo* imaging. *IEEE Transactions on IMAGE PROCESSING*, 2005. submitted to journal.

[51] A. Otero. Robust optical flow estimation. In *ICIIS '99: Proceedings of the 1999 International Conference on Information Intelligence and Systems*, page 370. IEEE Computer Society, 1999.

[52] S. Ch. Pei and L. G. Liou. Using moments to acquire the motion parameters of a deformable object without correspondences. *Image Vision Computing*, 12(8):475–485, 1994.

[53] R. Pless, T. Brodský, and Y. Aloimonos. Detecting independent motion: The statistics of temporal continuity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:768–773, 2000.

[54] R. Polana and R. C. Nelson. Low level recognition of human motion (or how to get your man without finding his body parts). In *IEEE Computer Society Workshop on Motion of Nonrigid and Articulate Objects*, 1994.

[55] W. K. Pratt. *Digital Image Processing*. Wiley, 1991. ISBN: 0-471-37407-5.

[56] B. Rieger, C. Molenaar, R.W. Dirks, and L.J. van Vliet. Alignment of the cell nucleus from labeled proteins only for 4d in vivo imaging. *Microscopy Research and Technique*, 64:142–150, 2004.

[57] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proceedings MICCAI'98*, volume 1496 of LNCS. Springer Verlag, 1998.

[58] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. Technical report, Institut National De Recherche En Informatique Et En Automatique, 1998.

[59] R. Rosales. Recognition of human action using moment-based features. Technical report, Boston University, 1998.

[60] R. Rosales and S. Sclaroff. Improved tracking of multiple humans with trajectory prediction and occlusion modeling. In *Proc. IEEE Workshop on Interpretation of Visual Motion*, 1998.

[61] R. Rosales and S. Sclaroff. 3d trajectory recovery for tracking multiple objects and trajectory guided recognition of actions. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1999.

[62] R. Rosales and S. Sclaroff. A framework for heading-guided recognition of human activity. *Comput. Vis. Image Underst.*, 91(3):335–367, 2003.

[63] A. Sarti, C. Ortiz de Solorzano, S. Lockett, and R. Malladi. A geometric model for 3-d confocal image analysis. *IEEE Transactions on Biomedical Engineering*, 47:1600–1609, 2000.

[64] J. Serra. *Image Analysis and Mathematical Morphology*. Book, 1982.

[65] V. Ulman. Image composition. Master's thesis, Faculty of Informatics, Masaryk Univerzity, 2003.

[66] V. V. Vinod and H. Murase. Video shot analysis using efficient multiple object tracking. *IEEE International Conference on Multimedia Computing and Systems '97. Proceedings*, pages 501–508, 1997.

[67] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, pages 137–154, 1997.

[68] P. A. Viola. *Alignment by Maximization Of Mutual Information*. PhD thesis, Massachuttes Institute of Technology, 1995.

[69] M. Šonka, V. Hlaváč, and R. Boyle. *Image Processing: Analysis and Machine Vision.* O'Reilly, 1999.

[70] W. Yu, G. Sommer, and K. Daniilidis. Multiple motion analysis: in spatial or in spectral domain? *Comput. Vis. Image Underst.*, 90(2):129–152, 2003.

[71] M. Zaffalon and M. Hutter. Robust feature selection by mutual information distributions. Technical report, IDSIA Switzerland, June 2002.

[72] D. S. Zhang and G. Lu. Segmentation of moving objects in image sequence: a review. *Circuits, Systems and Signal Processing (Special Issue on Multimedia Communication Services)*, 20:143–183, 2001.

[73] B. Zitová and J. Flusser. Image registration methods: a survey. *IVC*, 21(11):977–1000, October 2003.

Chapter 3

Goals of the thesis

3.1 Objectives

The main goal of the work is to develop a solution for correspondence establishment of selected objects in image data acquired from optical microscope. The solution as well as its implementation should be reasonably fast, accurate, reliable and also simple to use.

I would like to reach these properties by solving the following objectives:

- Enhance segmentation methods and modify them for the purpose of image data in live cell studies.

- Focus on the methods that use temporal segmentation besides the "classical" segmentation methods like thresholding, for example.

  ──────────

- Enhance the object tracking methods used in video surveillance applications and modify them for the purpose of image data in live cell studies.

- Focus mainly on methods that use temporal integration.

- Consider the movement prediction methods.

  ──────────

- Test and eventually enhance and modify the graph labeling methods to improve the correspondence establishing process.

  ──────────

- Try to propose some algorithms for automatic setting of parameters that influence the behavior of methods mentioned in previous items of this list.

Practical outcome of the work should be in the form of published algorithms as well as computer program.

The suggested solution will be tested on the data supplied from Laboratory of Optical Microscopy or from its collaborating institutions. The results will be compared with key solutions of the field and discussed in the Ph.D. thesis.

3.2   Time schedule proposal

- Defence of this Ph.D. thesis proposal in January 2006.

- Implemented tracking and temporal segmentation algorithms until June 2006.

- First publishing of tracking, temporal segmentation and/or optical flow algorithms until December 2006.

- Testing and comparing proposed methods in 2nd half of the year 2006.

- Working on graph labeling and movement prediction techniques in 2nd half of the year 2006.

- Both Master degree theses, that I'm supervising, should be defended until September 2007.

- Submission of the final version of thesis until September 2007.

3.3   Result presentations

I'm expecting to have at least one presentation about the enhancements and modifications of either segmentation or object tracking methods based on temporal integration. This presentation should be on some major conference in the field.

Besides, presenting at least once per year on an international conference, some technical report papers and presentations on seminars or sessions on local events are also expected. I would like to present my own results as well as present results from the two Master degree theses that I am supervising.

Chapter 4

Current study progress

4.1   Completed Courses

I have been focusing at image processing courses during the last three semesters. Attended courses are summarized below:

- Graphics algorithms and signal processing using DSP and FPGA (PA162) — spring 2004

- Advanced Computer Graphics (PA010) — autumn 2004

- IT-related English conversation (VB038) — autumn 2004, spring 2005

- Bionformatics I (IV107) – spring 2005

- Advanced Methods of Digital Image Processing (PA166) — spring 2005

I also passed the State doctoral exam in April 2005. The examination was in the fields of Discrete mathematics and algebra, Numerical methods and Computer graphics.

4.2   Publications and Posters

- V. Ulman. Quantitative Analysis of Intracellular Processes in Living Cells. In *Proceedings of 1st International Summer School of Mathematical Biology* Brno: Masaryk University. **In print**.

- M. Kozubek, Pe. Matula, Pa. Matula, M. Vařecha, J. Amrichová, V. Ondřej, E. Lukašová, S. Kozubek, V. Ulman and P. Krontorád. Automated confocal in vivo microscopy based on spinning disks. **Presented as poster on Digitální Zobrazování v Biologii a Medicíně 2005**.

4.3 Presentations

- 1st International Summer School of Mathematical Biology, Brno, 4–5th September 2005, Czech Republic

- 3rd European Medical & Biological Engineering Conference, IFMBE European Conference on Biomedical Engineering, Prague, 20–25th November, Czech Republic

- Digitální Zobrazování v Biologii a Medicíně 2005, České Budějovice, 12th May 2005, Czech Republic

- Local seminars and sessions at the Faculty of Informatics

4.4 Other Activities

- Teaching practicals in Introduction to the C language (IB071) — spring 2005

- Teaching practicals in Programming in C++ (PB161) — autumn 2005

- Software development in Laboratory of Optical Microscopy (`i3dlib` library)

- Supervising two Master degree thesis:

- Optical flow in image data of human *in vivo* cells (supervisor).

- Visualizations of events in human living cells (co-supervisor).

- Providing consultations to one Master degree thesis:

- The use of methods of linear and natural numbers programming in analysis of image data.