

**Interdisciplinary Project**  
**Cell nucleus segmentation and classification**

Ulrich Klank

April 14, 2006

Technische Universität München  
Department of Computer Science  
Chair for Computer Aided Medical Procedures and Augmented Reality

[campar.cs.tum.edu](http://campar.cs.tum.edu)



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background Information</b>	<b>3</b>
2.1	Technical Background	3
2.1.1	Fluorescence in Situ Hybridization	3
2.1.2	Different FISH methods	3
2.1.3	Image Data	3
2.2	Medical Background	4
2.2.1	The Benefits of FISH	4
2.2.2	Complex Diagnoses	8
<b>3</b>	<b>Methods</b>	<b>9</b>
3.1	Abstract	9
3.2	Segmentation of the Cells	9
3.2.1	Overview	9
3.2.2	Threshold and Edge Detection	9
3.2.3	Ellipse Detection	10
3.2.4	Other Approaches	11
3.3	Detection of Signals	11
3.3.1	Overview	11
3.3.2	Design	12
3.3.3	Learning Algorithm	13
3.4	Association of Signals	14
<b>4</b>	<b>Results</b>	<b>15</b>
4.1	Program	15
4.2	Statistics	15
4.2.1	Cell Detection	15
4.2.2	Signal Detection	17
<b>5</b>	<b>Conclusion</b>	<b>23</b>
	<b>Bibliography</b>	<b>25</b>



# 1 Introduction

Although automatization in the medical field is in progress, many tasks are still executed by humans, even though these tasks are annoying or rather simple. An example for such a task, which, however, is not so simple, is the counting of light spots, later called signals, on images of fluorescence marked cells under a microscopic. This text will provide information about this technique, how to generate such images, and its diagnostic possibilities by counting the signals. An approach to automatize this counting task will also be described. In particular, explanations will be given about how to generate images of cells by means of hybridization methods, such as FISH. Afterwards the counting task will be discussed in more detail and the algorithm behind the developed program will be introduced. This approach is especially interesting in the part relating to the detection of the border of different cells in the images. Finally an analysis of the results of this algorithm is performed, including a statistical evaluation of the rate of success. Particular attention is directed to reliability which is a feature of of utmost importance for medical systems



## 2 Background Information

### 2.1 Technical Background

#### 2.1.1 Fluorescence in Situ Hybridization

The idea of Fluorescence in situ hybridization (FISH) will be described briefly. The term "hybridization" refers to the connection of two compatible DNA strings. In order to detect and localize DNA sequences in cells, FISH uses probes marked with fluorescent dyes, which hybridize to compatible parts of the chromosomes in the cells. In Figure 2.1 the preparation steps for FISH are sketched. First a sequence is labeled with a fluorescent dye, for example a 4'-6-Diamidino-2-phenylindole (DAPI) dye or a Cyanin 3 (Cy3) dye. A method for labeling the probes is the so called "standard nick translation". Then these prepared probes are mixed with the denaturated target DNA of the patient to perform the hybridization. A possibility to get a specimen to work with the patient's cells is to use paraffin embedded tissues. How to remove the paraffin and perform the hybridization is described in [4]. Afterwards the specimen is prepared to be examined by a microscope. For every fluorescence dye there exists an excitation and an emission wavelength. An epifluorescence microscope makes separated photographs of the different excited dyes.

#### 2.1.2 Different FISH methods

The traditional FISH method is able to detect DNA-sequences from a length of about 5 million base pairs (Mb), and it is used in the metaphase, during which the chromosomes are packed in the well known X form. By simply employing the same technique in the interphase, where chromosomes are widely distributed, the resolution can be augmented up to 50 kb - 2 Mb, referring to [6]. This so called Interphase FISH is normally used for 3D imaging techniques, like the samples used for this project. An even higher resolution can be reached with the fibre FISH technique. Here the chromatin is reduced to the pure DNA by removing the proteins from it. The resolution will range from 1 kb to 500 kb.

#### 2.1.3 Image Data

The images which are used for this project as sample data are generated with multicolor deconvolution microscopy, as it is described in [4]. This means, images of a specimen, e.g. one entire cell layer of ovarian tissue, are sequentially taken by a camera on a microscope. There are taken up to 50 pictures with different focal planes of one specimen, where the focal plane is changed for each picture by 250 nm. In every focal plane one image is taken for every fluorochrome used. Which fluorochrome probes are used depends on the task. For example in the images 2.3, 2.5 and 2.4 a so-called centromere set of probes is used. The counterstain, shown in the Figure 2.4, is used to estimate the shape of the cell nuclei in the

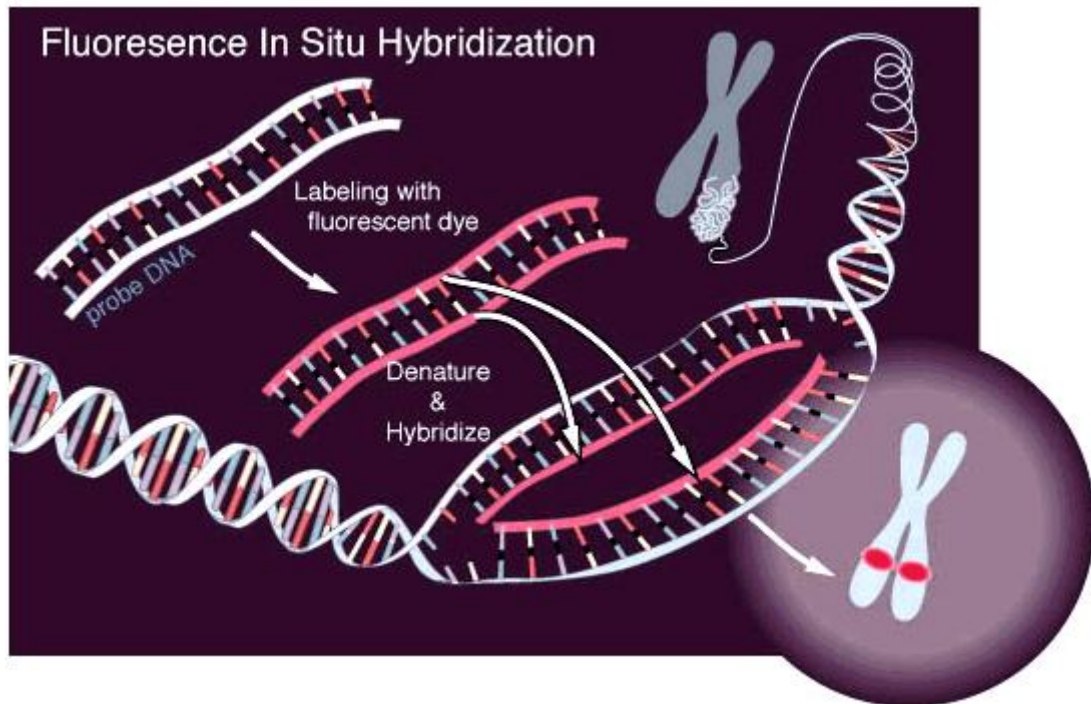


Figure 2.1: How to imagine fluorescence in situ hybridization

image. For the same specimen, Figure 2.5 shows a signal marking a special sequence of the DNA.

## 2.2 Medical Background

### 2.2.1 The Benefits of FISH

Mutation in cells is a major problem for the human body and therefore for human health in general. By monitoring the genetic material, many diseases can be predicted or explained. To keep up with the growing knowledge about the functionality of the human genes and the coherence between a mutation or other genetic disorders and a disease, it is necessary to find effective methods to use this knowledge and to help treating patients accordingly. An example for a benefit of the FISH technique is the diagnosis of a carrier of Duchenne Muscular Dystrophy (DMD), like described in [2]. Figure 2.6 shows an example for how the application of FISH supports a diagnosis. The most obvious symptom of the Duchenne syndrome is the lack of muscle force, caused by a malfunction of the dystrophy enzyme. This malfunction is caused by a deletion of a greater part of the dystrophy gene, which is located on the X chromosome. Since this effect is recessive, it normally does not affect women because of their second X chromosome. However, it is interesting to decide if a woman carries this gene defect or not, in order to predict the risk for their children. As the size of the part of the dystrophy gene which is mostly deleted by this disease is considerable, the difference of this affected gene to a FISH probe will be large enough to



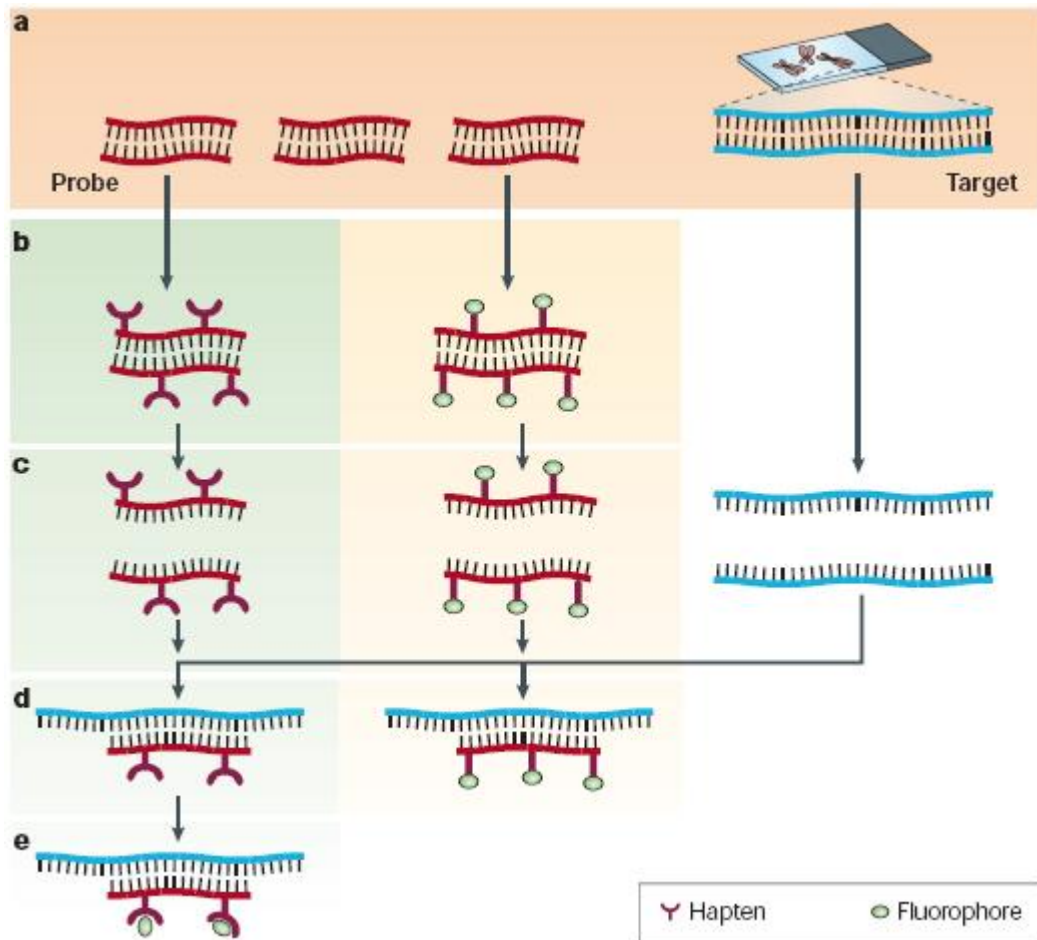


Figure 2.2: Another approach to explain FISH graphically, from [6]

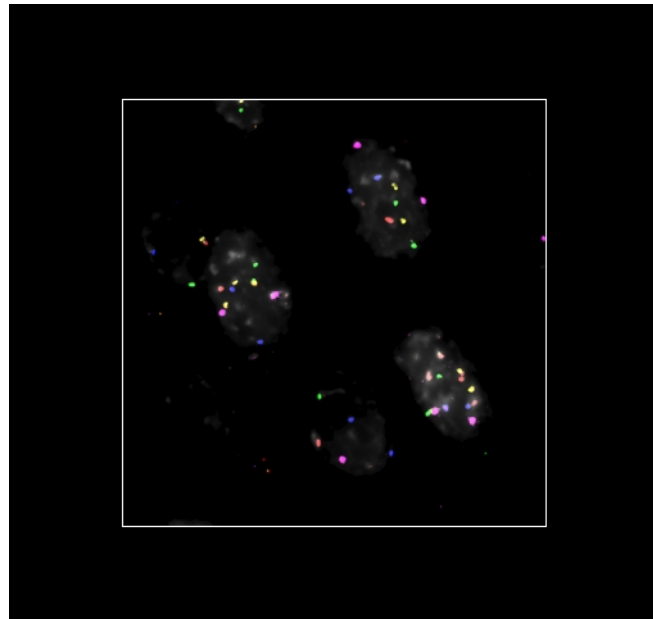


Figure 2.3: The overlay of all signals of a specimen in a image

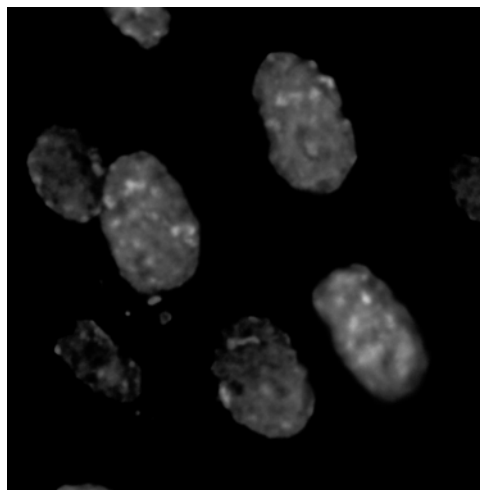


Figure 2.4: The counterstain in a central z-plane after deconvolution

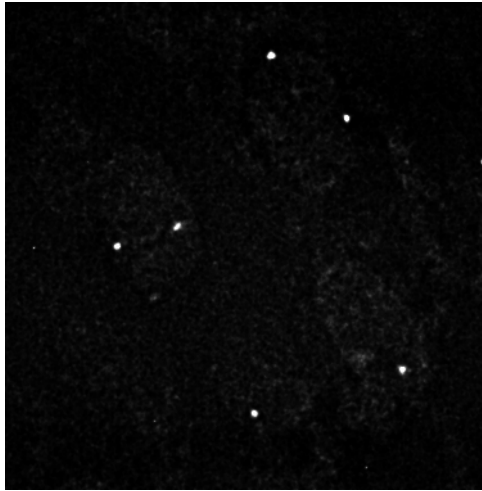


Figure 2.5: A central z-plane of a signal channel

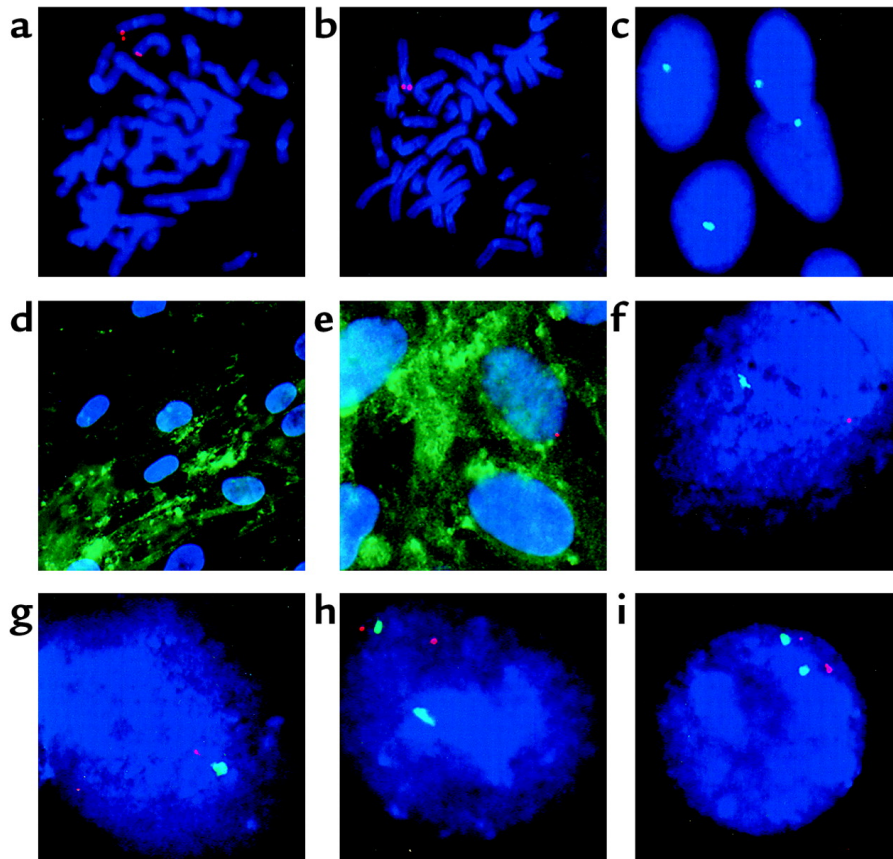


Figure 2.6: Images from [2], especially interesting are here the last four pictures (f,g,h,i) which show an example of how performing a diagnosis with fish images: images f and g are showing cells from a woman who is a carrier of the Duchenne syndrome, images h and i are of a not affected woman

exclude hybridization of the probe to the affected chromosome. By counting the resulting signals in a image of such a cell specimen, it can be decided if a woman is a carrier or not.

### 2.2.2 Complex Diagnoses

The most interesting thing about FISH and especially multiple FISH (mFISH) is that is possible to insert several different DNA probes in the same specimen. Hence, complex analyses, including tasks that depend on the relative location of different DNA sequences, can be executed. Complex analyses may include the detection of translocation between chromosomes, or multiple hit diseases, like solid tumors, that need normally more than one mutation in specific genes to break out. A specific example is the diagnosis of the malignancy of ovarian tumors in [4]. In a tumor cell, the number of chromosomes two cells will receive from their parental cell are distributed randomly and not by the expected fifty-fifty ratio, caused by already occurred mutations. These irregularities can be detected by a FISH analysis. While some other analyses will count only the number of chromosomes over all cells and obtain a nearly normal number of chromosomes, the FISH method detects the wrong spreading throughout the different cells. Most of the image samples analysed in this project contain multiple signals and are subject to a more complex diagnosis model.

## 3 Methods

### 3.1 Abstract

Solution of the counting task comprises the three main components: first the cell segmentation, then the signal detection and finally their association with the segmented cells. The signal detection is performed with a neural network and the cell segmentation by an ellipse detection. The association of the signals with the cells is conducted using the ellipse center location.

### 3.2 Segmentation of the Cells

#### 3.2.1 Overview

For segmentation of cells a kind of randomized Hough transformation is used. A similar approach is presented in [5]. In particular, for every image the edges are extracted with a threshold filter and a vertical and a horizontal edge filter. Afterwards, ellipse constructing points are selected randomly from the set of already extracted points, in order to create ellipses describing most of the extracted points.

#### 3.2.2 Threshold and Edge Detection

In order to reduce the number of points which have to be considered for the ellipse search, two steps are executed to select the most interesting points. First, the intensities are compared with the average intensity of current image. Each pixel with a intensity which is above the average intensity plus a constant value is saved. The best results were gained with a threshold of 9 over the average of the entire image. This specific value was determined on the base of a few sample images and was accepted as a good choice without further proof. For each stored pixel, instead of its intensity value, the value of 1 is saved, if its intensity is over the average intensity (which will always be the case for the stored pixel), or -1 in the other case. Still applying this condition, a value of 1 or -1 is stored for each of the eight pixels surrounding the saved pixel, in order to prepare the edge detection.

Using this information, a vertical and a horizontal edge filter is applied to the stored points to discard points that are not located on an edge. The horizontal filter characteristic for the point  $p_{x,y}$  is as follows :

$$c < \left\| \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} p_{x-1,y-1} & p_{x-1,y} & p_{x-1,y+1} \\ p_{x,y-1} & p_{x,y} & p_{x,y+1} \\ p_{x+1,y-1} & p_{x+1,y} & p_{x+1,y+1} \end{pmatrix} \right\|,$$

where  $c$  denotes an empirically chosen constant which indicates the presence of an edge. The vertical filter matrix is the transpose of this filter matrix.

In each sample image remain after these actions about 400 points per respective cell, which satisfied the three filtering conditions.

### 3.2.3 Ellipse Detection

Using the so detected set of points, the search for appropriate ellipses is performed. In a first step, two random points from the set are selected. In a second step a third point is selected that is linearly independent with respect to the two other points. With this three points an ellipse is defined, assuming that the first two points are a diameter of the ellipse, and the third is located somewhere on the ellipse. This procedure reduces the dependency of the ellipse detection from the number of cells in the image. An ellipse is uniquely defined by five points; however, five points always define a conic which only is in special cases an ellipse. So the test if the selected points describe an ellipse is slightly more complicated than only looking for linear independence of three points. An example of calculating the probability to find an ellipse that describes a cell will show the effect:

$n$  : number of cells in the image,  
 $p_i$  : probability of a point belonging to the border of the cell  $i$ ,  
 $r$  : probability of two points on the same edge forming a diameter.

$$\begin{aligned} p_i &= \frac{1}{n}; \\ r &\approx 0.05; \\ p(\text{get 5 points of the same cell } i) &= p_i^5; \\ p(\text{get 3 points of the same cell } i, \text{ wherein 2 define the diameter}) &= p_i^3 r; \\ &\Rightarrow \forall n > 4 \text{ the used method has a greater chance} \end{aligned}$$

The ellipse is calculated using the three points as follows: the middle of the line between the first and the second point defines the center of the new ellipse. This line is also used to define the angle of rotation of the ellipse with regard to the  $x,y$ -coordinate system and the scaling in this direction. The distance of the third point is used to define the scaling in the orthogonal direction. The center, the angle relative to the  $y$ -axis, and two scaling values  $a$  relative to the  $x$  axis and  $b$  relative to the  $y$ -axis are saved and describe an unique ellipse. More formally expressed, the 5 ellipse parameters  $x_{Center}, y_{Center}, \phi, a, b$  are calculated as follows from the three points  $\begin{pmatrix} p_x^1 \\ p_y^1 \end{pmatrix}, \begin{pmatrix} p_x^2 \\ p_y^2 \end{pmatrix}, \begin{pmatrix} p_x^3 \\ p_y^3 \end{pmatrix}$ :

$$x_{Center} = \frac{p_x^1 + p_x^2}{2}$$

$$y_{Center} = \frac{p_y^1 + p_y^2}{2}$$

$$a = \sqrt{(x_{Center} - p_x^1)^2 + (y_{Center} - p_y^1)^2}$$

$$\phi = \arccos \frac{x_{Center} - p_x^1}{\text{asign}(p_y^1 - y_{Center})}$$

$$b = \sqrt{(x_{Center} - p_x^3)^2 + (y_{Center} - p_y^3)^2}$$

With this information, the distance  $d$  of any point  $p$  to the border of the ellipse  $E$  can be calculated by the following formula:

$$d \left( E : (x_{Center}, y_{Center}, \phi, a, b), p : \begin{pmatrix} x \\ y \end{pmatrix} \right) = \left\| \begin{pmatrix} x - Center \\ y - y_{Center} \end{pmatrix} \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \frac{1}{a} & \frac{1}{b} \end{pmatrix} \right\|$$

For a result of  $d = 0$ , the point  $p$  is located on the surrounding border of the ellipse. A result of  $d = 1$  indicates the highest positive distance inside the ellipse, and outside,  $d > 1$  means that the point outside the double of the radius. All points for which a result of  $d = 0 \pm 0.05$  is calculated are assigned to this specific ellipse and increase the rating value of this ellipse. If the rating value of the ellipse exceeds a threshold, it will be checked, if there are already other ellipses with a center inside the just calculated ellipse or if the calculated ellipse's center is inside of any other already saved ellipse. Only the ellipse with the highest rating value is kept, and all others are discarded. All points that are defined by one of the ellipses are removed from the set of points used for generation of further random ellipses. If 80% of the points are defined by any ellipse, or after a time long enough to be sure there is no sense in continuing of the search, the algorithm terminates.

### 3.2.4 Other Approaches

During the first phase of the project, several other algorithms were tried to solve this cell segmentation problem. The first of the approaches was a threshold approach combined with neighborhood fusion of pixels to clusters. However, the problem of fusion of two cells which are connected to each other on the picture could not be solved with this method. Also clustering around a barycenter was difficult to stop at the right level. Further the problem of the fusion of two cells to one could not be solved adequately. The next development state of the algorithm included the reduction of the considered pixels to the edges. The first approach on this reduced set of points also failed: it detects edges of a cell following an edge in a more or less convex direction. This failed due to the lack of convexity of some cells. As the first working method, a Hough transformation was used, which was too slow without the random ellipse selection presented above.

## 3.3 Detection of Signals

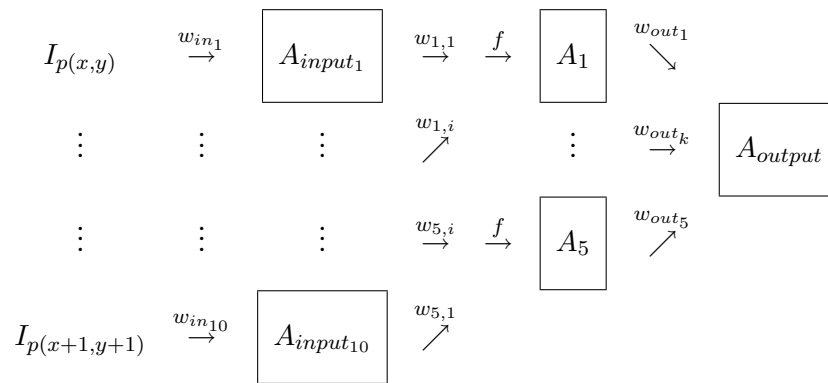
### 3.3.1 Overview

For detection of signals a Neural Network is applied. As input the Neural Network gets the average intensity and the next 3x3 pixels to decide if a pixel is part of a signal or not. In particular, a back-propagation network is used, to give the user of the program the possibility to adapt the recognition of the signals, and thereby bring in his expert knowledge. The evaluated signals have a size between 3 and about 30 pixel per image and normally are quite similar for the same kind of image, but differ a lot for another probe. In a last step all pixels of all images are merged to regions to define signals.

### 3.3.2 Design

The used neural network is a three layer feed forward network after [1] having an input layer with ten neurons, a hidden layer with five neurons and an output layer with one neuron. The ten input neurons are fed with the intensity of the eight surrounding pixels and the pixel itself and the average intensity of an area near this pixel, respectively. The hidden layer in a neural network is inserted to improve the learning effect. For a rather simple task like the segmenting of this kind of signals, a number of neurons that ranges between the number of inputs and intended outputs is a good choice. The output neuron should return a value greater than zero for a pixel which represents a part of a signal, otherwise a value less or equal zero.

The design is as follows, all abbreviations used will be explained later:



First, the activation function, marked with  $f$  in the design above, is explained. For the input and output layers there is no further need for an activation function, so the identity is used. For the hidden layer there is a need of an activation function which is:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The input vector  $v_{input}$  to the neural network for the point  $p(x, y)$  having the coordinates  $x$  and  $y$  with the normalized intensities  $I_{p(i,j)}$  of the point  $p(i, j)$ , and the average intensity of the Image  $I_{image}$  is as follows:

$$v_{input} = \begin{pmatrix} I_{p(x,y)} \\ I_{image} \\ I_{p(x-1,y-1)} \\ I_{p(x,y-1)} \\ I_{p(x+1,y-1)} \\ I_{p(x-1,y)} \\ I_{p(x+1,y)} \\ I_{p(x-1,y+1)} \\ I_{p(x,y+1)} \\ I_{p(x+1,y+1)} \end{pmatrix}$$

The elements of this input vector are multiplied by the weights  $w_{in_i}$  of the respective input neurons and passed to the hidden layer. For every neuron in the hidden layer exist



weights  $w_{j,i}$  for every input neuron. So there is further a threshold  $t_j$  which defines the inflection point of the activation function. Hence, the activation of the neuron  $j$  in the hidden layer  $A_j$  is calculated as follows:

$$A_j = f \left( -t_j + \sum_{i=1}^{10} v_{input} w_i w_{j,i} \right)$$

The final result of the calculation is the activation  $A_{output}$  of the output neuron. It is calculated from the hidden layer activations  $A_j$  and the output neuron weight  $w_{o_j}$  as follows:

$$A_{output} = \sum_{j=1}^5 w_{out_j} A_j$$

This value  $A_{output}$  decides if the pixel is part of a signal or not.

### 3.3.3 Learning Algorithm

The backpropagation algorithm is a form of supervised learning for a neural network. If the solution with the current weights is not exact enough for certain application, there may be an adaption of the network behavior. Enough learning samples which represent a class of signals and a known solution is available, it is possible to start a learning procedure with the neural network.

Since the backpropagation algorithm provides a method to learn information relatively fast and with a good stability of already learned facts, it has been chosen for the current problem. During the learning process, for every learning sample an output is decided and a supervisor decides that the output is not correct, all parameters are newly adjusted. The idea of backpropagation is to bring the error that is measured at the output neuron back through all the layers affecting all parameters with an decreasing influence, so the next output will be nearer at the expected value.

The version used herein for the neural network is presented shortly. The neurons of the Neuronal Network handle values around -1 and 1 and as final output activation -1 represents the decision against a signal and a 1 pro a signal. If the difference between the output of the neural network and the known solution is greater than 1, it's interpreted as an error.

The formula for the newly adjusted weights  $w_{out_j}^{new}$  connecting the output neuron and the hidden layer neuron  $j$  with the error  $\epsilon$  between the known output and the actual output is as follows:

$$w_{out_j}^{new} = w_{out_j}^{old} + 2\delta A_j \epsilon;$$

With  $\delta$  as a variable value below 1 that decides about the learning velocity. The lower the  $\delta$  value the more stable results are obtained and the higher a faster learning, but not as stable.  $A_j$  is the activation of the hidden neuron  $j$  as described in the section before.

The weights  $w_{hidden_{i,j}}^{new}$  connecting the input neurons  $i$  and the hidden layer neuron  $j$  will be changed according to the following formula:

$$w_{hidden_{i,j}}^{new} = w_{hidden_{i,j}}^{old} + 2\delta v_{input}(i) A_j (1 - A_j) \epsilon w_{out_j}^{old};$$

The threshold  $t_{hidden_j^{new}}$  of the hidden neurons is also changed:

$$t_{hidden_j^{new}} = t_{hidden_j^{old}} - 2\delta A_j(1 - A_j)\epsilon w_{out_j^{old}};$$

The weights of the input neurons are fixed, and are not adapted, but they are set by default to 10 for the first input neuron's weight and to 1 for the other 9 inputs. With this strategy a picture series can be learned after about 50 runs. In the provided learning environment 100 runs over all interesting points are performed to ensure obtaining an admissible result.

### 3.4 Association of Signals

For the association of a signal with a cell, a two step method is used. The first of 2 factors to assign a signal to a particular cell is the distance to the cell's surrounding ellipse's center. Only if the distance is less than the doubled radius, it can be associated with this specific cell. After the distances for all signals have been evaluated, the signals are associated with the respectively nearest cell provided that the corresponding cell has less than the expected number of signals. If no corresponding cell is available, the signal is associated with the "best" cell, not fulfilling the second condition, or it is dismissed as unassignable.

# 4 Results

## 4.1 Program

To offer easy handling of a plurality of images and managing existing knowledge about these images, a user interface is provided. The project's UI was programmed on the Microsoft .net 1.1 runtime environment in  $C^\#$ , because Windows is the operating systems on the computers of institute of human genetics. for this programming environment was finally decided by the fact that a graphical user interface is be an important feature, especially for future users. So the easy to program approach of  $C^\#$  won about faster solutions like  $C^{++}$ . The interface provides the possibility to visualize gained results in 2D and display a short textual note about the number of cells, signals and signals per cell. Also the referenced files are shown, categorized by a signal type. For a detailed information about the program and its functionality see [3].

## 4.2 Statistics

### 4.2.1 Cell Detection

As long as the images contain cells in elliptic shapes, the results are pretty good, but with some sort of images the results did not describe the cells good enough. Also the fact that the results depend from random values create some instability in the found cells. But overall the results are satisfactory.

Problems are caused by cell fragments at the borders of the images if they are not far away enough from other cells, which are completely in the image. In Figure 4.1 there is an overview of the success with testing data. So a ratio of over 90% is reached for successful cell segmentation. The false positive errors are represented by the number of cells with incorrect center, detected ellipses that do not have a corresponding cell. So the ratio of false positive is about 8%. Normally a false positive was caused by a not detected cell, which led to incorrect center. The false negative is the difference between the number of complete cells in the images and the number of detected ellipses. The ratio of false negative is about 9%. The ellipses mentioned in Figure 4.1 with incorrect size or shape did not affect the correct association of signals to this cell. Since the center was in the most cases still inside the cell, which is the crucial criteria to assign signals to this cell, this errors did not affect the final results.

Figure 4.2 shows a sample image with detected ellipses as an overlay. In this image all complete cells are detected correct. All complete cells have a blue cross inside, which describes the ellipse that defines the cell border. All points on the border, that work fine with this ellipse are marked in yellow green. An example for a more problematic sample is given in Figure 4.3, where in especially the sizes are badly analyzed, while the association

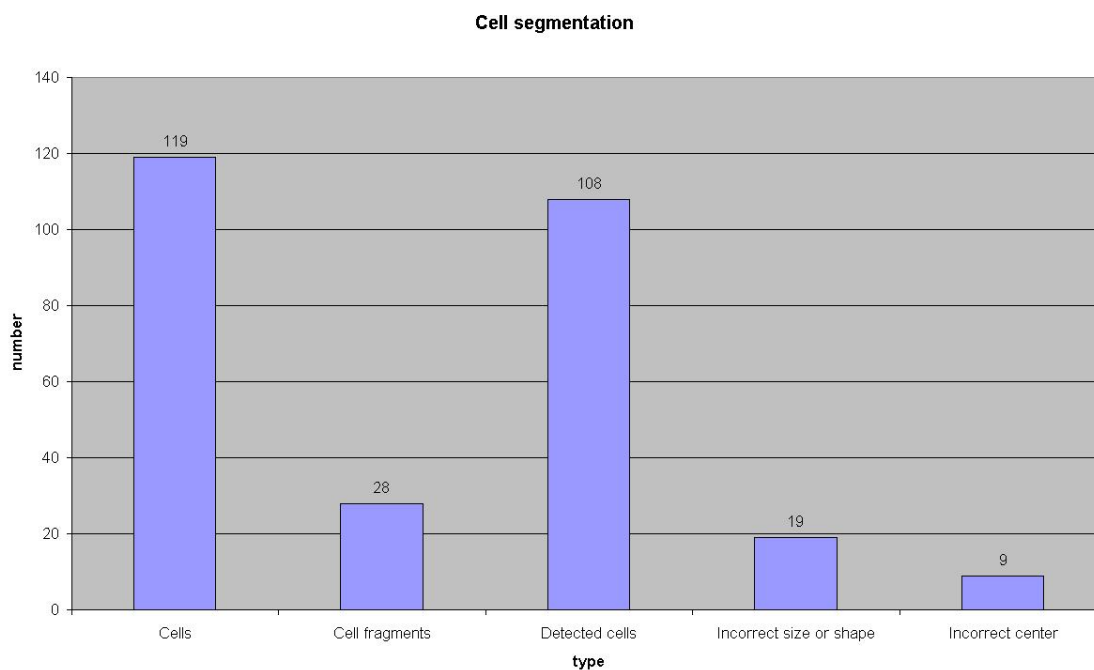


Figure 4.1: Cells in the test images and their detection. The first two bars represent the cells in the test images, the third one the number of cells detected by the algorithm, while the last bar represents the numbers of detected cells, that were not to be detected. The third bar represents the cells detected, while the fourth bar represents the part which were not perfectly detected and the last one useless cells.

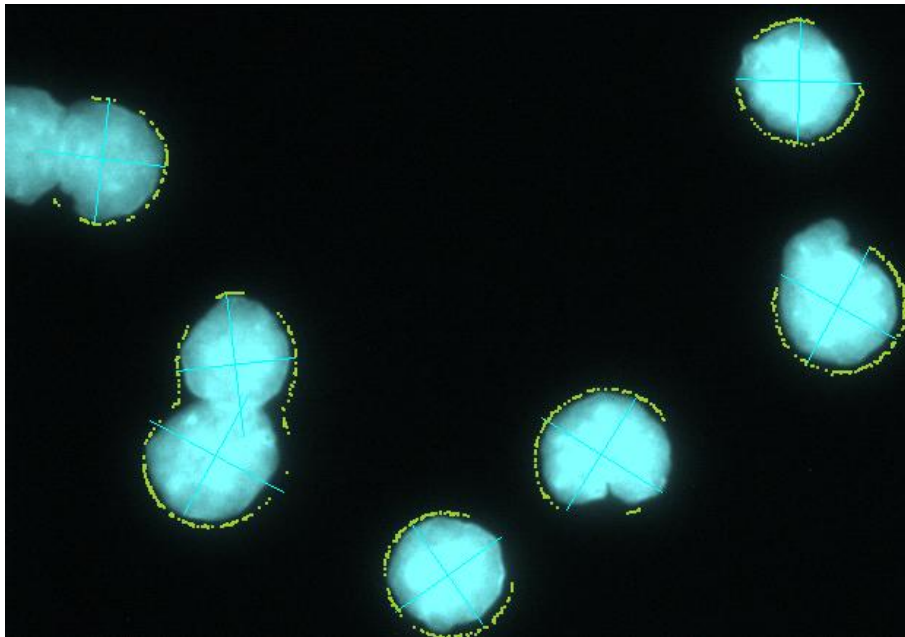


Figure 4.2: An image with correct cell detection, the blue lines inside the round bright blue circles describes the ellipses, and the pixel marked in yellow-green are the edges on that this ellipse is based

of the signals it still works. These are example for the cells with bad size referred in the statistics in Figure 4.1.

#### 4.2.2 Signal Detection

While the signal detection seemed in the beginning easier, it creates greater problems than expected, because most of the different signal types did not work with the same neural network, so there had to be more than one neural network in order to detect most of the signals. This causes problems with the usability of the program: the user is forced to know what kind of signals he is treating. Figure 4.4 shows the average success with the test data, but only for the signal types a neural network was trained for. The ratio of successfully detected signals in this class was over 94%, the ratio of false positive was below 1%, but 6% of the signals were not detected. Of the detected signals, 3% were associated with a wrong cell, mostly caused by unprecise centers of these detected cells. There are three main signal types in the proposed images used for this statistic. The first type, relatively large signals with low noise, normally gained by centromer probes, can be seen in Figure 4.5. All color channels with probes of this type could be segmented with one trained neural network. In Figure 4.6 there is a graphical representation of the result for Figure 4.5, all white spots that are marked with the yellow-green overlay are detected. A textual note would display, that there are 4 cells (see Figure 4.3) and 13 signals, while only 11 signals can be associated with these cells, so in the cells were irregularities. The second type, a sample can be seen in Figure 4.7, has a worse signal to noise ratio, but still relatively large signals. A sample of the third type can be seen in Figure 4.8. The most false negative cases were observed in

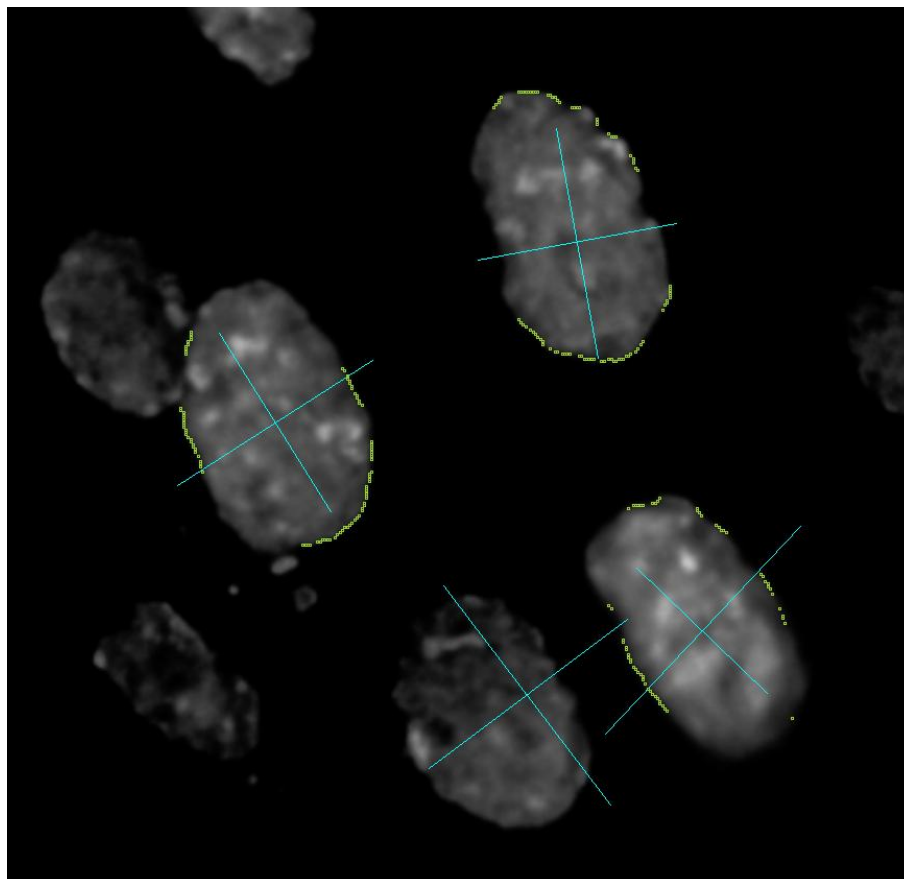


Figure 4.3: An image with problems in the cell segmentation. Some sizes are wrong and the center points are not correct, but still the number and more or less the position is usable.

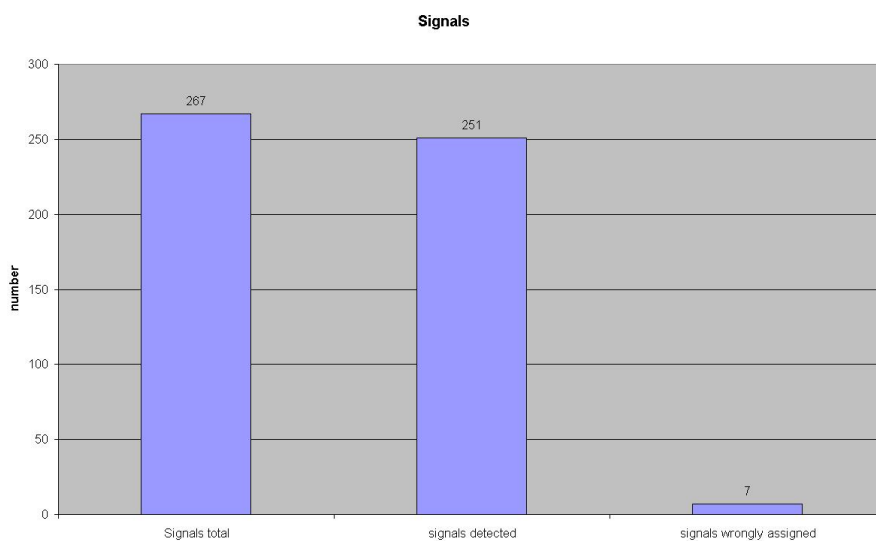


Figure 4.4: Signals on the test images and their detection

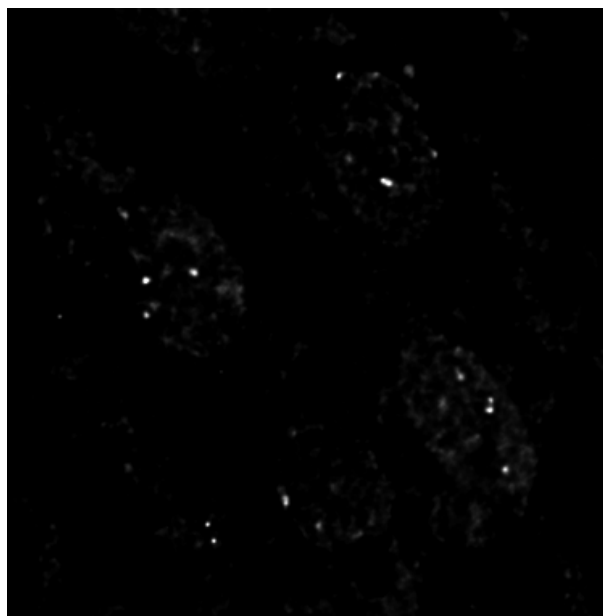


Figure 4.5: A sample for clear and large signals.

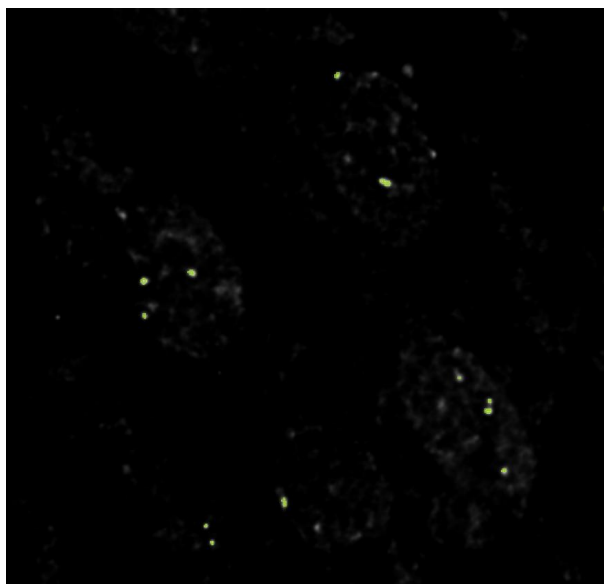


Figure 4.6: The same sample as in Figure 4.5, only here with detected signals as a yellow-green overlay. The gray noises are correctly not detected as signals.

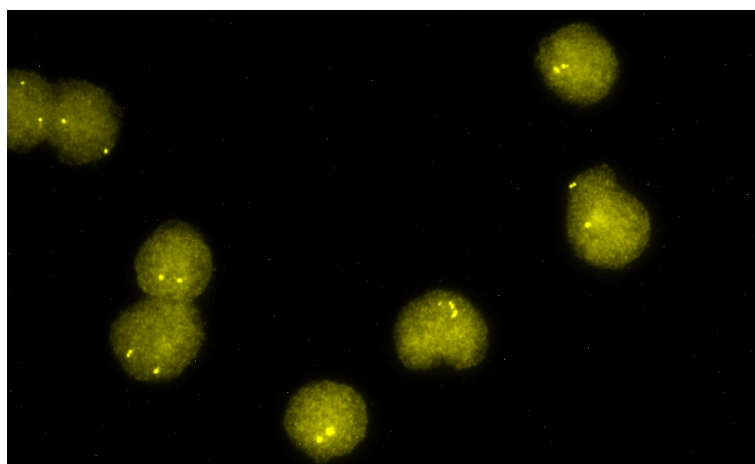


Figure 4.7: A sample for noisy, but large signals.



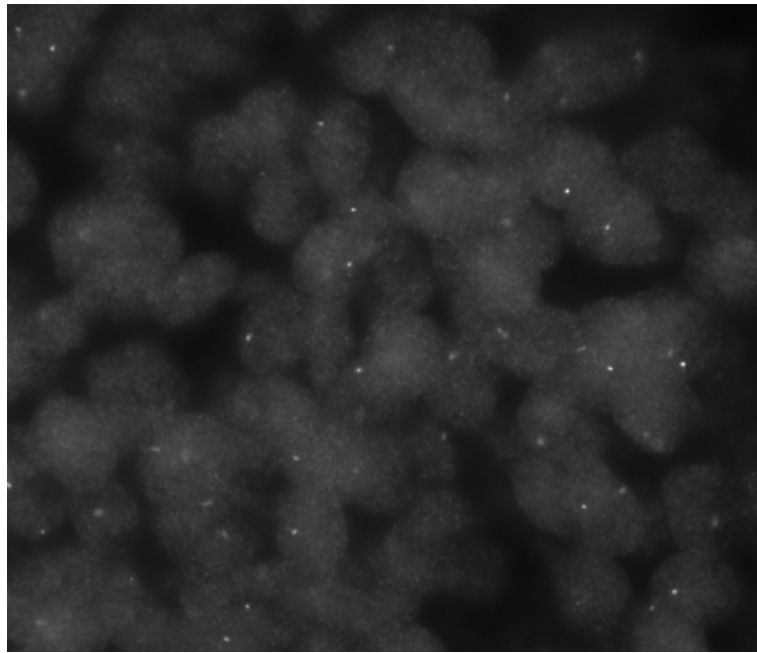


Figure 4.8: A sample for noisy and small signals.

images with signals of the third type.



## 5 Conclusion

In this project a program was developed which is able to count marked DNA-sequences in images of cells obtained by the FISH method. The algorithm uses three steps to solve this, first the segmentation of cells, then the detection of signals and the association of those two. The segmentation of the cells was solved by a randomized ellipse detection, based on a Hough transformation. The signal detection was performed by a neural network, that used only local information. The association of the signals to the cells uses the center of the ellipses which were gained by the segmentation.

This algorithm delivered satisfactory results, it was able to interpret "easy" images which comprise fewer cells and relative large signals. The percentage of errors was in an acceptable range. The usability of the program itself was an important point, and it is provided by an user interface handling a lot of tasks automatically, like for example ordering images to signal types by parsing the file name of these images.

Further work could be invested in the performance of the program. There is still a lot of space for optimization, especially in the cell segmentation algorithm. A second issue of improvements is the amount of training time the neural network needs to detect the more noisy signals. Another approach could involve providing more input parameter for each pixel.



# Bibliography

- [1] Burkhard Lenze. *Einführung in die Mathematik neuronaler Netze*. 1997.
- [2] Emanuela Gussoni, Richard R. Bennett, Kristina R. Muskiewicz, Todd Meyerrose, Jan A. Nolte, Irene Gilgoff, James Stein, Yiu mo Chan, Hart G. Lidov, Carsten G. Bönemann, Arpad von Moers, Glenn E. Morris, Johan T. den Dunnen, Jeffrey S. Chamberlain, Louis M. Kunkel, and Kenneth Weinberg. Long-term persistence of donor nuclei in a duchenne muscular dystrophy patient receiving bone marrow transplantation. 2002.
- [3] Ulrich Klank. Interdisziplinäres projekt: „cell nucleus classification segmentation“-dokumentation. 2006.
- [4] Christine Maierhofer, Rainer Gangnus, Joachim Diebold, and Michael R. Speicher. Multicolor Deconvolution Microscopy of Thick Biological Specimens. 2003.
- [5] S Tsuji and F Matsumoto - IEEE Transactions on Computers,. Detection of Ellipses by a Modified Hough Transformation. 1978.
- [6] Michael R. Speicher and Nigel P. Carter. The new cytogenetics: Blurring the boundaries with molecular biology. 2005.