3.3

Image processing

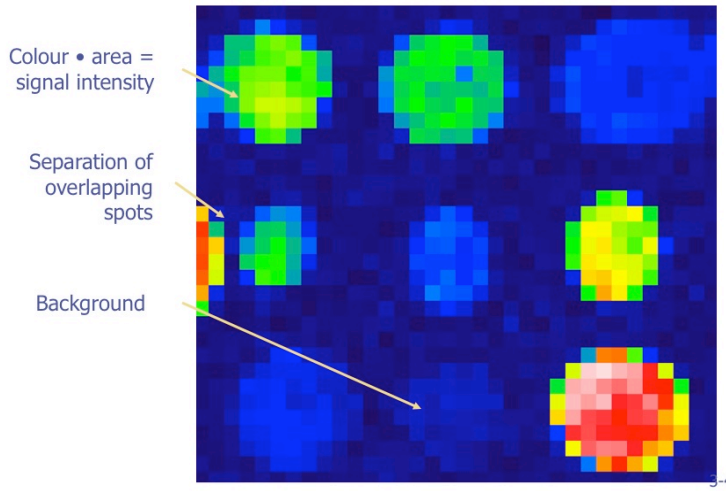
Steps in image processing

(1) Segmentation
- Distinguishing signals from background
- Distinguishing separate signals

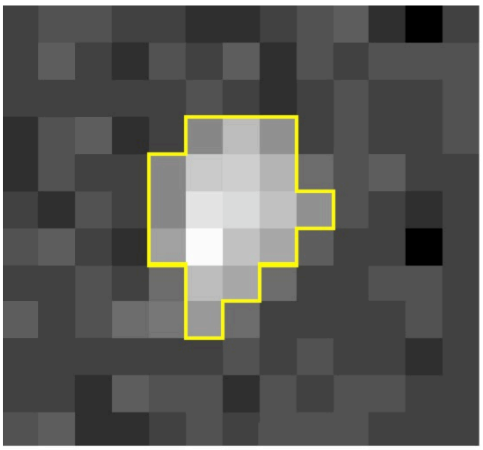
(2) Quantification
- Estimating signal intensity

(3) Background estimation/removal
- Subtracting non-uniform background signals
- Raw materials under ambient lighting of the excitation source/scanner light

Magnified subsection



Irregular segmentation



Segmentation

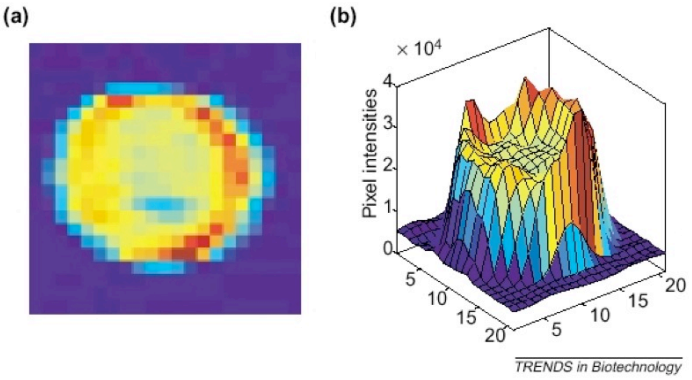
(1) Assigning each pixel to a particular class
- Here typically spot or background
- Each separate spot can be its own class

(2) Simple thresholding algorithm
- Given an intensity threshold t
- For all pixels (i, j) in ascending coordinate order:

intensity (i, j)	class (i, j-1)	class (i-1, j)	NEW class (i, j)
<=t			0
>t	0 or k	0 or k	k
	k	m	join k+m

(3) Problem: choosing a threshold

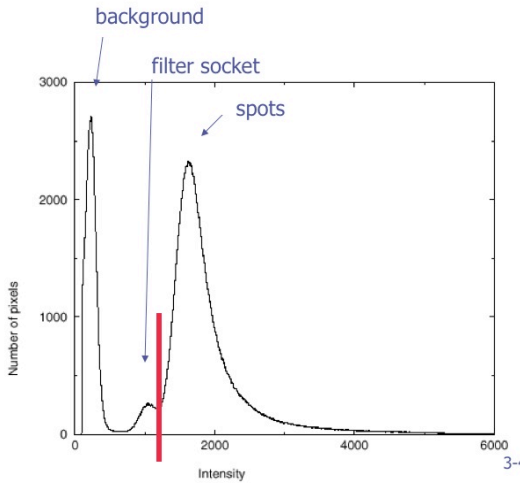
Spot detail



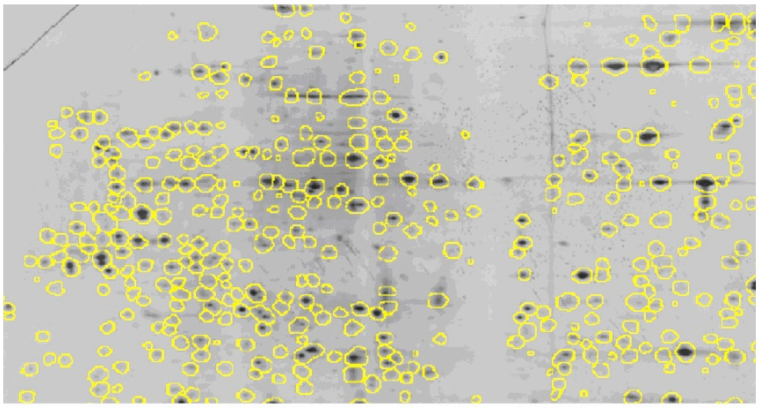
Pixel intensities coded as colours

Histogram thresholding

Finding and counting minima and maxima in the image histogram



Example segmentation



Quantification

(1) Average peak

- Average over all segmented pixels
- Does not take into account size variations, only for uniform spots
- Background-sensitive: non-uniform background will distort values

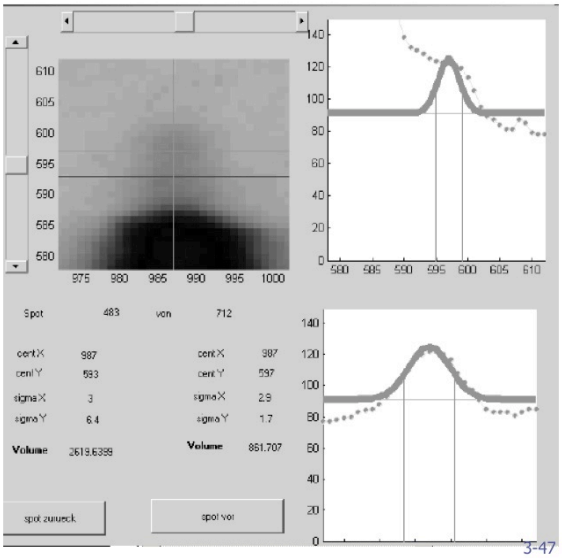
(2) Total intensity

- Summing up intensities of all segmented pixels
- Background-sensitive
- Segmentation-sensitive: unless spots have sharp borders, small signals vary considerably with segmentation threshold

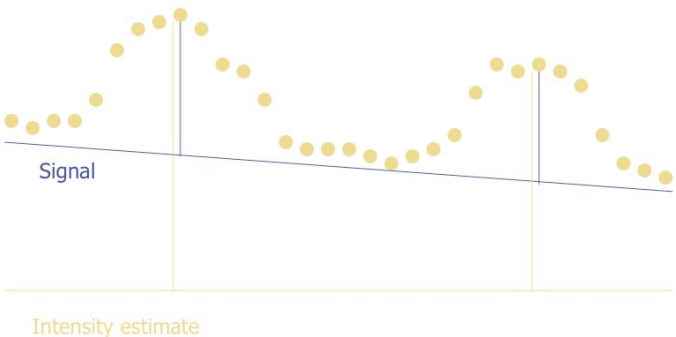
(3) Curve fitting

- Calculating model curve with smallest distance from segmented data
- Noise-sensitive: outliers can skew fitting

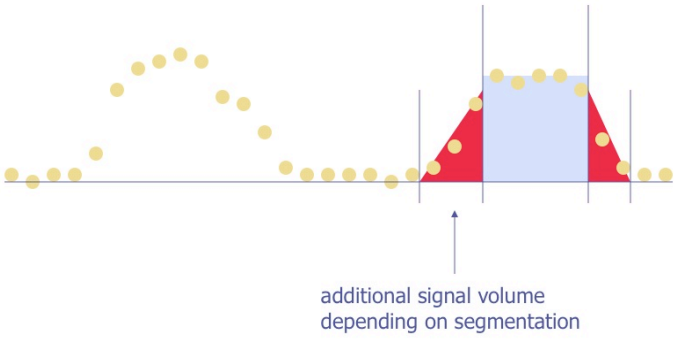
Spot shape



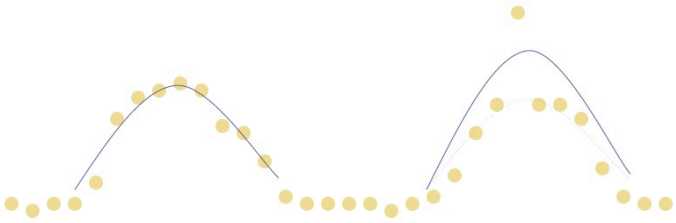
Background sensitivity



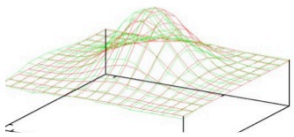
Segmentation sensitivity



Noise sensitivity



Gauss fitting



$M_R = \{p\ (x, y) \mid (x, y)\ \text{segmented}\}$

$M_V\ [H, B, c_x, c_y, s_x, s_y] = \{G\ (x, y) \mid (x, y)\ \text{segmented}\}$

$$G(x, y) = H \exp[-(x-c_x)^2/2s_x^2] \exp[-(y-c_y)^2/2s_y^2] + B$$

H	Height
B	Background
$c_x\ c_y$	Center
$s_x\ s_y$	Standard Deviation

⇐ **Model Instantiation**

Least-squares Fit of M_R and M_V with approximate start values

Background estimation/subtraction

(1) Background is not uniform
- Different lighting conditions in different part of the chips, optical non-uniformities
- No single threshold for the whole image!
- Not only relevant for threshold, but background (lighting) has to be subtracted as well to get real signal intensity

(2) Background estimation (after segmentation)
- Average of surrounding non-spot areas
- Fitting

(3) Background subtraction (before segmentation)
- Morphological operators

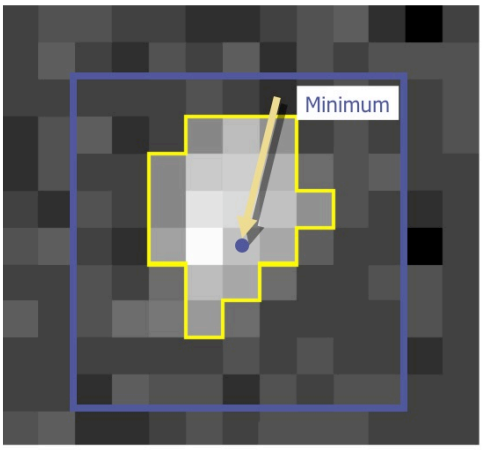
Morphological opening

(1) Pixel substituted by minimum over neighbourhood
- Shape of the neighbourhood is called the *structural element*
- Size of the structural element determines smoothness of operator

(2) Creating a background image
- Underestimating the background
- Structural element must be larger than spot size, then there will always be background pixel in the region

(3) Creates more reproducible quantitative results than local background estimation

Rectangular element



Registration to spot grid

(1) Orientation
- Manually by clicking on border spots
- Automatically (depending on particular features, like recognizable bord angles, markers, etc.)

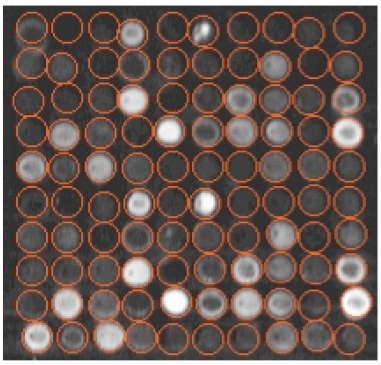
(2) Identification of raster positions
- Because of imprecise needle positions and depositing

(3) Comparatively simple because size of raster is known

Example registration

1. Seed points at theoretical grid points
2. Calculate center of gravity
(Sum of positions x intensity divided by total intensity)
3. Smallest circle above threshold

Mapping function assigns each circle to corresponding grid point



Gray level correlation

(1) Joint histogram
- p (a, b): Number of pixels where intensity a in the first image maps to intensity b in the second image
- Maximize correlation coefficient
- Requires same colour scale

(2) Joint entropy
- - sum (p (a,b) log p (a,b))
- Is higher if there are few pronounced entries in the joint histogram
- Works also with different colour scales (e.g. white matches with dark red, gray matches with light red, black matches with green)

(3) Optimize mapping (rigid, non-rigid) for maximum correlation

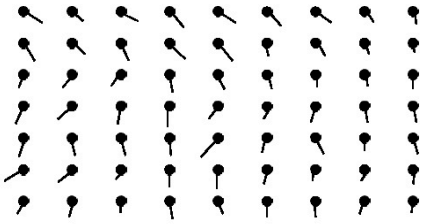
Example joint histogram

Unmatched

Matched

Optical flow (non-rigid mapping)

Vector field for each pixel defines mapping



Problem: Without constraint arbitrary matches are possible

Grid-based elastic matching

(1) Mapping function is a grid of displacement vectors
- Lower resolution than the image
- Displacement vectors of pixels are interpolated (linear or spline)
- Penalty for relative displacement compared to neighbours

(2) Iterative refinement
- Starting with rigid registration (= one vector)
- Subdividing grid
- Initializing subgrid with parent vectors and optimizing (with lower displacement)

(3) Problem: background areas
- Elastic matching can produce arbitrary results in areas without signals
- No clear optimization direction (noise maps to noise)

Point matching

(1) Registration between arbitrary point sets
- Each point has a coordinate
- Mapping function is assignment
- Outliers may be unassigned

(2) Grid constrains mapping
- For non-rigid mappings

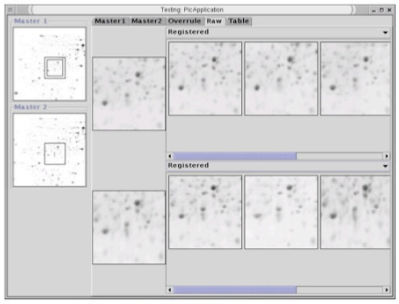
(3) Point attributes can be evaluated
- As a correlation function

Example: Proteomics data analysis

2D electrophoresis
- Quantitative protein detection as spots on a gel
- E.g. looking for cancer

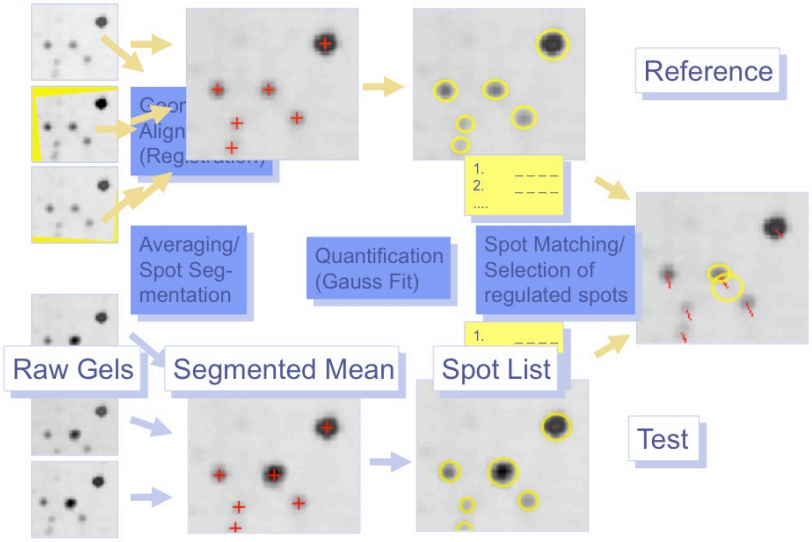
AgeLab Pharma

Aventis Research & Technologies



Intelligent data analysis
- Identification und quantification through statistical image processing
- Matching of different experiments (identity of spots unknown!)

Visual evaluation
- Presentation on natural background
- Pattern analysis, quality control and communication by the users



Registration Problems

(1) Assume identical samples and processing
- Fluctuations in the electrophoretic separation
- Staining variations
- Gel distortions

(2) Registration
- Based on mutual information (robust against staining and scanning differences)
- Allows non-rigid deformation, constrained mesh of control points

(3) Problems
- Some proteins vary their spot positions more than their neighbours
- Average of intensities smears spots

Problems of Quantification

(1) Multiple spots
- As many as 30% of the spots contain more than one protein
- Center may coincide or vary slightly (complex mixture of gaussians)

(2) Ridges
- Proteins that do not completely denaturate smear over a broad range
- Contribute to the background
- Use a local background for fitting, but sum based on global background

(3) Saturation
- Staining may saturate at high intensities
- Ignore those pixels in fitting

Matching different samples

(1) High variation
- Different runs, different processes
- Different biological conditions (protein concentrations may rise in minutes)

(2) Spot assignment solely based on position
- Calculation of an (affine/non-affine) mapping function and of a spot assignment by deterministic annealing

(3) Problems
- Multiple spots in one sample may separate in another
- No gold standard (without exhaustive identification via MS)

Spot assignment