Gustavo K. Rohde

Center for Bioimage Informatics
Dept. of Biomedical Engineering
Dept. of Electrical and Computer Engineering
Lane Center for Computational Biology
Carnegie Mellon University,
Pittsburgh, PA, USA 15232
gustavor@cmu.edu
www.andrew.cmu.edu/user/gustavor

Image Cytometry Course
Seattle, WA, USA
May 7th, 2010.
My goal

- Many methods/technologies have been developed, many of them in academia, to extract quantitative information from microscopy images
  - image restoration, pixel level segmentation, tracking, pattern recognition, etc.
- An overview of recent and old developments, with the aim of describing to a general scientific audience what is possible to extract from microscopy images now.
Roadmap

- **Digital images**
  - what are they, what limitations are imposed by a general optical setup.

- **Deconvolution & denoising (restoration)**
  - microscopy images are usually degraded, can we improve on these

- **Segmentation**
  - how to select target structures for further analysis

- **Tracking & registration**
  - how to analyze & segment cells in time

- **Pattern recognition**
  - What general tools are available for quantifying such information
Digital images

Optical system pipeline, point spread function, resolution, sampling, digitization, noise.
RGB = [Red, Green, Blue]

<table>
<thead>
<tr>
<th>[117, 134, 102]</th>
<th>[73, 86, 56]</th>
<th>[41, 45, 20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[110, 125, 94]</td>
<td>[69, 80, 50]</td>
<td>[40, 42, 18]</td>
</tr>
<tr>
<td>[103, 114, 84]</td>
<td>[63, 67, 40]</td>
<td>[40, 38, 15]</td>
</tr>
</tbody>
</table>
Optical imaging pipeline

Light pattern  →  Eg. fluorescence signal

System

- Blurring, convolution
- Sampling
- Digitization
- Noise

Digital Image
Degradation processes

original  blurred  sampling & dig.  noise
Blurring, point spread function

Point spread function

light pattern

original

blurred

\[ g(x) = \int f(u)h(x - u)du = f \ast h(x) \]
Sampling

Taking samples (usually at regular intervals), and storing only these.

Notion of resolution

Two point sources cannot be told apart visually if positioned closer than Rayleigh criterion.

\[ r_0 = \frac{0.61\lambda}{\text{N.A.}} \]

\[ \text{N.A.} = \frac{a}{2d} \]
How to sample?

(Nyquist, Shannon, Whitaker): Sample twice as fast as highest frequency present in the images (assume band-limited images).

Rayleigh & Nyquist:

Take samples at least at half of Rayleigh criterion for visually interpretable results.

\[ T = \frac{r_0}{2} \]
Example:

Undersampled: no visible structure
Pixels spaced too far apart.

Critically sampled: cell bundles visible

Digitization:

Each sample is stored with a finite precision (number of bits). If not enough bits are used, image may appear rough to the eyes.

Sometimes the error is modeled as additive noise. Assuming uniform distribution,

\[
\text{RMS} = \frac{Q}{\sqrt{12}}
\]
Image file formats

- RAW: saves the raw digitized data. Files may be large. No compression.
- GIF (graphics interchange format): 8-bit (256) colors. Supports animations. No compression.
- BMP (windows bit map): like raw. Usually no compression.
- TIFF (tagged image file format): 8 or 16 bit per color, sometimes compressed with lossless compression.
- JPEG (joint photographic experts group): usually 8-bits per color, with lossy compression (with discrete cosine transforms, applied in blocks).
Noise:

Repeating the acquisition of a given image, even under optimal conditions, will yield two slightly different images, because of several noise sources.

Photon counting:

Counting rare events, with some assumptions, yields a Poisson distribution. This is signal dependent noise.

\[ P(f) = \frac{\lambda f e^{-\lambda}}{f!} \]

Gaussian additive noise:

Thermal fluctuations in read out digitization devices. This is typically additive noise. Gaussian PDF.

\[ P(f) = \frac{1}{2\pi\sigma} e^{-\frac{(f-\mu)^2}{2\sigma^2}} \]

Quantization noise:

Usually assumed to be uniform distribution.

\[ \text{RMS} = \frac{Q}{\sqrt{12}} \]
Modern approaches

Rayleigh criterion is a notion of resolution particularly applicable for visual interpretation of the image data. Takes into account only PSF & wavelength.

Modern approaches rely on computer analysis of image data. Quantify resolution based on one’s ability to estimate distance between two point sources, given PSF, noise, wavelength, time integration, pixelation, etc.

\[ r_0 = \frac{0.61\lambda}{N.A} \]

A computational method may be able to decipher that there are two Gaussian functions here.
Modern approaches


Size of the error in estimating distance between two identical point source $d$ meters apart.

$$
\delta_d := \frac{1}{\sqrt{4\pi \cdot \Lambda_0 \cdot (t - t_0) \cdot \Gamma_0(d)}} \cdot \frac{\lambda}{n_a}
$$

Looks like Rayleigh, but includes stochastic (Poisson) nature of the experiment.
summary

- images available are invariably degraded by following processes:
  - blurring by PSF
  - sampling
  - digitization
  - noise

- these impose fundamental limitations (often not completely described) on operations that are to be described next.
Deconvolution & denoising

Blurring, PSF convolution

original scene

scene measured by microscope
Approximation:

Each pixel measurement is replaced by a weighted (PSF) local average of neighborhood pixels.
Mathematically:

Convolution problem: \[ g(x) = \int f(u)h(x - u)du \]
\[ = f * h(x) \]

\[ g = Hf \]

Solution: (such a solution often exists)
\[ f = H^{-1}g \]
Example: direct inversion

$f \ast h = g$

$\hat{f}$ RECOVERED
Example: direct inversion with noise

Highly unstable: small amount of variation (noise) in input image causes large variation in recovered image.
Wiener filtering (smoothing):

Image model: \[ g = Hf + w \]
Linear system, additive, independent noise.

Bayesian minimum squared error: minimize: \[ E \left\{ (f - \hat{f})^2 \right\} \]
random variable as well.

Solution (zero mean noise, signal & noise uncorrelated)

\[
\begin{align*}
\text{Covariance matrix of noise} & \quad \left( C_{ff}^{-1} + H^T C_w^{-1} H \right)^{-1} H^T C_w^{-1} g \\
\text{Covariance matrix of image} & \quad \Theta \quad \Theta 
\end{align*}
\]

Richardson-Lucy deconvolution:

Maximum likelihood:

\[ A^* = \arg \max_A P(\hat{x}; A) \]

likelihood function

Multiplicative iterative maximization, with Poisson probability model:

\[ f_{k+1}(x) = f_k(x) \left( \frac{g(x)}{f_k * h(x)} \right) * h(-x) \]

deciding when to stop can be problematic.

Least squares:

Penalized least squares:  \[ \min_f \|Hf - g\|^2 + \lambda^2 \|Lf\|^2 \]  (e.g. \( L = \nabla \))

SOLUTION:  \[ f = (H^T H + \lambda^2 L^T L)^{-1} H^T g \]

Sparsity enforcing terms:

\[ \min_f \|Hf - g\|^2 + \lambda^2 \sum_i |Lf(x_i)| \]

minimization usually iterative

Constrained least squares:  \[ \min_f \|g - Hf\|^2 \text{ with } f \geq 0 \]

quadratic minimization.
quick comparison ...
Combining image acquisition & deconvolution

Full scene $\rightarrow$ Deconvolution difficult
how many particles, where?

Signal emitted by single molecules

If signal emitted at different times
then problem simplifies significantly
$=$ finding center of mass!

Particle here!
Example:

Image Segmentation
Image segmentation:

Identify (contiguous) pixel coordinates that localize structures of interest.

Applications:

• Cells
• Tissues
• Nuclei
• Molecules
• Organelles
• …

Approaches discussed here:

• Intensity thresholding, morphological operations
• Region growing
• Model-based
Intensity thresholding:

- DAPI stained image
- Intensity histogram
- Segmented image
Otsu’s method

Find $T$ that minimizes within class variance:

$$\sigma_{\text{Within}}^2 = n_B(T)\sigma_B^2(T) + n_F(T)\sigma_F^2(T)$$


- $n_B(T) = \sum_{i=0}^{T} p(i)$  number of pixels in background class
- $n_F(T) = \sum_{i=T+1}^{N} p(i)$  number of pixels in foreground class
- $\sigma_B^2(T)$  variance of intensities in background class
- $\sigma_F^2(T)$  variance of intensities in foreground class
Threshold

Morphological operations

**Dilation**

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

**Dilation**

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

**Dilation**

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

Dilation

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

*Dilation*

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

Dilation

Expand binary image by union of structuring elements centered on each foreground pixel.
Morphological operations

**Erosion**

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
Morphological operations

Erosion

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
**Morphological operations**

**Erosion**

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
Morphological operations

**Erosion**

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
Morphological operations

**Erosion**

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
Morphological operations

**Erosion**

Shrink image by collecting the pixels where a structuring element fits in the foreground of the image.
Morphological operations

**Original**

**Opening**
Erosion followed by dilation.
Remove small structures

**Closing**
Dilation followed by erosion.
Close gaps.
Watershed methods

Idea: Consider grayscale image as a topographical surface.

- “Flood” surface from strategic points (seeds, local minima, etc.)
- When waters meet, build a dam
- Dams are watershed line segmentations

Can be applied on original intensity images, or transformed ones (e.g. edge filters, etc.).

Very fast, very flexible. Results can be very “noisy.” These can be improved somewhat with selective seeding, and other a priori constraints on the solutions.
example

raw image

watershed result on edge filter

image courtesy, Prof. Lance Davidson, Univ. of Pittsburgh
Model-based methods

Idea: introduce more a priori knowledge in an effort to discard erroneous solutions.

Example: cells/nuclei look like ellipsoids. Design an “ellipsoid detector.”

One example:
- apply filter to detect edges
- fit possible ellipses
- deform detected ellipses to better match cell contours

Active contours

**Idea:** Evolve an initial contour until some stopping criterion is satisfied.

**Stopping criteria:**
- edges
- piece-wise constant approximation
- intensity statistics

Flexible procedure for minimization, allows for splitting and merging of contours. Tend to be computationally expensive.


Example:

Speedup can be obtained by avoiding re-initialization by constraining level set function.

Tracking and registration
Motivation

Dynamics is crucial for understanding many biological processes:

• Tumor metastasis
• Wound repair
• Embryogenesis
• Cell cycle
• Microtubule/actin dynamics
• …

Discussed here:

• Active contour tracking of cell populations
• Particle tracking
• Registration-based methods
Simple particle tracking algorithm

Detect spots in each frame
- thresholding intensity, center of mass
- filtering

Assign correspondence
- nearest neighbors
- total transportation distance

Doesn’t work in practice: noisy images (spot detection difficult), many particles under motion (confusion in assignment), particles appear and disappear …
Making educated guesses

In addition to pixel intensity information at current frame, also weigh in information about the previous measurements.
Bayesian probabilistic framework

$X_t$ position

$Z_t$ pixel intensities

$$P(X_t|Z_{1:t}) = \frac{P(Z_t|X_t)P(X_t|Z_{1:t-1})}{\int P(Z_t|X_t)P(X_t|Z_{1:t-1})dX_t}$$

$$P(X_t|Z_{1:t-1}) = \int P(X_t|X_{t-1})P(X_{t-1}|Z_{1:t-1})dX_{t-1}$$

“Best” position estimate

$$\max_{X_t} P(X_t|Z_{1:t})$$
In practice

Estimating these integrals notoriously difficult. Two ideas:

**Idea 1:**
Assume Gaussian distributions; linear motion model;
**KALMAN FILTER**

**Idea 2:**
Estimate $P(X_t|Z_{1:t})$ by Monte Carlo simulation (must have motion model, distribution of noise).

These ideas can be extended to multiple particles; and hypothesis testing for multiple motion models.
Example (particles)


Uses multiple models: Brownian motion, first, and second order for tracks.
Example (cells):


Cell detection performed with level sets. Accounts for cell division, death, through a “track arbitrator.”
Tracking via registration:

**Idea:** Tracking can be done directly on pixel data, without needing to perform spot detection/cell segmentation.

**But:** Still need detection/segmentation for analyzing motion.

**Method:**

Find the spatial transformation (translation, rotation, deformation) that minimizes the intensity differences between two or more frames.
Example:

Courtesy: Prof. Lance Davidson,
University of Pittsburgh, Pittsburgh, PA,
USA.
Example:
Pattern Recognition
Purpose:

Aforementioned (deconvolution, denoising, segmentation, tracking, etc.) not the end goal. These are used for:

**EXTRACTING QUANTITATIVE INFORMATION ABOUT BIOLOGICAL PROCESSES**

More specifically:

**MAPPING QUANTITATIVE RELATIONSHIP BETWEEN STRUCTURE AND FUNCTION (CYTOMETRY)**

**IN ADDITION TO AUTOMATION, LARGE SCALE ANALYSIS.**
Structural measurements (morphometry):

Shape parameters:
- Area (A)
- Perimeter (P)
- Form factor (P^2/A)
- Equivalent diameter
- Excentricity
- Convex area
- Euler number
- Curvature distribution
- …
Texture measurements

Quantify spatial distribution of intensity values (often require rotation invariance):

- Average intensity
- Standard deviation
- Entropy
- Intensity histogram
- Haralick features
- Filtering
- …

Feulgen (DNA) stained nucleus
Haralick features

Co-occurrence matrix:

Probability that two neighboring pixels in an image have intensities \((i,j)\). These are computed by counting all pixels in an image.

Features:

- angular second moment
- correlation
- variance
- entropy
- …

\[
G = \begin{bmatrix}
p(1,1) & p(1,2) & \cdots & p(1,N_g) \\
p(2,1) & p(2,2) & \cdots & p(2,N_g) \\
\vdots & \vdots & \ddots & \vdots \\
p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g)
\end{bmatrix}
\]


Work well when many pixels are available from each structure of interest.
Filter derived features

Idea:
For each pixel, compute a weighted average with neighboring pixels.
Weights = filter.

Filters:
- Edges, derivative, laplacian
- Frequency components (Fourier)
- Multi-resolution (modulated Gaussians, wavelets)

Features:
- Statistics (average, std.) of different frequency bands
- Statistics on directional information (e.g. anisotropy)

Classification

“Separate” different populations
- automated, high throughput
- understanding differences

---

THE ANALYSIS OF CELL IMAGES*

Judith M. S. Prewitt and Mortimer L. Mendelsohn
Department of Radiology, University of Pennsylvania,
Philadelphia, Pennsylvania

(1966)
THE USE OF MULTIPLE MEASUREMENTS IN
TAXONOMIC PROBLEMS

Annals of Eugenics, 1936

By R. A. FISHER, Sc.D., F.R.S.

I. DISCRIMINANT FUNCTIONS

When two or more populations have been measured in several characters, \( x_1, \ldots, x_s \), special interest attaches to certain linear functions of the measurements by which the

**Idea:** find projection that maximizes

\[
\text{Mean separation} \quad \text{Sum of respective variances}
\]
To determine the class of unknown data:

- Measure distance of projected point to mean of each class
- Take shortest measurement
Classification with nearest neighbors

- Parameter 1
- Parameter 2

unknown data point

Count how many neighbors from different classes fall in the same vicinity

Take majority vote.

Extensions to fitting nonlinear separation boundaries; support vectors, etc.

Example:

Cancer detection and classification-based on nuclear structure.

Follicular adenoma of thyroid  Follicular carcinoma of thyroid
A software system

**Input:** Raw images from single tissue

**Segmentation**

**Feature extraction**

**Classification**

Classify new data, based on previously available training data.

**Output**

Decision:
- Normal
- FA
- FTC

segmentation
**results**

**thyroid lesions**

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>follicular adenoma</th>
<th>follicular carcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>follicular adenoma</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>follicular carcinoma</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

**liver lesions**

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>hepatoblastoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>hepatoblastoma</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
More details

- 125 features used (normalization, feature selection)
- Classification system tested for individual nuclei
- Classification of groups of nuclei by most common result (voting)

Related data:

- Around 50 nuclei necessary for perfect classification
Summary
Overview

- Digital images
  - pixels, color, bit resolution, image quality, file formats
- Image processing
  - Deconvolution & denoising (restoration)
    - direct inversion, Wiener & Richardson-Lucy filtering, model-based
  - Image segmentation
    - Thresholding, region growing, morphology, deformable models
- Tracking and registration
  - Particle detection, tracking, registration
- Pattern recognition
  - Parameter extraction
  - Classification
Software

- Do I need to program these?
  - Mostly, no. Researchers in the area often allow other researchers to use their code. Look up their references, contact them.

- Programming languages
  - C, Matlab, IDL, Java, Python …

- ImageJ (http://rsbweb.nih.gov/ij/)
  - Free downloadable Java program for analyzing images
  - Many plug-ins developed to implement most of the covered material.

- In addition to vendor software.
Thank you