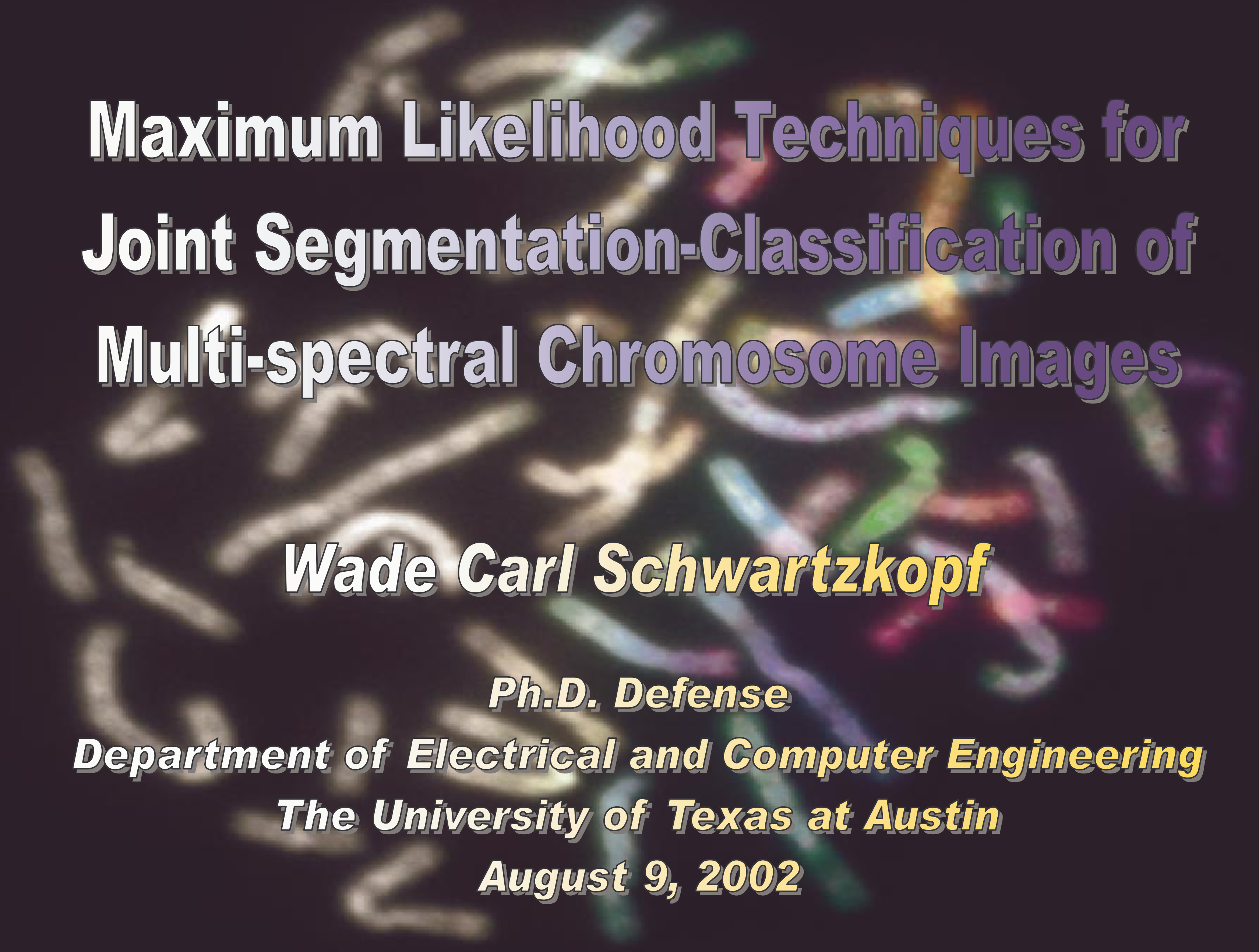


Maximum Likelihood Techniques for Joint Segmentation-Classification of Multi-spectral Chromosome Images

Wade Carl Schwartzkopf

Ph.D. Defense

Department of Electrical and Computer Engineering

The University of Texas at Austin

August 9, 2002

Outline

- Introduction
  - Motivation and Goals
  - Grayscale Chromosome Images
  - Multi-spectral Chromosome Images
- Contributions
- Results
- Conclusions

Motivation and Goals

- Chromosomes store genetic information
- Chromosome images can indicate genetic disease, cancer, radiation damage, etc.
- 325 clinical cytogenetic US labs perform over 250,000 diagnostic studies per year involving chromosome analysis
- Research goals:
  - Locate and classify each chromosome in an image
  - Locate chromosome abnormalities

Karyotyping

- 46 human chromosomes form 24 types
  - 22 different pairs
  - 2 sex chromosomes, X and Y
- Grouped and ordered by length except X and Y

Banding Patterns

Karyotype

Chromosome Abnormalities

- Abnormal number
  - Turner’s Syndrome (1 X, no Y chromosome)
  - Down’s Syndrome (3 of type 21)
- Translocations: Chronic myelogenous leukemia (type 9 and type 22)
- Deletions of genetic material: William’s Syndrome (gene missing in type 7)
- Research goals:
  - Locate and classify each chromosome in an image
  - Locate chromosome abnormalities

Denver Classifications

- Florescence microscopy
- Single dye, even stain
- Features
  - Length (2-10μm)
  - Relative centromere position
- Disadvantage: Only 7 distinguishable types [1960]

centromere

Banding Patterns

- Single dye, banding pattern staining [1969]
- Features
  - Length
  - Relative centromere position
  - Banding pattern
- All 24 types distinguishable
- Greatly improved manual chromosome analysis
- Disadvantage: Computer analysis difficult

Multi-spectral Chromosome Imaging

- Multiplex Fluorescence In-Situ Hybridization (M-FISH) [1996]
- Five color dyes (fluorophores)
- Each human chromosome type absorbs a *unique* combination of the dyes
- 32 (2^5) possible combinations of dyes distinguish 24 human chromosome types

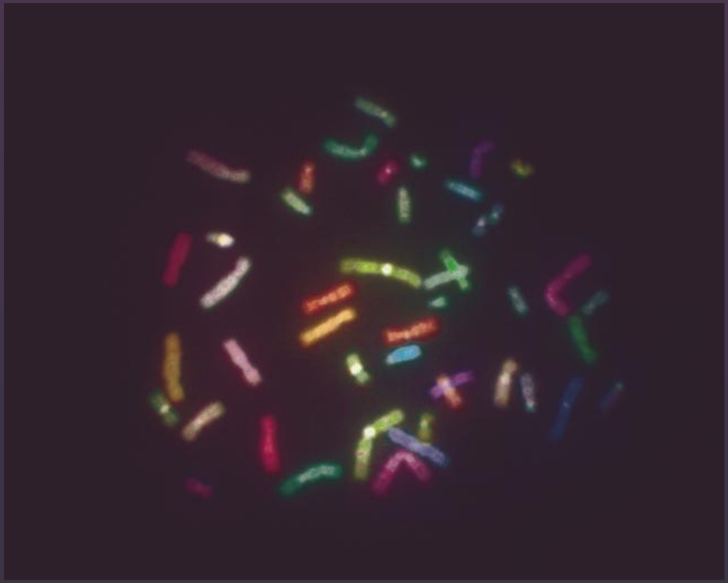
Healthy Male

M-FISH Images

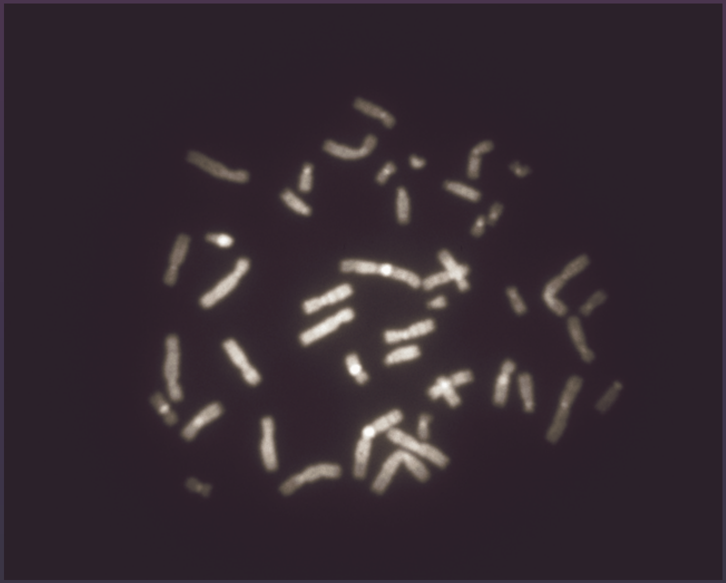
- Images of each dye obtained with appropriate optical filter
- Each pixel a five dimensional vector
- Each vector element gives contribution of a dye at pixel
- Chromosomal origin distinguishable at single pixel (unless overlapping)
- Unnecessary to estimate length, relative centromere position, or banding pattern

M-FISH Images

- 6th dye (DAPI) binds to all chromosomes



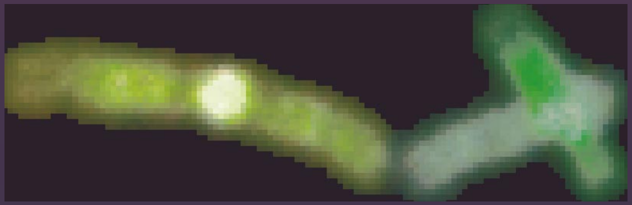
M-FISH Image
5 Dyes

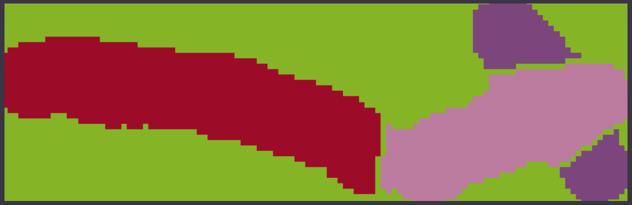


DAPI Channel
6th Dye

Traditional Chromosome Image Analysis

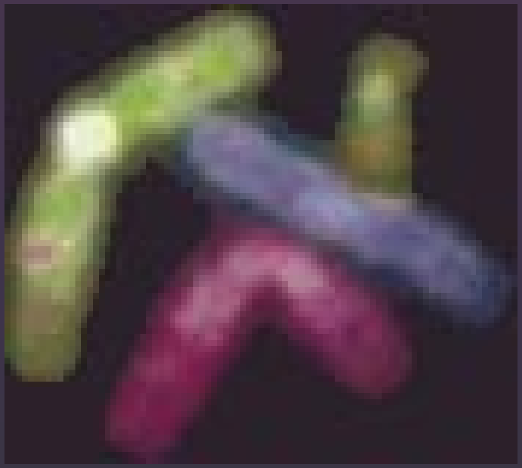
- Sequential: Segmentation then classification
- Segmentation
  - Background/foreground
  - Connected components
  - Cluster: groups of touching chromosomes
  - Cluster recognition/decomposition
    - Touching chromosomes
    - Overlapping chromosomes
- Classification





M-FISH Segmentation

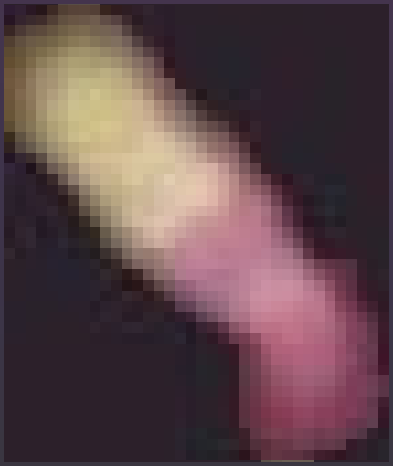
- Multi-spectral data has more segmentation information than boundary data alone
- Previous M-FISH methods
  1. Segment with DAPI (grayscale) channel
  2. Classify multi-spectral information



Raw Image

Translocations

- Exchange of material between two types
- More visible in M-FISH images
- Appear as objects with two different colors



M-FISH

Outline

- Introduction
- Contributions
  1. Chromosome segmentation using multi-spectral information
  2. Joint segmentation and classification
  3. Aberration scoring
- Results
- Conclusions

Problem Formulation

- C_i: the set of all pixels belonging to class i
- A_i^n: the set of pixels belonging to the n^{th} chromosome of class i
- $A_i^n \subseteq C_i$
- Segmentation-classification
  - Estimating each set A_i^n
  - Segmentation: Proposing a set of pixels, A'
  - Classification: Proposing a class, i

Proposed Approach

- Develop a measure of quality for segmentation-classification possibilities
  - Must be a function of both segmentation and classification
  - Measure is also a likelihood
- Choose a reasonable set of segmentation-classification possibilities
- Maximize the measure over the set of possibilities

Maximum Likelihood Formulation

- Proposed likelihood function for single candidate chromosome is a combination of several functions

$$\boxed{L(A',i) = L_{multi}(A',i)L_{size}(A',i)w(A')}$$

- $L_{\text{multi}}(\cdot,\cdot)$: multi-spectral likelihood function
- $L_{\text{size}}(\cdot,\cdot)$: size likelihood function
- $w(\cdot)$: weighting function
- i: class (classification)
- A': candidate chromosome (segmentation)

Multi-spectral Information

$$L_{multi}(A', i) = \frac{1}{|A'|} \sum_{m \in A'} p\left(\mathbf{m} \in C_i \middle| \mathbf{x}(\mathbf{m})\right)$$

- Average of individual pixel probabilities
- Bayesian pixel classifier returns probabilities for each class [Sampat, Castleman, and Bovik, 2002]
- C_i: set of all pixels of class i
- **m**: pixel
- **x**(**m**): multi-spectral image data at pixel **m**

Size Information

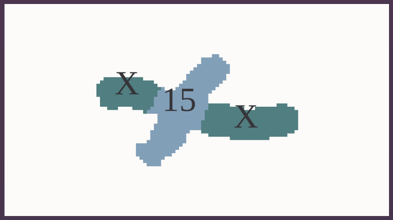
$$L_{size}\left(A',i\right)=e^{-\frac{\left(\frac{\left|A'\right|}{\sum_{n}\sum_{j}\left|A_{j}^{n}\right|}-\mu_{i}\right)^{2}}{2\sigma_{i}^{2}}}$$

- Likelihood function is a Gaussian that peaks at mean size of class *i*
- The size means and variances of the class *i* are μ_i and σ_i
- A_i^n: the set of pixels belonging to the nth chromosome of class

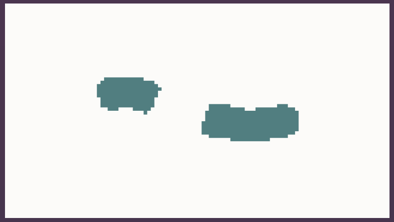
Weighting Function

- Measures the certainty of the likelihoods
- $w(\cdot)$ is the percentage of visible, or non-overlapped, pixels in the candidate chromosome
- Forces more certain non-overlapped chromosomes segments to be combined first
- Precludes possibility of a segment being left out of the middle of a chromosome

Estimating Area of Overlap

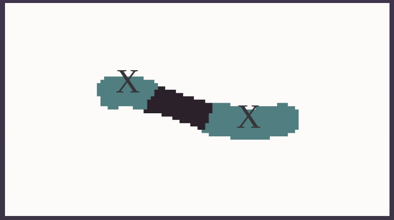


Overlapping Chromosomes



Chromosome Ends

Border Pixels (in black)



Overlapped area estimated (in black)

Weighting Function Example

- Yellow areas represent two possible segmentations for a single chromosome
- The function $w(\cdot)$ gives more weight to b

a) Incorrect Segmentation

b) Correct Segmentation

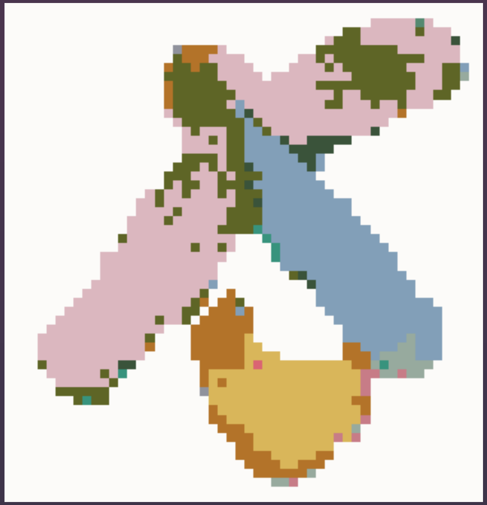
Segmentation Implementation

- Use multi-spectral information to determine segmentation possibilities
- Strategy: Oversegment and merge segments
  - Use pixel classification and post-processing to determine initial segments
  - Merge segments as long as the merging increases the proposed likelihood function

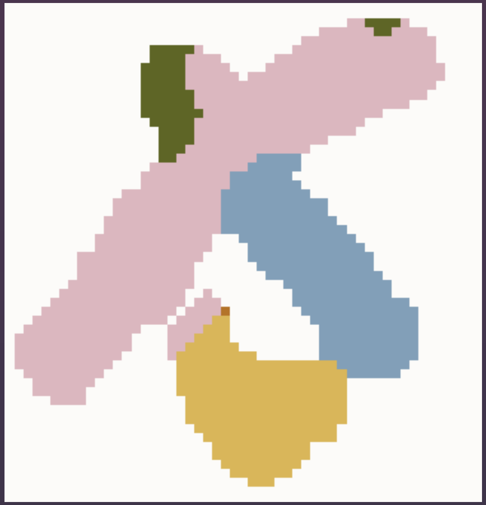
Example

Cluster to be decomposed

Determining Initial Segments



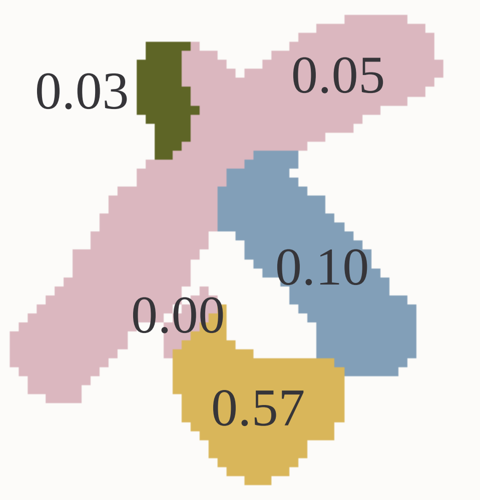
Pixel
Classification



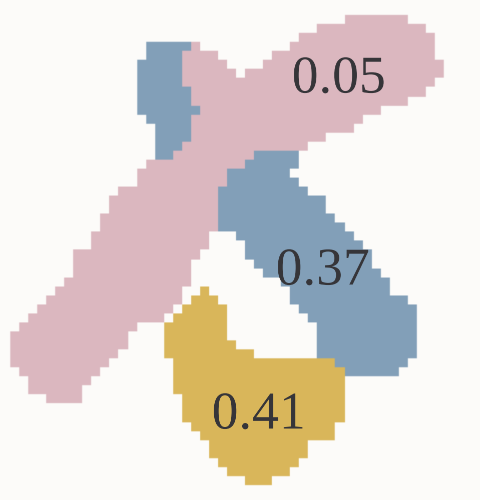
Majority
Filtering

Small Segment
Reclassification

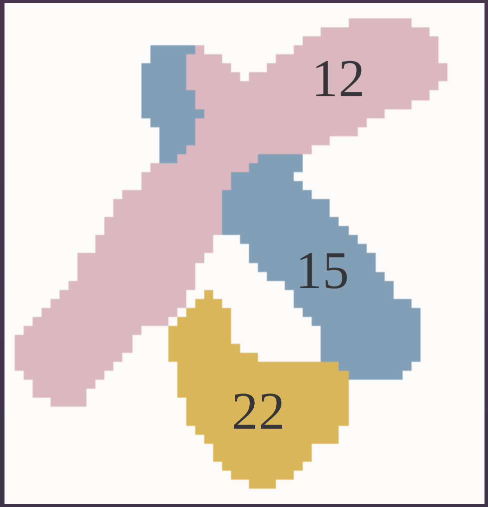
Merge Segments



Initial Likelihood Values



Merged Segments



Classification

Aberration Scoring

- Aberration scoring: assigning a value to the likelihood of abnormality
- Design of likelihood function has allowed for straightforward aberration scoring
  - Segments with low likelihood can be flagged as likely abnormalities
  - Low likelihood values also identify incorrect segmentation and classification
  - Likelihood values help direct user

Outline

- Introduction
- Contributions
- Results
  - Segmentation
  - Classification
  - Aberration Scoring
- Conclusions

Comparison

- Data: Advanced Digital Imaging Research M-FISH Chromosome Image Dataset
  - 200 M-FISH images
  - Wide variety of images
  - Many difficult images
  - Includes hand segmented ground-truth images
  - Freely available on web: `http://www.adires.com`
- Competition: Compare against user-guided Cytovision grayscale segmentation software [2001]

Segmentation Results

	Proposed Method	**Cytovision (Grayscale)**
Touches	77%	58%
Overlaps	34%	44%
Singles Oversegmented	0.8%	0.2%

More Results

Recognized as Clusters	**Proposed Method**	**Cytovision**
Clusters	95%	69%
Singles	6%	0.4%

	Proposed Method	**Only Pixel Classification**
Misclassified	8.1%	15%

Aberration Scoring with Proposed Method

	Normal Chromosomes	**Abnormalities**	
		Translocations	**Fragments**
Likelihood μ	0.44	0.12	0.02
Likelihood σ	0.24	0.10	0.02
< 0.1 likelihood	4.9%	49%	100%
< 0.3 likelihood	34%	96%	100%

V29 subset of ADIR dataset

15 images with 5 translocations each

(Cytovision does not perform aberration scoring)

Error Detection with Proposed Method

	< 0.1 likelihood
Abnormal	49.1%
Incorrect Segmentation	52.6%
Incorrect Classification	48.6%
Correct Segments	6.4%

Contributions

- Derived single, unified maximum likelihood hypothesis test framework
- Decomposed chromosome clusters using M-FISH multi-spectral data
- Combined segmentation and classification for increased accuracy in both
- Demonstrated effective aberration scoring
- Implemented joint segmentation-classification algorithm in C (2-3 minutes/image on 167MHz Unix machine)

Future Work

- Improvements in likelihood function
  - Shape
  - Number
- Pixel classification
- Overcome “greedy” algorithm difficulties
- Combine geometric, grayscale, and multi-spectral information for complete algorithm