Image Segmentation by Energy and Related Functional Minimization Methods

Eric Mason

July 12, 2006

University of Missouri – St. Louis

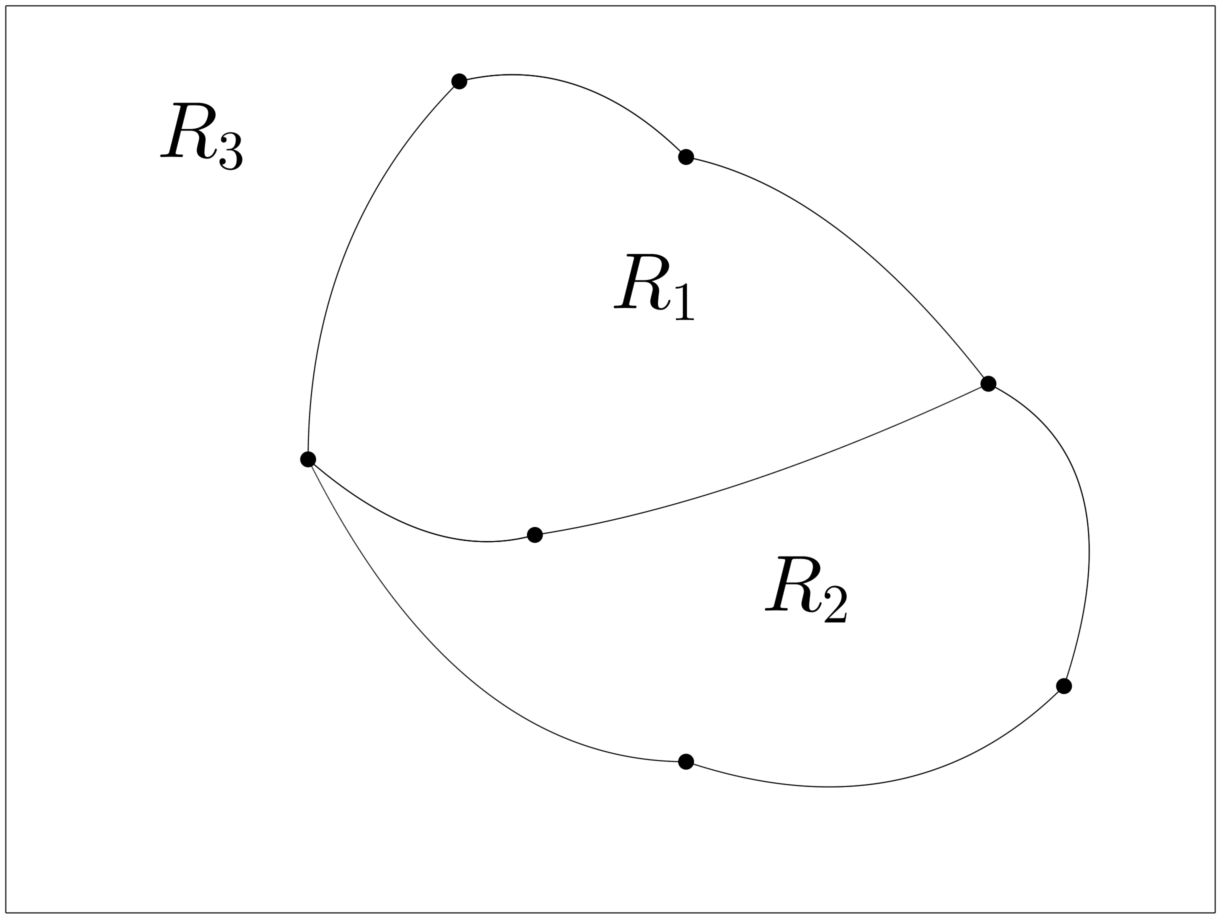
Presentation Outline

- Introduction
- Background
- Theoretical results
- Algorithm description
- Experimental results
- Conclusions and future work

The Mathematical Model of an Image

- $u\colon \Omega \to \mathcal{I}$
- $\Omega \subset \mathbb{R}^2$ bounded, simply connected
- $\mathcal{I} \subset \mathbb{R}$ is the intensity
- u is measurable

Segmentation - Intuitive Meaning . . .



. . . More Precisely

Definition[1;4] Let u_0 be a real valued observed image defined on a bounded, open, simply connected subset Ω of $\mathbb{R}^2$. A segmentation of u_0 is a finite collection of open subsets of Ω, $R_1, R_2, \ldots, R_n$, and a function u such that

1. The boundary ∂R_l of each R_l has finite length,

2. $\overline{\Omega} = \overline{R_1} \cup \overline{R_2} \cup \cdots \cup \overline{R_n}$,

3. $u\colon \Omega \to \mathbb{R}$ approximates u_0.

The Functional of Mumford and Shah

Given an observed u_0, minimize

$$E(u,K) = \mu^2 \iint\limits_{\Omega} (u-u_0)^2\,dxdy + \iint\limits_{\Omega\setminus K} |\nabla u|^2\,dxdy + \nu|K|$$

where K is the edge set that defines the regions in the segmentation and $|K|$ is its length.

What do the Terms Mean?

- $\mu^2 \iint\limits_{\Omega} (u - u_0)^2 \, dxdy$: is $u \approx u_0$?
- $\iint\limits_{\Omega \setminus K} |\nabla u|^2 \, dxdy$: is u smooth in the regions?
- $\nu |K|$: is the length of the boundary small?

Are all the Terms Necessary?

M&S write, "... drop any term and $\inf_u(E) = 0$:

- without the first, take $u = 0$, $K = \emptyset$;
- without the second, take $u = u_0$, $K = \emptyset$;
- without the third, take K to be a fine grid of N horizontal and vertical lines, $R_i = N^2$ small squares, $u =$ average of u_0 on each R_i.

... all three make E interesting."

On the Existence of Solutions

Little is known in general. M&S conjecture:

$$u_0 \in C(\Omega) \Rightarrow \exists\ (u^*, K^*)$$

$$\in \{(u, K) \colon |\nabla(u|R_i)| < \infty,$$

$$K \text{ a finite set of points joined by } C^1 \text{ arcs}\}$$

$$\text{such that } (u^*, K^*) \text{ minimizes } E.$$

Still unsolved, but there is a special case . . .

The Piece-wise Constant Model

Theorem[1;12] Let Ω be an open rectangle in $\mathbb{R}^2$, and let u_0 be continuous on $\Omega \cup \partial\Omega$. For every one dimensional $K \subset \Omega$ such that $K \cup \partial\Omega$ is made up of a finite number of C^1 arcs meeting only at end points and for all locally constant u on $\Omega \setminus K$ let

$$E_0(u, K) = \iint_{\Omega} (u - u_0)^2 \, dxdy + \lambda |K|.$$

Then there exists a u and a K that minimize E_0.

The 2-phase CV Algorithm*

$$u(x, y) = c_1 \, 1_{R_1}(x, y) + c_2 \, 1_{R_2}(x, y)$$

$$E_0(u, K) = \iint\limits_{R_1} (u - c_1)^2 \, dxdy + \iint\limits_{R_2} (u - c_2)^2 \, dxdy + \lambda |K|$$

---

* Chan and Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 2001.

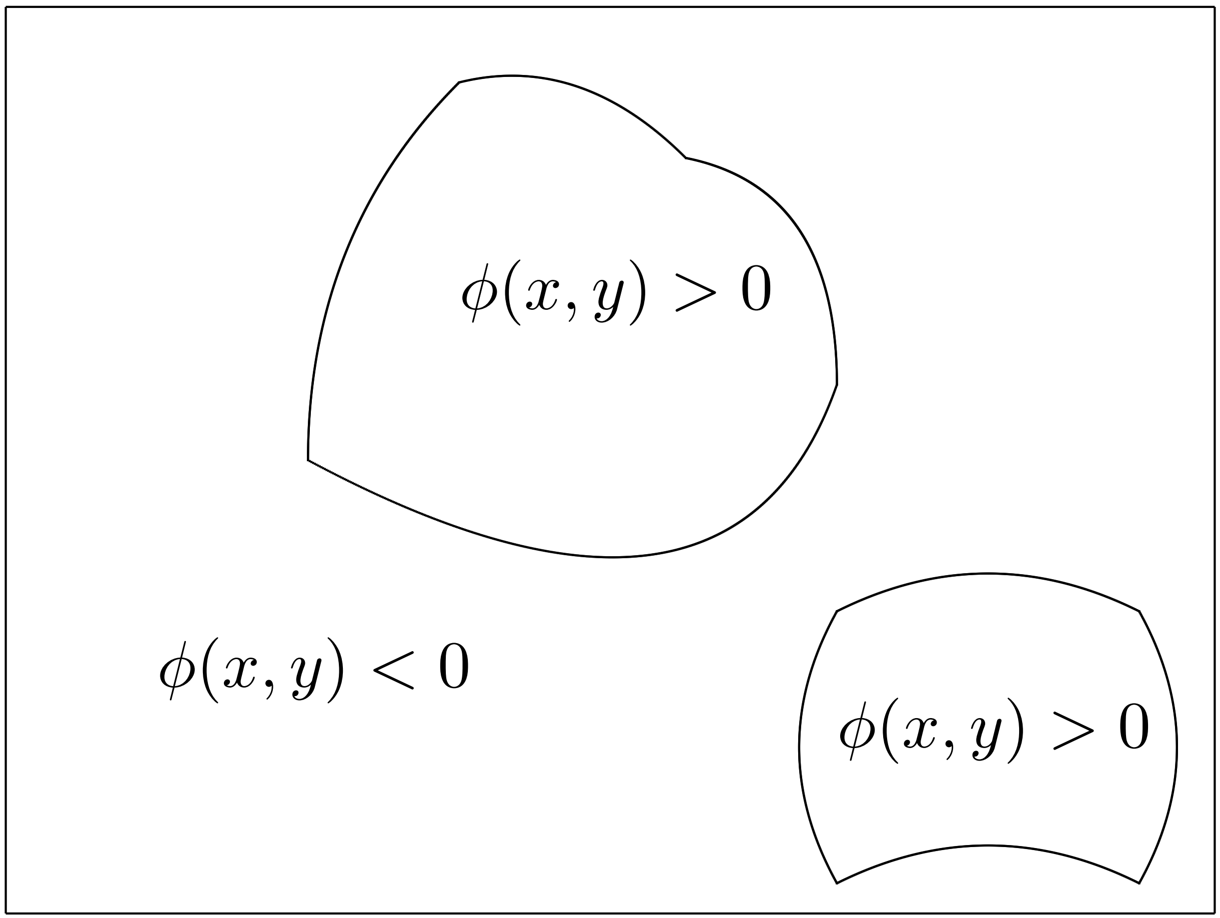
The 2-phase CV Algorithm - cont'd

Fix K then $c_i = \mathrm{mean}\{u(x,y)\colon (x,y) \in R_i\}$

- Use a level set function ϕ to define R_1 and R_2.
- Write $E_0(u,K)$ in terms of ϕ.
- Derive the Euler-Lagrange equation.
- Solve with gradient descent.

The Level Set Method

Region boundaries are the zero level sets of ϕ.



The Level Set Method-2

Characteristics of $\phi(x, y)$

- Satisfies a Lipschitz condition
- Typically $+/-$ distance to $\{(x, y) \colon \phi(x, y) = 0\}$
- Initial condition required
- Maintain value during evolution

The Method of Song and Chan*

Work with $E_0(u, K)$ directly and exploit

$$\overline{x}_{N+1} = \frac{N\overline{x} + x_{N+1}}{N+1}$$

$$S_{N+1} = S_N + \frac{N}{N+1}(x_{N+1} - \overline{x})^2.$$

Observe that only $\operatorname{sign}\big(\phi(x, y)\big)$ is needed.

---

* Song and Chan. A fast algorithm for level set based optimization. *UCLA CAM-02-68*, 2002.

The Method of Song and Chan - 2

(an example of a greedy algorithm)

- Pick ϕ_0, compute c_1 and c_2.
- For each pixel, how will $E_0(u, K)$ change if the pixel is moved to the other partition?
- If the energy will decrease then move the pixel and recompute c_1 and c_2.
- Repeat until an entire pass through the image moves no pixels from one partition to the other.

The Main Idea of the Thesis

Is there a solution path with other fidelity measures? Specifically, minimize

$$E_0^{(1)}(u,K) = \iint\limits_{\Omega} |u - u_0| \, dxdy + \lambda |K|$$
$$= \|u - u_0\|_1 + \lambda |K|.$$

- Classical methods don't apply directly
- Requires a regularization scheme

Why Use the L_1 Norm

- Sensitivity to noise
- Recent research in image restoration
- Sharper edges
- Challenging problem

The ROF Functional*

Developed to restore images corrupted by noise.

A different functional to minimize.

$$\iint\limits_{\Omega} (u - u_0)^2 \, dxdy + \lambda \iint\limits_{\Omega} |\nabla u| \, dxdy$$

---

* Rudin, Osher, Fatimi. Nonlinear total varation based noise removal algorithms. *Physica D*, 1992.

The ROF Model with L_1*

Minimize $\|u - u_0\|_1 + \lambda \| \sqrt{u_x^2 + u_y^2} \|_1$

- Contrast invariant
- More geometric
- Same as $E_0^{(1)}$ with binary images

---

* Chan, et al. Aspects of total variation regularized L_1 function approximation. *SIAM Jour. App. Math.*, 2005

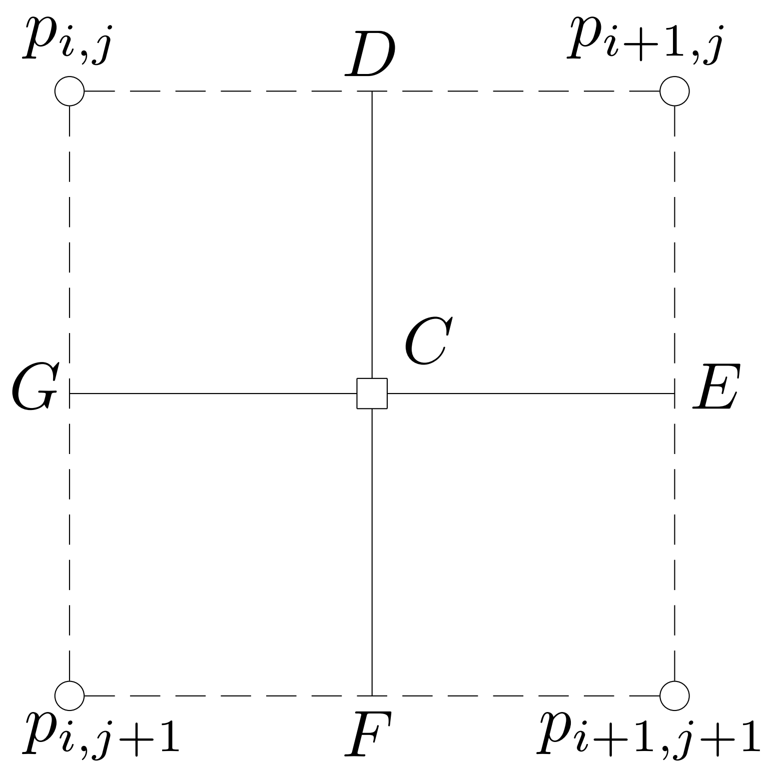
The Discretization of $E_0^{(1)}$

Assume integer lattice for image domain.

$$
\begin{aligned}
E_0^{(1)}(u, K) &= \|u - u_0\|_1 + \lambda |K| \\
&= \sum_{i=1}^{n} \iint_{R_i} |u - c_i| \, dxdy + \lambda |K| \\
&\approx \sum_{i=1}^{n} \sum_{(j,k) \in R_i} |u_{jk} - c_i| + \lambda |K|
\end{aligned}
$$

The result with l_2 fidelity is similar.

The Discretization of $|K|$

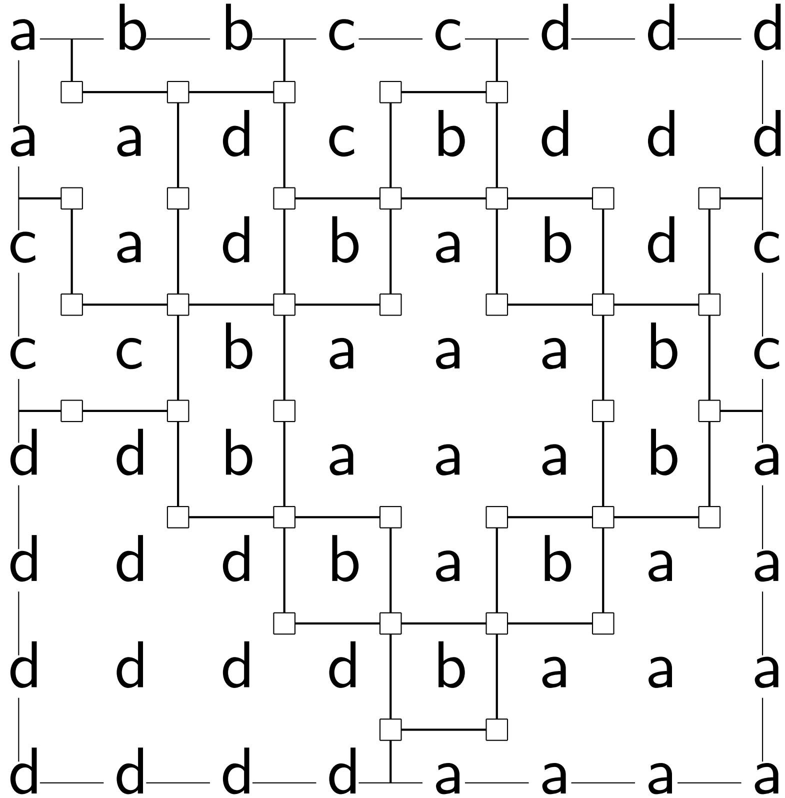


- pixels indicated by $p_{i,j}$
- each region contributes
- based on adjacent pixels
- segmentation components

The Boundaries in a Small Image

a	b	b	c	c	d	d	d
a	a	d	c	b	d	d	d
c	a	d	b	a	b	d	c
c	c	b	a	a	a	b	c
d	d	b	a	a	a	b	a
d	d	d	b	a	b	a	a
d	d	d	d	b	a	a	a
d	d	d	d	a	a	a	a

The Boundaries in a Small Image



- edges at component boundaries only
- possible configurations:
  - turn
  - continuation
  - T-junction
  - mosaic
- no ‘termination’

The Median not the Mean

- If K is fixed what is c_i?
- In the case of l_2, c_i is the mean
- Not true in the case of l_1
- The minimizing statistic is *any* median

Left and Right Medians

Definition[3;38] Let $X = \{x_1, x_2, \ldots, x_N\}^*$ be any finite set of real numbers with $x_1 \leq x_2 \leq \cdots \leq x_N$† and $N = 2m + p$ where $p \in \{0, 1\}$. The right median, M_X^+, is the value of x_{m+1}. The left median, M_X^-, is the value of x_{m+p}.

---

* There is no requirement that the elements be unique.

† These will be the pixel intensity values u_{jk}.

The Medians Bound the Minimizers

Lemma[1;38] If $X = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}$ the quantity

$$D_X^c = \sum_{i=1}^{N} |x_i - c|$$

is minimized by any c satisfying $M_X^- \le c \le M_X^+$.

A “well” known fact from statistics. e.g., G. U. Yule.

An Introduction to the Theory of Statistics. Griffin, 1911.

The Total Deviation

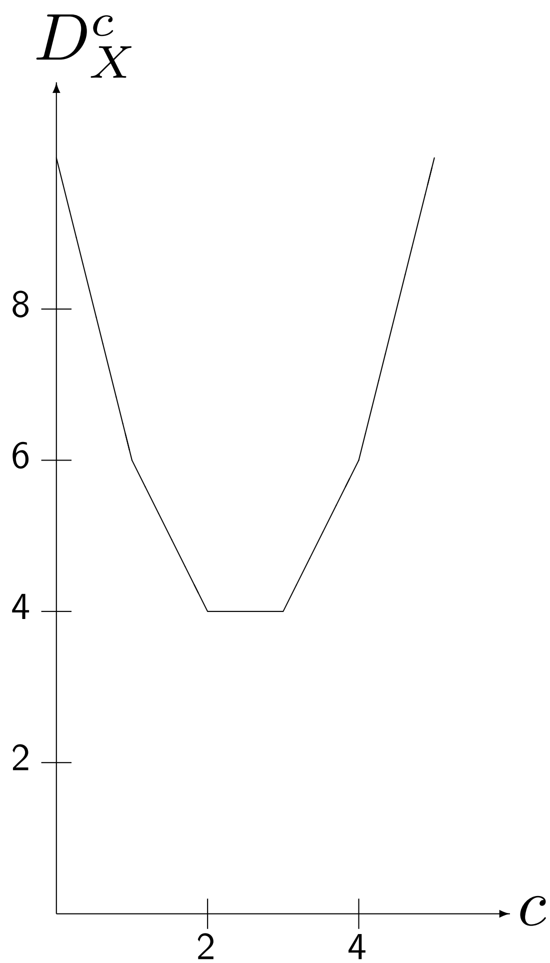
Definition[4:40] The minimum value of D_X^c given by the lemma is called the total deviation of X and is represented by the symbol D_X.

Note:

The value of c is not unique, but . . .

the quantity D_X is well defined.

Global Minimizers are not Unique



- e.g., $X = \{1, 2, 3, 4\}$.
- If K is fixed, then . . .
- Minimum $E_0^{(1)}(u, K)$ is unique.
- $u(x, y)$ is not unique.

Some Facts about D_X

Lemma[2;40] $x_1 \le x_2 \le \cdots \le x_N,\ N = 2m+p \Rightarrow$

1. $D_X = \sum_{i=m+1+p}^{N} x_i - \sum_{i=1}^{m} x_i$

2. $X' = X \cup \{y\} \Rightarrow D_{X'} \ge D_X$ and

$$D_{X'} - D_X = \begin{cases} y - M_X^+ & \text{if } y \ge M_X^+; \\ M_X^- - y & \text{if } y \le M_X^-; \\ 0 & \text{otherwise.} \end{cases}$$

Some Facts about D_X - cont'd

3. $X' = X \setminus \{x_k\} \Rightarrow D_{X'} \leq D_X$ and

$$D_{X'} - D_X = \begin{cases} x_k - M_X^+ & \text{if } x_k \leq M_X^-; \\ M_X^- - x_k & \text{if } x_k \geq M_X^+; \\ 0 & \text{otherwise.} \end{cases}$$

The lemma is analogous to formulas for mean and l_2 used with Song and Chan's algorithm

Partitions and Ordered Partitions

Definition[2;37] Let $X = \{x_1, x_2, \ldots, x_N\}$ be a nonempty finite set of real numbers. A partition of X is a collection of pairwise disjoint nonempty subsets of X, say $\{X_1, X_2, \ldots, X_n\}$, such that

$$\bigcup_{i=1}^{n} X_i = X.$$ The partition is *ordered* if

$$i < j \Rightarrow \max(X_i) \leq \min(X_j).$$

X_i and X_j are *adjacent* if $|i - j| = 1$.

The 'Closeness' of a Simple Partition

Theorem[3;46] Let $X = \{x_1 \leq x_2 \leq \cdots \leq x_n\}$. Suppose X is partitioned into two subsets Y and Z. Let $S = D_Y + D_Z$. If S is the minimum value over all such partitions then Y and Z can be written as $Y = \{x_1, x_2, \ldots, x_s\}$ and $Z = \{x_{s+1}, x_{s+2}, \ldots, x_N\}$ with $1 \leq s < N$ and $x_s \leq \frac{1}{2}(M_Y^+ + M_Z^-) \leq x_{s+1}$.

Theorem applies directly to the 2-phase case.

The Mapping $S(s)$

Partitions of the form

$$Y_s = \{x_1, x_2, \ldots, x_s\}$$
$$Z_s = \{x_{s+1}, x_{s+2}, \ldots, x_N\}$$

characterize the minimum value of S, so define

$$S\colon \{0, 1, 2, \ldots, N\} \to \mathbb{R}^+ \text{ by } S(s) = D_{Y_s} + D_{Z_s}.$$

The smallest possible S is a value of $S(s)$.

Generalization to Larger Partitions

Theorem[4;49] Let P_n be the set of all partitions of X of size n, and let $P_n^* \subset P_n$ be the set of all ordered partitions of X of size n. For $p \in P_n$ let $S_p = D_{X_1} + D_{X_2} + \cdots + D_{X_n}$.

1. $S_n = \min\limits_{p \in P_n}(S_p),\ S_n^* = \min\limits_{p \in P_n^*}(S_p) \Rightarrow S_n^* = S_n$

2. $p^* \in P_n^*,\ S_{p^*} = S_n,\ \text{adjacent } X_i, X_{i+1} \in p^* \Rightarrow$

$$\max(X_i) \leq \frac{1}{2}(M_{X_i}^+ + M_{X_{i+1}}^-) \leq \min(X_{i+1})$$

Practical Significance

Search for minimum partition simplified

- $|P| = \mathcal{S}_N^{(n)} = O(n^N)$ - exponential growth*
- $|P^*| = O(N^{n-1})$ - polynomial for fixed n
- Provides an initial condition for segmentation

---

* $\mathcal{S}_N^{(n)}$ is a Stirling number of the second kind.

Application to Segmentation

Recall l_1 version of M&S functional.

$$E_0^{(1)}(u,K) = \sum_{i=1}^{n} \sum_{(j,k)\in R_i} |u_{jk} - c_i| + \lambda |K|$$

- The regions induce a partition of the pixel values
- Use Theorem 4 to select initial condition
- Use Lemma 2 in a greedy algorithm

Some Background on Theorem 4

Is this a new result?

- Closely related to cluster analysis
- Fisher proved part 1 in the case of l_2*
- Hwang proved part 1 in the case of l_p†
- Part 2 appears to be new.

---

* On grouping for maximum homogeneity. *JASA*, 1958

† Optimal partitions. *Journ. Opt. Theory and App.*, 1981

Application Principles

- Direct use of theorem 4 requires a sort.
- Therefore, use a histogram of $u(R_i)$.
- Only compute S for distinct values of u.
- Apply lemma 2.

Data Structures

For a set 'X' define 'H' with these components:

H.hist - the histogram of the values in X

H.chist - the cumulative values of H.hist

H.m - the floor of $N/2$

H.p - is either 0 or 1 so that $N=2*H.m + H.p$

H.MXM $=$ X(m+p) when X is sorted in ascending order

H.MXP $=$ X(m+1) with 1 as the first index

H.TD $=$ the total deviation from the median of X

The Algorithm

$n =$ the desired number of segmentation components.

$\mathcal{A} = \{r_1, r_2, \ldots, r_n\}$, the component identifiers.

$u_0 =$ the observed image.

$H_0 = \mathsf{InitH}(u_0)$.

$[c_1, c_2, \ldots, c_{n-1}] =$ split points for best fidelity.

$R =$ the segmentation defined by $[c_1, c_2, \ldots, c_{n-1}]$.

$[H_1, H_2, \ldots, H_n] =$ histograms of initial components.

The Algorithm - cont'd

repeat

for each pixel p

$r_p = R(p)$

for each $r_l \in \mathcal{A} \setminus \{r_p\}$

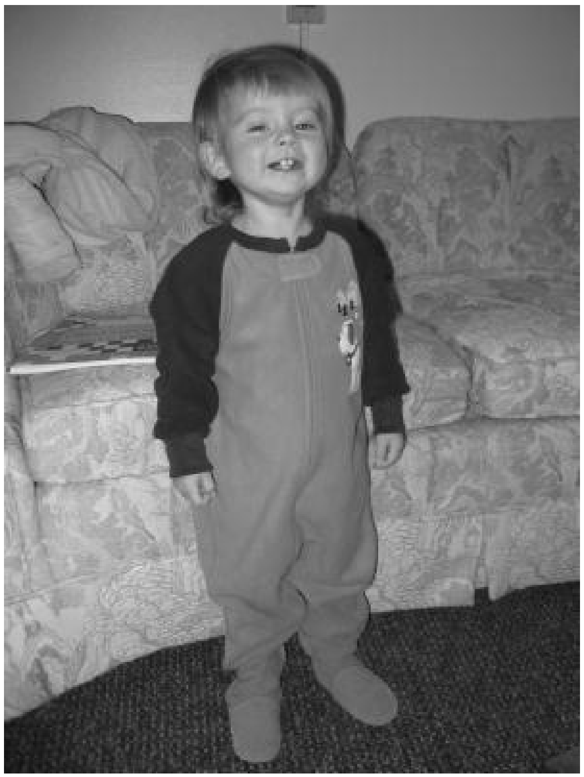
$D_l =$change in $E^{(1)}$ if $r_p = r_l$

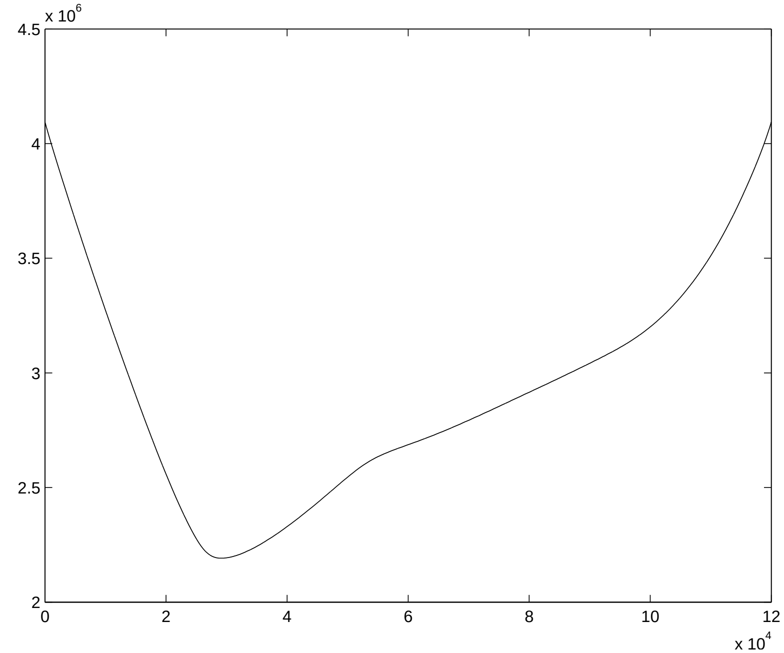
$D_{\min} = \min(D_l)$

if $D_{\min} < 0$ set $R(p) = r_{l_{\min}}$

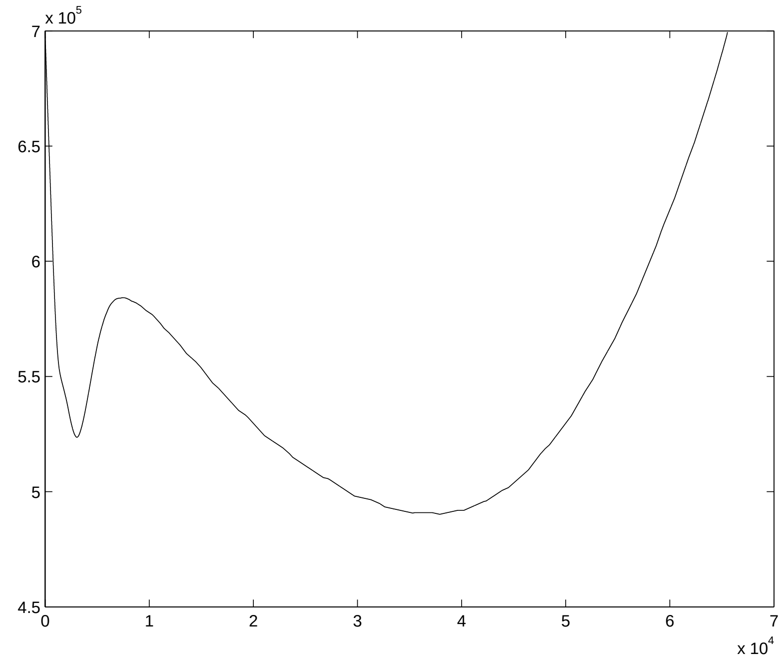
until no pixels are moved.

The S Curve of an Image

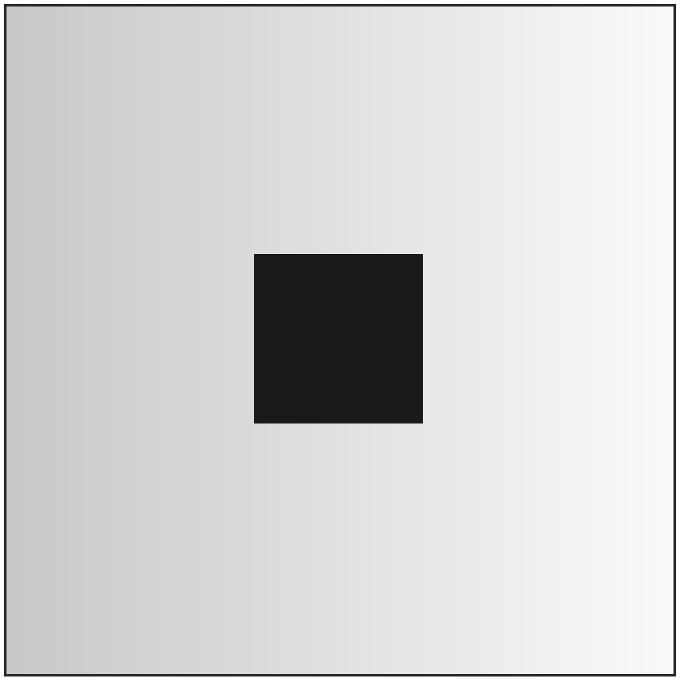


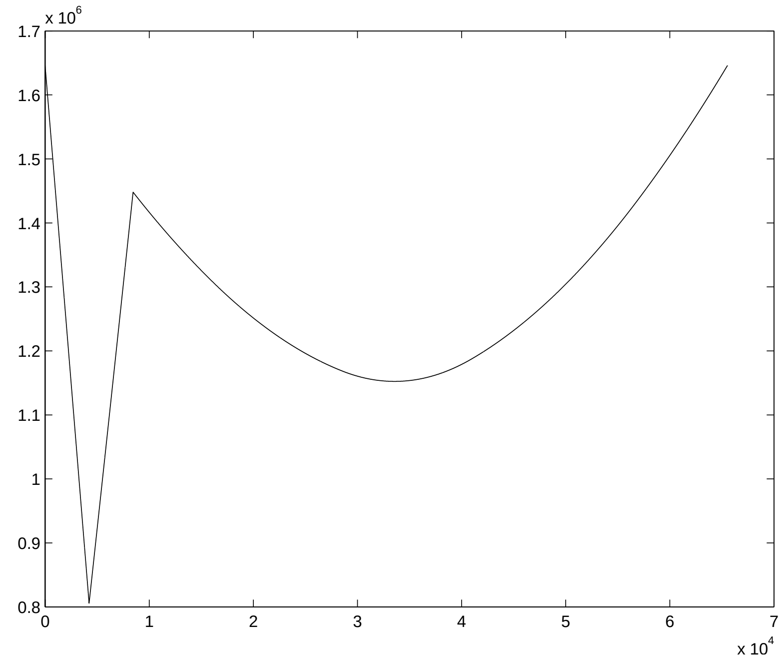


Local Minima in Fidelity

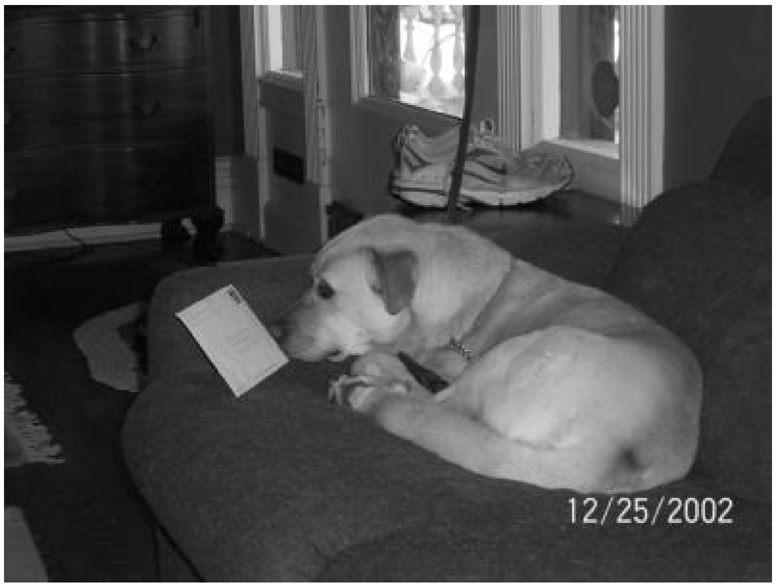


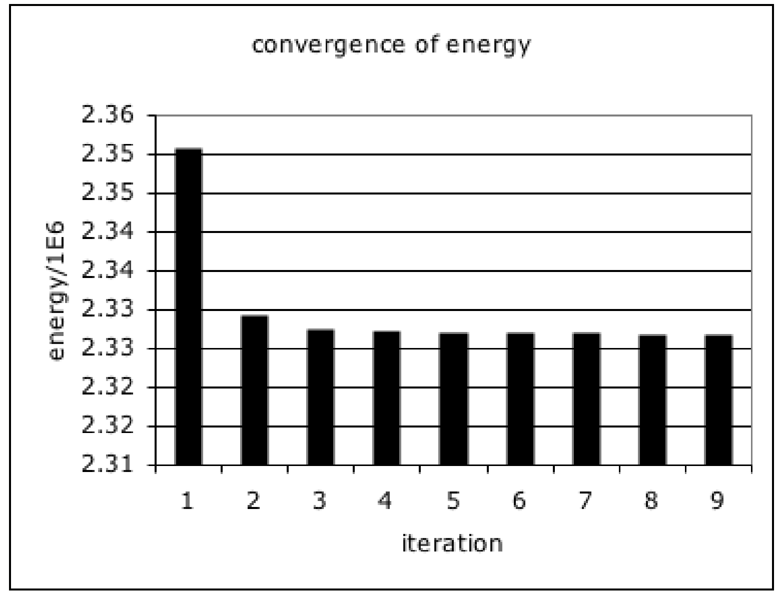
A Synthetic Image with Local Minima



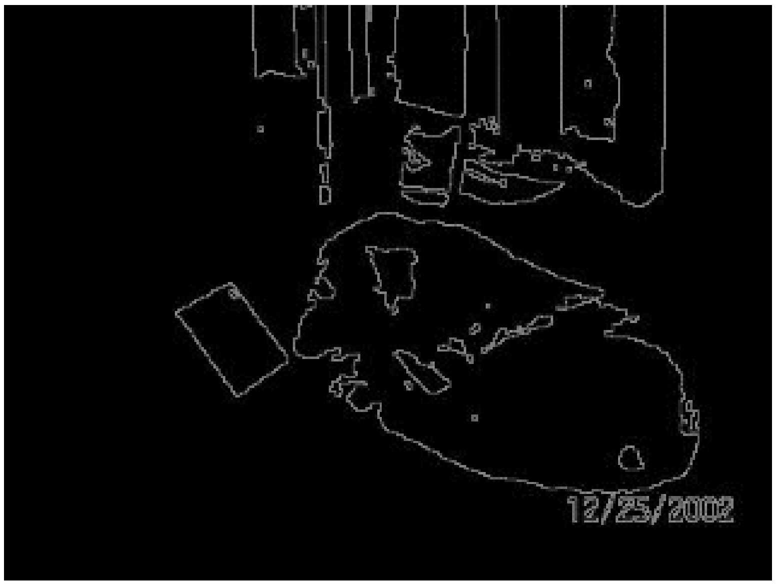


Convergence Behavior

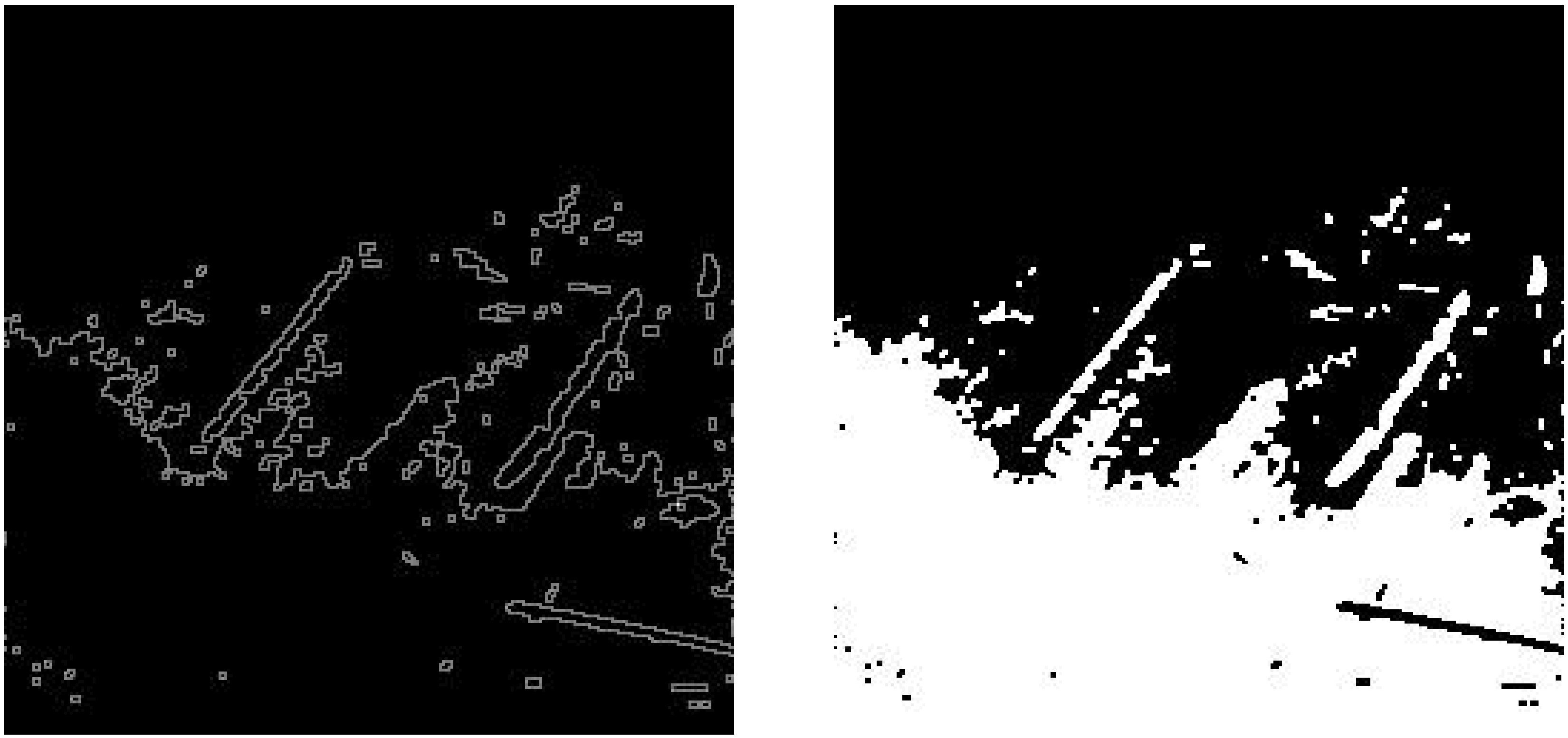




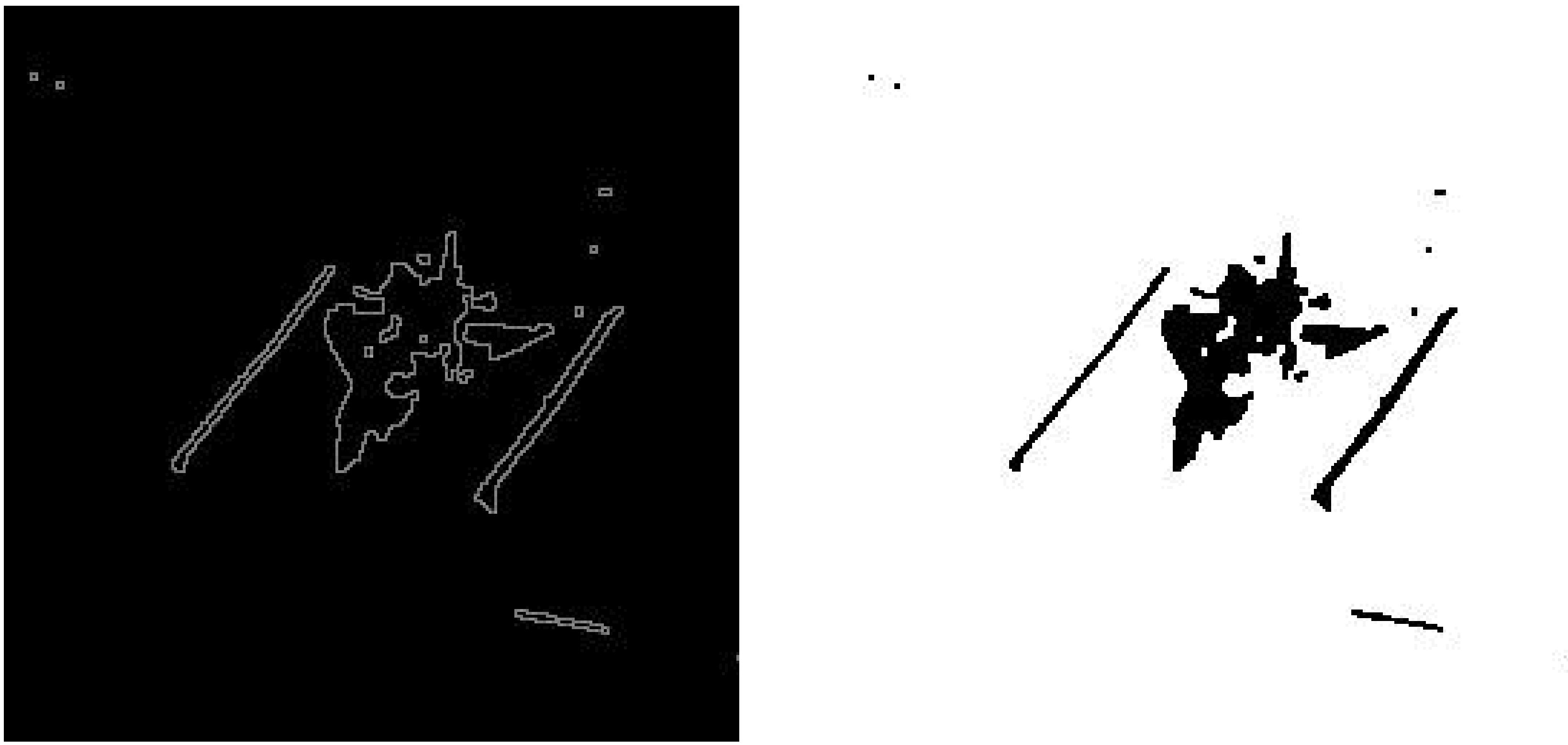
Segmentation Result



The Effect of Local Minima



Using the Correct Initial Condition



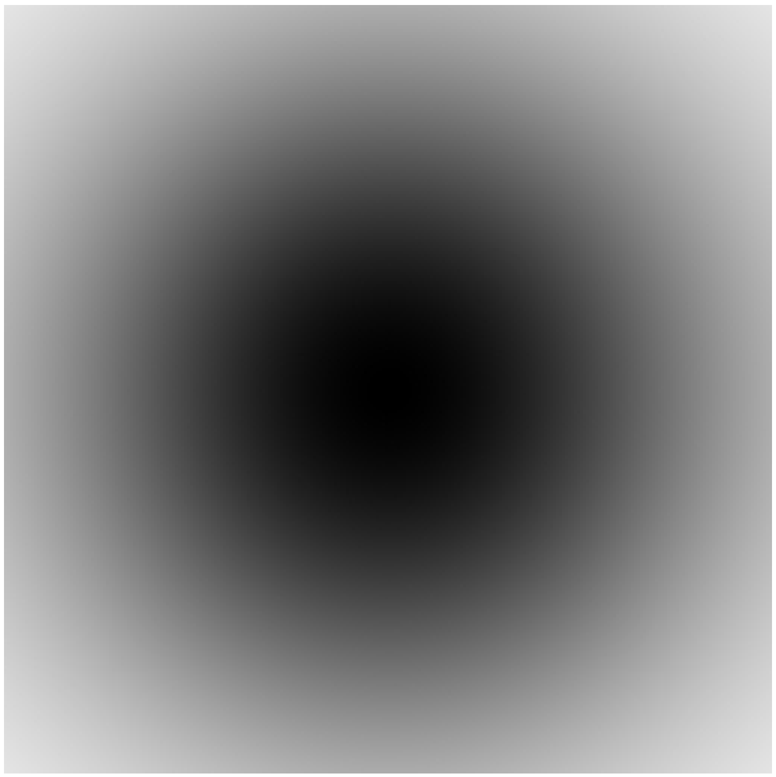
Benefit of Data Driven Initial Condition

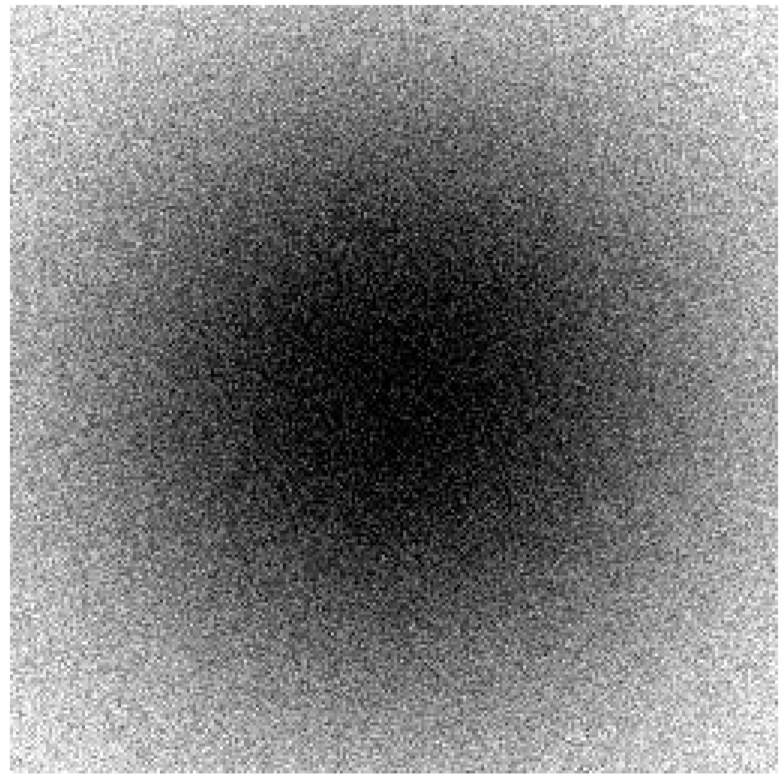
Final Approximating Image

Converging Edges - 1

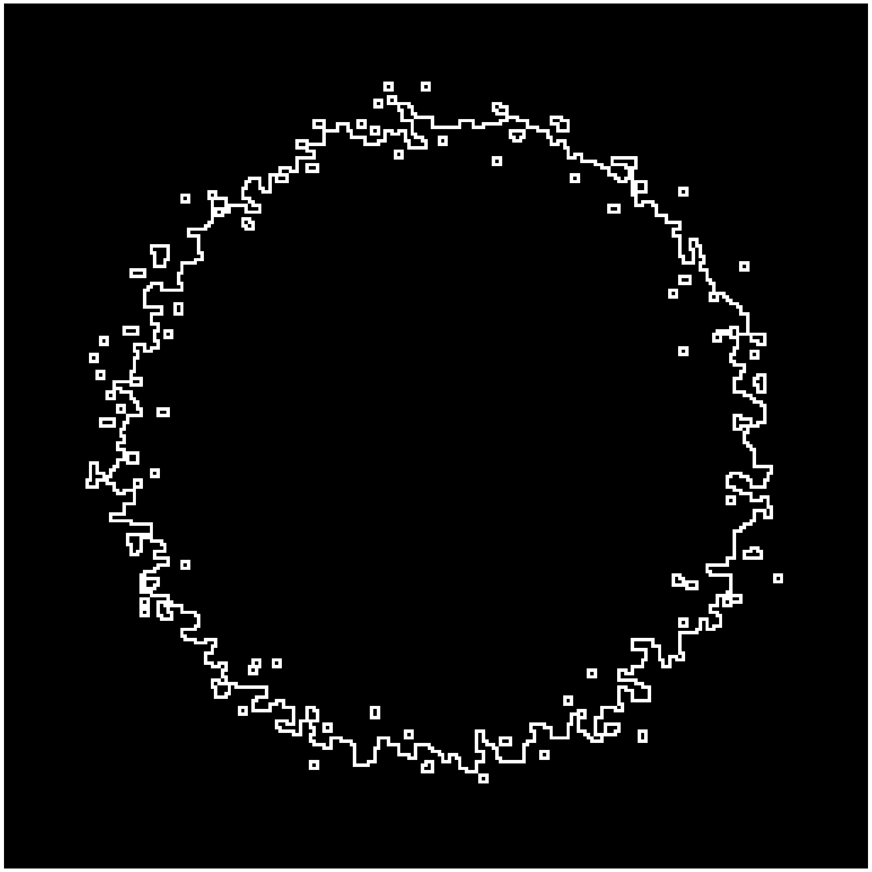
Converging Edges - 2

Synthetic Image with Noise



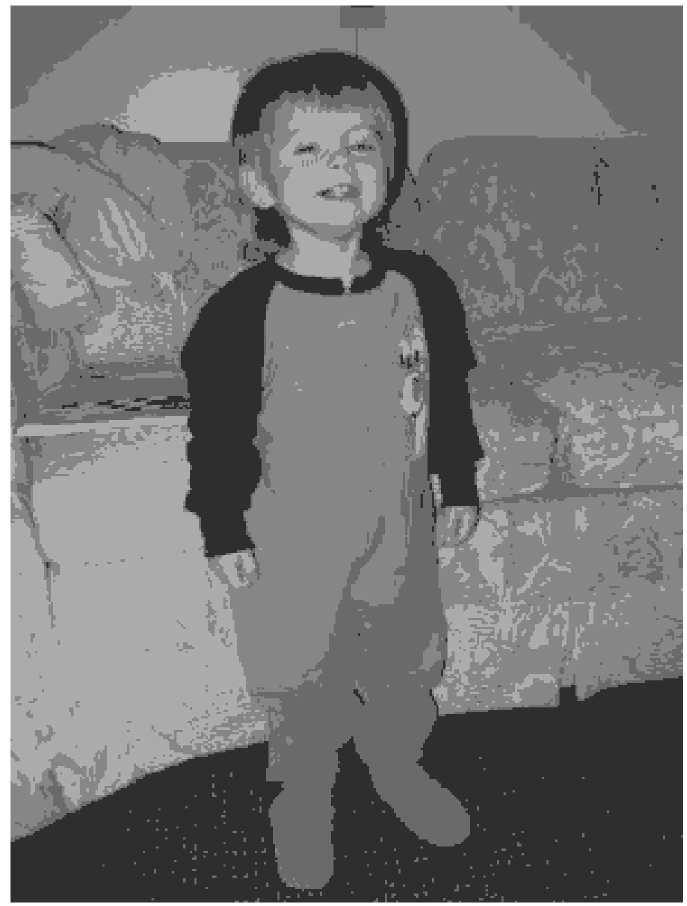


Edges with Noise



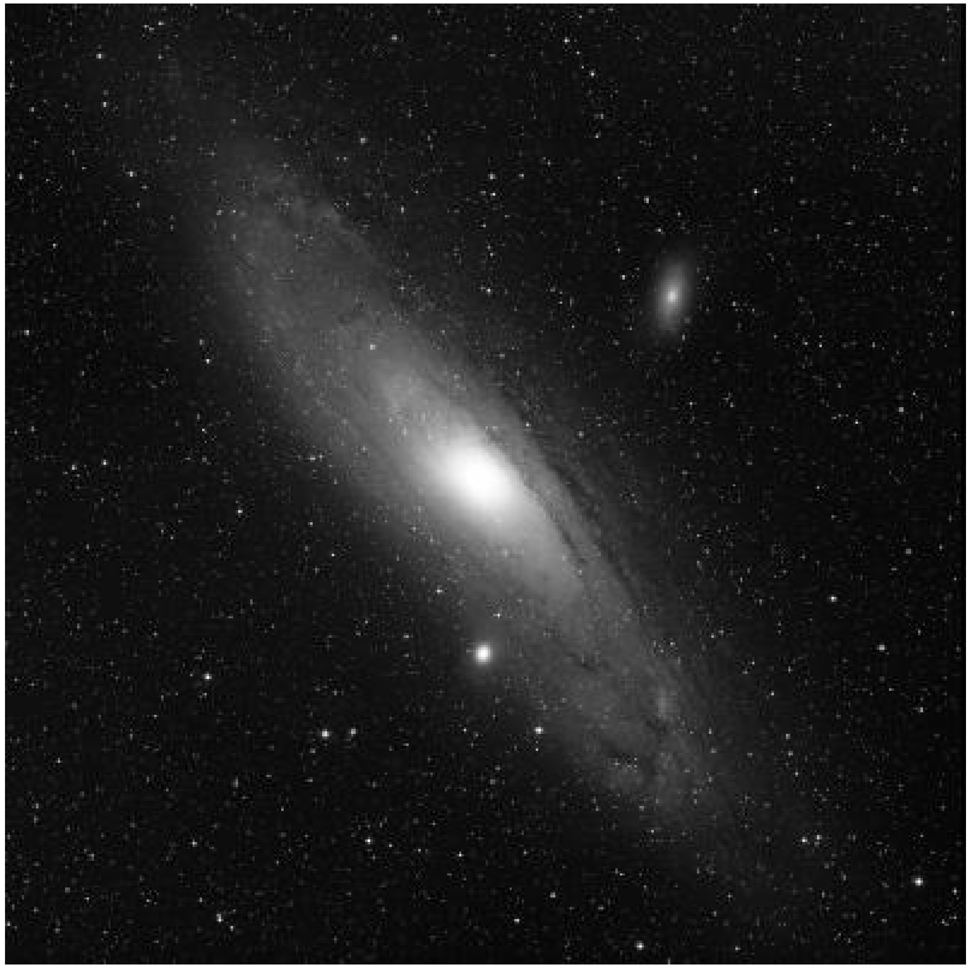
Natural Image with Noise

4 Phase vs. 2 Phase

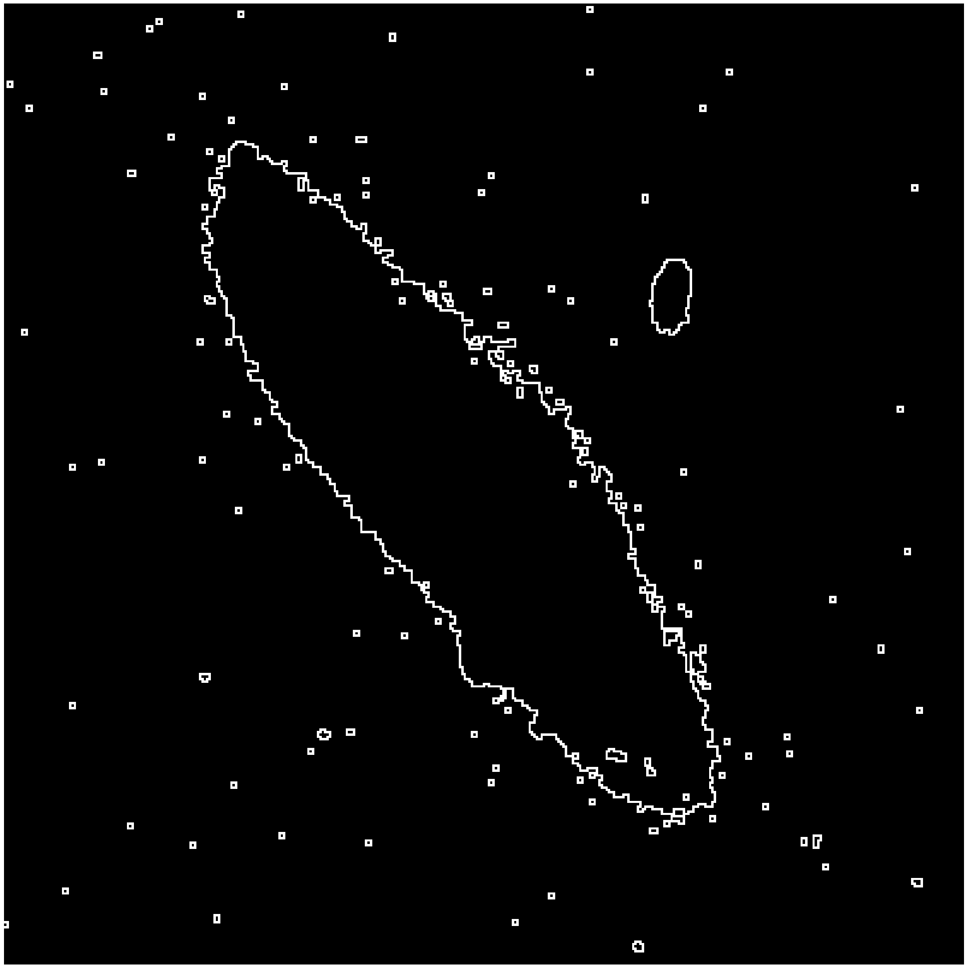


4 Phase Segmentation Edges

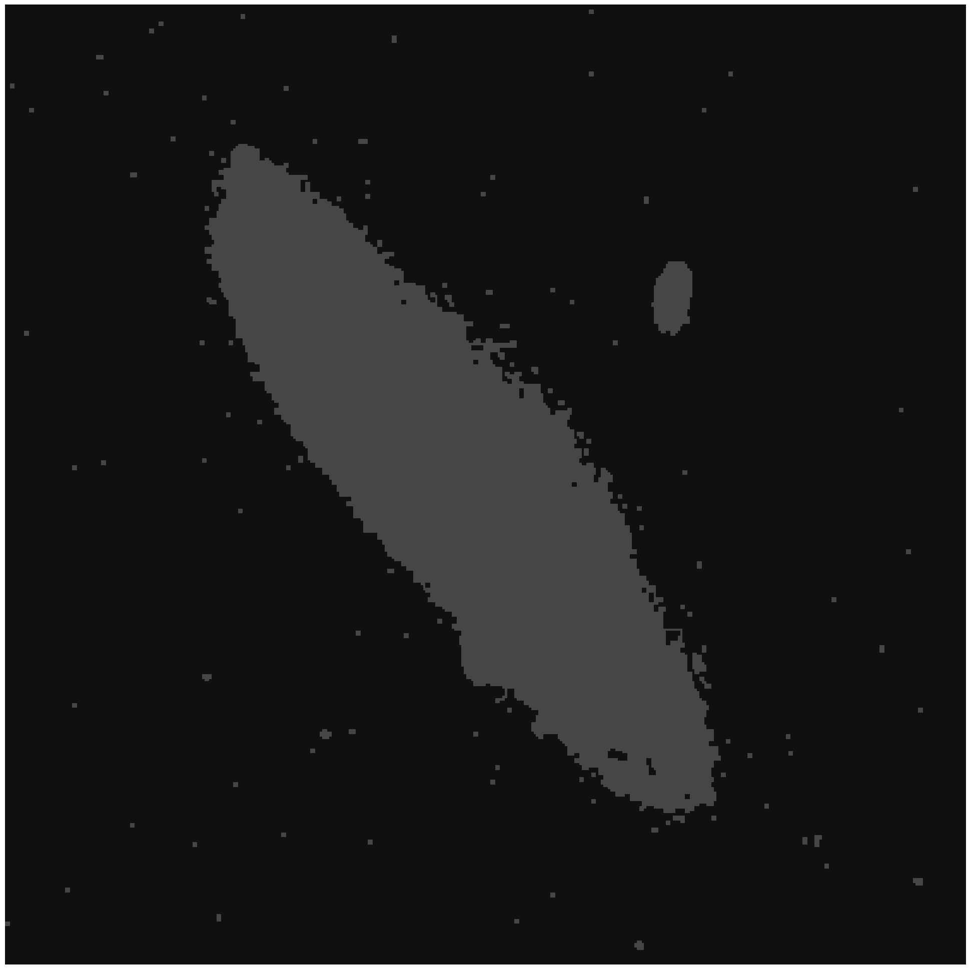
Andromeda - Image



Andromeda - Edges



Andromeda - Image



Conclusion

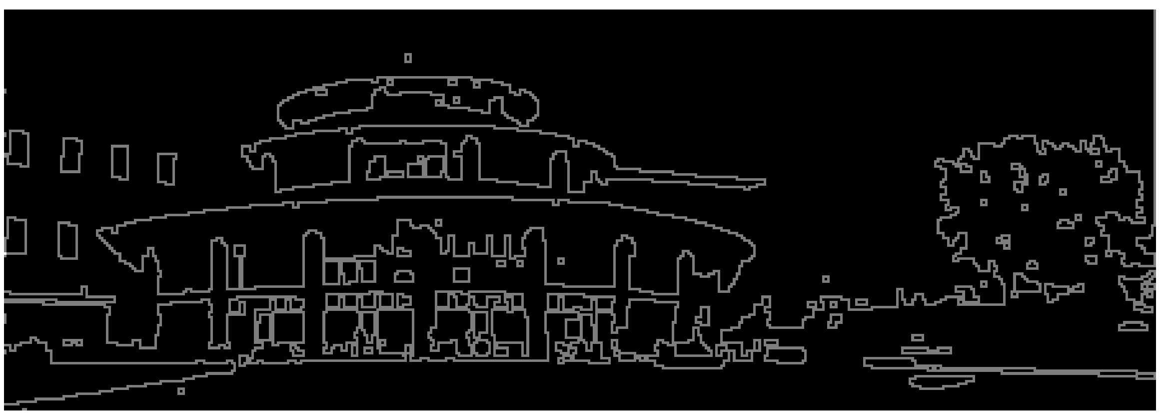
Major Contributions:

- M&S Model with l_1 norm for fidelity
- Data driven selection of initial condition
- Work directly with functional
- Avoid local minima traps

Future Work

- Explore the significance of the S curve
- Opportunities for parallelization
- Computational tools for l_p norms
- Data driven choice of number of phases

Questions?



Backup Slides

This Slide Blank

Level Set CV model - 1

Recall that $\phi(x, y)$ defines R_1 and R_2.

$$E(c_1, c_2, \phi) = \iint_{\phi(x,y)>0} (u_0 - c_1)^2 \, dxdy$$

$$\iint_{\phi(x,y)<0} (u_0 - c_2)^2 \, dxdy$$

$$+ \lambda \cdot \text{Length}(C)$$

Level Set CV model - 2

Use the Heaviside function $H(z) = 1_{z \geq 0}(z)$

$$
\begin{aligned}
E(c_1, c_2, \phi) = \int_\Omega (u_0 - c_1)^2 H(\phi(x,y))\, dxdy \\
+ \iint_\Omega (u_0 - c_2)^2 (1 - H(\phi(x,y)))\, dxdy \\
+ \lambda \cdot \iint_\Omega |\nabla H(\phi(x,y))|\, dxdy
\end{aligned}
$$

Level Set CV model - 3

Write E-L for fixed c_1 and c_2.

$$\delta(\phi(x,y))\Big[\lambda\,\mathrm{div}\left(\frac{\nabla\phi(x,y)}{|\nabla\phi(x,y)|}\right) - (u_0(x,y)-c_1)^2 - (u_0(x,y)-c_2)^2\Big] = 0$$

$\delta(\phi(x,y))$ is a Dirac delta function.

Evolve ϕ; recompute c_1 and c_2; repeat.

Level Set CV model - 4

Need a regularization of $H(z)$. C&V use the C^∞

$$H_\epsilon(z) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{z}{\epsilon}\right)\right)$$

with $\epsilon = 1$. They claim the non-compact support helps avoid local minima.

The Proof of Lemma 1

Consider $x_n \le c \le c' \le x_{n+1}$ and write

$$\begin{aligned}
D_X^{c'} - D_X^{c} &= \sum_{i=1}^{n}(c' - x_i) + \sum_{i=n+1}^{N}(x_i - c') \\
&\quad - \Big(\sum_{i=1}^{n}(c - x_i) + \sum_{i=n+1}^{N}(x_i - c)\Big) \\
&= n(c' - c) - (N - n)(c' - c) \\
&= (2n - N)(c' - c).
\end{aligned}$$

The Proof of Lemma 2

Many cases to consider - all computational. e.g., if N is odd and $y \geq M_X^+$ write

$$
\begin{aligned}
D_{X'} &= y + \sum_{i=m+2}^{N} x_i - \sum_{i=1}^{m+1} x_i \\
&= y - x_{m+1} + \sum_{i=m+2}^{N} x_i - \sum_{i=1}^{m} x_i \\
&= y - M_X^+ + D_X.
\end{aligned}
$$

The Proof of Theorem 3(2)

Assume Y and Z yield the minimum S. Suppose $y = \max(Y) > \frac{1}{2}(M_Z^- + M_Y^+) \to Z$. Use Lemma 2 to compute new S.

$$\begin{aligned} M_Z^- - y - (y - M_Y^-) &= M_Z^- + M_Y^- - 2y \\ &\leq M_Z^- + M_Y^+ - 2y \\ &= 2(\frac{1}{2}(M_Z^- + M_Y^+) - y) < 0 \end{aligned}$$

contradicting minimality of S.

The S Curve bounds S_p

S contains the minimum and the maximum.

$$
\begin{aligned}
S_p &= D_{X_1} + D_{X_2} + \cdots + D_{X_n} \\
&= \sum_{x_i \in X_1} |x_i - M^+_{X_1}| + \cdots + \sum_{x_i \in X_n} |x_i - M^+_{X_n}| \\
&\leq \sum_{x_i \in X_1} |x_i - M^+_{X}| + \cdots + \sum_{x_i \in X_n} |x_i - M^+_{X}| \\
&= D_X
\end{aligned}
$$