united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR1237/6

School on Mathematical Problems in Image Processing

(4 - 22 September 2000)

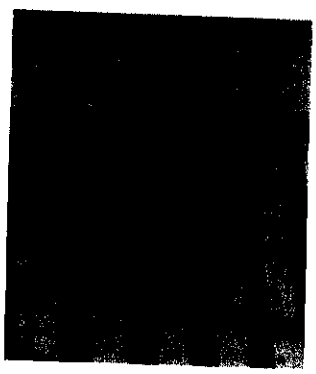
Visual Learning and Gibbs Reaction-Diffusion

S.C. Zhu
Computer and Information Science Department
Ohio State University
Columbus, OH 43210-1277
United States of America

Visual Learning and Gibbs Reaction-Diffusion

Song-Chun Zhu
Computer and Information Science Department
Ohio State University

Motivation



The input image The restored image

- How to compute or restore the background image?
- How to represent clutter–structured noise?
- Applications: Image segmentation, ATR ...

Why Visual Learning?

We assume:

$$\mathbf{I}^{\mathrm{obs}} = \mathbf{I} + \mathbf{I}^n$$

$\mathbf{I}^{obs}$ is the input image, $\mathbf{I}^n$ is the noise image, and $\mathbf{I}$ is the image to recover.

By Bayes Theory:

$$\begin{aligned} p(\mathbf{I}|\mathbf{I}^{\mathrm{obs}}) &\propto p(\mathbf{I}^{obs}|\mathbf{I})p(\mathbf{I}) \\ &= p(\mathbf{I}^n|\mathbf{I})p(\mathbf{I}) \\ &= p(\mathbf{I}^n)p(\mathbf{I}) \end{aligned}$$

We assume $p(\mathbf{I}^n|\mathbf{I}) = p(\mathbf{I}^n)$ for a given environment.

An example of buildings

An example of trees

What Is Visual Learning?

Given a set of training images, such as trees,

$$\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_M$$

We assume these images are random samples from an unknown distribution $p_{true}(\mathbf{I})$. The goal of viusal learning is to estimate a distribution $p(\mathbf{I})$ which we call the **MODEL** so that $p(\mathbf{I})$ is as close to $p_{true}(\mathbf{I})$ as possible.

Questions:

- How to estimate $p(\mathbf{I})$?
- How to measure the distance between $p(\mathbf{I})$ and $p_{true}(\mathbf{I})$?

How to estimate $p(\mathbf{I})$?

The estimation iterates in three steps:

1. Select a set of features $S = \{F_1, F_2, ..., F_n\}$, we can measure the statistics $\mu_1^{obs}, \mu_2^{obs}, ..., \mu_n^{obs}$ extracted by $F_1, F_2, ..., F_n$ and averaged over $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_M$
2. $p(\mathbf{I})$ is constrained to reproduce the observed statistics so that as far as the features $\{F_1, F_2, ..., F_n\}$ are concerned, we cannot distinguish $p(\mathbf{I})$ from $p_{true}(\mathbf{I})$.
3. The distance between $p(\mathbf{I})$ and $p_{true}(\mathbf{I})$ is measured by summing the distance between μ_K^{obs} and μ_K^{syn} for those $F_K \notin S$.

The whole process is guided by a minimax entropy theory.

Gibbs distribution of images

The Minimax entropy model – (Zhu, Wu, Mumford 1996)

$$p(\mathbf{I}; \Lambda, S) = \frac{1}{Z} e^{-U(\mathbf{I}; \Lambda, S)}$$

with potential function

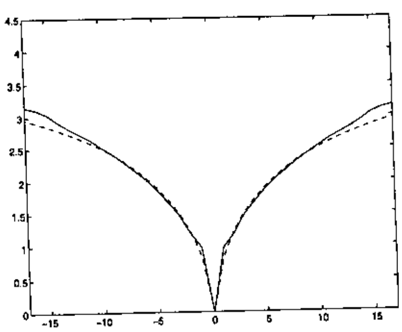
$$U(\mathbf{I}; \Lambda, S) = \sum_{i=1}^{n} \sum_{(x,y)} \lambda_i(F_i * \mathbf{I}_{(x,y)})$$

- $S = \{F_1, F_2, ..., F_n\}$
- $\Lambda = \{\lambda_1(), ..., \lambda_n()\}$, $\lambda_i()$ is approximated by a piecewise-constant function – a vector.

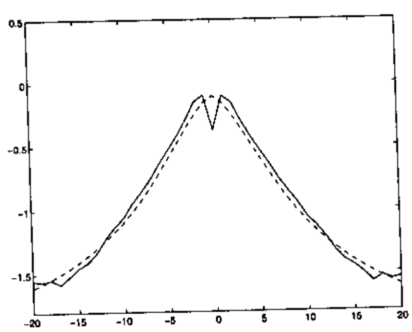
Gibbs distribution of images

The learned potential functions,

$$a\left(1 - \frac{1}{1 + (|\xi|/b)^{\gamma}}\right)$$



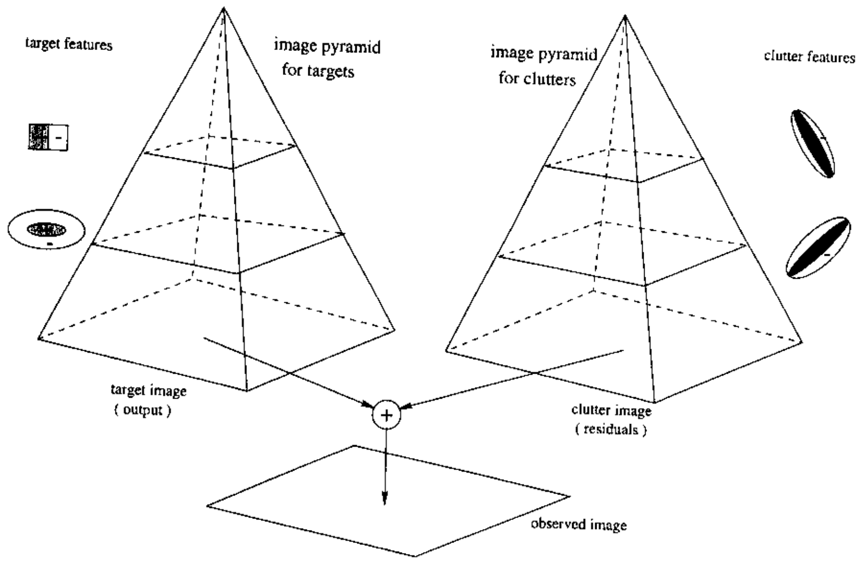
$\phi()$ $a > 0$



$\psi()$ $a < 0$

$$U(\mathbf{I}; \Lambda, S) = \sum_{i=1}^{n_d} \sum_{(x,y)} \phi_i(F_i * \mathbf{I}_{(x,y)}) + \sum_{i=n_d}^{n} \sum_{(x,y)} \psi_i(F_i * \mathbf{I}_{(x,y)}).$$

Diagram for removing clutters.



Computation

Minimizing the potential by gradient descent,

$$\mathbf{I}_t = \sum_{i=1}^{n_d} F_i^- * \phi_i'(F_i * \mathbf{I}) + \sum_{i=n_d+1}^{n} F_i^- * \psi_i'(F_i * \mathbf{I})$$

$$\mathbf{I}_0 = \mathbf{I}^{obs}, \quad F_i^-(x,y) = -F_i(-x,-y).$$

Generalizing the divergence

$$\text{div} = F_1^- * \ + F_2^- * \ + \cdots + F_n^- *$$

$$\vec{V}_{(x,y)} = (\phi_1'(), \dots, \phi_{n_d}'(), \psi_{n_d+1}'(), \dots, \psi_n'())$$

$$\mathbf{I}_t = \text{div}(\vec{V})$$

Gibbs Reaction-diffusion

Stochastical diffusion–the Langevin equation

$$d\mathbf{I}_t = -\nabla U(\mathbf{I})dt + \sqrt{2T(t)}dw_t$$

$$w(x, y, t) \sim N(\mu(x, y), |t|), \qquad dw_t = \sqrt{dt}N(0, 1).$$

Proposition

Under mild conditions on U, the Langevin equation approaches a global minimum of U at a low temperature.

Previous work

1. Reaction-diffusion for pattern formation.

— Turing 1952, Murray 1981, Turk 1991, Witkin and Kass 1991

Let $A(x, y, t), B(x, y, t)$ be chemical concentration,

$$\begin{aligned} \frac{\partial A}{\partial t} &= D_a \Delta A + R_a(A, B) \\ \frac{\partial B}{\partial t} &= D_b \Delta B + R_b(A, B) \end{aligned}$$

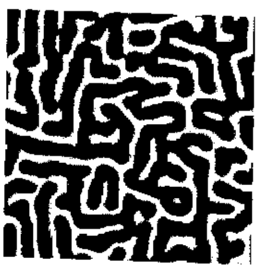
- $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is for diffusion.
- $R_a(A, B), R_b(A, B)$ are for the reaction. e.g.,

$$\begin{aligned} R_a(A, B) &= A * B - A - 12 \\ R_b(A, B) &= 16 - A * B \end{aligned}$$

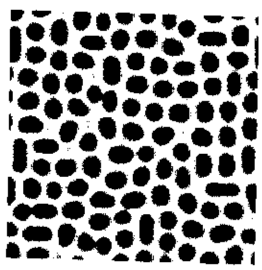
Pattern formation

Pattern formation by Gibbs reaction-diffusion

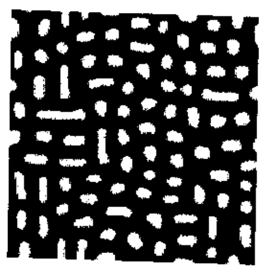
synthesized leopard blobs and zebra stripes.



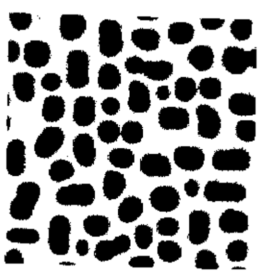
2 filters



2 filters



2 filters



3 filters

2 filters

3 filters

Previous work

2. **Anisotropic diffusion**

— Perona and Malik 1990, Nitzberg and Shiota 1992.

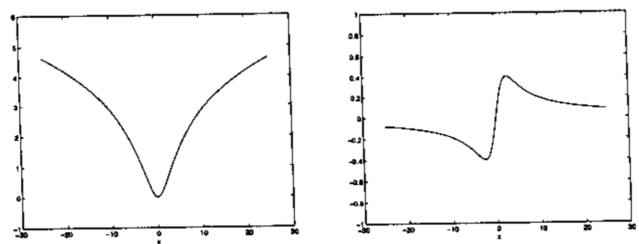
$$\mathbf{I}_t = \mathrm{div}(c(x,y,t)\nabla \mathbf{I}), \qquad \mathbf{I}(x,y,0) = \mathbf{I}_o.$$

e.g.

$$\mathbf{I}_t = \nabla_x\left(\frac{1}{1+(\mathbf{I}_x/b)^2}\mathbf{I}_x\right) + \nabla_y\left(\frac{1}{1+(\mathbf{I}_y/b)^2}\mathbf{I}_y\right)$$

It minimizes

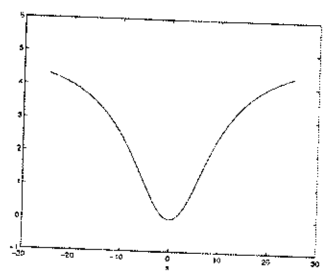
$$U(\mathbf{I}) = \int\int \psi(\nabla_x \mathbf{I}) + \psi(\nabla_y \mathbf{I})\; dxdy,$$



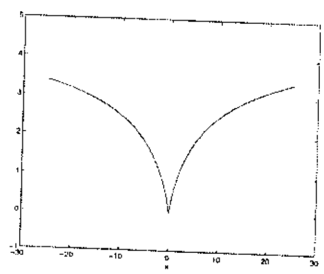
$\psi(\xi) = \log(1+(\xi/b)^2)$ $\quad$ $\psi'(\xi) = \frac{\xi}{1+(\xi/b)^2}$

The potential function for diffusion,

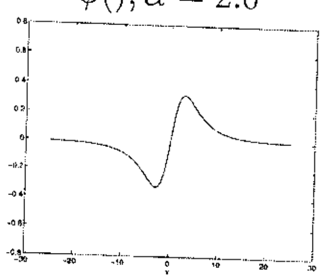
$$\phi(\xi) = a(1 - \frac{1}{1 + (|\xi|/b)^{\alpha}})$$



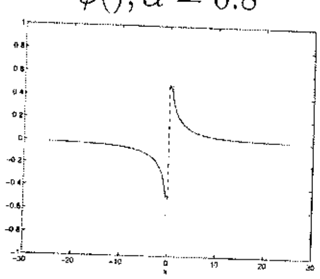
$\phi(), \alpha = 2.0$



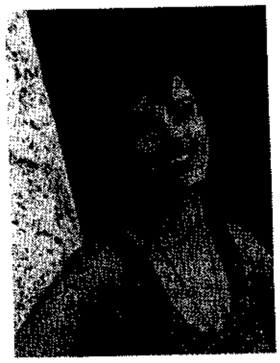
$\phi(), \alpha = 0.8$



$\phi'(), \alpha = 2.0$

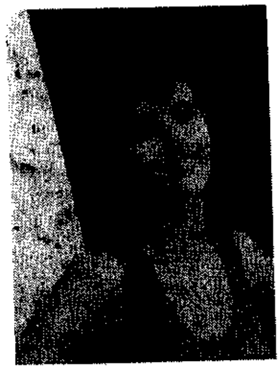


$\phi'(), \alpha = 0.8$



a

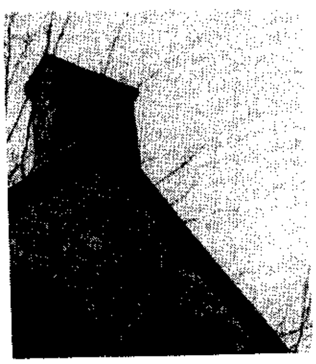
b



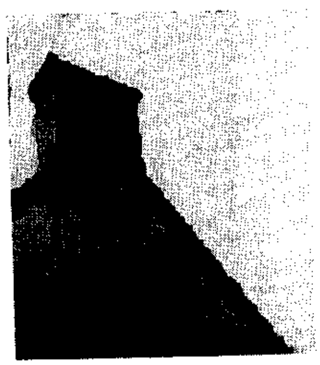
c

d

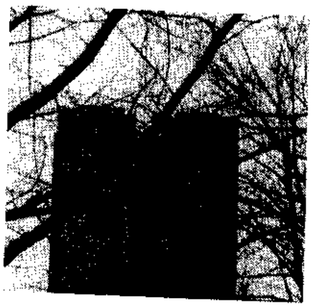
Clutter Removal by GRADE



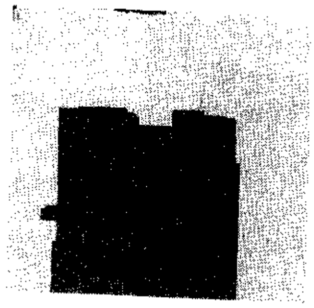
the input image



the restored image



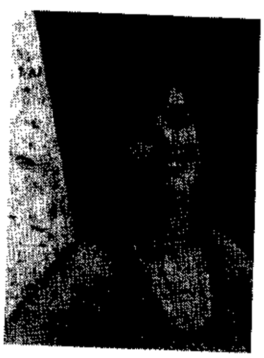
the input image



the restored image

Image Restoration by Anisotropic Diffusion

Perona-Malik anisotropic diffusion



(t=50)

(t=100)

(t=200)

Discussion

Problems with previous methods:

- Why applying the chemical/physical equations to image processing?
- The design and choice of the equations is subjective.
- The reaction terms are unstable/unbounded.

Advantages:

- Reducing PDE design to statistical modeling.
- Learning and verifying the PDE for a given application.