united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR1237/7

School on Mathematical Problems in Image Processing

(4 - 22 September 2000)

Statistical and Computational Theories of Vision

(Lectures 1, 2 and 3)

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

These are preliminary lecture notes, intended only for distribution to participants

Statistical and Computational Theories of Vision

--- Modeling, Learning, and Sampling

Ten Lectures at
The Abdus Salam International Centre for Theoretical Physics
By

Song Chun Zhu
www.cis.ohio-state.edu/~szhu
And
Ying Nian Wu
www.stat.ucla.edu/~ywu

September 11-15, 2000
Trieste, Italy

List of Lectures

1. Seeing as Statistical Inference:
   --- an overview of computer vision.
2. Minimax Entropy Principle and Texture Modeling.
   --- from Markov random fields to FRAME
3. Julesz Ensemble: a Mathematical Definition of Texture.
   --- towards a "trichromacy" theory of texture.
4. Ensemble Equivalence and Its Implications.
   --- a learning paradigm and fast algorithms
5. Statistics of Natural Images and Generic Prior Models.

List of Lectures

6. PDE and Gibbs Reaction-Diffusion Equations.
7. Gestalt Laws and Shape Modeling
8. A General Theory of Visual Learning:
   ---- from descriptive to generative models.
9. Stochastic Computing by Markov Chain Monte Carlo.
10. Data Driven MCMC In Computer Vision
   ---- Image Segmentation and Recognition.

SCTV Lecture I.

Seeing as Statistical Inference
--- an overview of computer vision

Theoretic Streams of Vision

Although philosophers' curiosities about visual perception date back to Aristotle, the field of computer vision only emerged in the late 1970s when researchers brought their expertise from many areas:

- Human vision
2. Pattern recognition
- Engineering and computer science
4. Information theory and ecology
5. Mathematics and Physics

Area 1: Human Vision

- Optics
  --- e.g. color theory.
- Neurophysiology
  --- single cell recording: e.g. V1 receptive fields, Gabor filters, and face/hand cells in IT etc.
- Psychophysics
  --- e.g. stereo vision, texture, apparent motion etc.
- Visual perception and cognition
  --- e.g. shape representation, Gestalt psychology, etc.

--- Recently, multi-channel spike train and fMRI are new tools for both neurophysiology and psychology.

Area 2: Pattern Recognition

1. Bayes decision theory,
2. Non-metric methods (tree structured)
3. Neural networks
- Syntactic (grammatical) methods.
5. Statistical learning theory.

Area 3: Engineering and Computer Science

1. Photogrammetry and geometric modeling
   --- e.g. 3D reconstruction, camera calibration etc.

2. Artificial intelligence
   --- e.g. search and relaxation algorithms, Bayesian networks

3. image compression
   --- e.g. image pyramid and wavelets, over-complete bases.

Area 4: Information Theory and Ecology

This theory views vision systems as information encoding devices that encode information in images by reducing redundancy, and it argues that brain functions couldbe understood by studying the statistics in natural images.

- Statistics of natural images and neuron adaptation.

2. Information measurements:
   --- e.g. entropy, mutual information.

- Performance bound analysis.
  --- e.g. Cramer-Rao, Fishers, Sanov, Chernoff.

Area 5: Mathematics and Physics

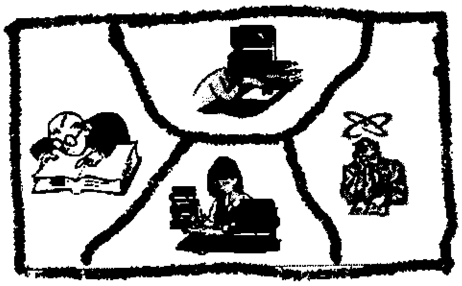
- Statistics Physics
  --- e.g. Markov random field, MCMC simulation, mean field theory.
- Spatial Statistics
  --- e.g. spatial stochastic processes, shape theory.
- Pattern Theory
  --- deformable templates,

4. Variational Methods and PDEs

Vision Triology

Phase I : 1970s and 1980s

Studying individual visual cue/module and human vision phenomenon using theories and techniques in various areas.



---- Six blind men and an elephant.
---- Schools, and religions.

Vision Triology

Phase II: 1990s and early 2000s.

Integration of visual modules and unification of theories.

- Different schools start communicate in the same wavelengths
2. Mutual understanding and respects between different views.

Vision Triology

Phase III: the next 20 years

All theories find their roles in visual learning, modeling and computing.
All modules find their position in a general purpose vision system.
Almost all human visual perception phenomena/experiments can be explained

--- A grand unified theory is emerging: a symphony

What is Vision

Vision is to solve image rendering equations:

Image formation(shape, surface, lighting, viewpoint) = image I

Solving the equations seems to include two main parts:

1. Scene reconstruction.
2. Scene understanding. Both shape and surface appearance patterns were generated through various stochastic processes, The patterns may be described by the history of these processes.

The formulation and accuracy of the image formation functions and shape/surface models depend not only on physics, but also on biologic vision and tasks.

What is Vision cont.

What we see is the solution to a computational problem, our brains compute the *most likely causes* from the photon absorptions within our eyes ---- Helmholtz, 1866.

Vision is a *process* that produces from images of the external world *a description* that is *useful* to the viewers and not cluttered with irrelevant information. ---- Marr, 1976.

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

What is Vision cont.

Because vision has a purpose, we divide the contents of a scene into two categories:

1. The stochastic patterns that we care about in a vision task and represent them by a vector of random variables W.

2. The other stochastic patterns are considered as background clutter.

Bayesian Statistical Formulation

The optimal solution maximizes the posterior :

$$W^* = \text{argmax}\; p(W \mid I)$$

$$= \text{argmax}\; p(I \mid W)\, p(W)$$

Bayes

A set of possible solutions (causes, interpretations)

$$(W_1, W_2, \ldots, W_k) \sim p(W \mid I)$$

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Vision Problems

Three main problems to solve:

- Representational: Stochastic Modeling.
- Computational: Stochastic Inference.
- Performance: Bounds of Performance and Convergence Rate

No, nothing else!

What are the criteria?

Two axioms that *bother* computer vision .

Axiom I: a bag of tricks seems enough.

We have plentiful theories and algorithms, so that for a small set of images, there exists a piece of theory or algorithm, which works well on these images.

Axiom II: all theories are rational.

We have plentiful images, so that for given theory or algorithms, there exists a small set of images, on which the theory or algorithm does what it is supposed to do.

Models of English (by Shannon 1948)

We look at a sequence of English models and their random samples

- A uniform model $p(l)$ for the English letters and space.

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

2. $p(l)$ that accounts for letter frequency

OCRO HLI RGWR NMIELWIS EU LL NBBESEBYA TH EEI ALHENHTTPA OO BTTV

Models of English (by Shannon 1948)

- $P(l_i \mid l_{i-1})$ --- the bi-gram statistics

  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILON ASIVE TUCOOWE FUSO TIZIN ANDY TOBE SECE CTISBE

- $P(l_i \mid l_{i-1}, l_{i-2})$ --- the tri-gram statistics

  IN NO IST LAY WHEY CRATICT FROURE BERS GROCID PONDENOME OF DEMOSTRUES OF THE REPTAGIN IS RECOACTINA OF CRE

5. $P(l_i \mid l_{i-1}, l_{i-2}, l_{i-3})$

  THE GENERATED JOB PROVIDUAL BETTER TRAND THE DISPLAYED CODE ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IN TO HOCK BOTHE

Models of English (by Shannon 1948)

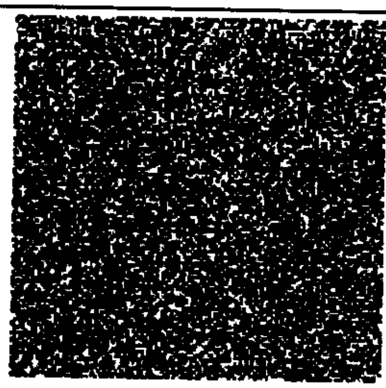
- Using lexicon $p(w)$ that accounts for word frequency.
  REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GARY COME
- $p(w_i \mid w_{i-1})$
  THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITTER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE.
- Stochastic context free grammar …
- Constrained context free grammar…

10. Model of discourse …

Samples of Real World Images

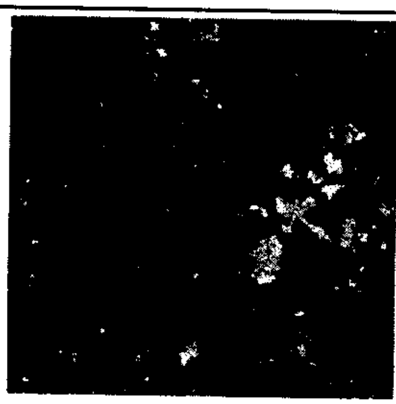
Our daily observation are trajectories in the image universe Ω^*

Level 0: Uniform Distribution



The size of image universe is in the order of $256^{256 \times 256}$, if computer keeps sampling from this space, it may not hit any realistic images even if it runs for years!
Real world images have almost zero volume in the space. Think about the probability that you hit a whale when you dive in the pacific ocean from a helicopter.

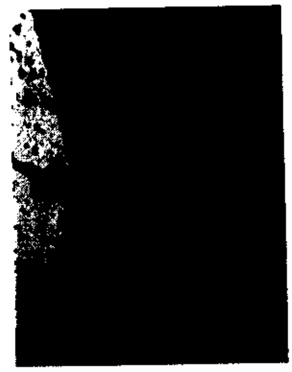
Level 1: Stochastic Generic Prior



The image is a random sample from a generic prior (MRF) model which accounts for the frequency gradient filters at 4 scales and 2 orientations. It has scale invariant properties. (Zhu and Mumford, PAMI nov. 1997)

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

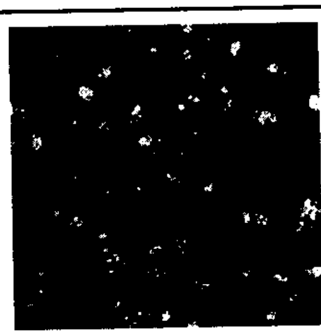
Applications of Level 1 Models



De-noising and image enhancement (zhu and Mumford, PAMI Nov. 1997)

Level 2: Stochastic Texture Modeling

observed

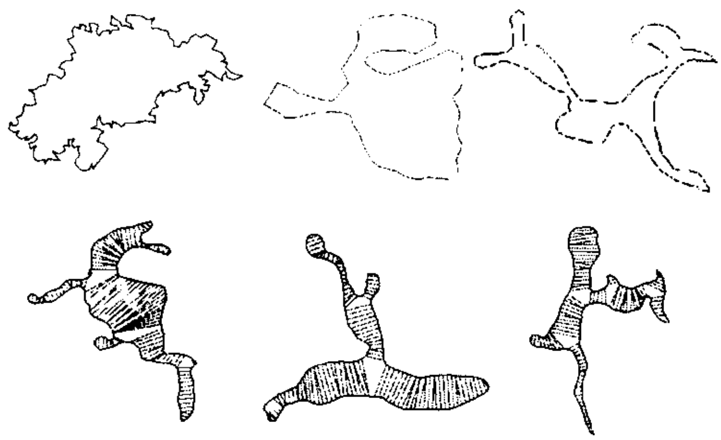


synthesized

Given a texture image as training data, a FRAME model is learned from which a randomly sampled image is synthesized.

(Zhu, Wu, Mumford, Neural Computation, Nov. 1997)

Level 2: Stochastic Shape Modeling

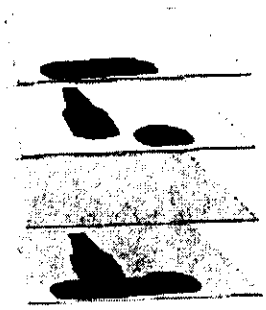


Random 2D shapes sampled from a Gibbs model. (Zhu, PAMI, 1999)

Level 2: Image Segmentation

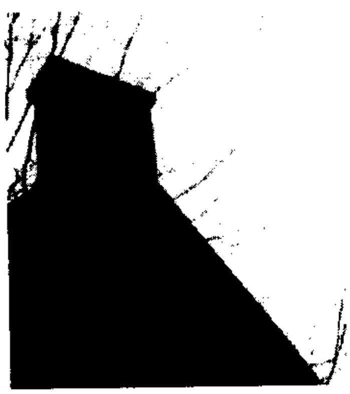
A region competition algorithm that integrate shape and texture Models.
(Zhu and Yuille, PAMI 1996).

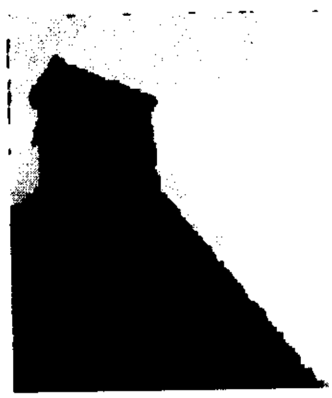
Level 3: Layers of Stochastic Processes



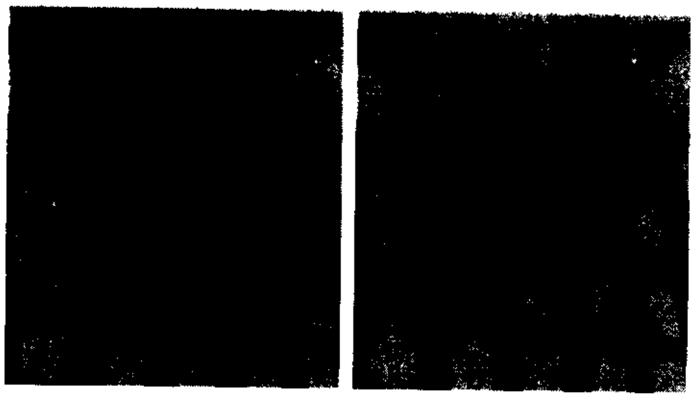
Layers of stochastic shape models (Nitzberg and Mumford, 1995)
Other examples include layers of motion (Wang and Adelson, 1996),
and textons process (Zhu and Guo, 2000)

Level 3: an u + v model



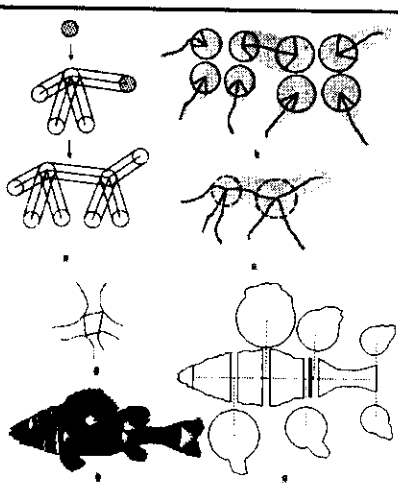


Level 3: an u + v model



Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

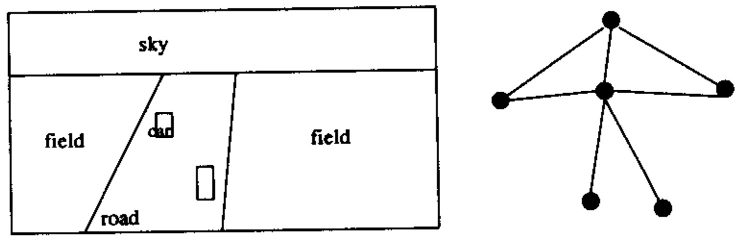
Level 4: Stochastic Generative model of Object Shapes



(Zhu and Yuille, IJCV 1996)

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Level 5: Stochastic Scene Models



(Modestino and Zhang, PAMI, 1992)

The Era of Stochasticity

"The true logic of this world is in the calculus of probability"
---- J. C. Maxwell.

"For over two millennia, Aristotle's logic has ruled over thinking, ... in the third millennium, stochastic modeling and inference now emerge as better foundations of thinking, as essential ingredients of theoretical mathematics, even the foundations of mathematics itself."
---- D. B. Mumford, 1999.

The Era of Stochasticity has come !!

I will discuss the details and the ideas will become
clear as the lecture proceeds.

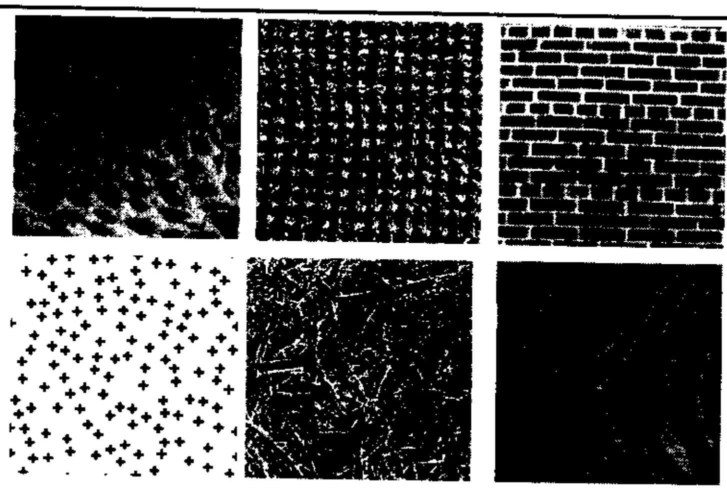
SCTV Lecture 2

Minimax Entropy and Texture Modeling

--- From Markov Random Fields to FRAME

- FRAME is an acronym for Filters, Random fields And Minimax Entropy
- Ref: Zhu, Wu, and Mumford, CVPR 1996, Neural Computation 1997, IJCV 1998.

Texture Patterns



Objective: A Common Model of Texture?

Textures are generated by diverse physical and chemical processes in natural. One way to study texture phenomena is to understand these processes, such as PDEs etc. However, it is hardly believable that our brain cells reconstruct these physical and chemical processes in texture perception. Instead it is likely that we may achieve a common texture theory that can account for all texture phenomena.

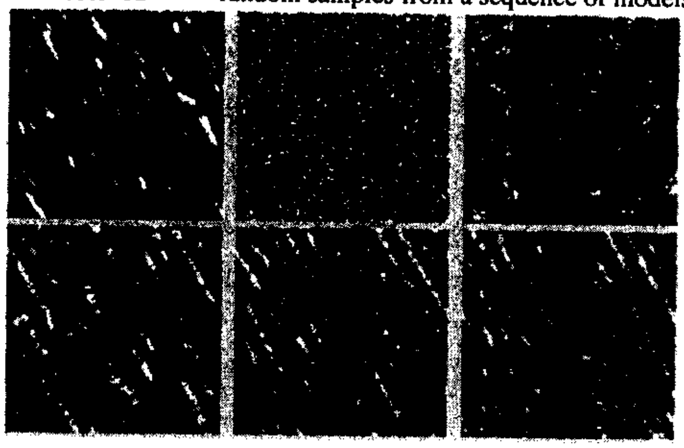
We will answer this question in about four lectures.

1. A mathematical model of texture.
2. A mathematical definition of texture.
3. The link between model and definition.
4. Fast algorithms for texture analysis and synthesis.

An Simple Example of 1D Signal

See transparencies for an example that motivates texture modeling.

An Example of Texture Modeling



Problem formation

Input: a set of images

$$S = \{ I_1, I_2, \ldots, I_n \} \sim f(I)$$

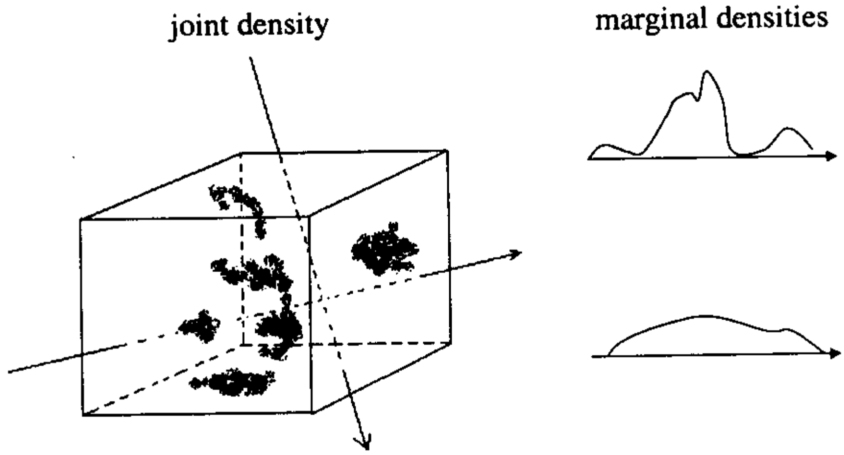
Output: a probability model

$$p(I) \rightarrow f(I)$$

Thus it is posed as a statistical inference problem.

Here, f(I) represents the ensemble of images in a given domain, we shall discuss the relationship between ensemble and probability later.

Probability Density Estimation



Feature Extraction

extracting image features/statistics as transforms

$$H_1(\mathrm{I},\ z),\ H_2(\mathrm{I},\ z),\ \ldots\ ,H_k(\mathrm{I},\ z)$$

For example:

histograms of Gabor filter responses.

$$\mathrm{H}_i(\mathrm{I},z)=\sum_{(x,y)}\delta(z-F_i*\mathrm{I}(x,y)) \qquad z\in \mathrm{R}$$

Other features/statistics: Gabors, geometry, Gestalt laws, faces.

Modeling by Maximum Entropy

Among all model p that satisfy the constraints, we choose one that has maximum entropy.

$$p = \arg\max - \int p(\mathrm{I}) \log p(\mathrm{I}) \, d\,\mathrm{I}$$

Subject to:

$$E_p[H_i(\mathrm{I}, z)] = E_f[H_i(\mathrm{I}, z)] \approx \frac{1}{n}\sum_j H_i(\mathrm{I}_j, z), \quad \forall z, i$$

$$\int p(\mathrm{I}) \, d\,\mathrm{I} = 1$$

Remark: p and f have the same projected marginal statistics.

A Gibbs Model of Texture

Solving this constrained optimization problem yields:

The FRAME model (Zhu, Wu, Mumford, 1996)

$$p(\mathrm{I}; \beta, \mathrm{F}) = \frac{1}{Z} \exp\{ - \sum_{j=1}^{k} \sum_{(x,y)} \beta_j(\mathrm{I}_j(x, y)) \}$$

$$\beta = (\beta_1(), \beta_2(), \ldots, \beta_k())$$

$$\mathrm{F} = \{ \mathrm{F}_1, \mathrm{F}_2, \ldots, \mathrm{F}_k \}$$

Remark: all known exponential models are from maxent., and maxent was proposed in Physics (Jaynes, 1957). The nice thing is that it provides a parametric model integrating features.

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Model Learning

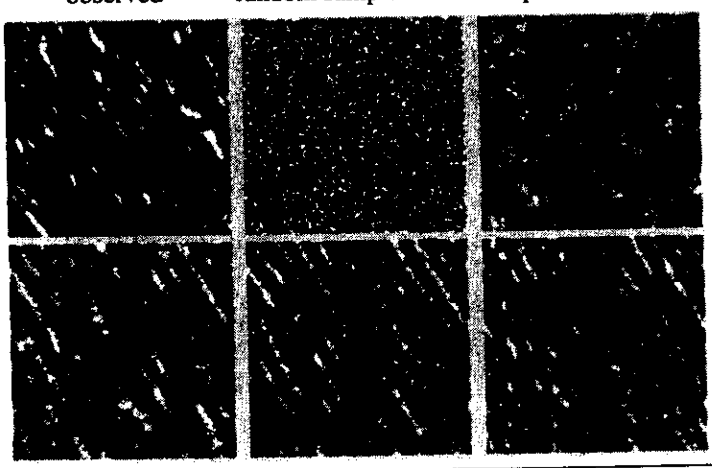
Two learning phases:

1. Choose information bearing features
   -- augmenting the probability family.
   $\Omega_1, \Omega_2, \ldots, \Omega_n \rightarrow \Omega^*$

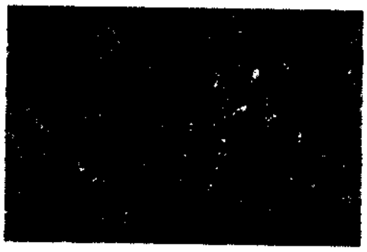
2. Compute the parameter Λ by MLE
   -- learning within a family.

An Example of Texture Modeling

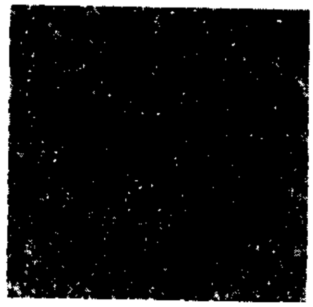
observed random samples from a sequence of models



Example: texture modeling



Observed

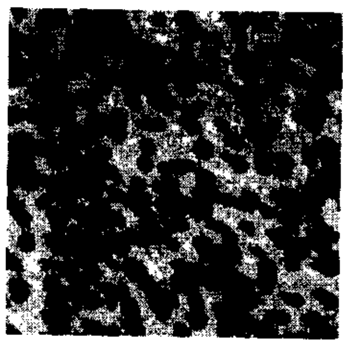


MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Example: texture modeling

Observed

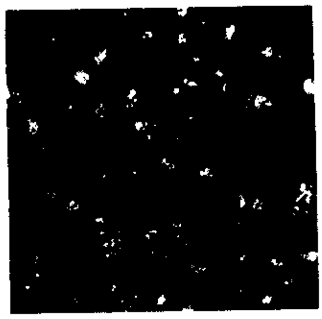


MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Example: texture modeling

Observed



MCMC sample

Summary of MaxEnt

Choosing a set of features F and their statistics H, we have a Sufficient statistics description of the image H(I). This yields A set of probabilities p that satisfy the constraints:

$$\Omega_{f,F} = \{ \ p : E_p[H(\mathrm{I})] \ = \ E_f[H(\mathrm{I})] \ \}$$

This space depends on both f and F, from which we chose

$$p(\mathrm{I};\, \beta, \mathrm{F}) = \frac{1}{Z} \exp\{ - \sum_{j=1}^{k} \sum_{(x,y)} \beta_j(\mathrm{I}_j(x,y)) \ \}$$

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Minimum Entropy Principle

The goodness of a model p is decided by the feature F and statistics H, and the selection of F (H) is to minimize a Kullback-Leibler divergence between f and p.

$$F^* = \arg\min D(f \,\|\, p)$$

$$= \arg\min \int f(\mathrm{I}) \log \frac{f(\mathrm{I})}{p(\mathrm{I};\beta,\mathrm{F})} d\,\mathrm{I}$$

$$= \arg\min E_f[\log p] - E_f[\log p]$$

$$= \arg\min \; entropy(p) - enrtopy(f)$$

Minimax Entropy Learning

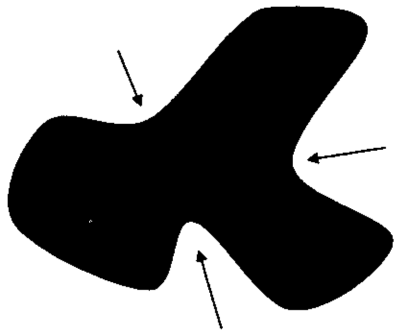
$$p^* = \arg\min_{p\in\Omega} D(f \,\|\, p) = \arg\min_{p\in\Omega} E_f[-\log p]$$

For a Gibbs (max. entropy) model p, this leads to the minimax entropy principle (Zhu,Wu, Mumford 96,97)

$$p^* = \arg\min_{F} \{ \max_{\beta} \text{entropy}(p(\mathrm{I}\,;\beta)) \}$$

Minimax Entropy Learning (cont.)

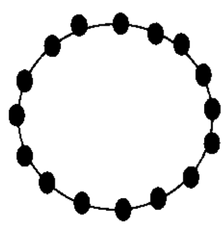
Intuitive interpretation of minimax entropy.



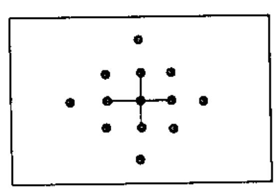
Markov Random Fields

- Ising model (1920s)

  typical vs most probable



2. Besag (1973,74)

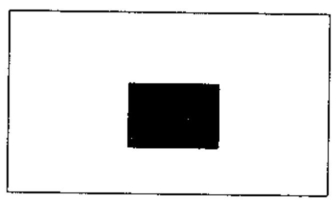


Comparison

FRAME model extends traditional MRF models in two aspects:

- It replaces cliques by information bearing features, such as Gabor filters, wavelets. In this sense, some effective image representation is embedded into Gibbs models
- The forms of potential functions are learned from data in a non-parametric way. It is manually chosen.

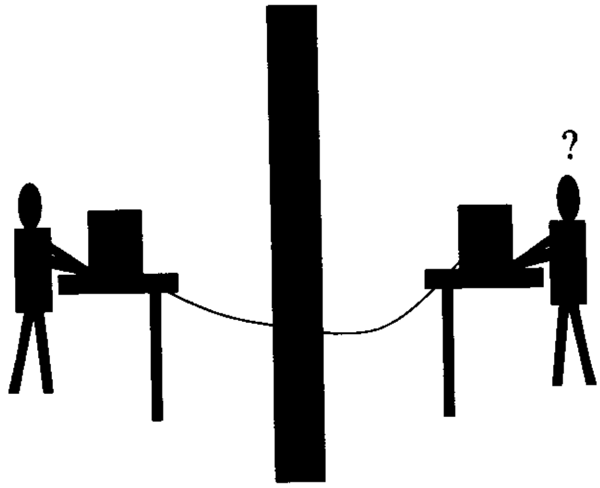
What is a Texture Model?



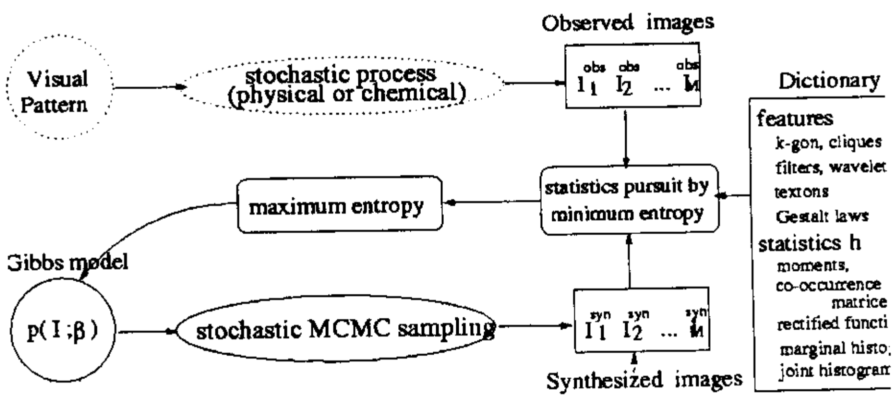
In computer vision, we need a model for an arbitrary image patch A.

$$p(\mathrm{I}_{A} \mid \mathrm{I}_{\partial A})$$

Learning Problem Revisit



A Minimax Entropy Learning Scheme

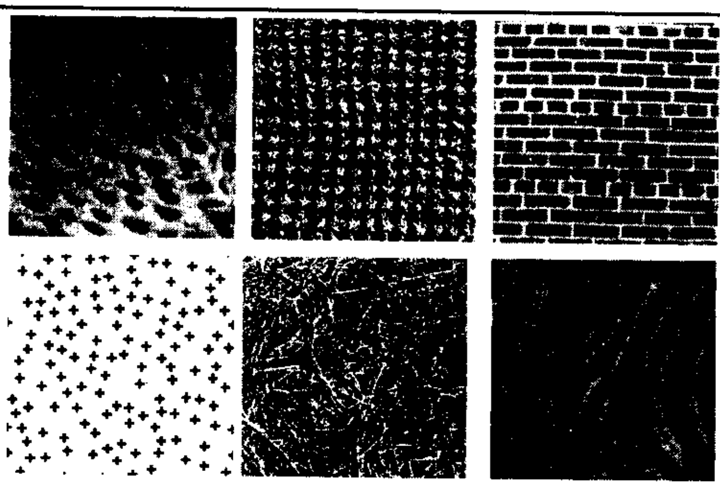


SCTV Lecture 3

Julesz Ensemble:
a Mathematical Definition of Texture
--- towards a "trichromacy" theory of texture.

Ref: Zhu, Liu, and Wu, PAMI 2000, and reference therein.

Examples of Texture Pattern



Objectives

We are interested in two questions.

- What is a texture?
  --- a mathematical/physical definition.
- How should we represent a texture in a computer?
  --- a legitimate mathematical model.

To achieve a theory as clean as color theory.

1. Electro-magnetism: a color is defined by a wavelength λ.
2. Tri-chromacy theory: a visible color is represented by (r,g,b).

Doubts

Two doubts for the existence of a mathematical definition of textures:

- Natural textures have diverse origins
  --- physical, chemical, biologic processes.
- Texture is a spatial phenomenon
  --- it is not defined on one point, unlike color, and its statistics properties vary over locations. E.g. if it is not defined on n x m lattice, then it is not defined in (n+1) x (m+1).

Rationales

Rationales for the existence of a mathematical definition:

1. Texture perception is a psychological phenomenon, and we say a set of texture images belong to "a texture" if they look the "same" to human vision systems. The latter are not interested in the differences within the set. Our definition of texture needs to be consistent with human perception.

2. It is unlikely that human visual systems reconstruct the natural stochastic processes, instead it is just a "percept". We may achieve a texture definition once we know:

what the visual cortex is computing.

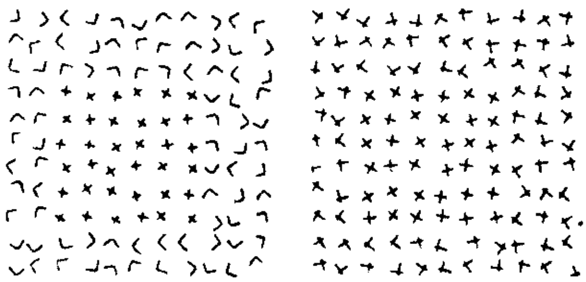
Julesz Quest

Dr. Julesz is a well known psychologist, once at the Bell Labs, and then a professor at Rutgers. He is best known for his work on random dot stereo and texton theory. Though both theories were not really correct, his work inspired many thinkers, including David Marr for establishing the field of computational vision.

"What features and statistics are characteristics of a texture pattern, so that texture pairs that share the same features and statistics cannot be told apart by pre-attentive human visual perception?"

---1960--80s

Psychophysics Experiments



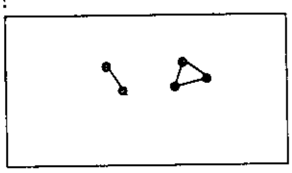
Texture discrimination in early vision (0.1-0.4sec).

Major Technical Difficulties

- What kind of features and statistics should we search for?

2. A mathematical tool for mixing the exact amount of statistics for a given recipe.

Julesz and his school focused on k-gon statistics (k=2,3), which limited the scope of search. Unfortunately, our visual cortices do not compute k-gon statistics!

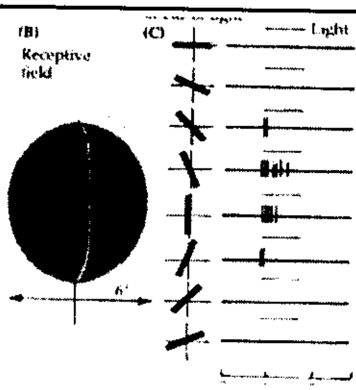


Neurophysiology Experiment

Huber and Weissel 1960s — Tai Sing's Monkey

Single neuron recording in the V1 area in cats and monkeys.

What are V1 cells Computing ?



This leads to many explanations: edge detection, texton, and Gabor filters in the 1980s. It also inspired wavelets in thinking.

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Psychology Again

- Bergen and Adelsen 1991.
- Chubb and Landy 1991.
- Karni and Sagi 1991.

A conjecture:
"A texture pair cannot be told apart if they share the same histograms for a bank of Gabor response"

A reckless experiment by Heeger and Bergen 1995.

A Mathematical Problem Again

A key is to find a mathematical tool:

Given an image I^{obs} and its extracted statistics $h(I^{obs})$, how do we create a random image that shares the same statistics – not mixed or biased by other statistics?

Julesz ensembles

Given a set of normalized statistics, $h = (h^{(i)} : i = 1, 2, \ldots, k)$
for images on a finite lattice Λ*, we define an equivalence class*

$$\Omega_\Lambda(H) = \{ \mathbf{I} : h(\mathbf{I}) \in H \}$$

H is an open set centered at h.

A Julesz ensemble $\Omega(h)$ ***is the limit of*** $\Omega_\Lambda(H)$ ***as*** $\Lambda \to Z^2$
under some boundary conditions.

Julesz Ensemble

As image lattice goes to infinity in the 2D plane, statistical fluctuations diminishes and thus we obtain a deterministic set.

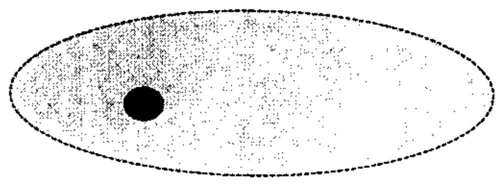


image space on Z^2

Draw random samples from the ensemble by Markov chain Monte Carlo method, such as Gibbs sampler (Geman and Geman 1984).

Texture on Z^2

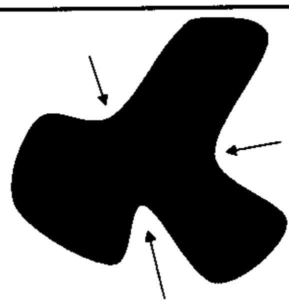
A Julesz ensemble is associated with a uniform distribution.

$$q(\mathrm{I};\, \Phi) = \begin{cases} 1/\,|\Omega(\Phi)| & \text{for } \mathrm{I} \in \Omega(\Phi), \\ 0 & \text{else.} \end{cases}$$

The image universe

Feature Pursuit Process

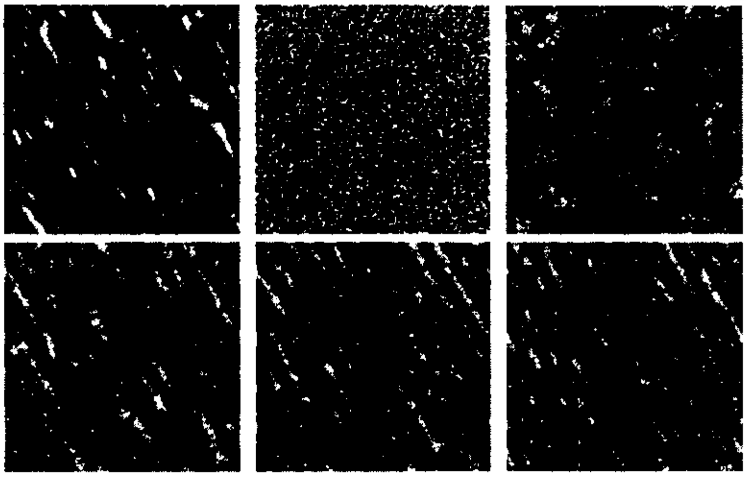
Revisit to minimax entropy:



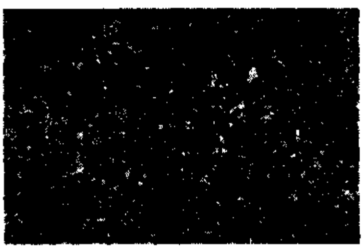
- Each new feature/statistics brings additional constraints, and thus shrinks the volume of the Julesz ensemble. Since logarithm of the volume is entropy. A feature that cut the volume the most is minimizing the entropy.

2. Within the Julesz ensemble, we adopt a uniform distribution that has the maximum entropy.

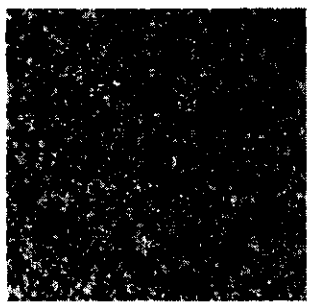
Example



Example: texture modeling



Observed



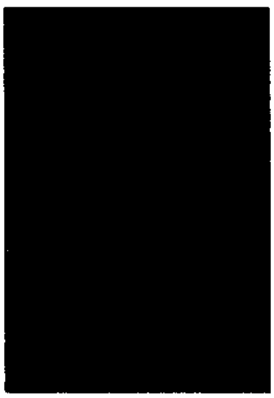
MCMC sample

Example: texture modeling

Observed MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

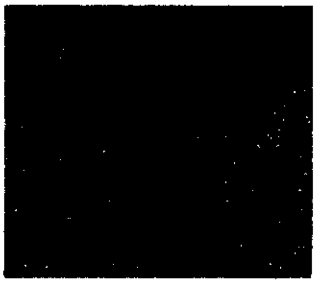
Example: texture modeling



Observed MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Example: texture modeling

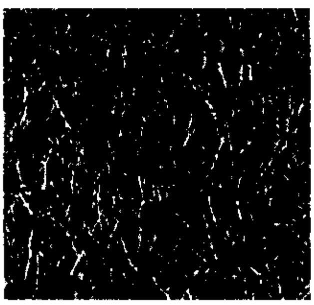


Observed

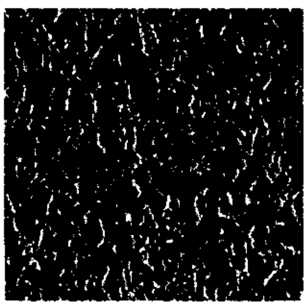
MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Example: texture modeling



Observed

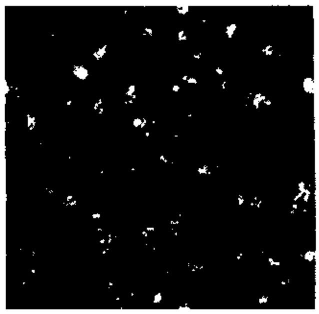


MCMC sample

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000

Example: texture modeling

Observed



MCMC sample

A Definition of Texture on Z^2

A texture pattern in human visual perception is a Julesz ensemble $\Omega(h)$ on Z^2, where h is the sufficient and necessary set of statistics.

Here we want to identify the minimum statistics h used in human perception. Unfortunately, it is hard to achieve this goal, because:

1. h changes with experience and is adaptable (Karni and Sagi 1991)
2. Many different combination of h may yield similar effects.

Nevertheless, we define such a definition as a mathematical concept that is helpful in thinking.

Ten SCTV lectures at ICTP, Trieste, Italy, by S.C. Zhu, 09/2000