BAYESIAN INFERENCE AND MRF MODELS

Lecturer: Sang Hwa Lee

Goals

- Introduce the basic theory of Bayesian inference and MRF models for image processing
- Understand the elementary concepts in MRF models
- Understand why energy minimization methods are needed
- Understand how to use MRF models in the computer vision

Contents

- Labeling problems
- MRF model
- Energy minimization
  - Simulated annealing (Monte Carlo Simulation)
  - Belief propagation
  - Graph cut
- Applications of MRF models in computer vision
  - Stereo matching
  - Super resolution

LABELING PROBLEM

Labeling Problem

- What is labeling problem?
  - To assign a label from the label set to each of the sites
- Labels
  - Variables to discriminate the sites
  - Random (probabilistic) vs non-random (deterministic)
  - Discrete vs continuous
  - Types
    - Class symbol: object detection (ex: +1, -1)
    - Integer number: disparity values (ex: 0-10)
    - Vectors: motion vector, color, parameter set of function
- Sites
  - Independent locations or objects to have labels
  - Pixel, block, feature vector, image
  - Regular (pixel) vs irregular (feature point) structure

Examples of Labeling Problem

Research subject	Site type	Label type	Result of labels
Stereo matching	Pixel, block, pre-segment region	Integer number	Disparity map
Motion estimation	Block, pixel, pre-segment region	Integer 2-D vector	Motion vector
Segmentation, Binarization (document image)	Pixel, block, pre-segment region	Integer, symbol, color	Segmentation map
Image restoration, Image inpainting	Pixel, block, pre-segment region	Gray-scale (0-255), color	Recovered image
Classification, Recognition	pre-segment region, image	Class (+1, -1), symbol, ID number	Object detection, Classification, Recognition
Vector quantizer, Super resolution	Pixel, block	Vector	Codebook Magnified block
Curve fitting	Feature point	Vector, integer	Parameters of function

Extension of labeling Problem

- When we define what to estimate in a site system, any kinds of labeling problems can be established.
  - Define the label set and the site system
  - Application to many problems in computer vision
- But, a labeling problem does not include any methodology and algorithms.
  - The methods and solutions are open to you!
  - You can define new labeling problems and their solutions by our own algorithms!

Combinatorial Problem (1)

- Combinatorial problem ?
  - The solution is a combination of possible labels.
- Labeling problem is a kind of combinatorial problem
  - The labels of sites are the combination of labels.
- Difficulty
  - **D1**: When there are many number of labels,
    - There are innumerable combinations of labels
  - **D2**: When the labels of sites are not independent,
    - The label of a site can be changed with respect to different combinations of labels in other sites.

Combinatorial Problem (2)

- How can you find the combination of labels in those difficulties?
  - Too many possible combinations for one solution
  - Effect on the labels of other sites
- Need of efficient way to find the combinatorial solutions
- The methodology is
  - ***Simulated annealing (Monte-Carlo simulation)***
  - ***Belief propagation***
  - ***Graph cut***
  - ***Stochastic (non-linear) diffusion***

BAYESIAN INFERENCE FOR MRF MODELS

Introduction (1)

- MRF: A mathematical framework to find the combinatorial solutions in the difficulties (especially focused on D2).
  - For D2: MRF models
  - For D1 and D2: Energy minimization methods
- MRF model is a general framework to use the correlation of labels in the neighborhood.
  - MRF models deal with D2 (Not independent!).
  - MRF models → *Prior model*

Introduction (2)

- MRF defines the problems in the probabilistic domain
  - Consider the labels as random variables
    - Label $\rightarrow$ *field*
  - Maximum a posteriori (MAP) estimation
    - Maximize $p(\text{labels} \mid \text{observations})$
    - Ex: stereo matching: $p(\text{disparity} \mid \text{stereo images})$
- It's difficult to deal with probability domain.
  - PDF of labels should be known or assumed in advance.
- We will transform the probability domain into the energy domain by Gibb's random field.

Bayesian Estimation (1)

- Maximum Likelihood (ML) Estimator
  - Given labels: p(observations | labels)
  - Ex: Block matching motion estimation
  - Full search methods

- Maximum A Posteriori (MAP) Estimator
  - Given observations: p(labels | observations)
  - Use prior information of label by Bayesian Decomposition

Likelihood model Prior model

$$p(L \mid O) = \frac{p(O \mid L)\, p(L)}{p(O)}$$

Constant for labels

Bayesian Estimation (2)

- MRF-MAP estimator generates two energy terms.
  - Likelihood model
    - To measure how well the observations are matched
    - Observations are directly related to image values.
    - Ex: block matching error
  - Prior model
    - To exploit the correlations of neighboring fields
    - Prior model is related to the label field that you want to estimate.
    - Ex: smoothness condition of labels
- How can we minimize the joint energy function against D1 and D2?
  - Sub-optimum problems

Bayesian Estimation (3)

- Energy minimization methods find the optimal (actually sub-optimum) solution against D1, D2.
  - Simultaneous minimization of joint energy functions, likelihood and prior terms
- Two types of energy minimization
  - E1: Minimizing energy function directly
  - E2: Stabilizing energy states
- E1: To find labels in minimizing energy function
  - Simulated annealing, graph cut, gradient descent
- E2: To propagate (or diffuse) the energy space first, and find labels by winner-takes-all
  - Belief propagation, non-linear (or stochastic) diffusion

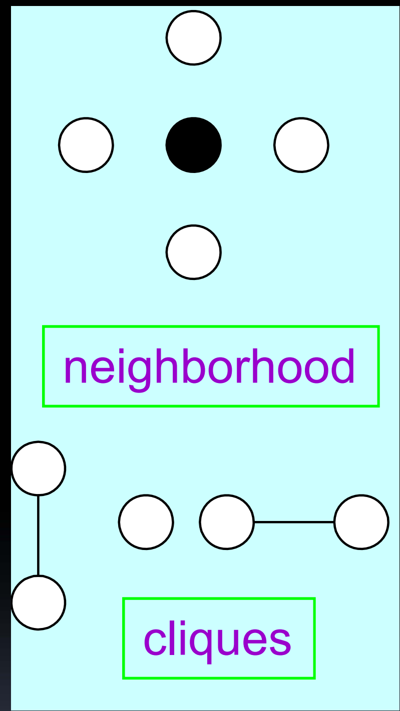
Mathematical model of MRF

- Definition of MRF

$$p(x_i \mid \forall x_j \neq x_i) = p(x_i \mid x_j \in N_{x_i}),$$

$$p(x_i \mid x_j) > 0 \ \ (\neq 0) \text{ and } \sum_i p(x_i \mid x_j) = 1$$

- Neighborhood system
  - To define the neighboring sites
  - Clique: subset of neighborhood to have interactions between sites.
  - Neighborhood system is open to you.
    - First-order MRF window is sufficient for any applications.



1st MRF window

Mathematical model of GRF

- Definition of Gibb's Random Field (GRF)

$$p(x) = \frac{1}{Z} \exp\left\{-\frac{U(x)}{T}\right\}$$

  - Z: normalization factor
  - T: Cooling temperature (Thermodynamics)
    - T varies as the iterative cooling (energy minimization) process.
  - $U(x)$: Gibb's potential (energy function)
- To transform the energy $U(x)$ of random variable x into the probability $p(x)$

Equivalence of MRF and GRF

- Equivalence of MRF and GRF

$$p(x) = \frac{1}{Z} \exp\left\{ -\frac{1}{T} \sum_{c \in C} V_x(x, y \mid y \in c) \right\},$$

$$U(x) = \sum_{c \in C} V_x(x, y \mid y \in c)$$

- Sum of all clique potentials $V_x(.)$ of MRF model is equivalent to Gibbs potential $U(x)$.
  1. Define neighborhood system and cliques
  2. Define clique potentials → MRF models (prior terms)
  3. Apply the sum of clique potentials to Gibb's potential
  4. Find the optimal solution with minimum energy

General MAP-MRF Framework

- Labeling problem with MAP-MRF framework
  - General tool for prior modeling of label fields
    - To exploit the correlation between neighboring fields
    - Label set: random variables
    - MRF models: prior knowledge of labels
  - Not energy minimization
- MAP-MRF framework for stereo matching
  - To find disparity map to maximize $P(\mathrm{L}|\mathrm{r}, \mathrm{l})$
  - Bayesian decomposition
    - Derive likelihood and prior terms
    - Maximize $P(\mathrm{L}|\mathrm{r},\mathrm{l})$ = maximize $P(\mathrm{r}|\mathrm{l},\mathrm{L})\ P(\mathrm{L}|\mathrm{l})/\ P(\mathrm{r}|\mathrm{l})$
  - By MRF - GRF equivalence
    - Minimize $U(\mathrm{L}|\mathrm{r},\mathrm{l})$ = minimize $\mathrm{U}(\mathrm{r}|\mathrm{l},\mathrm{L}) + \mathrm{U}(\mathrm{L}|\mathrm{l}) - \mathrm{U}(\mathrm{r}|\mathrm{l})$

Likelihood Model

- To measure how well the observations are matched using the estimated labels.
- Observations in images
  - related to image values (intensity, color, texture…)
  - At least, two observations are required to evaluate how well the labels match the observations
    - Ex: two stereo images, two video frames…
- Examples of likelihood models
  - Block (feature) matching:
    - stereo matching, motion estimation, segmentation, super resolution…
  - Shape (geometric) matching:
    - Curve fitting, shape (pattern) classification

Prior Model

- To exploit the correlations of neighboring fields
  - Prior model is related to the labels that you want to estimate.
  - Correlation: prior knowledge of labels
    - Smoothness, regular patterns, convex set
  - MRF models in the form of clique potentials
- MRF modeling is a kind of regularization process of the estimated labels.
  - Constrained problems: hard regularization
    - Assign the labels within the constraints
    - Ex: POCS (projection onto convex set)
  - MRF problems: soft regularization
    - Assign the labels with respect to MRF models (interaction)

How to use Bayesian MRF

1. Define what to estimate from your observations
   - Select research subjects
2. Define likelihood model
   - Using the label fields to be estimated
3. Design prior model
   - using the prior knowledge of label fields
   - MRF models in clique potentials
4. Apply one of energy minimization schemes
   - Joint energy minimization
5. Optimize the parameters in the energy function
6. Evaluate the results
   - Final results in MRF models should be better than those only in likelihood model.

APPLICATIONS OF BAYESIAN INFERENCE AND MRF MODELS

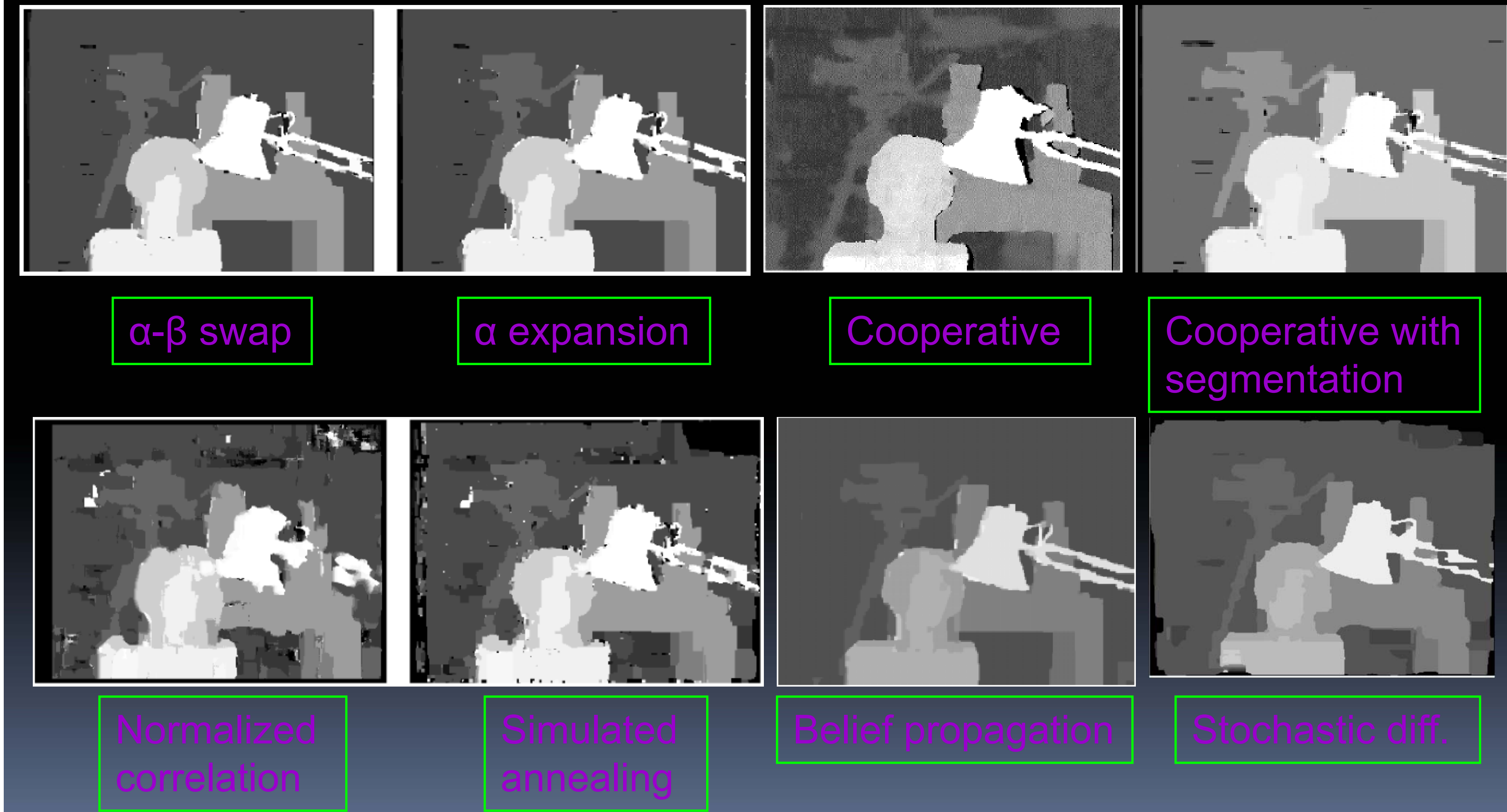
Stereo Matching (1)

- Most general problem with MRF modeling
  - Label set: disparities (depths)
  - MRF models of disparity field: prior model
- Energy functions
  - Likelihood model:
    - how well the disparities match the corresponding sites in the sense of block (pixel) difference of intensity or color
  - Prior model:
    - smooth and discontinuity-preserving energy function of the disparity

$$E(f) = \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q) + \sum_{p \in \mathcal{P}} D_p(f_p),$$

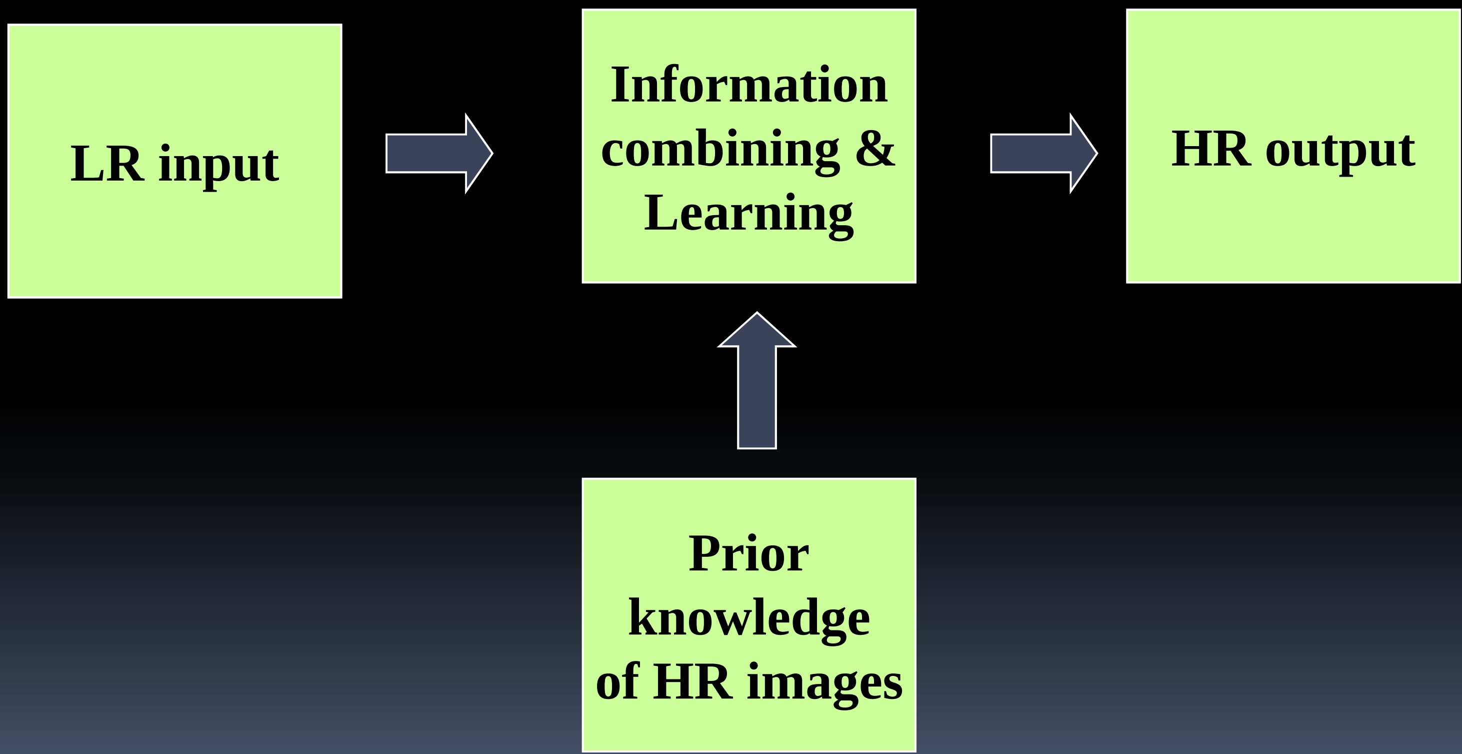
Stereo Matching (2)

- Comparison (Taxonomy of …, 2002. 6, IJCV)



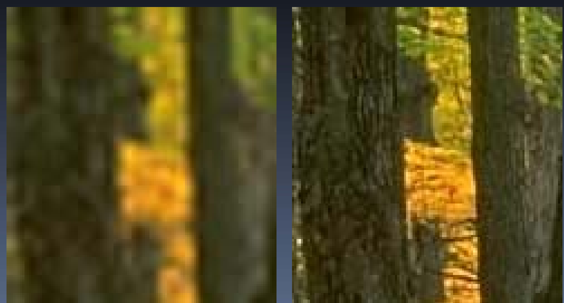
Learning-based SR (1)

- SR as recognition process



Learning-based SR (2)

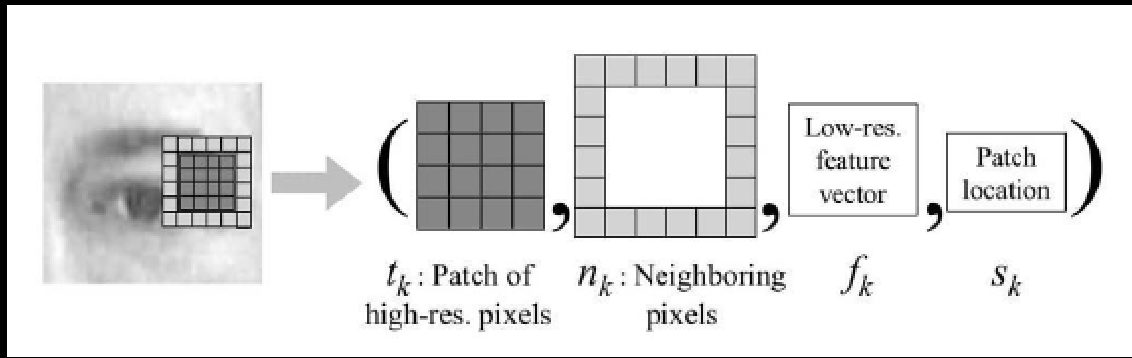
- NYU: image analogies (*SIGGRAPH 2001*)
  - Training pairs



Learning-based SR (3)

Learning-based SR (4)

- Data Entry for energy function



- Likelihood model

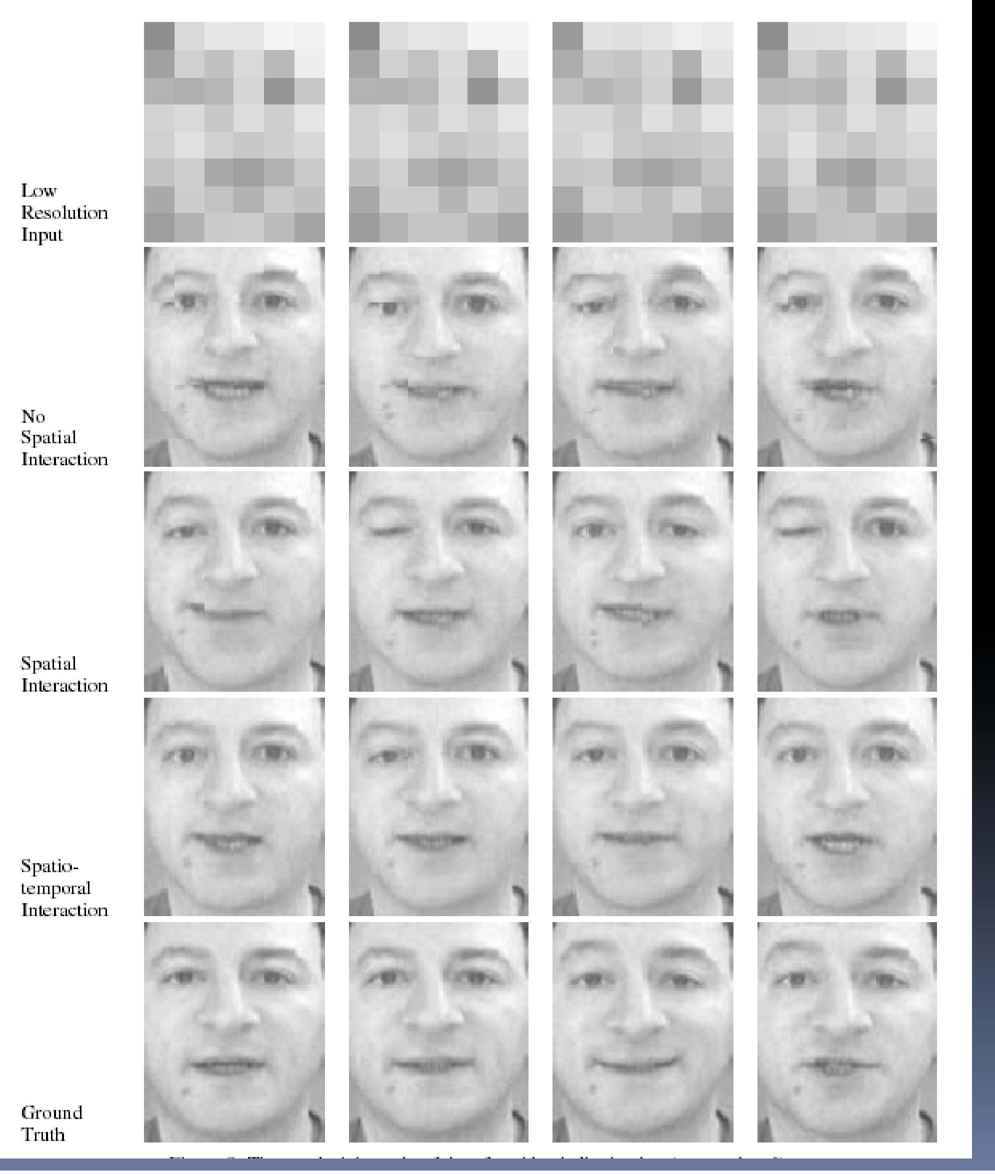
$$P(L \mid H) = \prod_{l=1}^{N} \frac{1}{\sigma_L \sqrt{2\pi}} \, exp\Big(- \frac{(L(l) - (AH)(l))^2}{2\,\sigma_L{}^2} \Big).$$

- Prior model

$$\phi(T_p = t_k, T_q = t_l) \propto exp\Big(- \sum_{overlap} \big(t_k(u) - n_l(v)\big)^2 - \sum_{overlap} \big(n_k(u) - t_l(v)\big)^2 \Big)$$

Learning-based SR (5)

- 결과
  - 8배 확대



ENERGY MINIMIZATION METHODS

Simulated Annealing (1)

- Literature
  - J.Besag, *J.Royal stat. soc*. B, 1974.2.
  - S.Geman and D.Geman, *IEEE PAMI* 1984.6
- Classical energy minimization
  - Monte Carlo simulation
  - Representative method for labeling problem
    - Energy minimization via MRF modeling
    - Prior model is defined by clique potentials
  - Origin from thermodynamics
    - Annealing
    - Cooling procedure
  - Random sampling of the label field
  - 2D potential plane

Simulated Annealing (2)

- Metropolis algorithm

1. Set $i = 0$ and $T = T_{max}$. Choose an initial $\mathbf{u}^{(0)}$ at random.
2. Generate a new candidate solution $\mathbf{u}^{(i+1)}$ at random.
3. Compute $\Delta U = U(\mathbf{u}^{(i+1)}) - U(\mathbf{u}^{(i)})$.
4. Compute the probability P from

$$P = \begin{cases} \exp(-\frac{\Delta U}{T}) & if \Delta U > 0 \\ 1 & if \Delta U \leq 0 \end{cases} \tag{3.9}$$

5. If $P = 1$, accept the perturbation; otherwise draw a random number that is uniformly distributed between 0 and 1. If the number is less than P, accept the perturbation.
6. Set $i = i + 1$. If $i < I_{max}$, go to 2.
7. Set $i = 0$ and $\mathbf{u}^{(0)} = \mathbf{u}^{(I_{max})}$. Reduce T according to a temperature schedule. If $T > T_{min}$, go to 2; otherwise, terminate.

Simulated Annealing (3)

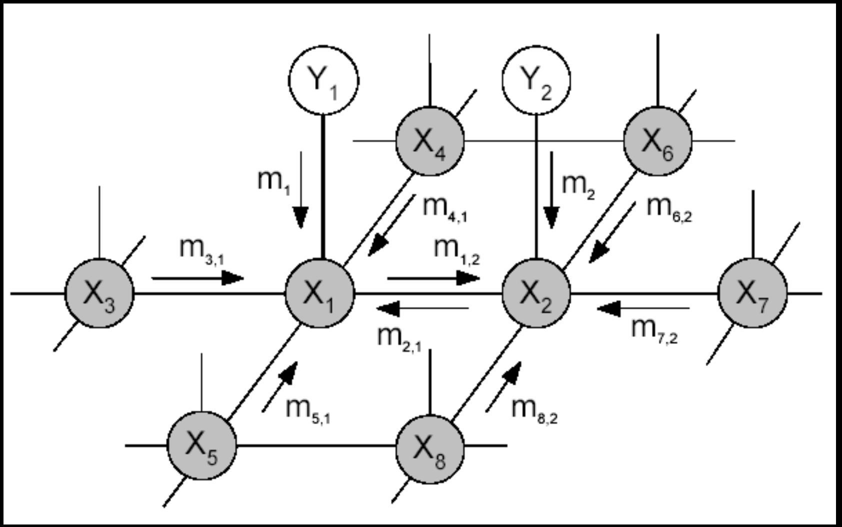
- Sampling and decision methods
  - Metropolis method
    - Soft decision (global optimal)
  - Iterated conditional mode (ICM)
    - Hard decision (sub-optimal)
  - Gibbs sampler
    - Conditional sampling based on neighborhood distribution
- Features
  - Guarantees the global optimality theoretically
  - Very slow

Belief propagation (1)

- Literature
  - J.Sun, H.Y.Shum, *ECCV-2002*
- Features
  - MAP-MRF framework
  - Graph-based optimization
    - Propagation of messages via Bayesian network
    - Max-product algorithm
  - Robust energy measure
  - Use of multiple cues
    - Occlusion, line, color segmentation information

Belief propagation (2)

- Max-product algorithm
  - A pixel (y) is connected to all labels (x)
  - {y} is a vector where each element is the matching cost given different label.



Message weights

$$\psi_{st}(x_s, x_t) = \exp(-\rho_p(x_s, x_t)))$$

$$\psi_s(x_s, y_s) \propto \exp(-\rho_d(F(s, x_s, I)))$$

Message update

$$m_{1,2}^{new} \leftarrow \kappa \max_{x_1} \psi_{12}(x_1, x_2) m_1 m_{3,1} m_{4,1} m_{5,1}$$

Belief propagation (3)

- Max-product algorithm

1. Initialize all messages as uniform distributions
2. Update messages iteratively for i=1:T

$$m_{st}^{i+1}(x_t) \leftarrow \kappa \max_{x_s} \psi_{st}(x_s, x_t) m_s^i(x_s) \prod_{x_k \in N(x_s) \setminus x_s} m_{ks}^i(x_s)$$

3. Compute beliefs

$$b_s(x_s) \leftarrow \kappa m_s(x_s) \prod_{x_k \in N(x_s)} m_{ks}(x_s)$$

$$x_s^{MAP} = \arg\max_{x_k} b_s(x_k)$$

Graph cut (1)

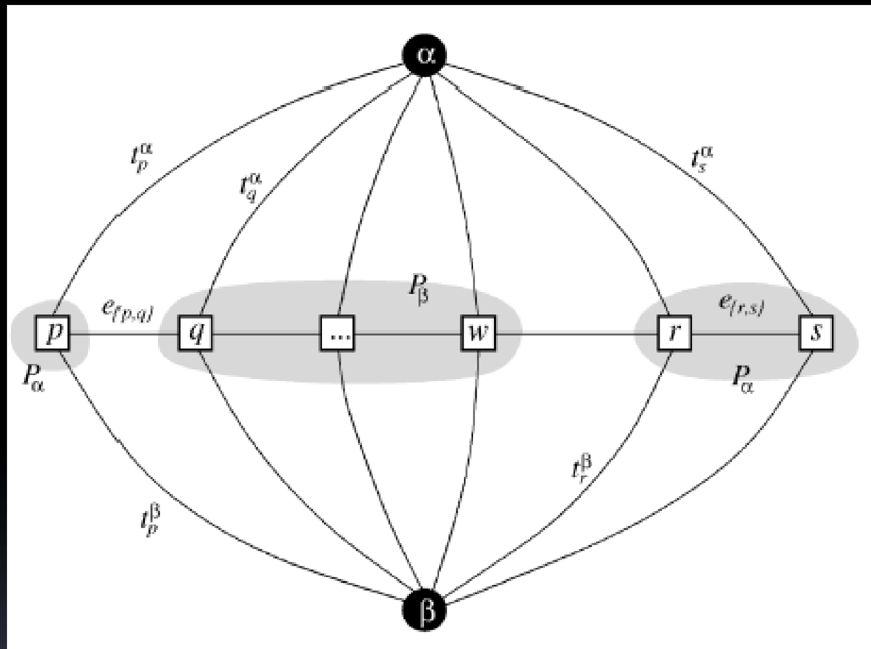
- Literature
  - Y.Boykov, O.Veksler, R.Zabih, *IEEE PAMI*, 2001 Nov.
  - V.Kolmogorov and R.Zabih, *CVPR-01*, *ECCV-2002*
- Graph-based labeling and energy minimization
  - Suitable for combinatorial problem and MRF modeling
    - Two energy terms: data and prior models
    - Multi-dimensional energy function
  - Multiple-move of labels
    - α-β swap:
    - α expansion
  - Local optimization of 2D potential plane with the bound
  - Fast and good performances

Graph cut (2)

- Applications of graph cut
  - Combinatorial problems
  - Image restoration
  - Correspondence estimation
    - Motion and disparity
    - Occlusion
    - Multiple-view
  - Voxel occupancy
- What energy functions can be minimized via graph cut?
  - V.Kolmogorov, *ECCV-2002*
  - Energy function is in the metric space
    - Triangular inequality

Graph cut (3)

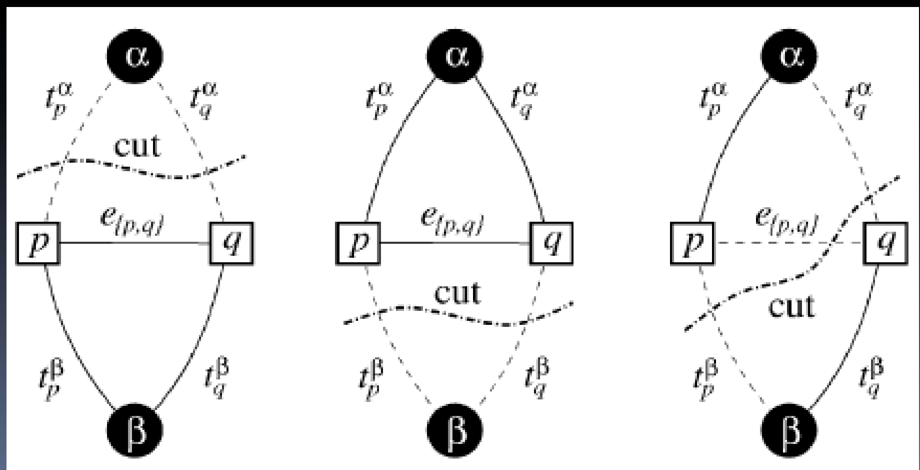
- α-β swap
  - Move only in two labels
  - Vertices
    - pixels
    - labels (terminal)
  - Edges
    - t-link: terminal-pixel
    - n-link: neighboring pixel-pixel



Graph cut (4)

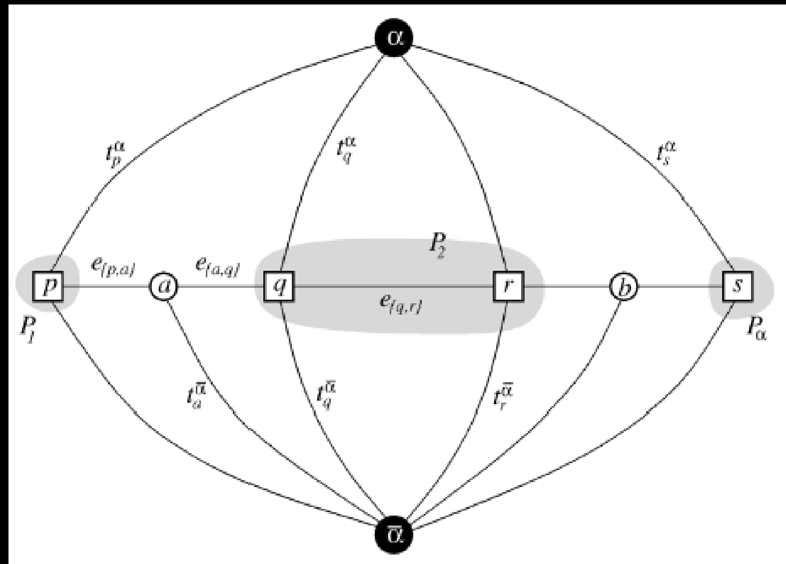
- α-β swap
  - Energy function
    - Edge weights
    - Data term and prior model
    - MRF model
  - Graph cut
    - Minimal edge cut

edge	weight	for
t_p^{α}	$D_p(\alpha) + \sum_{q \in \mathcal{N}_p,\, q \notin \mathcal{P}_{\alpha\beta}} V(\alpha, f_q)$	$p \in \mathcal{P}_{\alpha\beta}$
t_p^{β}	$D_p(\beta) + \sum_{q \in \mathcal{N}_p,\, q \notin \mathcal{P}_{\alpha\beta}} V(\beta, f_q)$	$p \in \mathcal{P}_{\alpha\beta}$
$e_{\{p,q\}}$	$V(\alpha, \beta)$	$\{p,q\} \in \mathcal{N}$, $p,q \in \mathcal{P}_{\alpha\beta}$



Graph cut (5)

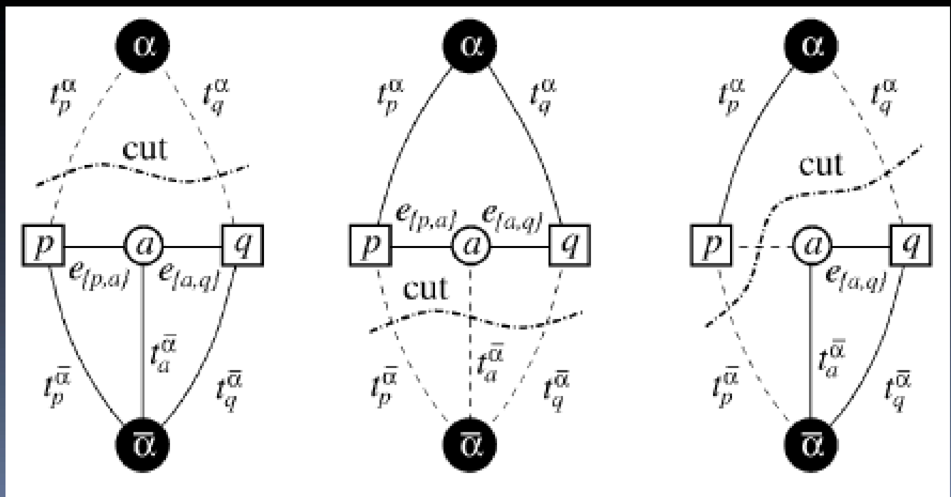
- α expansion
  - Move only in a label
  - Vertices
    - pixels
    - labels (terminal)
    - auxiliary nodes (between separated pixels)
  - Edges
    - t-link: terminal-pixel
    - n-link: neighboring pixel-pixel, pixel-auxiliary node



Graph cut (6)

- α expansion
  - Energy function
    - Edge weights
    - Data term and prior model
    - MRF model
  - Graph cut
    - Minimal edge cut

edge	weight	for
$t_p^{\bar{\alpha}}$	∞	$p \in \mathcal{P}_\alpha$
$t_p^{\bar{\alpha}}$	$D_p(f_p)$	$p \notin \mathcal{P}_\alpha$
t_p^{α}	$D_p(\alpha)$	$p \in \mathcal{P}$
$e_{\{p,a\}}$	$V(f_p, \alpha)$	$\{p,q\} \in \mathcal{N},\ f_p \neq f_q$
$e_{\{a,q\}}$	$V(\alpha, f_q)$	
$t_a^{\bar{\alpha}}$	$V(f_p, f_q)$	
$e_{\{p,q\}}$	$V(f_p, \alpha)$	$\{p,q\} \in \mathcal{N},\ f_p = f_q$



Graph cut (7)

- Algorithms: α-β swap (top), α expansion (bottom)

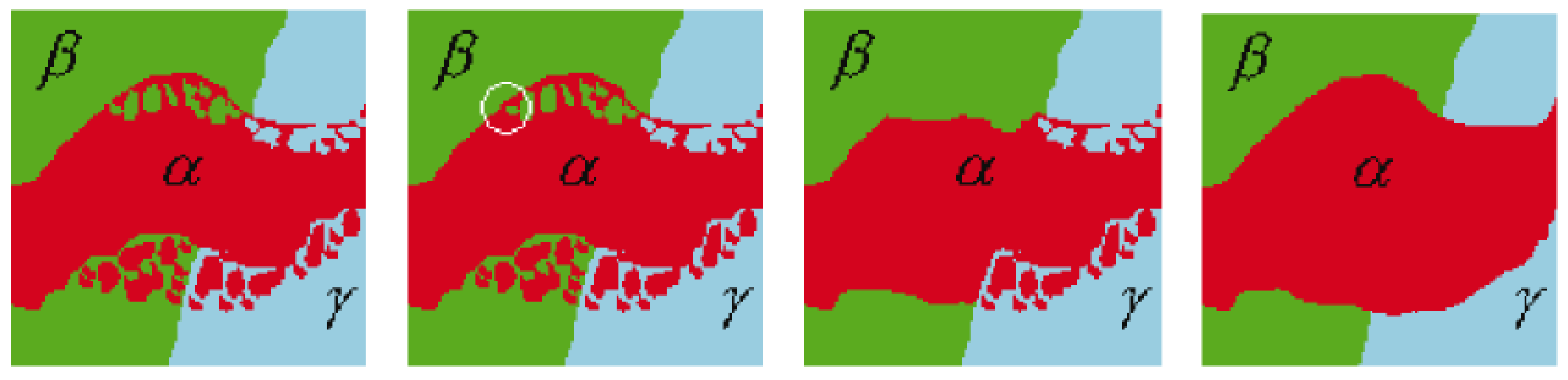
1. Start with an arbitrary labeling f
2. Set success := 0
3. For each pair of labels $\{\alpha, \beta\} \subset \mathcal{L}$
   - 3.1. Find $\hat{f} = \arg\min E(f')$ among f' within one α–β swap of f
   - 3.2. If $E(\hat{f}) < E(f)$, set f := $\hat{f}$ and success := 1
4. If success = 1 goto 2
5. Return f

---

1. Start with an arbitrary labeling f
2. Set success := 0
3. For each label $\alpha \in \mathcal{L}$
   - 3.1. Find $\hat{f} = \arg\min E(f')$ among f' within one α-expansion of f
   - 3.2. If $E(\hat{f}) < E(f)$, set f := $\hat{f}$ and success := 1
4. If success = 1 goto 2
5. Return f

Graph cut (8)

- α-β swap and α expansion: comparison
  - α expansion guarantees the optimality generally
  - α expansion is better and faster



Single-move α-β swap α expansion

CONCLUSIONS

Some Comments (1)

- The results are mostly dependent on the MRF models, not energy minimization.
  - The performances of energy minimization methods are critically related to MRF models and image characteristics.
    - Need of elaborate optimization
  - Energy minimization methods effect mainly on the properties of speed, convergence, and stability.
  - We can’t tell which method is better in any problems.
- The best performance results from the optimization of MRF models, parameters, and applied energy minimization method.

Some Comments (2)

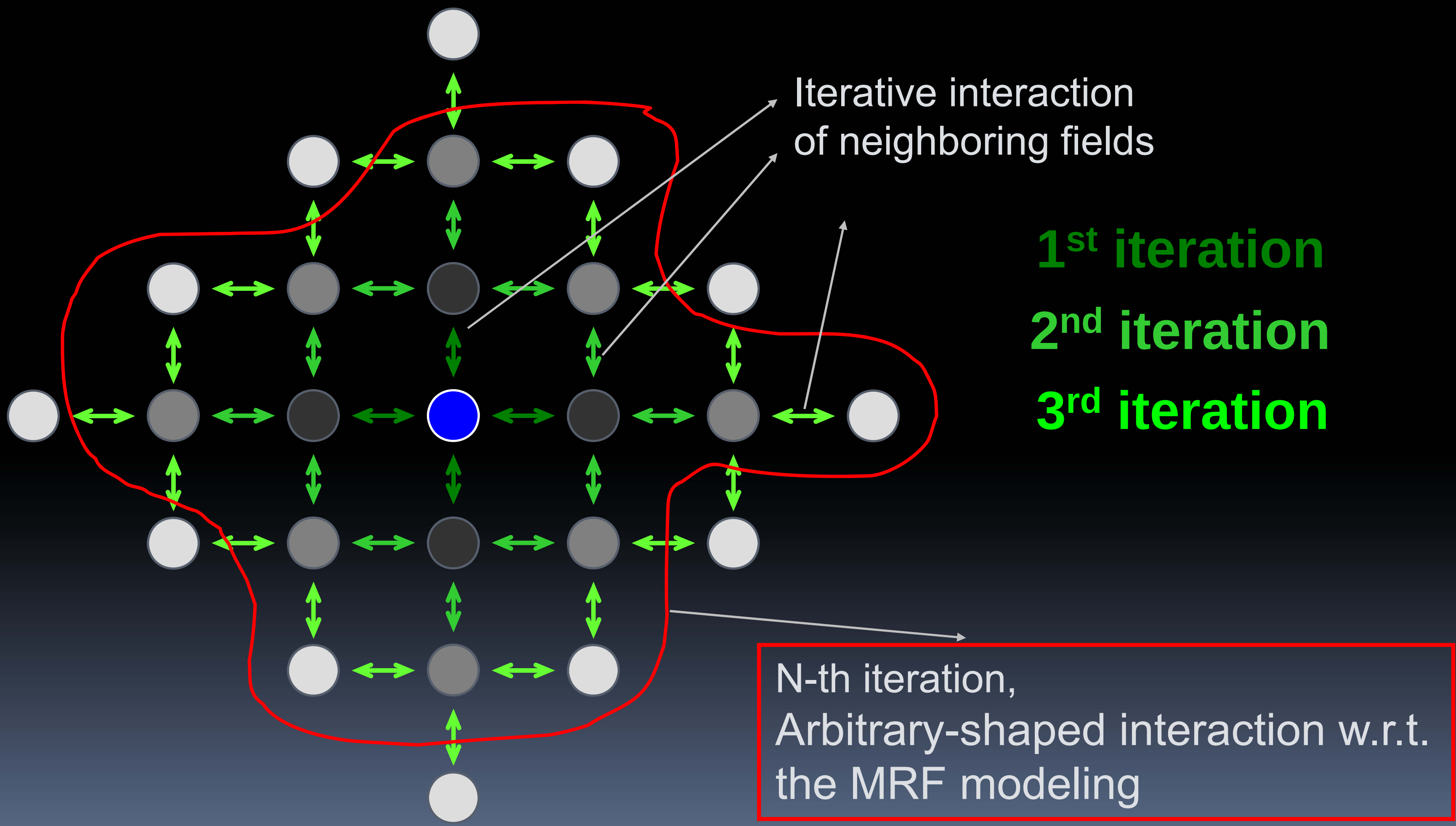
- Theoretically, guaranteeing the global minimization dose not always obtain the global solution in the real implementation.
  - You just find a sub-optimum.
  - You don’t know exactly the solution you have found is globally optimal.
  - The sub-optimum depends on the parameters, initial conditions (likelihood model).
  - Examples:
    - Simulated annealing
    - Binary graph cut with Pott’s model
- You can find the best sub-optimum in your work.

Some Comments (3)

- The global optimum is not always the best solution as you expect.
  - The global optimum depends on the energy function that you have designed.
  - The solution is the best only in your MRF models and energy function.
  - Global optimum is mainly decided by the prior models.
    - Likelihood model
    - Types of energy measure functions
    - Energy minimization methods
- You should design the proper prior models to obtain the correct results as you expect.

Some Comments (4)

- 1^{st}-MRF window is sufficient in pixel-wise dense fields.



Some Comments (5)

Convergence of iterative energy minimization

- Ex: stereo matching