Presented at IIT Bombay, India (January 2006)

# Energy Minimization Methods in Image Segmentation

Zoltan Kato

Institute of Informatics

University of Szeged

Hungary

Zoltan Kato has been partially supported by the Janos Bolyai Research Fellowship of the Hungarian Academy of Science, Hungarian Scientific Research Fund - OTKA T046805, National University of Singapore: NUS research grant R252-000-085-112, CWI Amsterdam: ERCIM postdoctoral fellowship, Hungarian – French Bilateral Fund

# Overview

- Probabilistic approach
  - Markov Random Field (MRF) models
  - Markov Chain Monte Carlo (MCMC) sampling
- Variational approach
  - Shape priors for variational models

# Why MRF Modelization?

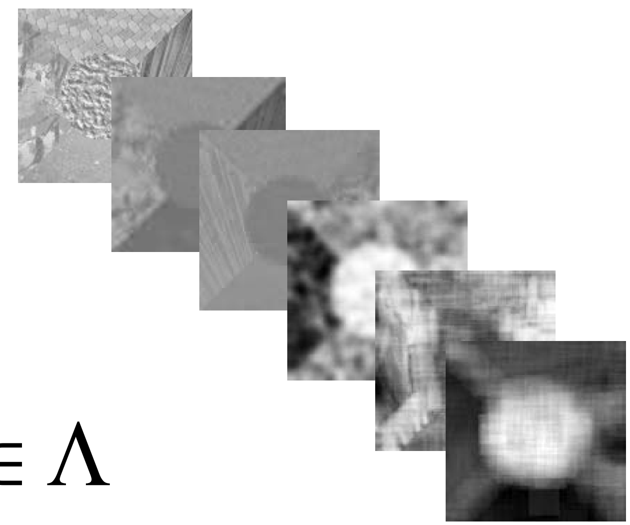
- In real images, regions are often homogenous; neighboring pixels usually have similar properties (intensity, color, texture, …)
- Markov Random Field (MRF) is a statistical model which captures such contextual constraints
- Well studied, strong theoretical background
- Allows MCMC sampling of the (hidden) underlying structure.

# What is MRF

- To give a formal definition for Markov Random Field, we need some basic building blocks
  - Observation Field and Labeling Field
  - Pixels and their Neighbors
  - Cliques and Clique Potentials
  - Energy function
  - Gibbs Distribution

# Overview of MRF Approach - Labelling

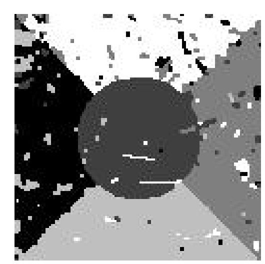
1. Extract features from the input image
   - Each pixel *s* in the image has a feature vector $\vec{f}_s$
   - For the whole image, we have



$$f = \{\vec{f}_s : s \in S\}$$

2. Assign each pixel *s* a label
   - For the whole image, we have $\omega_s \in \Lambda$

$$\omega = \{\omega_s, s \in S\}$$



# Probability Measure, MAP

- For an *nXm* image, there are $(n \cdot m)^{\Lambda}$ possible labelings.
  - Which one is the right segmentation?
- Define a *probability measure* on the set of all possible labelings and select the most likely one.
- $P(\omega \mid f)$ measures the probability of a labelling, given the observed feature $f$
- Our goal is to find an optimal labeling $\hat{\omega}$ which *maximizes* $P(\omega \mid f)$
- This is called the *Maximum a Posteriori* (MAP) estimate:

$$\hat{\omega}^{MAP} = \arg\max_{\omega \in \Omega} P(\omega \mid f)$$

# Bayesian Framework

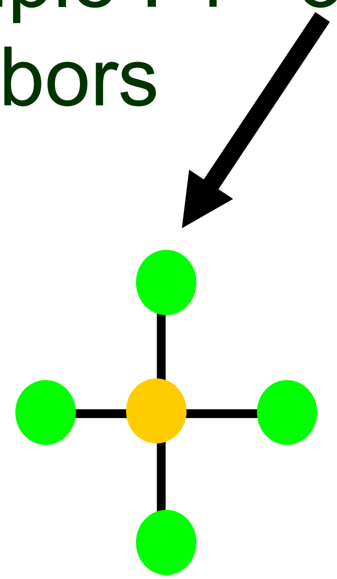
- By Bayes Theorem, we have

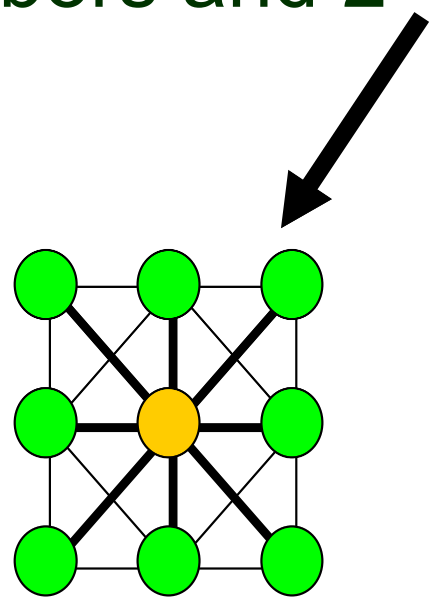
$$P(\omega \mid f) = \frac{P(f \mid \omega)P(\omega)}{P(f)}$$

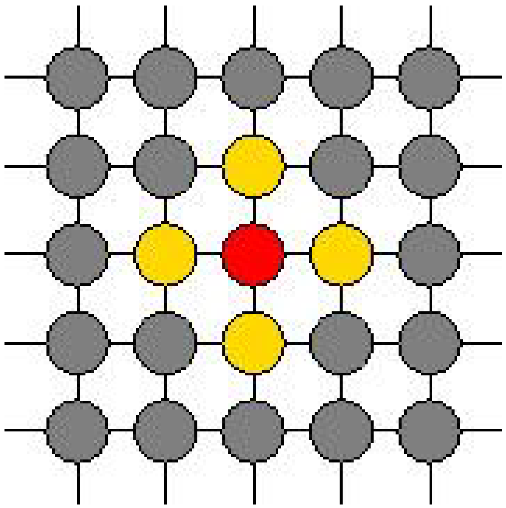
- $P(f)$ is constant
- We need to define $P(\omega)$ and $P(f \mid \omega)$ in our model

# Definition – Neighbors

- For each pixel, we can define some surrounding pixels as its neighbors.
- Example : 1st order neighbors and 2nd order neighbors





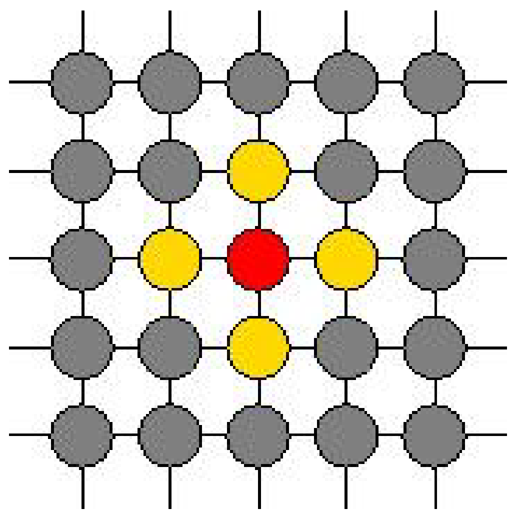


# Definition – MRF

- The labeling field *X* can be modeled as a Markov Random Field (MRF) if
  1. For all $\omega \in \Omega : P(\mathrm{X} = \omega) > 0$
  2. For every $s \in S$ and $\omega \in \Omega$ ,

     $$P(\omega_s \mid \omega_r, r \neq s) = P(\omega_s \mid \omega_r, r \in N_s)$$

  $N_s$ denotes the neighbors of pixel *s*



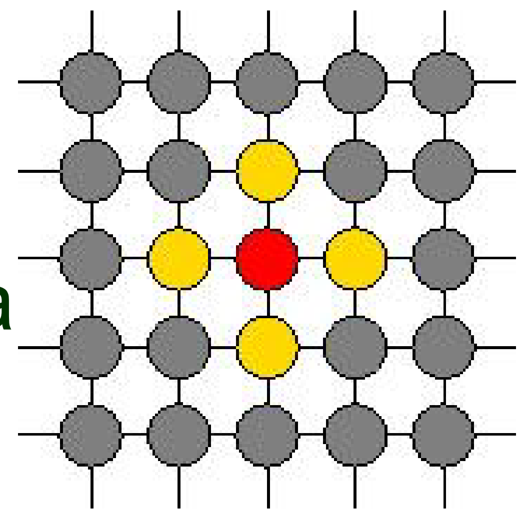
# Hammersley-Clifford Theorem

- The Hammersley-Clifford Theorem states that a random field is a MRF if and only if $P(\omega)$ follows a Gibbs distribution.

$$P(\omega) = \frac{1}{Z}\exp(-U(\omega)) = \frac{1}{Z}\exp(-\sum_{c \in C} V_c(\omega))$$

- where $Z = \sum_{\omega \in \Omega} \exp(-U(\omega))$ is a normalization constant

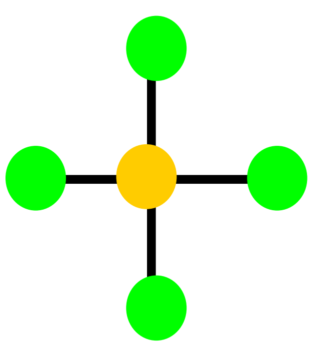
- This theorem provides us an easy way of defining MRF models via *clique potentials*.

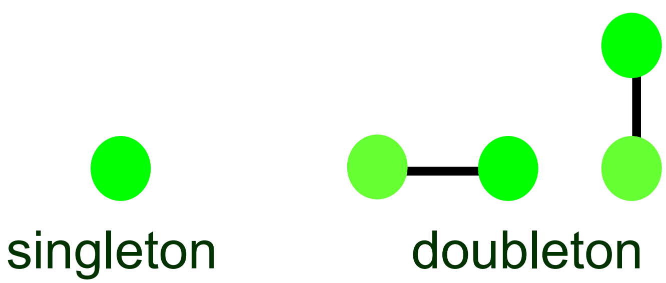


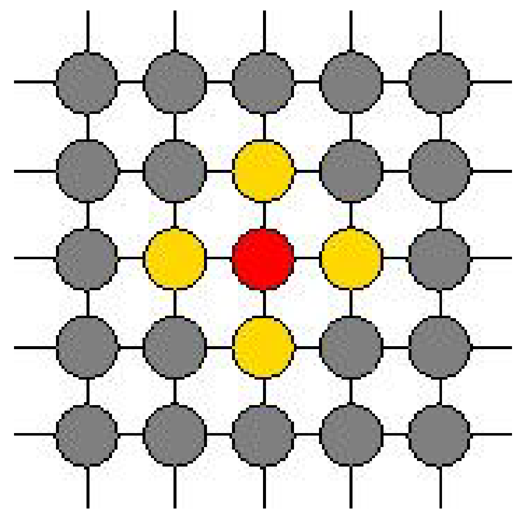
# Definition – Clique

- A subset $C \subseteq S$ is called a *clique* if every pair of pixels in this subset are neighbors.
- A clique containing *i* pixels is called *$i^{th}$ order clique, denoted by* $C_i$ .
- The set of cliques in an image is denoted by

$$C = C_1 \cup C_2 \cup \ldots \cup C_n$$



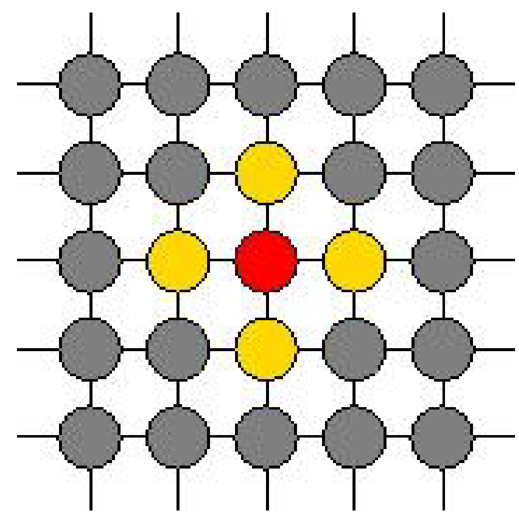


# Definition – Clique Potential

- For each clique *c* in the image, we can assign a value $V_c(\omega)$ which is called *clique potential* *of* c, where $\omega$ is the configuration of the labeling field
- The sum of potentials of all cliques gives us energy $U(\omega)$ of the configuration $\omega$

$$U(\omega) = \sum_{c \in C} V_C(\omega) = \sum_{i \in C_1} V_{C_1}(\omega_i) + \sum_{(i,j) \in C_2} V_{C_2}(\omega_i, \omega_j) + \ldots$$



# Current work in MRF modeling

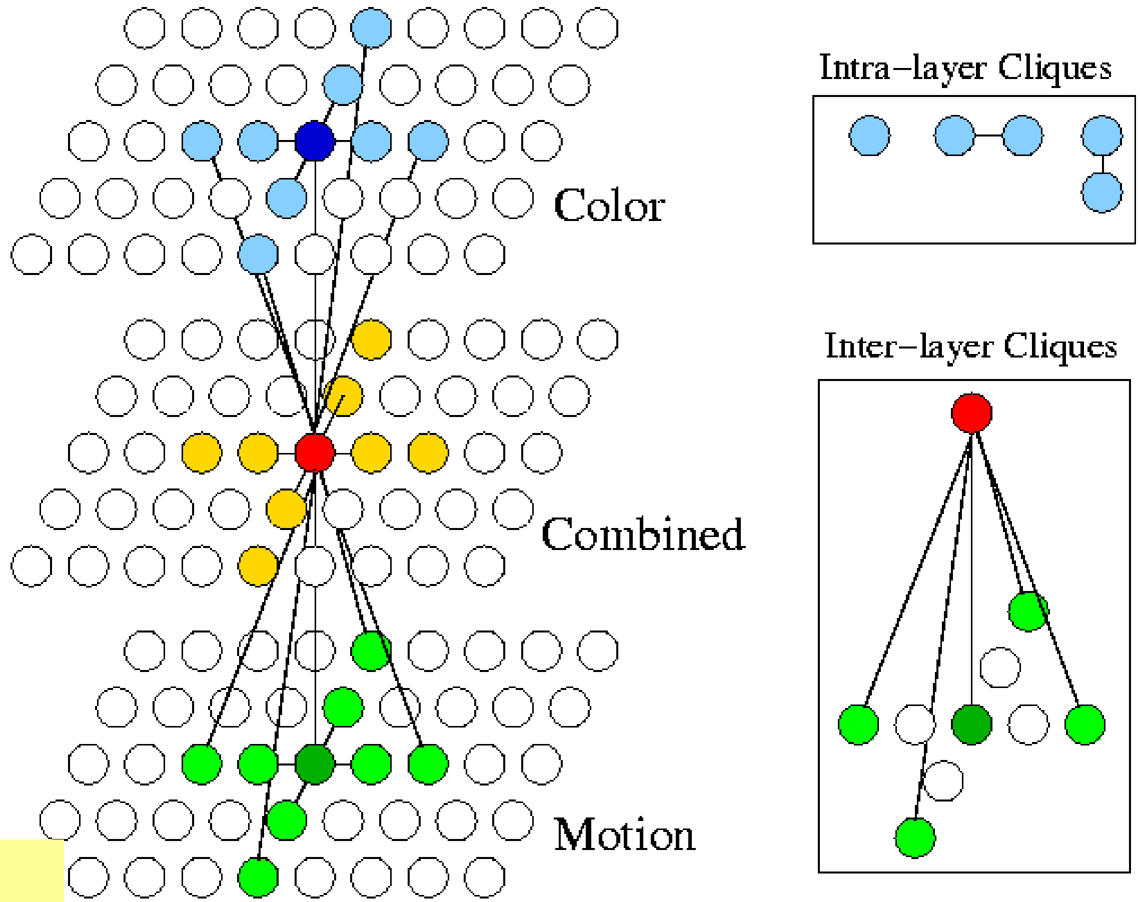
- Multi-layer MRF model for combining different segmentation cues:
  - Color & Texture [ICPR2002,ICIP2003]
  - Color & Motion [HACIPPR2005, ACCV2006]
  - …?

# Project Objectives

- Multiple cues are perceived simultaneously and then they are integrated by the human visual system [Kersten *et al. An. Rev. Psych.* 2004]
  - Therefore different image features has to be handled in a parallel fashion.
- We attempt to develop such a model in a Markovian framework
  - Collaborators:
    - Ting-Chuen Pong from HKUST – Hong Kong

# Multi-Layer MRF Model: Neighborhood & Interactions

- $\omega$ is modeled as a **MRF**
  - Layered structure
  - "Soft" interaction between features
- ➔ *P($\omega$ | f)* follows a **Gibbs distribution**
  - Clique potentials define the local interaction strength
- MAP ⇔ **Energy minimization** (*U($\omega$)*)



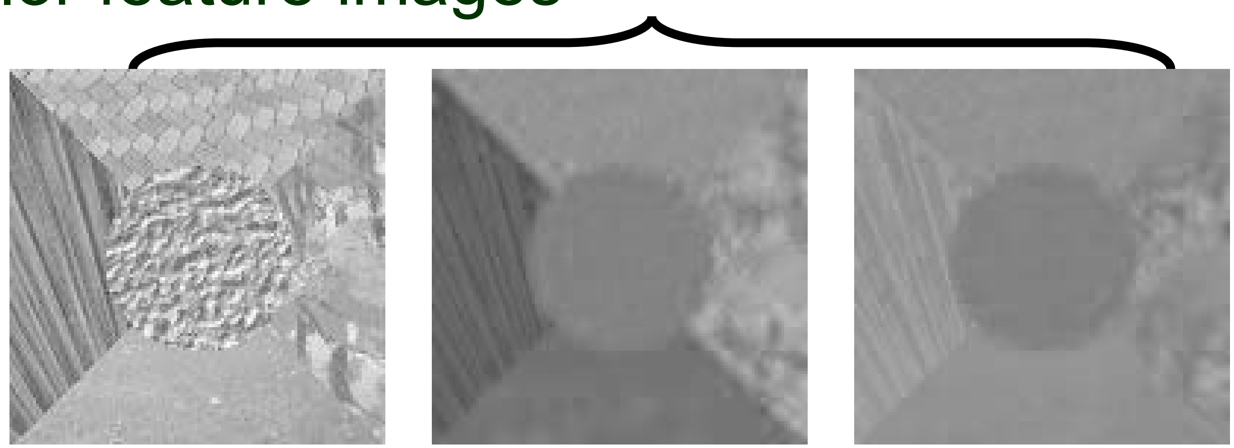

Hammersley-Clifford Theorem:

$$P(\omega) = \frac{\exp(-U(\omega))}{Z} = \frac{\exp(-\sum_{C} V_C(\omega))}{Z}$$

**Model ⇔ Definition of clique potentials**

# Extract Color Feature

- We adopt the CIE-L*u*v* color space because it is ***perceptually uniform***.
  - Color difference can be measured by Euclidean distance of two color vectors.
- We convert each pixel from RGB space to CIE-L*u*v* space ➔
  - We have 3 color feature images

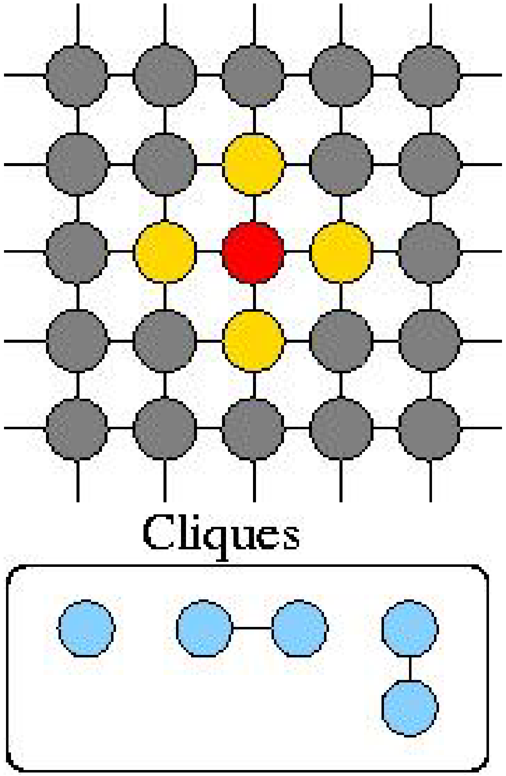


$L^*$ $u^*$ $v^*$

# Color Layer: MRF model

- Pixel classes are represented by multivariate Gaussian distributions:

$$P(f_s \mid \omega_s) = \frac{1}{\sqrt{(2\pi)^n \mid \Sigma_{\omega_s} \mid}} \exp(-\frac{1}{2}(\vec{f}_s - \vec{u}_{\omega_s})\Sigma_{\omega_s}^{-1}(\vec{f}_s - \vec{u}_{\omega_s})^T)$$




- **Intra-layer** clique potentials:
  - ***Singleton***: proportional to the likelihood of features given $\omega$: *log(P(f |* $\omega$*)).*
  - ***Doubleton***: favours similar classes at neighbouring pixels – ***smoothness prior***

$$V_{c_2}(i,j) = \begin{cases} -\beta & if \quad \omega_i = \omega_j \\ +\beta & if \quad \omega_i \neq \omega_j \end{cases}$$

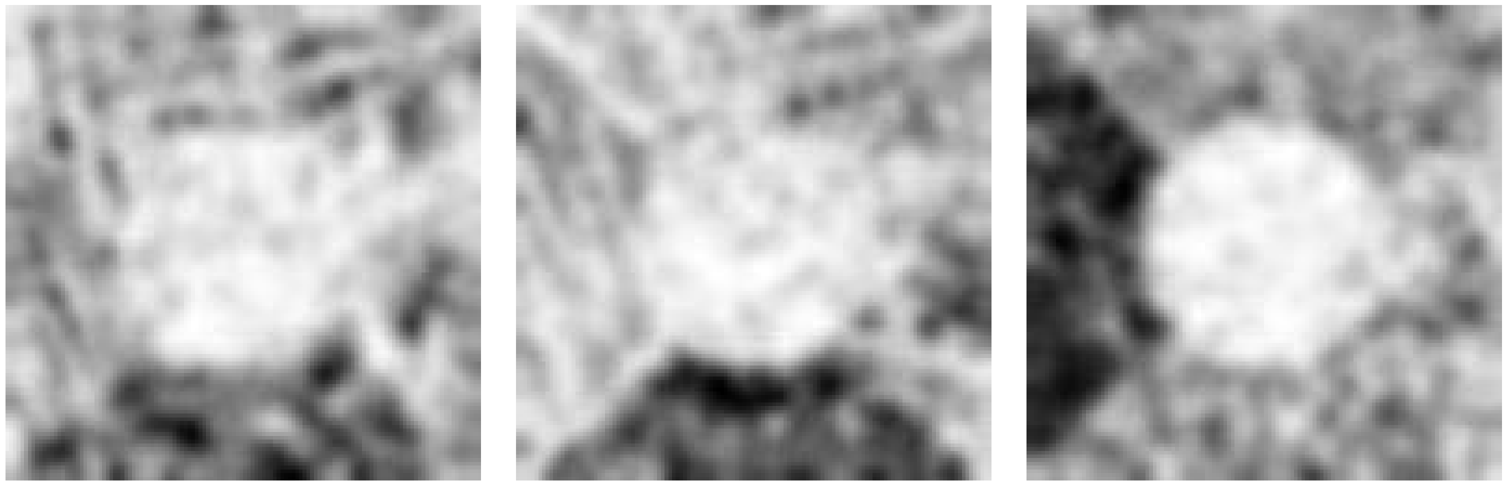
- [+ Inter-layer potentials (later…)]

# Texture Layer: MRF model

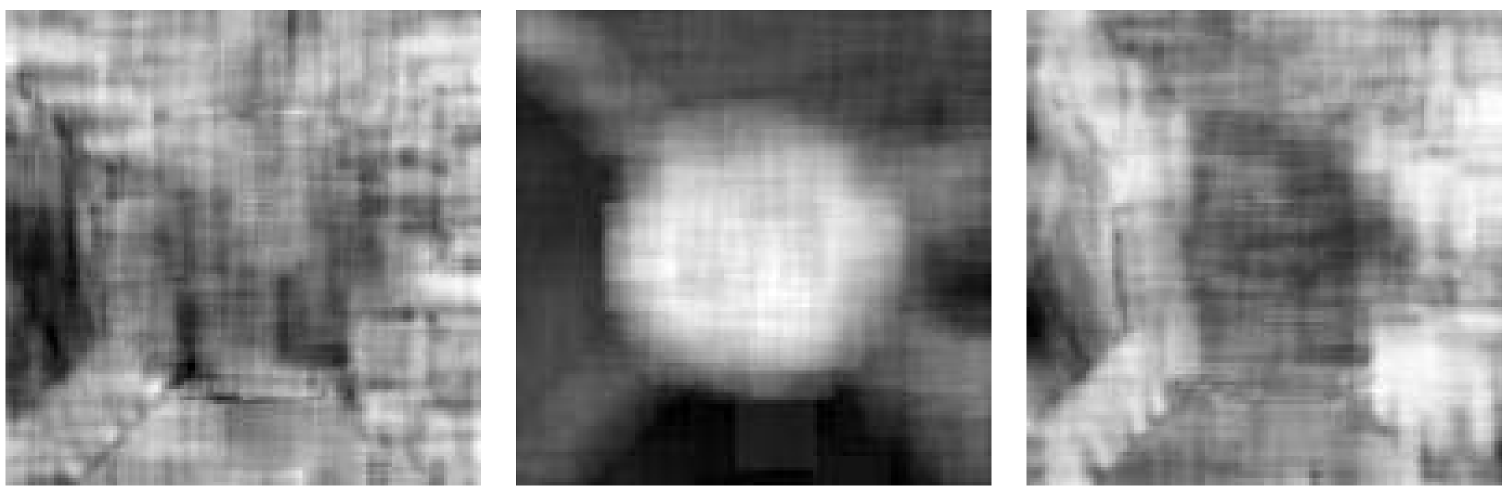
- We extract two type of texture features
  - *Gabor feature* is good at discriminating strong-ordered textures
  - *MRSAR feature* is good at discriminating weak-ordered (or random) textures
  - The number of texture feature images depends on the size of the image and other parameters.
    - Most of these doesn't contain useful information →
  - Select feature images with high discriminating power.
- MRF model is similar to the color layer model.

# Examples of Texture Features

**Gabor features:**



**MRSAR features:**



# Combined Layer: Labels

- A label on the combined layer consists of a pair of color and texture/motion labels such that $\eta_s = <\eta_s^c, \eta_s^m>$, where $\eta_s^c \in \Lambda^c$ and $\eta_s^m \in \Lambda^m$
- The number of possible classes is $L^c \times L^m$
- The combined layer selects the most likely ones.

# Combined Layer: ***Singleton*** potential

- Controls the number of classes:

$$V_s(\eta_s) = R\left((10 N_{\eta_s})^{-3} + P(L)\right)$$

  - $N_{\eta_s}$ is the percentage of labels belonging to class $\eta_s$
  - *L* is the number of classes present on the combined layer.
- Unlikely classes have a few pixels ➔ they will be penalized and removed to get a lower energy
- $P(L)$ is a log-Gaussian term:
  - Mean value is a guess about the number of classes,
  - Variance is the confidence.

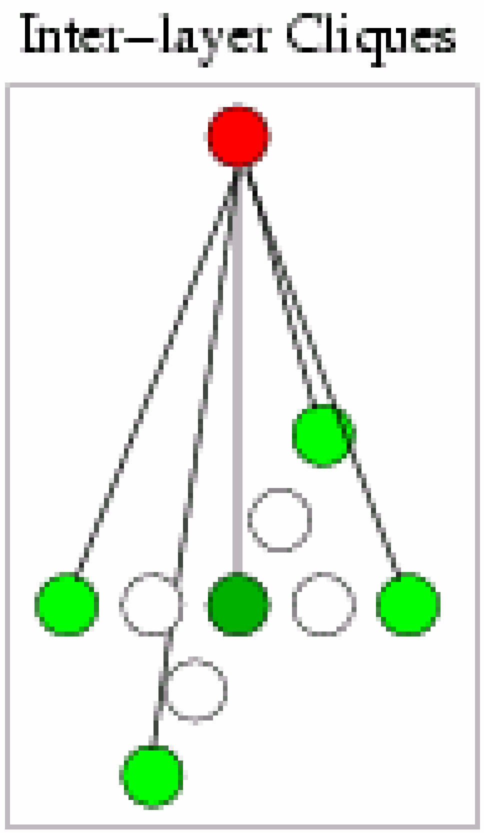
# Combined Layer: ***Doubleton*** potential

- Preferences are set in this order:
  1. Similar color and motion/texture labels
  2. Different color and motion/texture labels
  3. Similar color (resp. motion/texture) and different motion/texture (resp. color) labels
     - These are contours visible only at one feature layer.

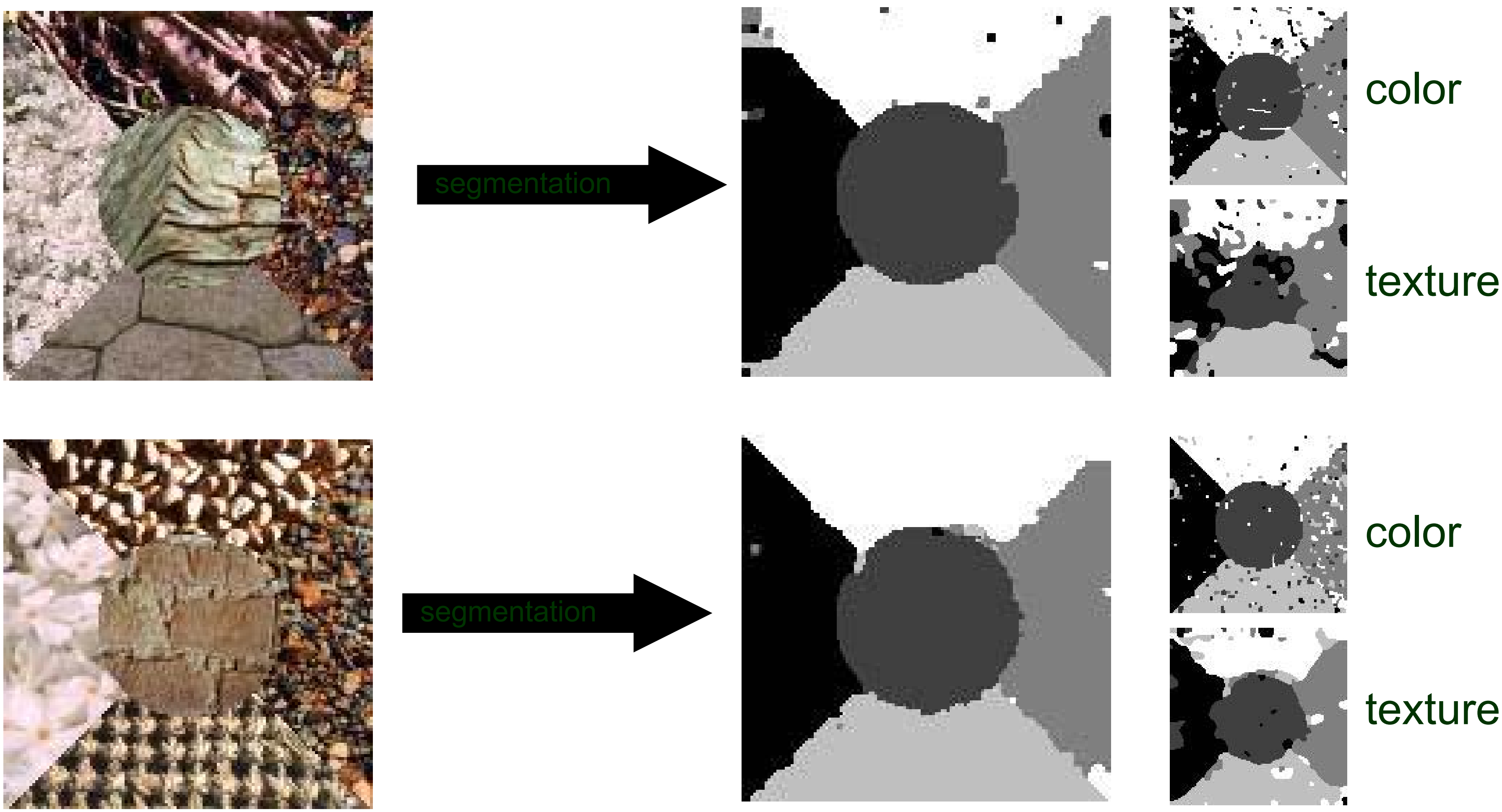
$$\delta(\eta_s, \eta_r) = \begin{cases} -\alpha & \text{if } \eta_s^c = \eta_r^c, \eta_s^m = \eta_r^m \\ 0 & \text{if } \eta_s^c \neq \eta_r^c, \eta_s^m \neq \eta_r^m \\ +\alpha & \text{if } \eta_s^c \neq \eta_r^c, \eta_s^m = \eta_r^m \\ +\alpha & \text{if } \eta_s^c = \eta_r^c, \eta_s^m \neq \eta_r^m \end{cases}$$

# Inter-layer clique potential

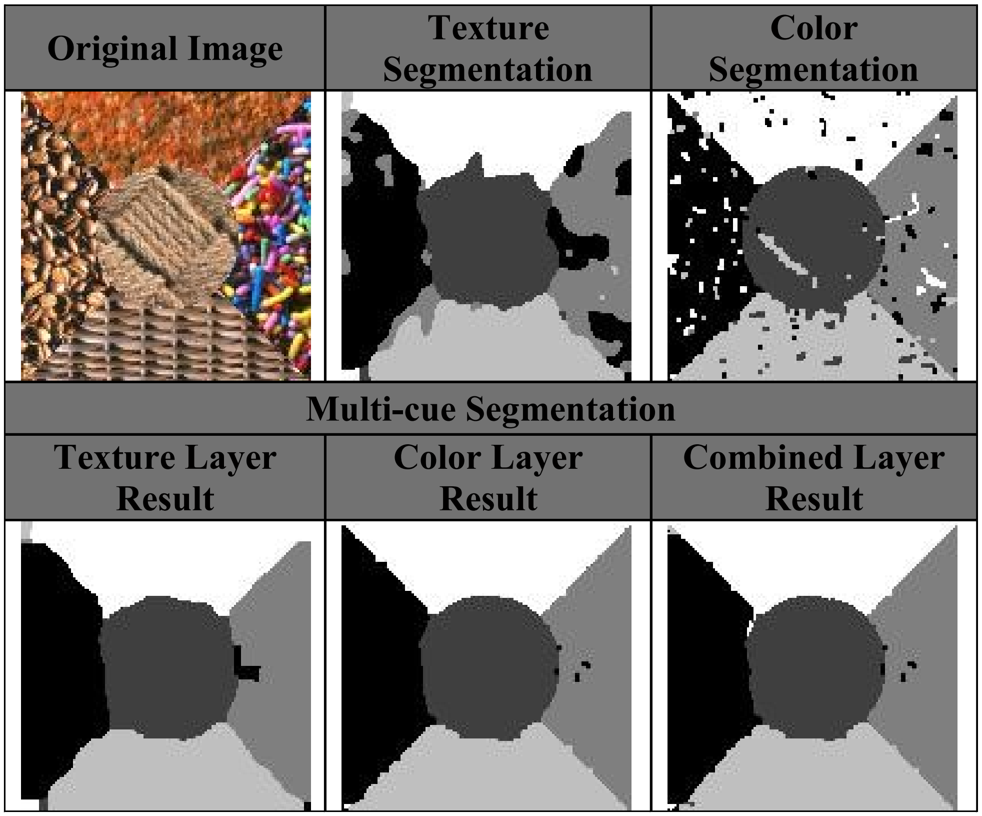
- Five pair-wise interactions between a feature and combined layer
- Potential is proportional to the difference of the singleton potentials at the corresponding feature layer.
  - Prefers $\omega_s$ and $\eta_s$ having the same label, since they represent the labeling of the same pixel
  - Prefers $\omega_s$ and $\eta_r$ having the same label, since we expect the combined and feature layers to be homogenous

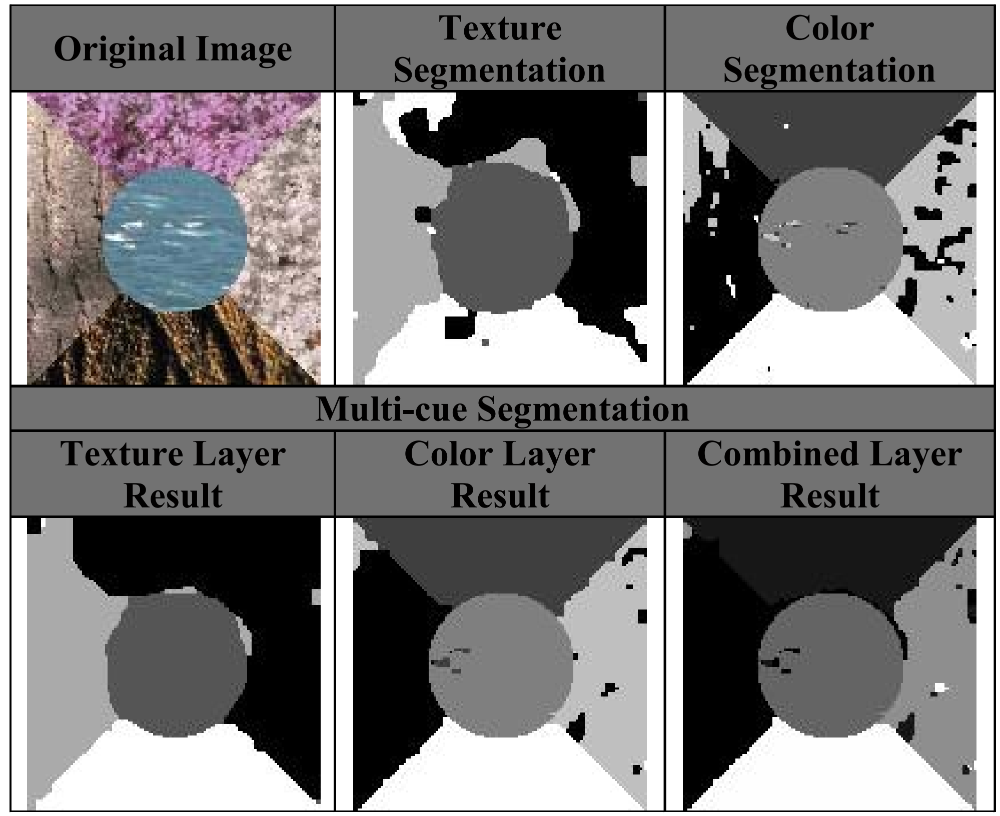


# Color Textured Segmentation

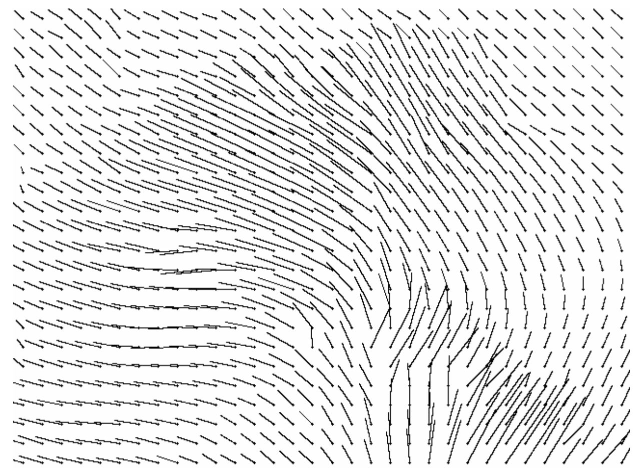


# Color Textured Segmentation

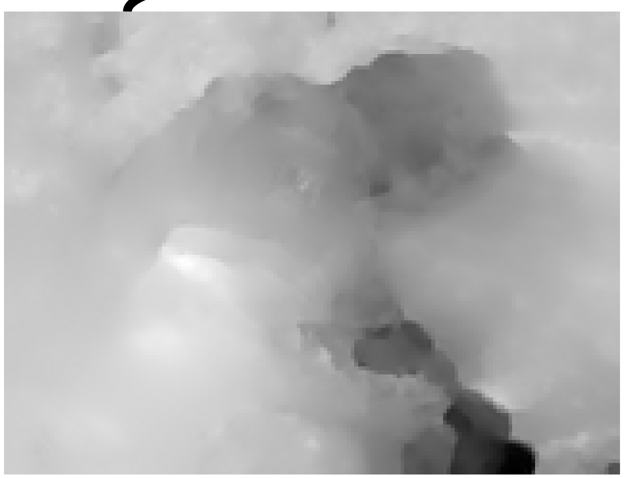


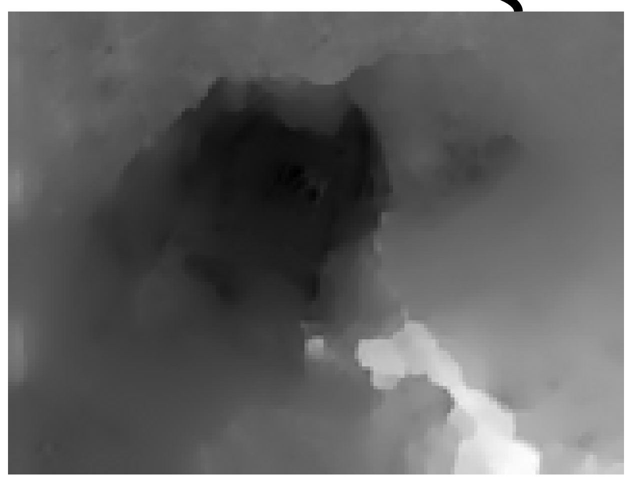



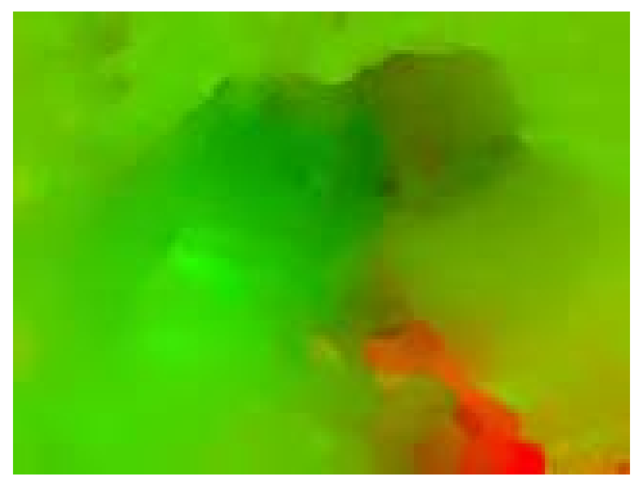
# Motion Layer:
# 1. Flow-based model



- Compute optical flow data which characterizes the visual motion of the pixels
  - Proesmans *et al.* [ECCV 1994]
  - 2D vector field ➔ we have 2 motion feature images







- Then a similar MRF model can be applied at the motion layer as for the color layer.
  - Note that the Gaussian likelihood implies a ***translational motion model***

# Motion Layer:
# 2. Motion compensated model

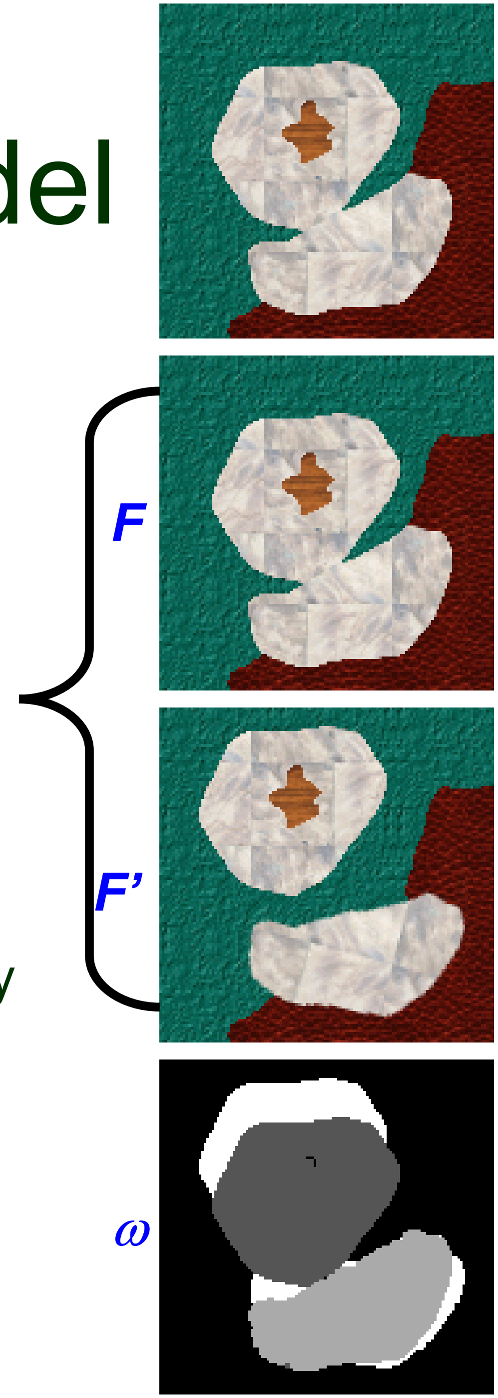
- Each motion-label is modeled by an affine motion model:

$$\underbrace{\mathbf{u}_s}_{\text{motion}} = \underbrace{\mathbf{A}(\omega_s)}_{\text{affine matrix}}\, s + \underbrace{\mathbf{T}(\omega_s)}_{\text{translation}}$$

  - $\mathbf{u}_s$ gives the motion at $s$ assuming label $\omega_s$
    - Given 2 successive frames ***F*** and ***F'*** and
    - assuming brightness / color constancy
  - we have the ***singleton*** potential

$$\|F(s)-F'(s+u_s)\|^2$$

- A special label is assigned to ***occluded*** pixels
  - Occluded pixels will have a high color difference for any motion label
  - ➔ occluded ***singleton*** potential is a constant penalty lower than these differences.
- ***Doubleton*** potential is the usual ***smoothness prior***.
- [+ Inter-layer potentials]



# Color & Motion Segmentation