Lecture 3. MRF's Modeling Vision

Note Title 1/17/2010

This lecture will introduce two types of Markov Random Field (MRF) models. These will be undirected graphs.

$$P(\underline{\omega}|I) = \frac{1}{Z[I]} e^{-\left\{\sum_{i \in V} \phi_i(\omega_i, I) + \sum_{ij \in E} \psi_{ij}(\omega_i, \omega_j)\right\}}$$

unary term depends only on one state ω_i

binary term depends on pairs of states

can include I but typically doesn't

Note: the first lecture described models with unary terms only.

Application — Image Labelling.

This extends the model from lecture 1. The binary term is used to impose spatial context. E.G. to make neighbouring pixels more likely to have the same label.

This is a type of weak spatial smoothness — smoothness is encouraged for the labeling (e.g. we encourage neighboring pixels to have the same label) but we allow smoothness to be broken at boundaries of regions.

Note: this model only imposes limited context using limited knowledge about real world images.

E.G. This model allows there to be many small "grass" regions while, in practice, grass usually forms a single contiguous region. Also in reality there are spatial relationships between regions — 'sky' must be above 'grass'.

Technically, $\psi_{ij}(\omega_i, \omega_j) = -A\,\delta(\omega_i, \omega_j) + B$ where $A, B > 0$

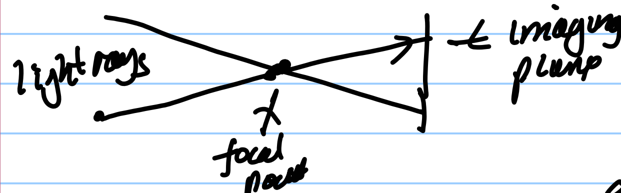
Models later in the course will relax some of these limitations. But limited models can work well, at least on restricted datasets.

Note: $\delta(\omega_i, \omega_j) = 1$, if $\omega_i = \omega_j$
$= 0$, otherwise.

(2) <u>Binocular Stereo</u> → Another application of this model!

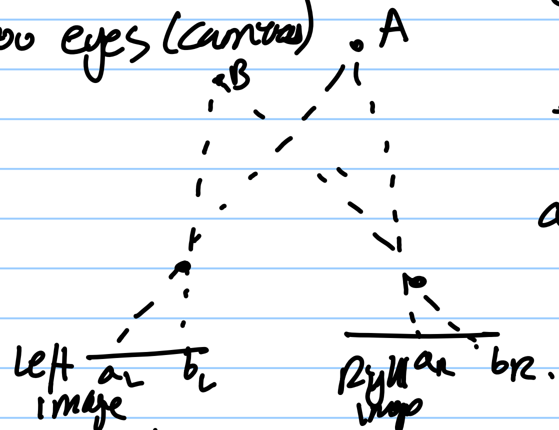
<u>Stereo = Depth</u>

Assume pinhole camera model of eye (or camera)

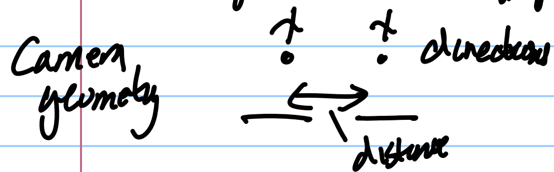


For more realistic models see Forsyth & Ponce. <u>But</u> the pinhole object is good to first order.

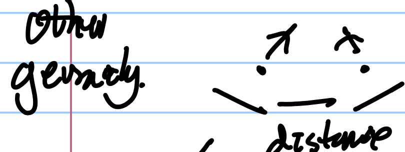
Two eyes (cameras)

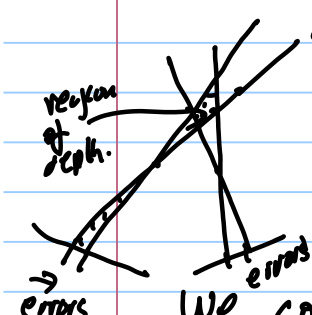


<u>If</u> we can match a_L to a_R and b_L to b_R then we can estimate the positions A & B. if we know the camera geometry.



<u>To</u> learn about camera geometry — read Forsyth & Ponce or Hartley & Zisserman.





In practice, there will be errors in positions a_L, a_R, causing errors in depth — hence depth estimates are poor if objects are at great distance compared to distance between cameras.

<u>We</u> concentrate on the 'correspondence problem' how to match a_L to a_R and b_L to b_R — instead of a_L to b_R and a_R to b_L, which gives the wrong depth.

<u>Disparity</u> is the key concept.

If point i in the left image matches a point $i + d(i)$ in the right image, then $d(i)$ is the disparity of i. The depth is inversely proportional to the disparity.

(3) Unary Term:

let $(f * I^L)_i$ be a filter applied to the left image and evaluated at i.

$(f * I^R)_i$ same for the right eye.

then, for example, $\phi_i(\omega_i, I_L, I_R) = | f(* I^L)_i - (f * I^R)_{i+\omega_i} |$

<u>note:</u> f can be vector valued, e.g. we could set

$$(f * I^L)_i = (I^L_{i-2}, I^L_{i-1}, I^L_i, I^L_{i+1}, I^L_{i+2})$$

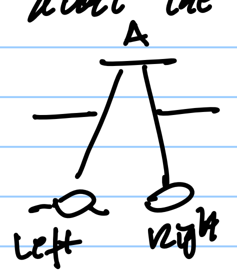
<u>note:</u> we are defining disparity relative to the left image – we could do it relative to the right instead

<u>Binary Term:</u> $\psi_{ij}(\omega_i, \omega_j)$ imposes weak smoothness on the disparity.

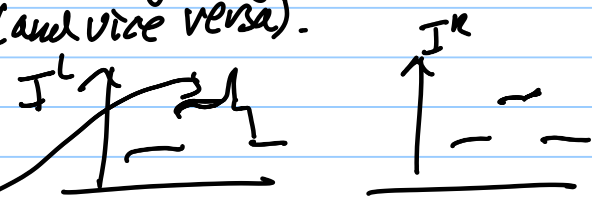
This corresponds to assuming that the depth is usually smooth but occasionally has a large discontinuity.

<u>Note:</u> weak smoothness can be justified from the statistics of real world images, (later in the course)

<u>Note:</u> this model only uses limited knowledge about the world. For example, consider 'half-occlusions'



parts of object A will be visible to the right eye and invisible to the left (and vice versa).



so unmatched points in the left eye can gives clues for the presence of discontinuities in depth (known to Leonardo da Vinci).

'Labeling' and 'Binocular Stereo' can both be modeled by the MRF on page 1.

Technically, this is a Conditional Random Field (CRF) since it specifies the distribution $P(W|I)$ only

(4) Another Model → Classic Geman & Geman.
Blake & Zisserman.

This model assumes that the observed image is a noisy version of an 'ideal image'.

$$P(I|\omega) = \prod_{i \in D} \sqrt{\frac{\tau}{\pi}}\, e^{-\tau(I_i - \omega_i)}$$

— generative noise model is additive Gaussian noise

$\tau = \frac{1}{2\sigma^2}$ ↗ precision, variance

The prior imposes smoothness but with line process terms which break smoothness e.g.

$$P(\omega, y) \propto \exp(-E(\omega, y))$$

$$E(\omega, y) = A \sum_{(i,j) \in \mathcal{E}} (\omega_i - \omega_j)^2 (1 - y_{ij}) + B \sum_{(i,j) \in \mathcal{E}} y_{ij}$$

$y_{ij} \in \{0, 1\}$

If $y_{ij} = 1$, then we break the smoothness between pixels i & j. If $y_{ij} = 0$, then we encourage smoothness.

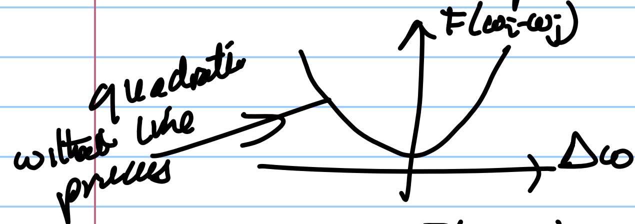
The $\{y_{ij}\}$ are latent variable, they determine where the smoothness is broken.

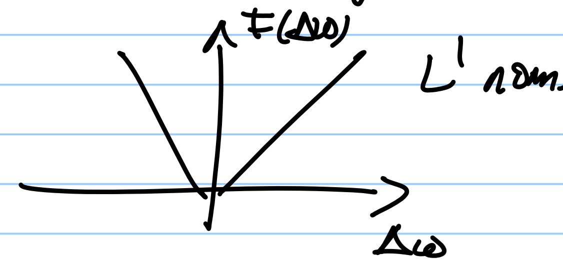
This model is called the 'weak membrane' model. Like weak smoothness, but membrane has a stronger meaning because smoothness could include higher order terms involving more than two pixels

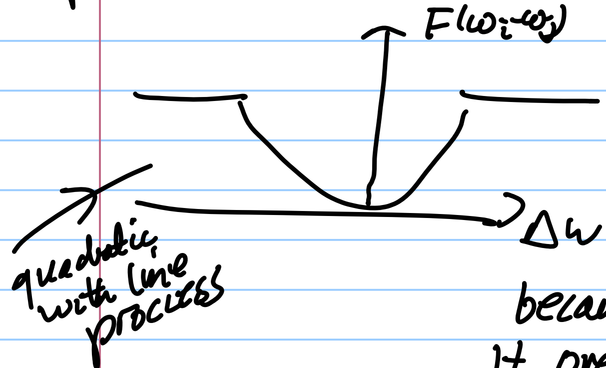
(5) What types of weak smoothness?

The intuition is that neighbouring pixels have similar states → whether the states are labels, disparity/depth, or intensity values.

There are some potentials: assumed to depend only on the difference $\Delta\omega = \omega_i - \omega_j$





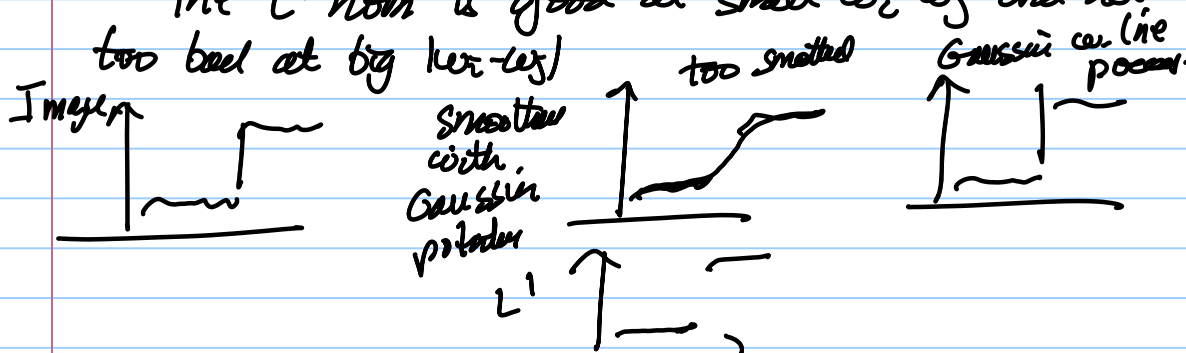


Comment (i) the quadratic norm is bad at large $\omega_i-\omega_j$ because it penalizes this quadratically. It oversmooths → quadratic = Gaussian implies non-robust.

Also, the Gaussian is bad at small $\omega_i-\omega_j$ because it doesn't penalize small fluctuations very much.

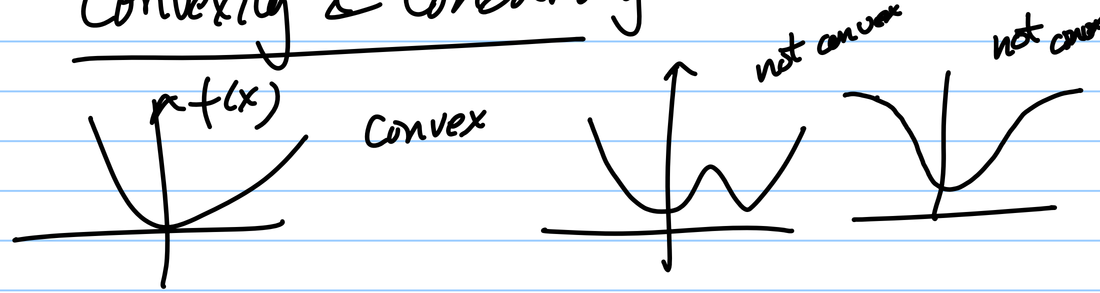
The Gaussian with line process is good at large $\Delta\omega$ because the line process cuts the smoothness.

The L^1 norm is good at small $\omega_i-\omega_j$ and not too bad at big $|\omega_i-\omega_j|$



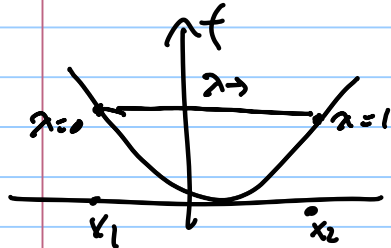
The L^1 norm also has the advantage of being convex which helps for inference (see next lecture).

(6) Convexity & Concavity



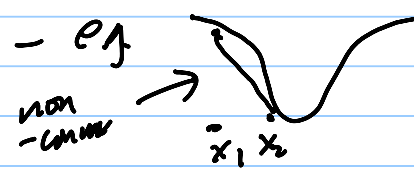
Convex if $f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$

for any x_1, x_2 & $\lambda \in [0,1]$



If $f(.)$ is a continuous differentiable function then convexity is equivalent to requiring that the Hessian $\frac{\partial^2 f}{\partial x \partial x}$ is positive definite.

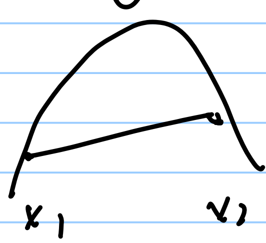
A convex function has at most one minimum. But there are functions which are not convex which have a single minimum - eg



Convexity is maintained if we impose linear constraint e.g. $\lambda f(x) = \alpha$, for constant α

If a function is convex then we can usually find an algorithm that can find its minimum..
→ see books of convex optimization.

Concavity is the opposite of convexity



$f(\lambda x_1 + (1-\lambda) x_2)$
$\geq \lambda f(x_1) + (1-\lambda) f(x_2)$

If $f(.)$ is convex then $-f(.)$ is concave, and vice versa.