Professor Anil Kokaram

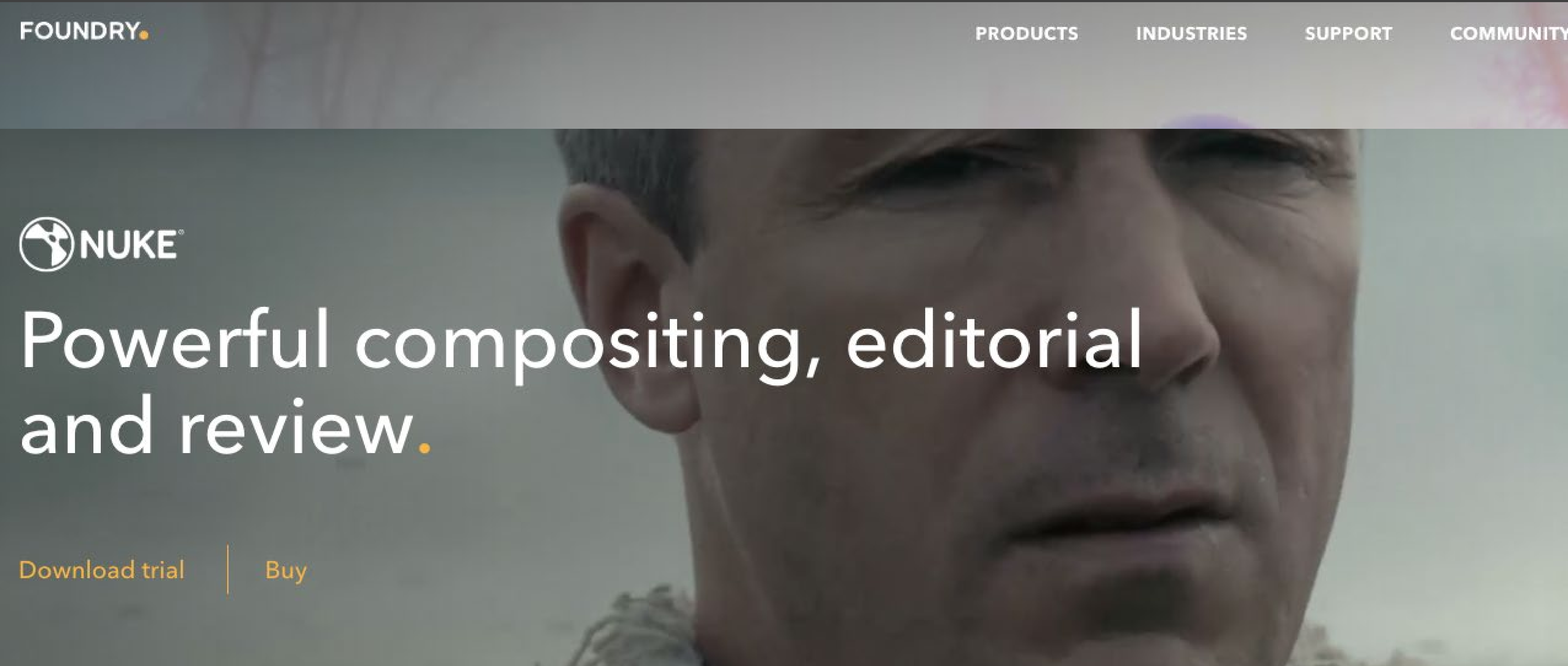
Anil.Kokaram@tcd.ie

5C1 Motion Picture Engineering

Digital Video Engineering

Engineering for Moving Pictures

Last time you had a tiny taste of NUKE



Some points to remember

- In Cinema Postproduction pixel depth is HUGE
  - NUKE uses FLOATS scaled 0 – 1
  - Typical to find 2 byte integers (0-65535)
  - A pixel stored as RED GREEN BLUE , each number being floats or 2 byte integers
  - Modern picture sizes
    - 720p (1280 x 720)
    - 1080p (1920 x 1080)
    - 4K (3840 x 2160)
    - 8K, 16K
- In streaming media pixel depth is SMALL!
  - 1 BYTE PER PIXEL PER COLOUR CHANNEL
  - High Dynamic Range pixels use 10 bits per pixel
- Matlab uses the top left corner of an image as the origin of the coordinate system
- NUKE uses the bottom left corner

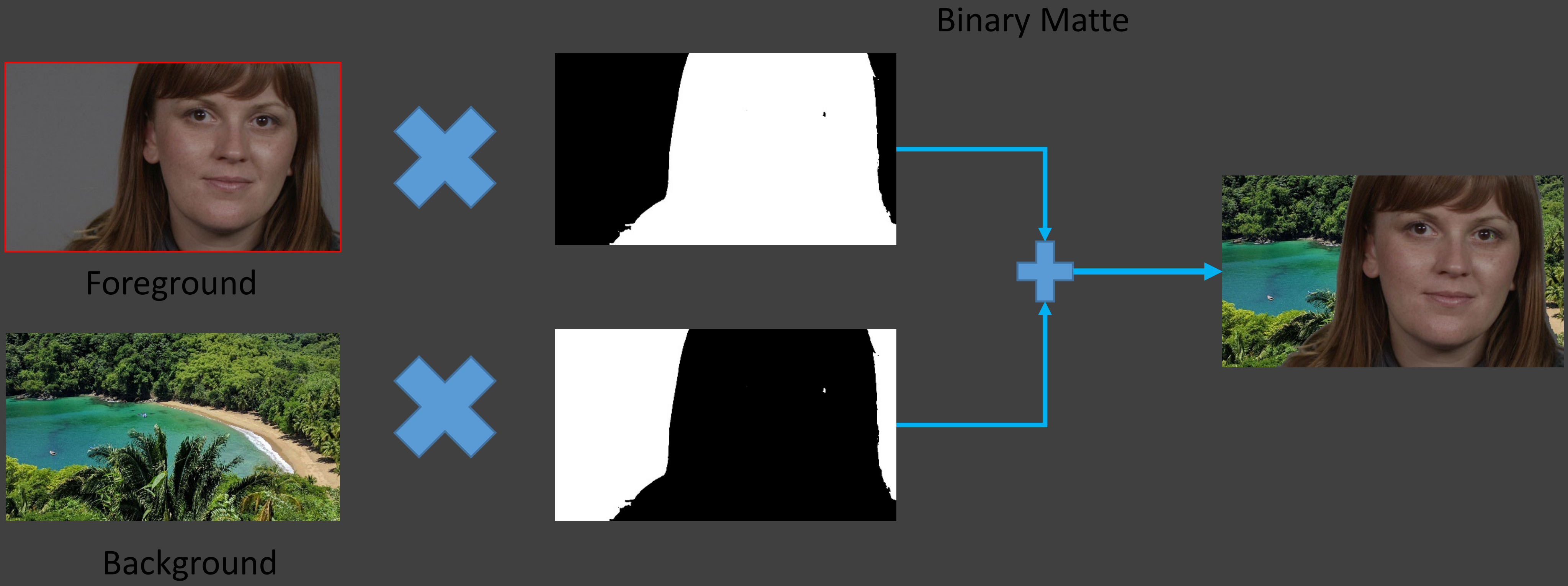
Markov Random Fields

Capturing the idea of smoothness in our solutions

$$\mathbf{Y} = \mathbf{F} \times \alpha + \mathbf{B} \times (1 - \alpha)$$

$\mathbf{x}$ is the position of the pixel with value $I(\mathbf{x})$ Let $\alpha(\mathbf{x})$ be the binary matte value at that site. Foreground is indicated by $\alpha = 1$

Colour Keying



Colour segmentation/keying is about discovering $\alpha(\mathbf{x})$ based primarily on colour alone.

To be clearer….

$$\boldsymbol{I}(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x}) \times \alpha(\boldsymbol{x}) + \boldsymbol{B}(\boldsymbol{x}) \times (1 - \alpha(\boldsymbol{x}))$$

The pixel of interest is at site $\boldsymbol{x} = [h, k]$ and it has the value $\boldsymbol{I}(x)$ in the COMPOSITE image.
The foreground and background pixels at that same site have values $\boldsymbol{F}(\boldsymbol{x})$ and $\boldsymbol{B}(\boldsymbol{x})$
The matte at the same site has the value $\alpha(x)$ and is either 0 or 1 in this case

Let us consider $p(\alpha(\boldsymbol{x})|I(\boldsymbol{x}))$

We want to choose α which is most probable GIVEN the observed value of a pixel $I(\mathbf{x})$ at that same site.

If $p(\alpha = 1|I) > p(\alpha = 0|I)$ then $\alpha = 1$ and that site must be foreground

If $p(\alpha = 0|I) > p(\alpha = 1|I)$ then $\alpha = 0$ and that site must be background

The implicit algorithm here is that we are making decisions on a per-pixel basis. Visting each site in turn asking "Does this pixel colour look more like the background colour than the foreground colour?" If it does then it must be the background then.

$$p(\alpha(\mathbf{x})|I(\mathbf{x})) = \frac{\overbrace{p(I(\mathbf{x})|\alpha(\mathbf{x}))}^{\text{Likelihood}}\,\overbrace{p(\alpha)}^{\text{Prior}}}{\underbrace{p(I(\cdot))}_{\text{Evidence}}}$$

$$p(\alpha|I) = \begin{cases} \dfrac{p(I|\alpha{=}1)p(\alpha{=}1)}{p(I(\cdot))} & \alpha = 1 \\[2ex] \dfrac{p(I|\alpha{=}0)p(\alpha{=}0)}{p(I(\cdot))} & \alpha = 0 \end{cases}$$

MARKOV RANDOM FIELD

Kindof obvious intuition about how variables iteract in a local region

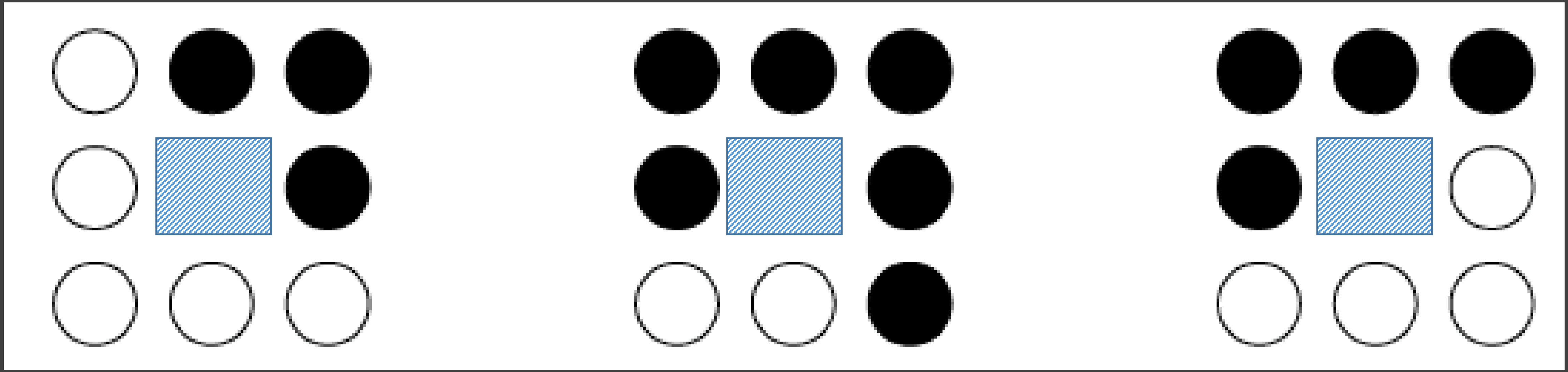
MRF

- Captures this idea: To find out things about my current *site*, I only need to know what is happening in the *neighbourhood* of that site and not anything else.

$$p(b(\mathbf{x})|\mathbf{B}) = p(b(\mathbf{x})|\mathbf{B}(\mathbf{x} \in \mathcal{N})) \tag{1}$$

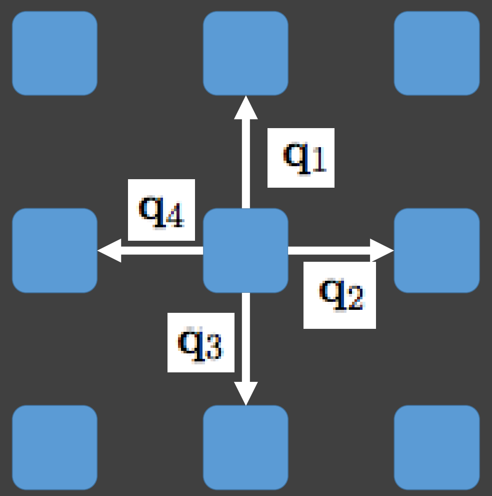
- Multimensional version of a Markov Processes
- Established in Signal Processing (actually Control) since 1972 by John Woods and heavily used in image coding schemes,
- Popularised by the Geman brothers in 1987 as Gibbs Fields together with their work on Simulated Annealing.
- Leads to well understood optimisation schemes and behaviours
- Tends to go hand in hand with Bayesian inference as it is most useful as a prior

$p(\alpha = 0|\mathcal{N}_\alpha) > p(\alpha = 1|\mathcal{N}_\alpha)$ $p(\alpha = 1|\mathcal{N}_\alpha) > p(\alpha = 0|\mathcal{N}_\alpha)$ $p(\alpha = 1|\mathcal{N}_\alpha) = p(\alpha = 0|\mathcal{N}_\alpha)$



A Markov Random Field (MRF) has this property. So let's use it as a prior for the Matte

The Gibbs Energy Function as MRF



$$\mathbf{q}_1 = [0 \ -1]$$
$$\mathbf{q}_2 = [1 \ 0]$$
$$\mathbf{q}_3 = [0 \ 1]$$
$$\mathbf{q}_4 = [-1 \ 0]$$

$$p(\alpha(\mathbf{x})|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z} \exp -\Lambda \left[\sum_{k=1}^{4} \lambda_k |\alpha(\mathbf{x}) \neq \alpha(\mathbf{x}+\mathbf{q}_k)| \right]$$

If $\alpha \in \pm 1$ can use Ising model

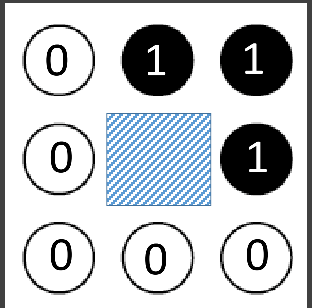
$$p(\alpha(\mathbf{x})|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z} \exp -\Lambda \left[\sum_{k=1}^{4} \lambda_k |\alpha(\mathbf{x})\alpha(\mathbf{x}+\mathbf{q}_k)| \right]$$

Hammersley-Clifford theorem posits that if we define local energy functions like this, then the field of variables over the whole image is an MRF

Some examples…

8-neighbourhood

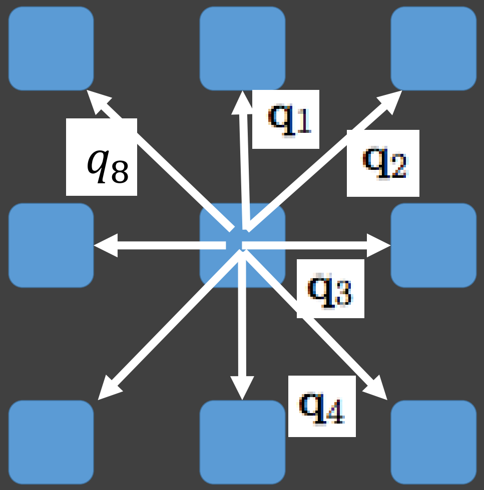
$$p(\alpha(\mathbf{x})|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z}\exp -\Lambda\left[\sum_{k=1}^{4}\lambda_k|\alpha(\mathbf{x}) \neq \alpha(\mathbf{x}+\mathbf{q}_k)|\right]$$



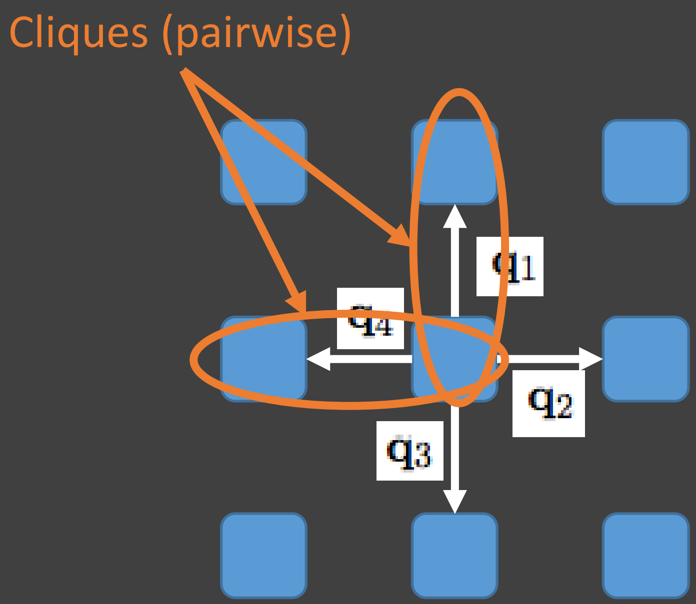
$$p(\alpha(\mathbf{x})=0|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z}\exp -\Lambda\left[\sum_{k=1}^{8}\lambda_k|0 \neq \alpha(\mathbf{x}+\mathbf{q}_k)|\right]$$
$$= \frac{1}{Z}\exp -\Lambda\left[1+1+1+0+0+0+0+0\right]$$
$$= \frac{1}{Z}\exp -\Lambda\left[3\right]$$

$$p(\alpha(\mathbf{x})=1|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z}\exp -\Lambda\left[\sum_{k=1}^{8}\lambda_k|1 \neq \alpha(\mathbf{x}+\mathbf{q}_k)|\right]$$
$$= \frac{1}{Z}\exp -\Lambda\left[0+0+0+1+1+1+1+1\right]$$
$$= \frac{1}{Z}\exp -\Lambda\left[5\right]$$

$$p(\alpha(\mathbf{x})=0|\cdot) > p(\alpha(\mathbf{x})=1|\cdot)$$



Some jargon



$$\mathbf{q}_1 = [0\ \ -1]$$
$$\mathbf{q}_2 = [1\ 0]$$
$$\mathbf{q}_3 = [0\ 1]$$
$$\mathbf{q}_4 = [-1\ 0]$$

$$p(\alpha(\mathbf{x})|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z} \exp -\Lambda \left[\sum_{k=1}^{4} \lambda_k |\alpha(\mathbf{x}) \neq \alpha(\mathbf{x}+\mathbf{q}_k)| \right]$$

$$p(\alpha(\mathbf{x})|\mathcal{N}_\alpha(\mathbf{x})) = \frac{1}{Z} \exp -\Lambda \left[\sum_{k=1}^{4} V(\alpha(\mathbf{x}), \alpha(\mathbf{x}+\mathbf{q}_k)) \right]$$

Potential Energy Function

Putting it together

$$p(\alpha|I,\mathcal{N}_\alpha) \propto p(I|\alpha,\mathcal{N}_\alpha)p(\alpha|\mathcal{N})$$

$$\propto \frac{1}{Z}\exp-\left[(\mathbf{I}-\bar{\mathbf{I}}_\alpha)^T\mathbf{R}_\alpha^{-1}(\mathbf{I}-\bar{\mathbf{I}}_\alpha)\right]\exp-\left[\sum_{k=1}^{4}V(\alpha(\mathbf{x}),\alpha(\mathbf{x}+\mathbf{q}_k))\right]$$

But we only really know this for the background! i.e. when \alpha = 0!

From last time

$$p(I|\alpha=0)\propto\exp-\left[\frac{(B_y-\bar{B}_y)^2}{2\sigma_y^2}+\frac{(B_u-\bar{B}_u)^2}{2\sigma_u^2}+\frac{(B_v-\bar{B}_v)^2}{2\sigma_v^2}\right]$$

Putting it together

$$\propto \frac{1}{Z} \exp - \left[(\mathbf{I} - \bar{\mathbf{I}}_\alpha)^T \mathbf{R}_\alpha^{-1} (\mathbf{I} - \bar{\mathbf{I}}_\alpha) \right] \exp - \left[\sum_{k=1}^{4} V(\alpha(\mathbf{x}), \alpha(\mathbf{x} + \mathbf{q}_k)) \right]$$

$$-\ln\left(p(\alpha | I, \mathcal{N}_\alpha)\right) = \ln(Z) + \left[(\mathbf{I} - \bar{\mathbf{I}}_\alpha)^T \mathbf{R}_\alpha^{-1} (\mathbf{I} - \bar{\mathbf{I}}_\alpha) \right] + \left[\sum_{k=1}^{4} V(\alpha(\mathbf{x}), \alpha(\mathbf{x} + \mathbf{q}_k)) \right]$$

Maximum a-Posteriori (MAP) Estimate

$$-\ln\left(p(\alpha|I,\mathcal{N}_\alpha)\right) = \ln(Z) + \left[(\mathbf{I}-\bar{\mathbf{I}}_\alpha)^T \mathbf{R}_\alpha^{-1}(\mathbf{I}-\bar{\mathbf{I}}_\alpha)\right] + \left[\sum_{k=1}^{4} V(\alpha(\mathbf{x}),\alpha(\mathbf{x}+\mathbf{q}_k))\right]$$

$$E(0) = \ln(Z) + \left[(\mathbf{I}-\bar{\mathbf{I}}_\alpha)^T \mathbf{R}_\alpha^{-1}(\mathbf{I}-\bar{\mathbf{I}}_\alpha)\right] + \left[\sum_{k=1}^{4} V(0,\alpha(\mathbf{x}+\mathbf{q}_k))\right]$$

$$E(1) = \ln(Z) + E_t + \left[\sum_{k=1}^{4} V(1,\alpha(\mathbf{x}+\mathbf{q}_k))\right]$$

$$\alpha = \begin{cases} 0 & \text{if } E(0) < E(1) \\ 1 & \text{Otherwise} \end{cases}$$

We assume that the foreground colour is sampled from a uniform distribution

But wait …

- This is only the CONDITIONAL probability of one alpha value at one site GIVEN the current state of all the other alphas
- How can we maximise the JOINT probability for all the variables (\alpha) at all sites?

ICM Solution

J. R. Statist. Soc. B (1986)
48, *No.* 3, *pp.* 259–302

On the Statistical Analysis of Dirty Pictures

By JULIAN BESAG

University of Durham, U.K.

[*Read before the* Royal Statistical Society, *at a meeting organized by the* Research Section *on Wednesday, May 7th, 1986,* Professor A. F. M. Smith *in the Chair*]

- Solution at each site affects all the other sites
- Can ”Pick the best”, pick the best value for alpha at one site then with that new value, visit the next site. Then iterate.
- This is “Iterated Conditional Modes”, we are choosing the variable which maximises the local conditional density, and then doing the same at every site.
- Converges to a local minimum
- ICM was first proposed by Besag in 1986
- Since then more optimal techniques have been presented.

MAP Algorithm

1. Measure Gaussian parameters for background somewhere in the example image. Get $\bar{B}_y, \sigma_y^2$, etc.
2. Set some penalty E_t for setting $\alpha = 1$ at a site
3. For `iteration = 1 :  N`
   - For every pixel site $\mathbf{x}$

     (a) Measure $E_l, E_s(0), E_s(1)$

$$E_l = \frac{(B_y - \bar{B}_y)^2}{2\sigma_y^2} + \frac{(B_u - \bar{B}_u)^2}{2\sigma_u^2} + \frac{(B_v - \bar{B}_v)^2}{2\sigma_v^2}$$

$$E_s(0) = \sum_{k=1}^{4} V(0, \alpha(\mathbf{x} + \mathbf{q}_k))$$
$$E_s(1) = \sum_{k=1}^{4} V(1, \alpha(\mathbf{x} + \mathbf{q}_k))$$

     (b) Set $E(0) = E_l + E_s(0)$ and $E(1) = E_t + E_s(1)$

     (c) If $E(0) < E(1)$ set $\alpha(\mathbf{x}) = 0$ Otherwise $\alpha(\mathbf{x}) = 1$

Maximum Likelihood Algorithm

1. Measure Gaussian parameters for background somewhere in the example image. Get $\bar{B}_y, \sigma_y^2$, etc.
2. Set some threshold E_t for deciding whether a site is background or not
3. At every pixel site measure

$$E(\mathbf{x}) = \frac{(B_y - \bar{B}_y)^2}{2\sigma_y^2} + \frac{(B_u - \bar{B}_u)^2}{2\sigma_u^2} + \frac{(B_v - \bar{B}_v)^2}{2\sigma_v^2} \quad (2)$$

4. If $E(\cdot) < E_t$ set $\alpha = 0$ Otherwise $\alpha = 1$

Visitation Schemes

- Sequential / raster scan
  - Forward/Backward
- Checkerboard : in-place updating
- Synchronous Vs in-place updating
- Sychronous : you update all the pixels into another array and ping pong between two storage arrays
- In place : update using checkerboard pattern

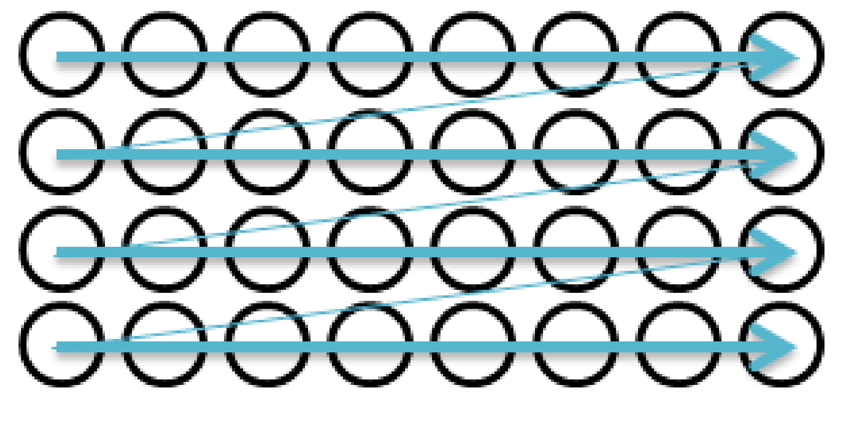
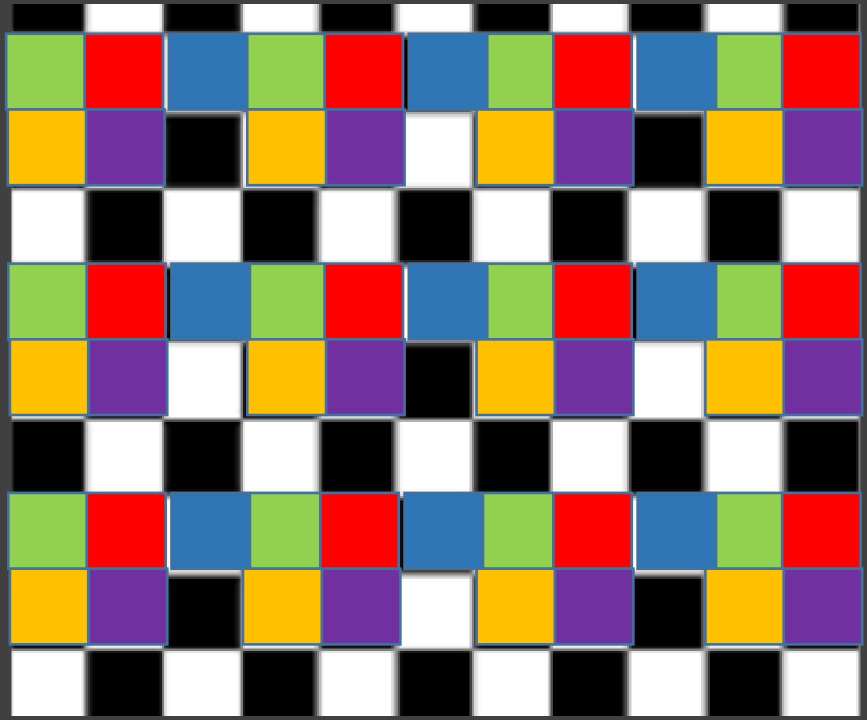


Figure 1



Manually set the location of this area
And measure the mean and variance of each colour plane

$$E(\mathbf{x}) = \frac{(B_y - \bar{B_y})^2}{2\sigma_y^2} + \frac{(B_u - \bar{B_u})^2}{2\sigma_u^2} + \frac{(B_v - \bar{B_v})^2}{2\sigma_v^2}$$


E_t = 40
\lambda = 10.0
Iteration = 1


E_t = 40
\lambda = 10.0
Iteration = 2


E_t = 40
\lambda = 10.0
Iteration = 3


E_t = 40
\lambda = 10.0
Iteration = 4


E_t = 40
\lambda = 10.0
Iteration = 10


E_t = 40
\lambda = 0.0
Iteration = all
This is ML


MAXIMUM LIKELIHOOD

MAP

The first paper to present Matting in a Bayesian formulation

A Bayesian Approach to Digital Matting

Yung-Yu Chuang[1] Brian Curless[1] David H. Salesin[1,2] Richard Szeliski[2]

[1]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195
[2]Microsoft Research, Redmond, WA 98052
E-mail: {cyy, curless, salesin}@cs.washington.edu szeliski@microsoft.com
http://grail.cs.washington.edu/projects/digital-matting/

Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR 2001), Vol. II, 264-271, December 2001

Key terms in Modern Matting

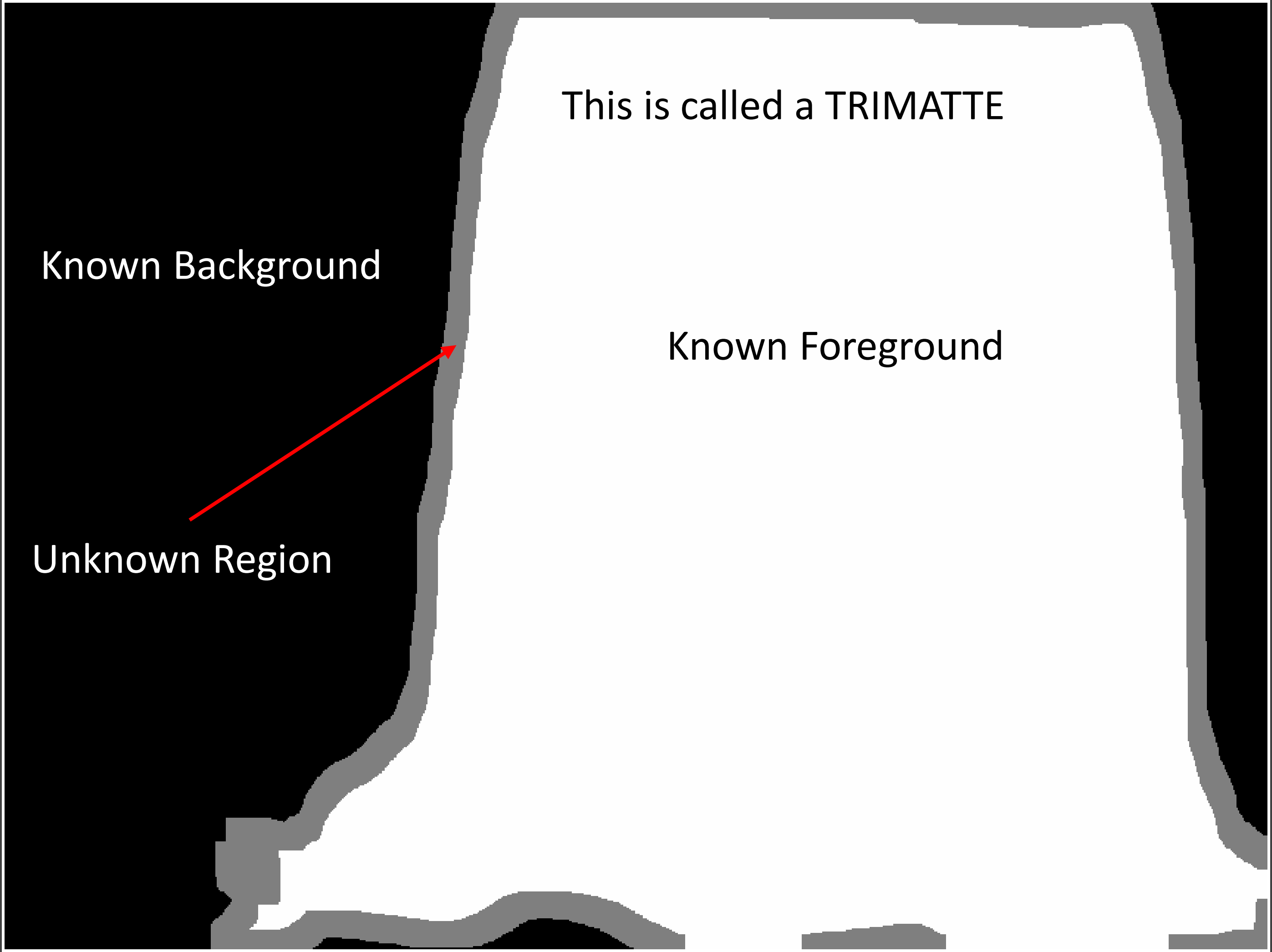
Composite Image

$$C = \alpha F + (1 - \alpha) B\ ,$$

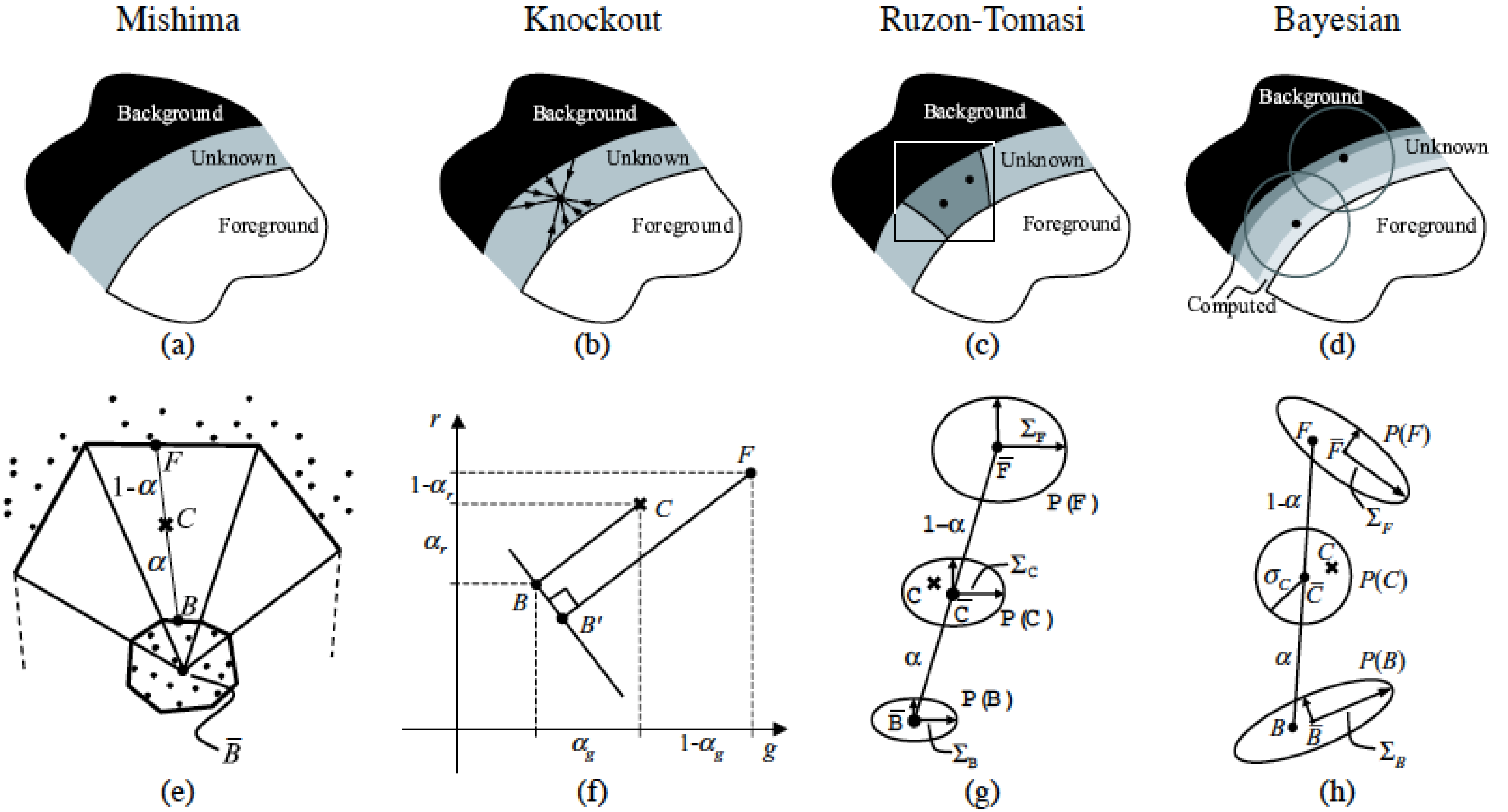
Premultiplied Foreground

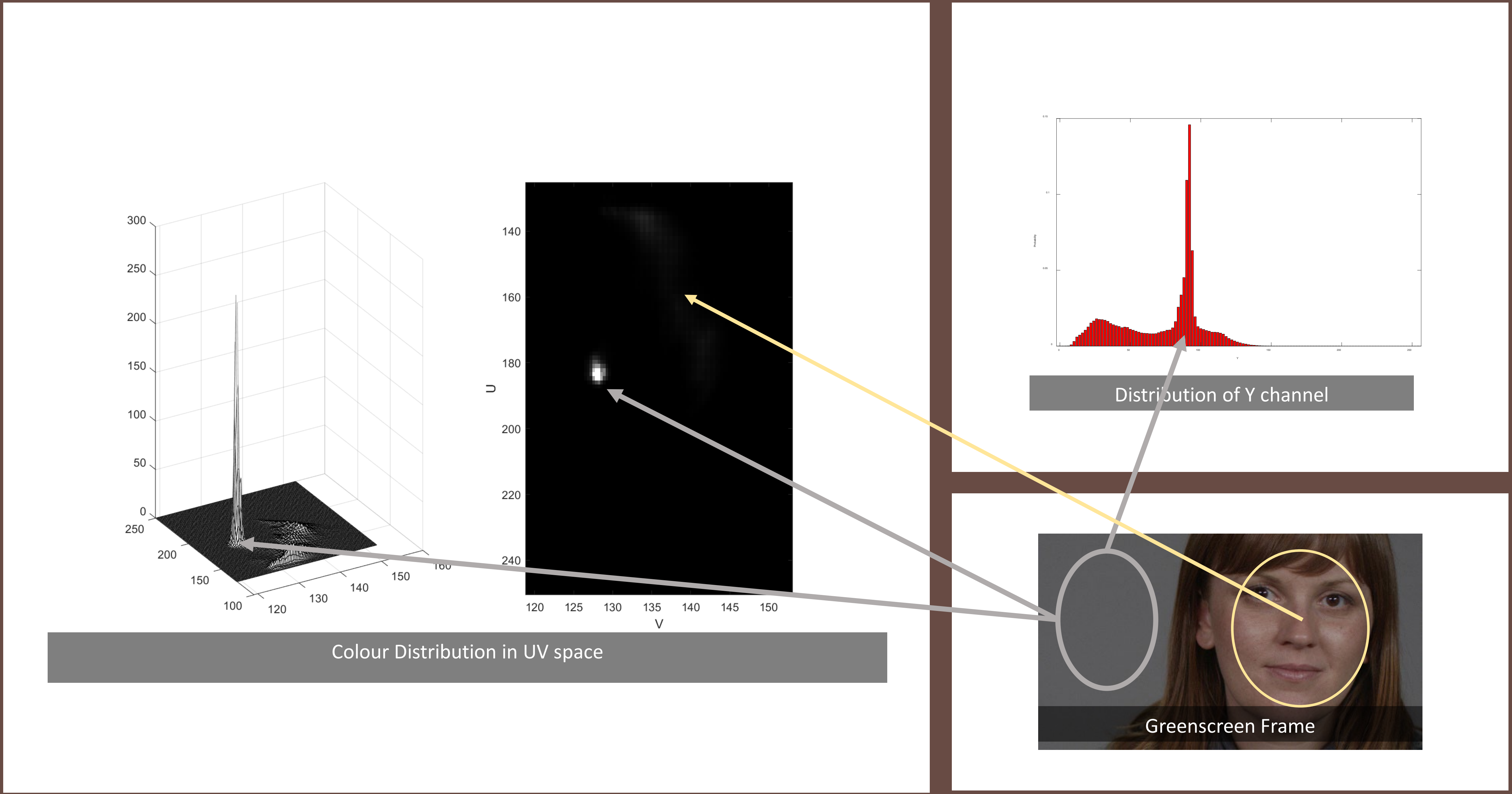
Premultiplied Background

In fact the Premulitplied foreground and background are both actually the F and B composited against black (i.e. 0) respectively.



$$C = \alpha F + (1 - \alpha) B ,$$




Colour Distribution in UV space
Distribution of Y channel
Greenscreen Frame
U
V
300
250
200
150
100
50
0
140
160
180
200
220
240
120
125
130
135
140
145
150

A big note

The compositing equation is actually this

$$\mathbf{C}(h,k) = \alpha(h,k)\mathbf{F}(h,k) + (1-\alpha(h,k))\mathbf{B}(h,k) + \epsilon(h,k)$$

But we drop the (h,k) pixel indices for simplicity and in many published papers authors simply ignore Matrix convention and don't bother to make the vectors or Matrices bold. In the Computer Vision literature they also don't bother to make a distinction between a matrix and a vector as far as the font is concerned. THIS IS VERY BAD! IT MAKES PAPERS WAY HARDER TO READ THAN THEY SHOULD BE. The DSP guys didn't do this .. but they have been swamped by the CV people. Vectors are supposed to be written as common case, bold; matrices as capital letters in bold.

Let’s go deeper

The likelihood again

In the compositing equation, $\epsilon(\cdot,\cdot)$ is the error between the model and the actual observed composite image. This is typically due to noise, so a decent enough model is $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathcal{I})$, where I am using $\mathcal{I} = [1\ 0\ 0; 0\ 1\ 0; 0\ 0\ 1]$. That means that we can rewrite the equation like so

$$\epsilon = \mathbf{C}(h,k) - \alpha\mathbf{F} - (1-\alpha)\mathbf{B}$$

Which allows us to write the likelihood as follows.

$$p(\mathbf{C}|\alpha, \mathbf{F}, \mathbf{B}) \propto \exp\left(-\frac{(\mathbf{C} - \alpha\mathbf{F} - (1-\alpha)\mathbf{B})^T \mathcal{I} (\mathbf{C} - \alpha\mathbf{F} - (1-\alpha)\mathbf{B})}{\sigma_\epsilon^2}\right)$$

$$C = \alpha F + (1-\alpha) B ,$$

$$\begin{aligned} &\arg\max_{F,B,\alpha} P(F, B, \alpha \,|\, C) \\ &= \arg\max_{F,B,\alpha} P(C \,|\, F, B, \alpha)\, P(F)\, P(B)\, P(\alpha) \,/\, P(C) \\ &= \arg\max_{F,B,\alpha} L(C \,|\, F, B, \alpha) + L(F) + L(B) + L(\alpha) , \end{aligned}$$

$C = \alpha F + (1-\alpha) B + \varepsilon$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$

$$L(C \,|\, F, B, \alpha) = -\|C - \alpha F - (1-\alpha) B\|^2 / \sigma_C^2 .$$

$$L(F) = -(F - \overline{F})^T \, \Sigma_F^{-1} \, (F - \overline{F}) \,/\, 2 .$$

Ping-pong between estimating alpha and F/B

Because of the multiplications of α with F and B in the log likelihood $L(C \mid F, B, \alpha)$, the function we are maximizing in (4) is not a quadratic equation in its unknowns. To solve the equation efficiently, we break the problem into two quadratic sub-problems. In the first sub-problem, we assume that α is a constant. Under this assumption, taking the partial derivatives of (4) with respect to F and B and setting them equal to 0 gives:

$$\begin{bmatrix} \Sigma_F^{-1} + I\alpha^2/\sigma_C^2 & I\alpha(1-\alpha)/\sigma_C^2 \\ I\alpha(1-\alpha)/\sigma_C^2 & \Sigma_B^{-1} + I(1-\alpha)^2/\sigma_C^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} = \begin{bmatrix} \Sigma_F^{-1}\overline{F} + C\alpha/\sigma_C^2 \\ \Sigma_B^{-1}\overline{B} + C(1-\alpha)/\sigma_C^2 \end{bmatrix}, \tag{9}$$

where I is a 3×3 identity matrix. Therefore, for a constant α, we can find the best parameters F and B by solving the 6×6 linear equation (9).

Optimal solutions

- Belief Propagation (Judea Pearl)
- Graph Cuts (Boykov, Zabih)

Understanding Belief Propagation and its Generalizations

Jonathan S. Yedidia
MERL
201 Broadway
Cambridge, MA 02139
yedidia@merl.com

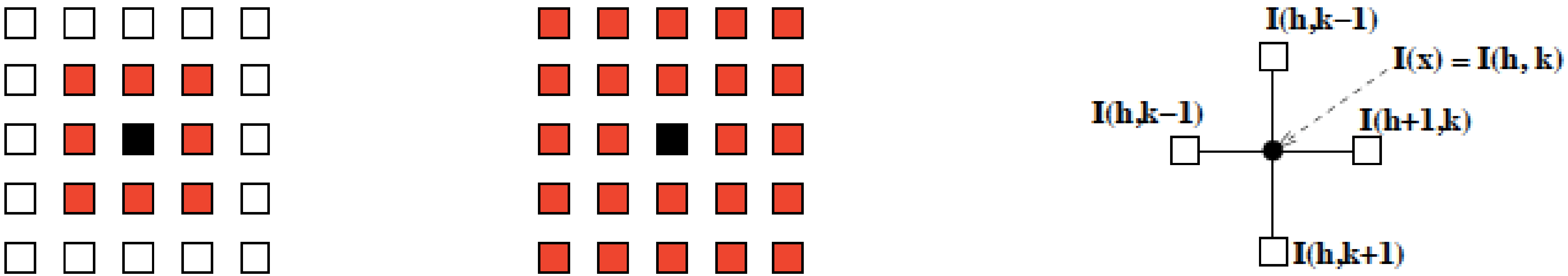
William T. Freeman
MERL
201 Broadway
Cambridge, MA 02139
freeman@merl.com

Yair Weiss
School of Computer Science and Engineering
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
yweiss@cs.huji.ac.il

Fast Approximate Energy Minimization via Graph Cuts

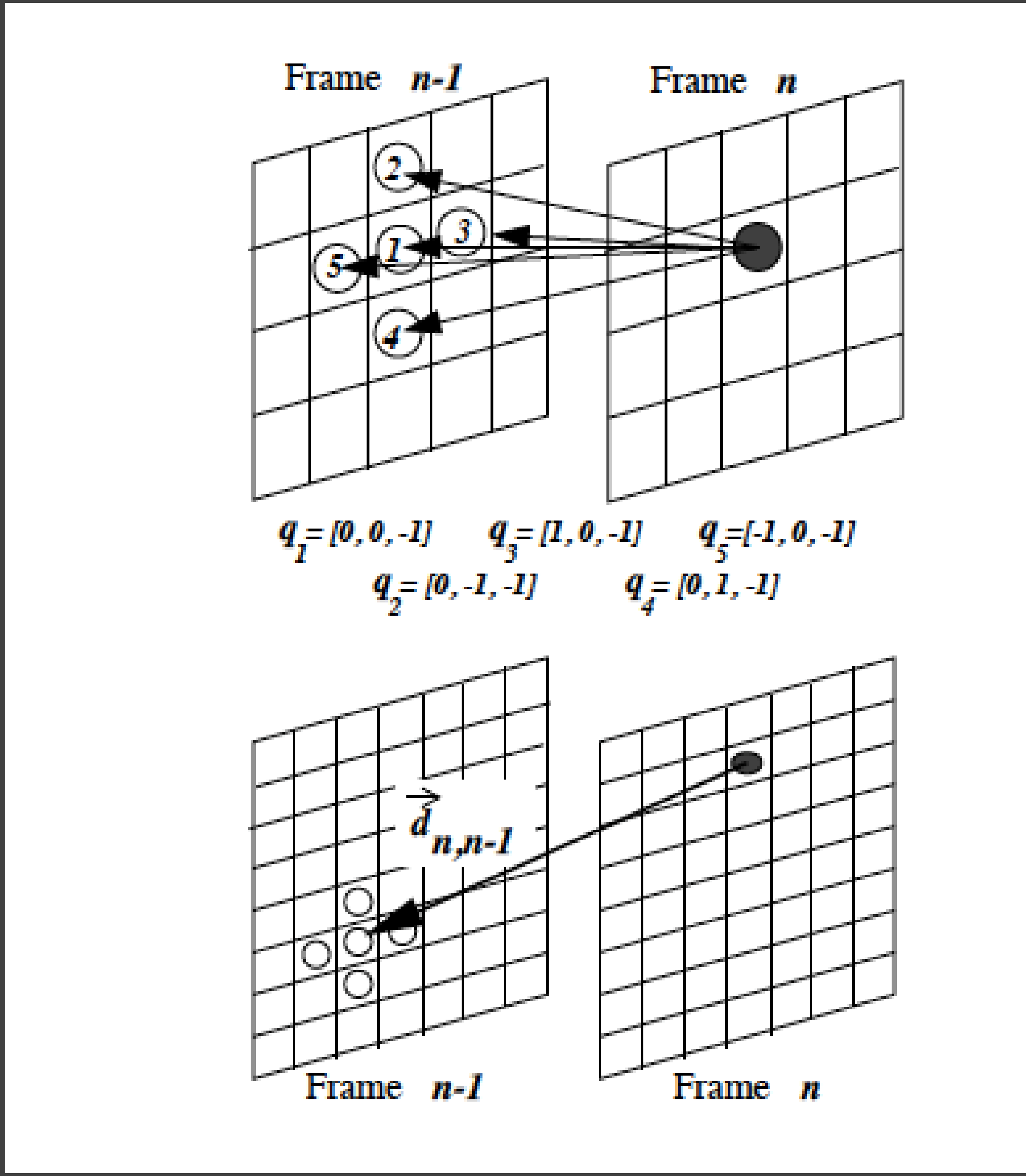
Yuri Boykov Olga Veksler Ramin Zabih
Computer Science Department
Cornell University
Ithaca, NY 14853

Other kinds of Markov Fields : Autoregressive process



$$p(I(\mathbf{x})|\mathbf{I}) \propto \begin{cases} \exp - \left(\frac{[I(\mathbf{x}) - \sum_{k=1}^{P} a_k I(\mathbf{x}+\mathbf{q}_k)]^2}{2\sigma_e^2} \right) & \text{2DAR} \\ \exp - \left(\Lambda \sum_{k=1}^{4} \lambda_k (I(\mathbf{x}) \neq I(\mathbf{x}+\mathbf{q}_k))^2 \right) & \text{GMRF} \end{cases}$$

MRFs for Video



Need to know about motion !
Need to know how to measure success!

FIN