Professor Anil Kokaram

Anil.Kokaram@tcd.ie

5C1 Motion Picture Engineering

Digital Video Engineering

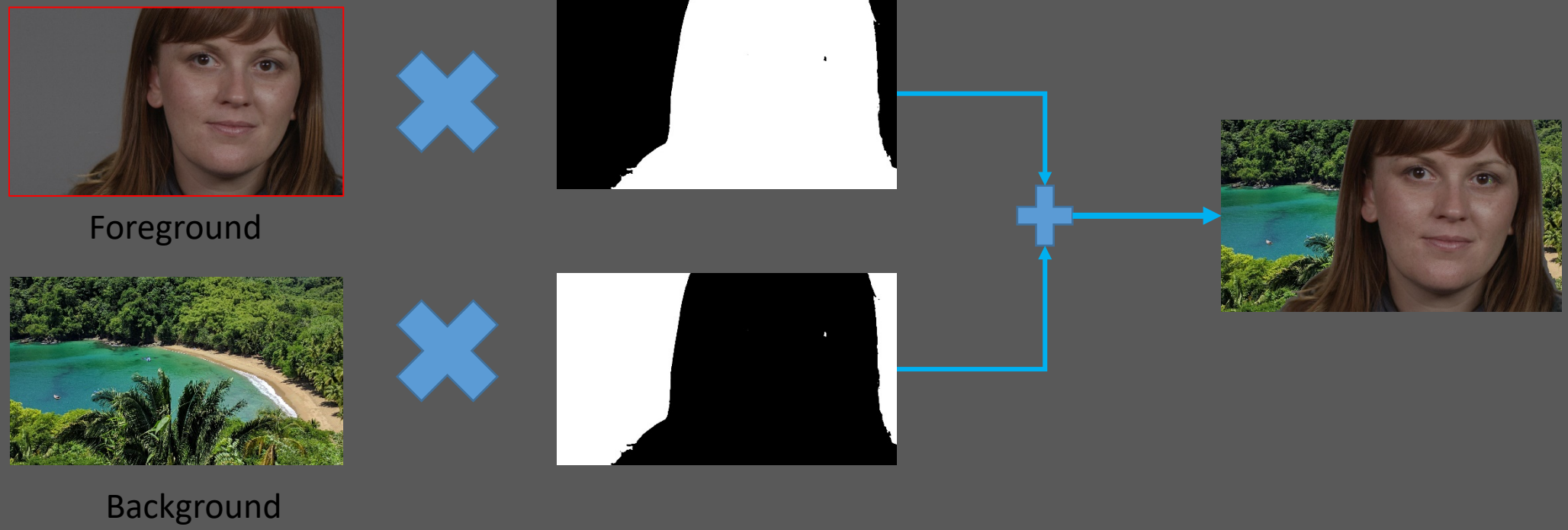
Engineering for Moving Pictures

Colour Keying

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

$\mathbf{x}$ 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 $\alpha = 1$

Binary Matte



Foreground

Background

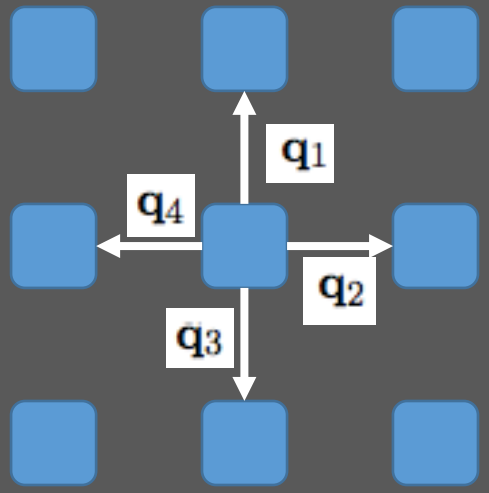
Colour segmentation/keying is about discovering $\alpha(\mathbf{x})$ based primarily on colour alone.

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 = 0.0
Iteration = all
This is ML

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


E_t = 40
\lambda = 10.0
Iteration = 10


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

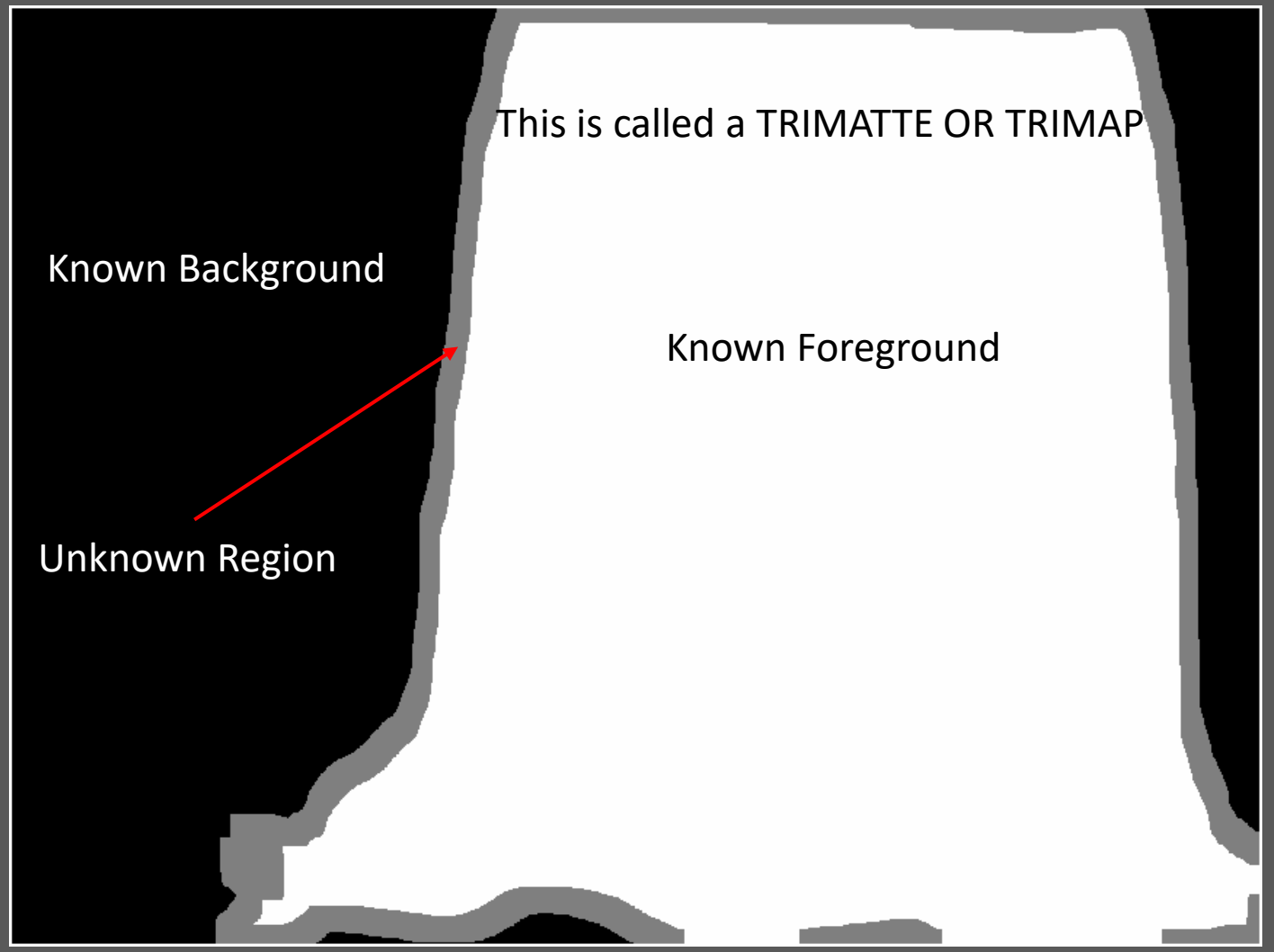
Composite Image

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

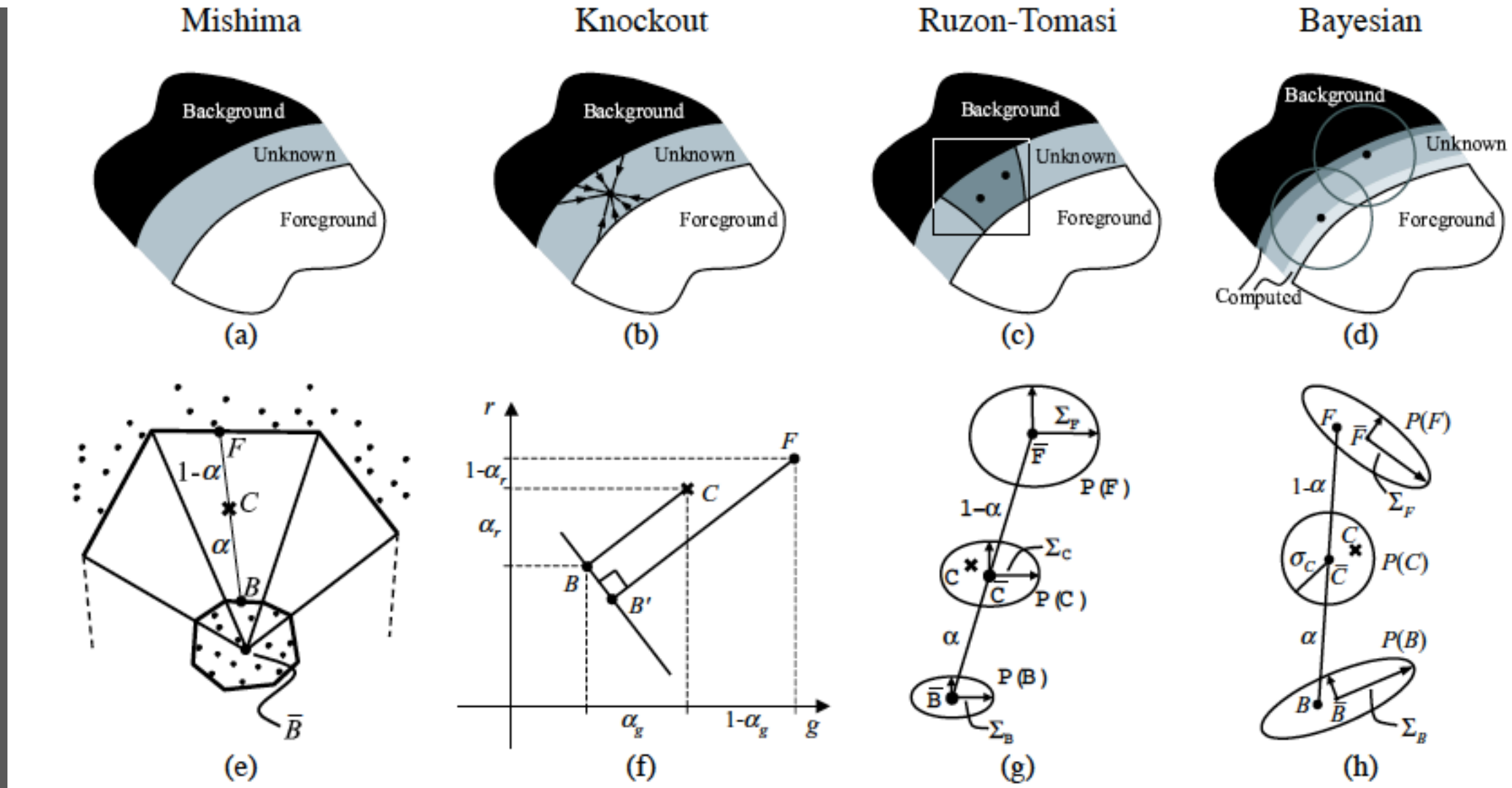
Premultiplied Foreground

Premultiplied Background

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



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



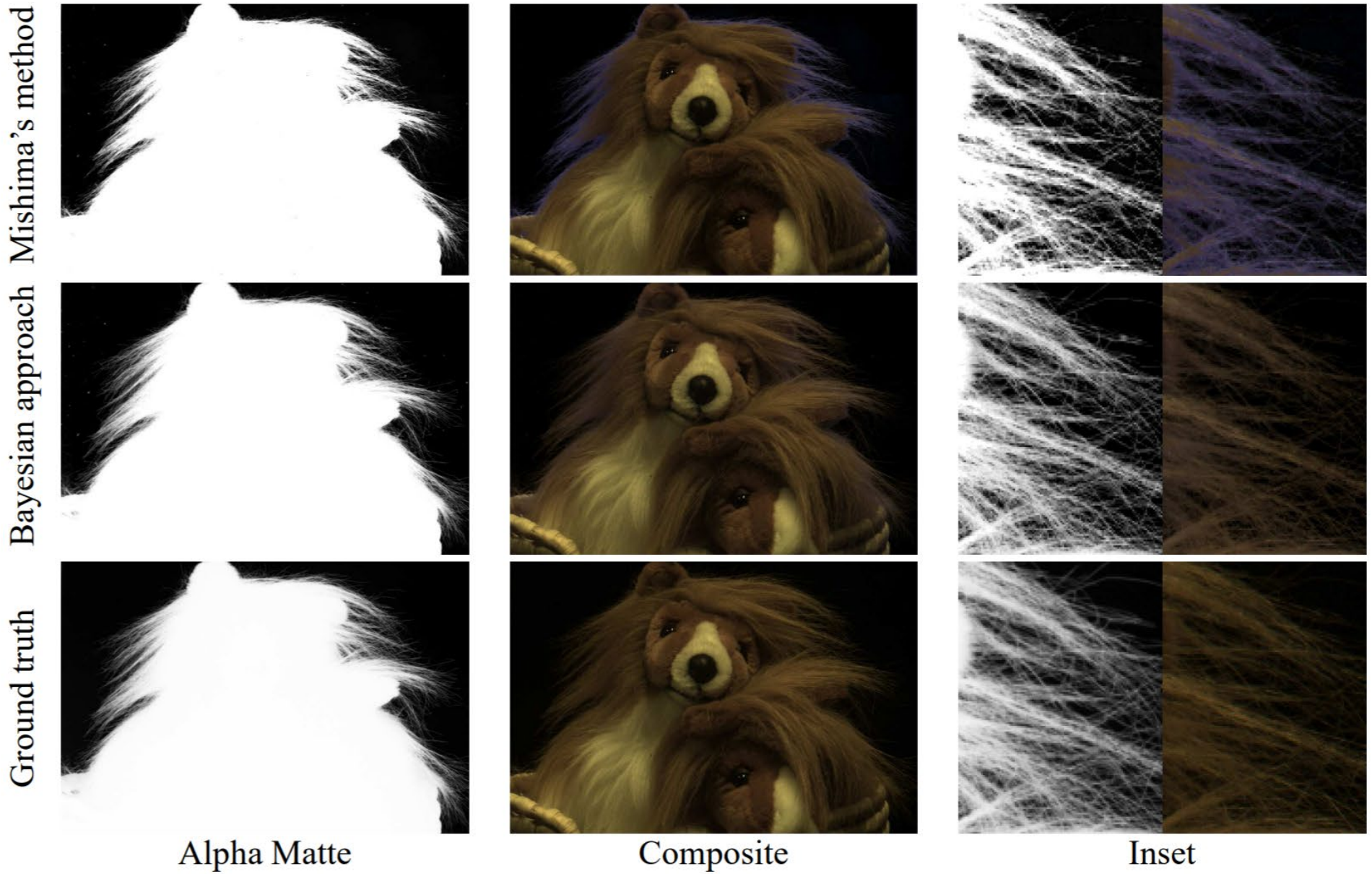


Figure 3 Blue-screen matting of lion (taken from leftmost column of Figure 2). Mishima's results in the top row suffer from "blue spill." The middle and bottom rows show the Bayesian result and ground truth, respectively.

Optimal solutions

- Belief Propagation (Judea Pearl)
- Graph Cuts (Boykov, Zabih)
- These solve the “labelling” problem ith MRF priors optimally

Understanding Belief Propagation and its Generalizations

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

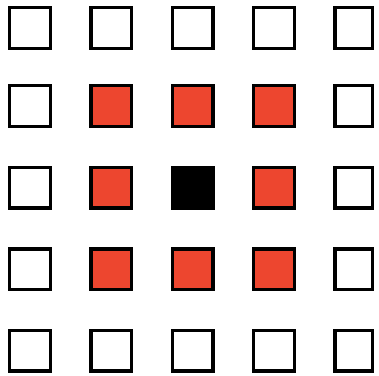
William T. Freeman
MERL
201 Broadway
Cambridge, MA 02139
freeman@merl.com

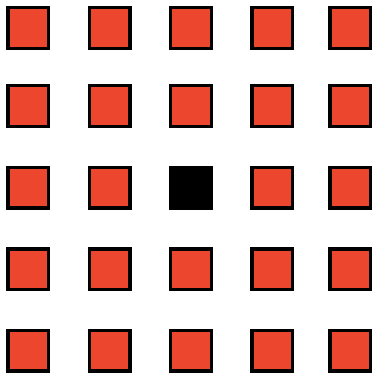
Yair Weiss
School of Computer Science and Engineering
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
yweiss@cs.huji.ac.il

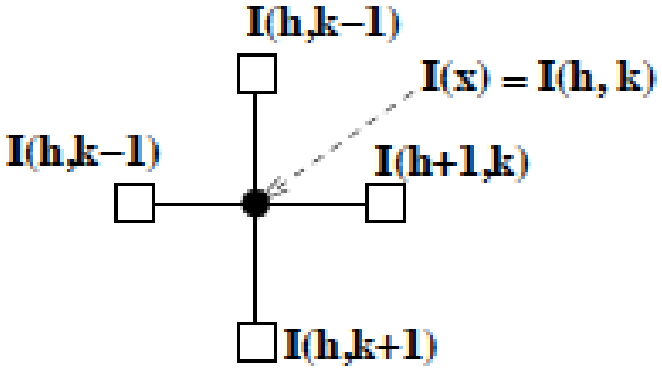
Fast Approximate Energy Minimization via Graph Cuts

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

Other kinds of Markov Fields : Autoregressive process

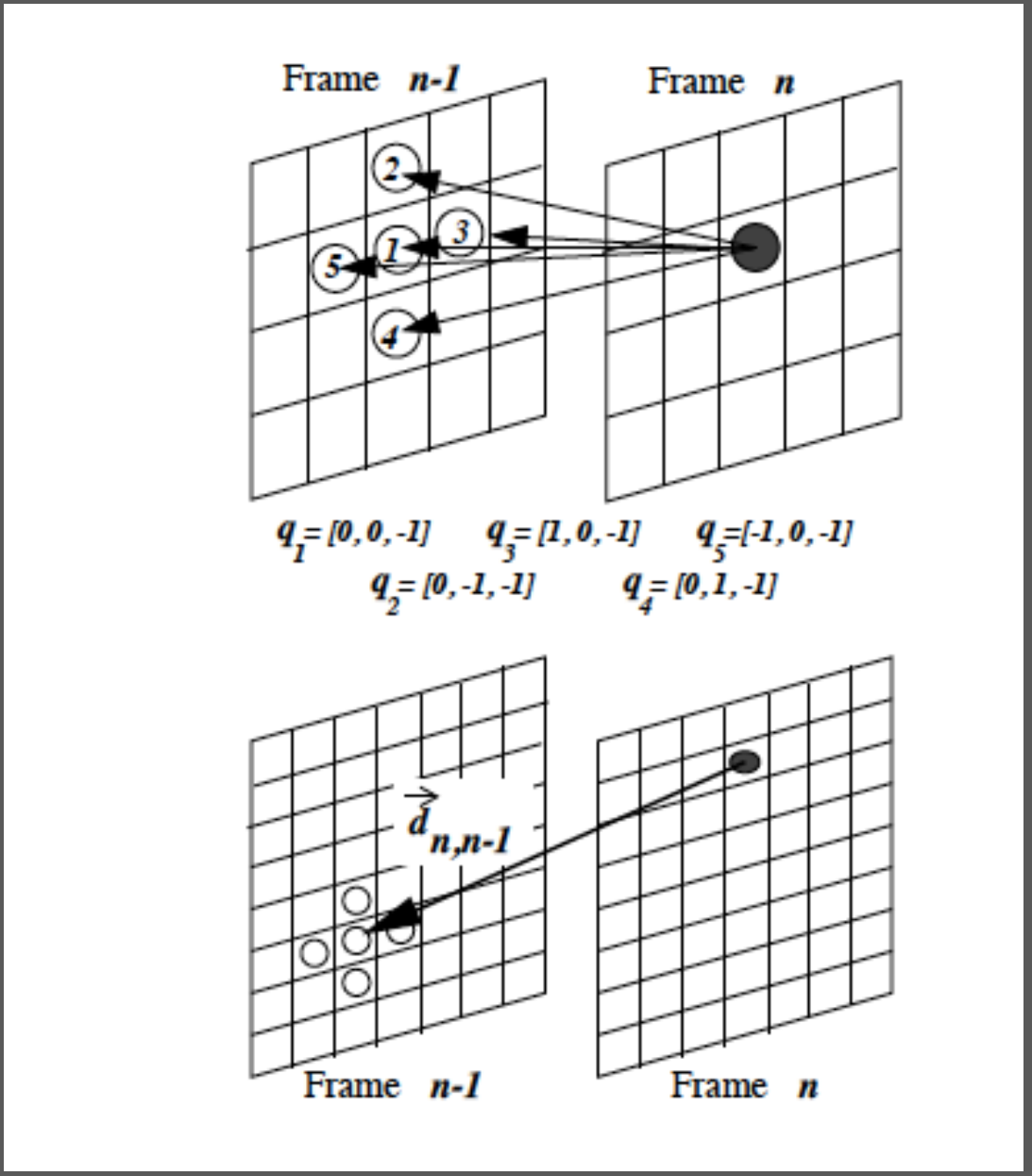






$$p(I(\mathbf{x})|\mathbf{I}) \propto \begin{cases} \exp - \left(\frac{[I(\mathbf{x}) - \sum_{k=1}^{P} a_k I(\mathbf{x}+\mathbf{q}_k)]^2}{2\sigma_e^2} \right) & \text{2DAR} \\ \exp - \left(\Lambda \sum_{k=1}^{4} \lambda_k (I(\mathbf{x}) \neq I(\mathbf{x}+\mathbf{q}_k))^2 \right) & \text{GMRF} \end{cases}$$

MRFs for Video



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

In 2009 IEEE Conference on Comp Vision and Pattern Recognition http://www.alphamatting.com/index.html

A Perceptually Motivated Online Benchmark for Image Matting

Christoph Rhemann[1]*, Carsten Rother[2], Jue Wang[3], Margrit Gelautz[1], Pushmeet Kohli[2], Pamela Rott[1]

[1]Vienna University of Technology, Vienna, Austria
[2]Microsoft Research Cambridge, Cambridge, UK
[3]Adobe Systems, Seattle, USA

- Idealised synthetically created composite images created with GT mattes
- New metrics introduced : Gradient difference; Connectivity measure
- Good mattes are “as smooth as” the ground truth matte and “as conncted as” the grond truth matte

Anat Levin in 2008 produced a work of genius

- A. Levin, D. Lischinski and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, Feb. 2008, doi: 10.1109/TPAMI.2007.1177.

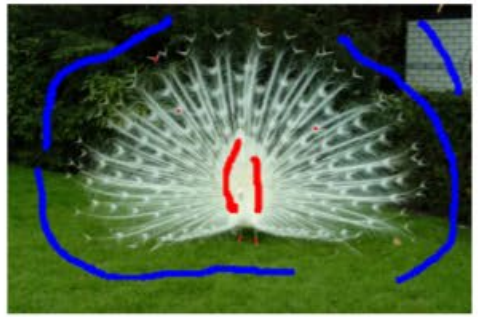
A Closed Form Solution to Natural Image Matting

Anat Levin Dani Lischinski Yair Weiss
School of Computer Science and Engineering
The Hebrew University of Jerusalem
{*alevin,danix,yweiss*}*@cs.huji.ac.il*

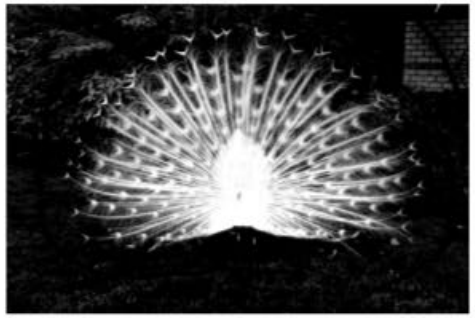
Transform our 3 variable problem into 1

- Assume F and B are locally constant in a small window
- Instead of estimating α, F, B from the compositing equation $C = \alpha F + (1-\alpha)B$ JUST ESTIMATE α
- $C = \alpha(F - B) + B$
- $\Rightarrow \alpha = \frac{C}{F-B} - \frac{B}{F-B}$
- $\Rightarrow \alpha = aC + b$
- Now our non-linear problem for estimating α, F, B becomes a linear least squares problem because α is dependent on C and B through two linear coefficients a, b
- She then went on to remove a,b from the solution entirely
- Very fast algorithm … looks a lot like filtering

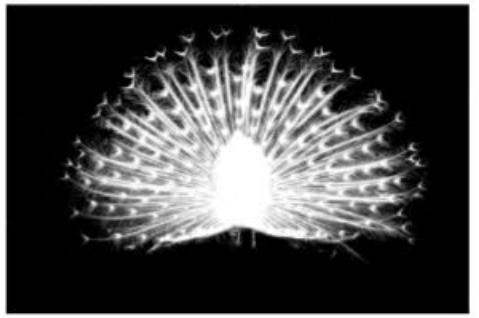
Amazing results



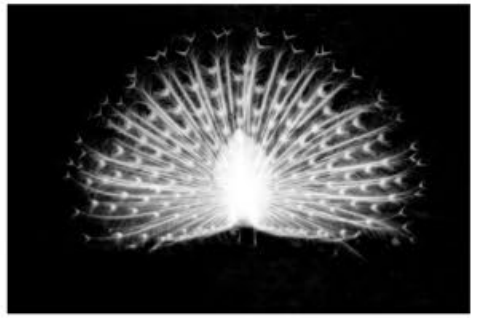
(a) Peacock scribbles



(b) Poisson from scribbles

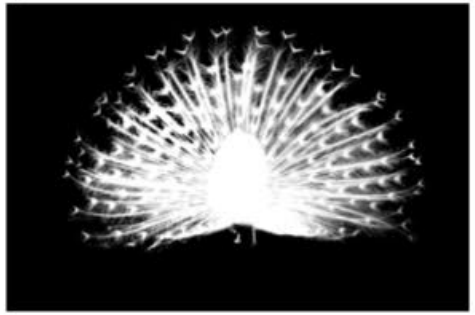


(c) Wang-Cohen

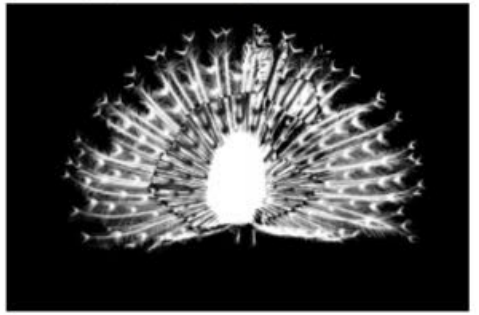


(d) Our result

(e) Peacock trimap



(f) Poisson from trimap

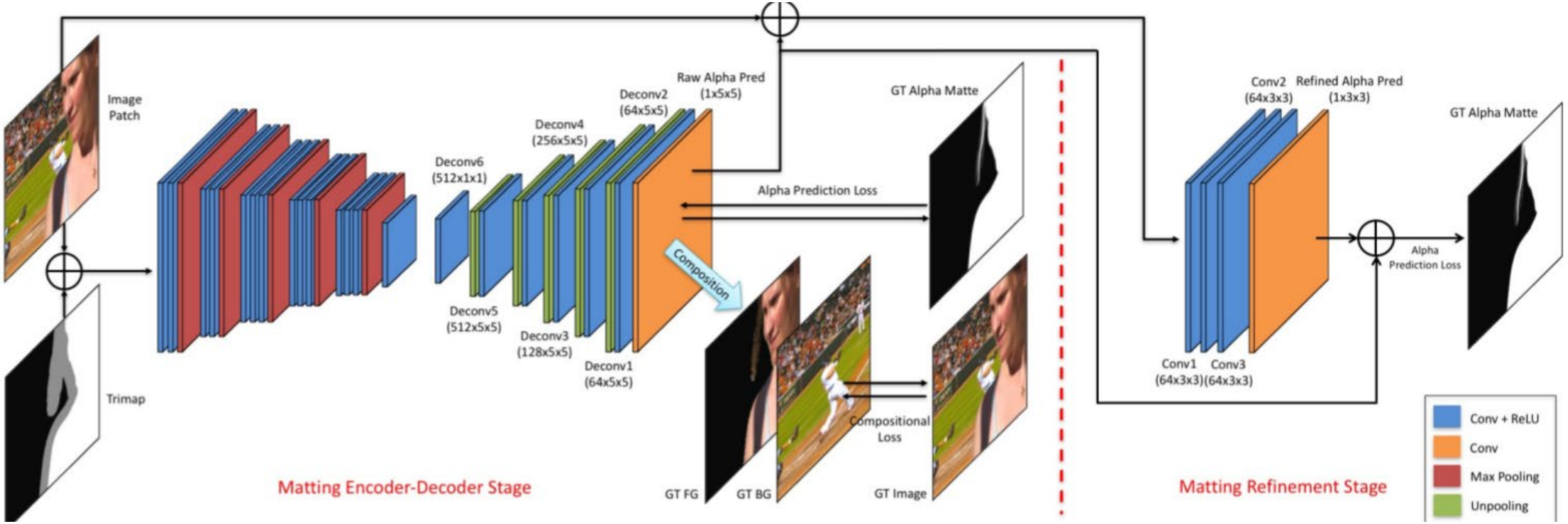


(g) Bayesian

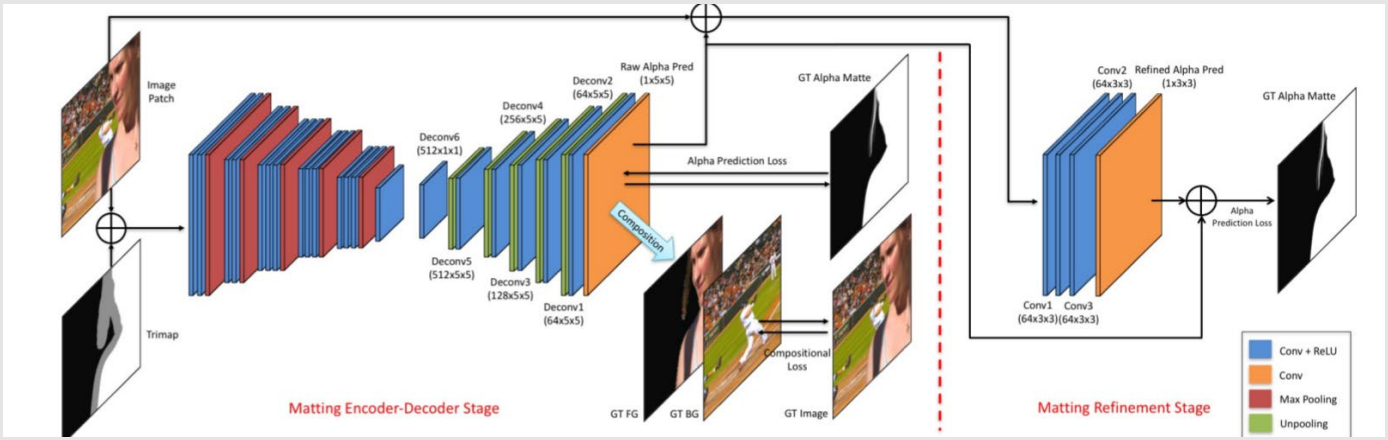
(h) Random walk

Then in 2017 .. Along came Deep Nets

- N. Xu, B. Price, S. Cohen and T. Huang, "Deep Image Matting," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 311-320



Key Ideas



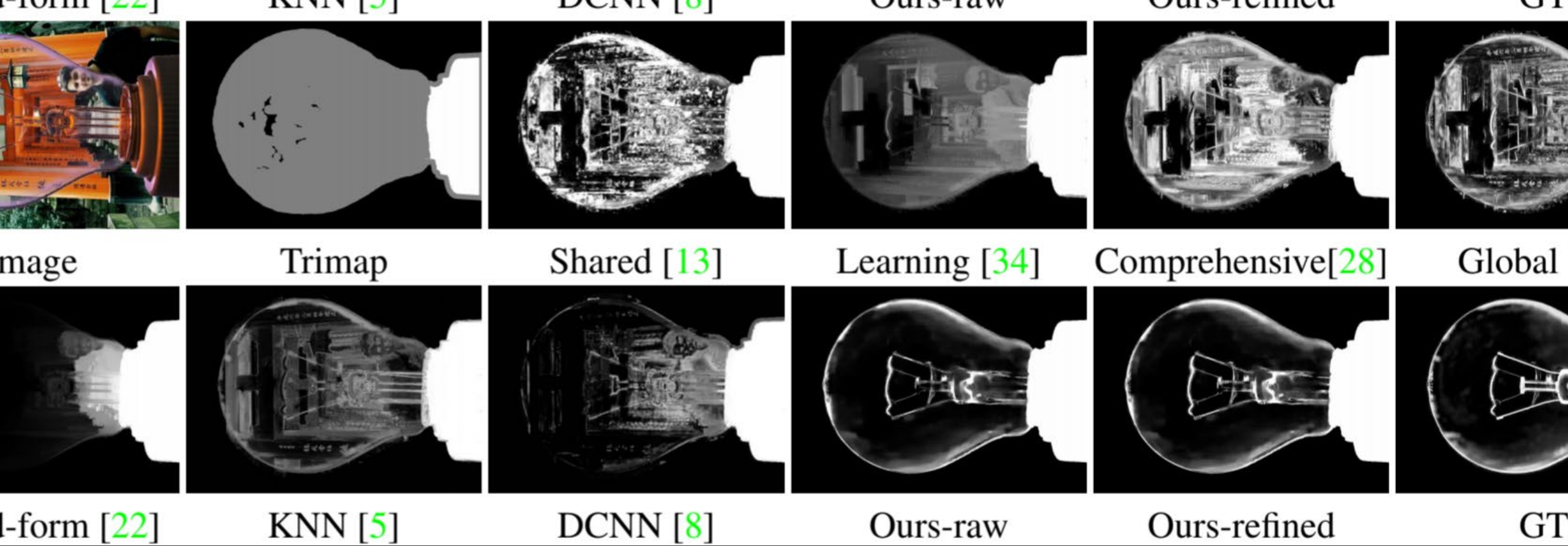
- Input is picture and associated trimatte or trimap
- Two stages : Encoder Decoder followed by fully connected "Refinement" stage. Train E-D first then train "refinement" stage
- Loss function combines α-matte loss with composite image loss
- $L_\alpha = \sqrt{\{\alpha - \alpha_g\}^2 + \epsilon}$; $L_C = \sqrt{(C - C_g)^2 + \epsilon}$
- $L = w_l L_\alpha + (1 - w_l) L_C$
- So first network is trained to produce BOTH the alpha-matte AND the observed composite image
- Second network has NO DOWNSAMPLING so acts to refine

Results etc

Table 1. The quantitative results on the Composition-1k testing dataset. The variants of our approaches are emphasized in italic. The best results are emphasized in bold.

Methods	SAD	MSE	Gradient	Connectivity
Shared Matting [13]	128.9	0.091	126.5	135.3
Learning Based Matting [34]	113.9	0.048	91.6	122.2
Comprehensive Sampling [28]	143.8	0.071	102.2	142.7
Global Matting [16]	133.6	0.068	97.6	133.3
Closed-Form Matting [22]	168.1	0.091	126.9	167.9
KNN Matting [5]	175.4	0.103	124.1	176.4
DCNN Matting [8]	161.4	0.087	115.1	161.9
Encoder-Decoder network (single alpha prediction loss)	59.6	0.019	40.5	59.3
Encoder-Decoder network	54.6	0.017	36.7	55.3
Encoder-Decoder network + Guided filter[17]	52.2	0.016	**30.0**	52.6
Encoder-Decoder network + Refinement network	**50.4**	**0.014**	31.0	**50.8**

- 320 x 320 image crops : Can’t process 720 for instance.
- 49,300 images
- Encoder initialised with first 14 layers of VGG-16



Levin’s success can explain why a DNN can work

Levin and DNN

- Levin reposed the mating problem as essentially a filtering problem over pixels
- The nonlinearity was removed by introducing “latent” variables “a,b”
- The DNN is learning to do the same thing PLUS it is injecting semantics into the problem … learning to spot objects and what their “coherency” should be.

Why have I bothered to show you anything before 2017?

- In postproduction : DNNs are still not being used
- They are unpredictable in the sense that they are good with things they have seen before but not well behaved with “new” material
- In video temporal coherency is EVERYTHING. Video matting is still trying to sort this out. DNNs are only now emerging (2020) which address temporal coherency
- DNNs are massively computationally heavy. Imagine trying to do all of this with 8K plates! (In postproduction they call a frame a PLATE)
- Hybridisation of these two genres of techniques is the way forward DNNs + Levin
- Remember 95% success is still not good enough for post-production!

M. Forte, B. Price, S. Cohen, N. Xu and F. Pitié, "Interactive Training And Architecture For Deep Object Selection," *2020 IEEE International Conference on Multimedia and Expo (ICME)*, London

Francois Pitie (sigmedia.tv)

Working to push deep matting into a usable state

See ICME paper from 2020 which won a prize

FIN