An MRF model for Binarization of Natural Scene Text

Anand Mishra, Karteek Alahari and C. V. Jawahar

IIIT Hyderabad, India

Natural Scene Text: Recent Interest

Detecting Text in Image

Natural scene text detection competitions at ICDAR 2003, 2005 and 2011

Stroke Width Transform based text detection (Boris Epshtein et al., CVPR 2010)

Natural Scene Text: Recent Interest

Detecting text in a street view

More challenging Street View Text (SVT) dataset

Words are treated as "objects" (Kai Wang and Serge Belongie, ECCV 2010)

Natural Scene Text: Recent Interest

Text Recognition

Word spotting in wild (Kai Wang and Serge Belongie, ECCV 2010)

Enforcing similarity constraints for better scene text recognition (David Smith *et al.*, CVPR 2011)

Char 74k dataset (TE de Campos *et al.*, VISSAP 2009)

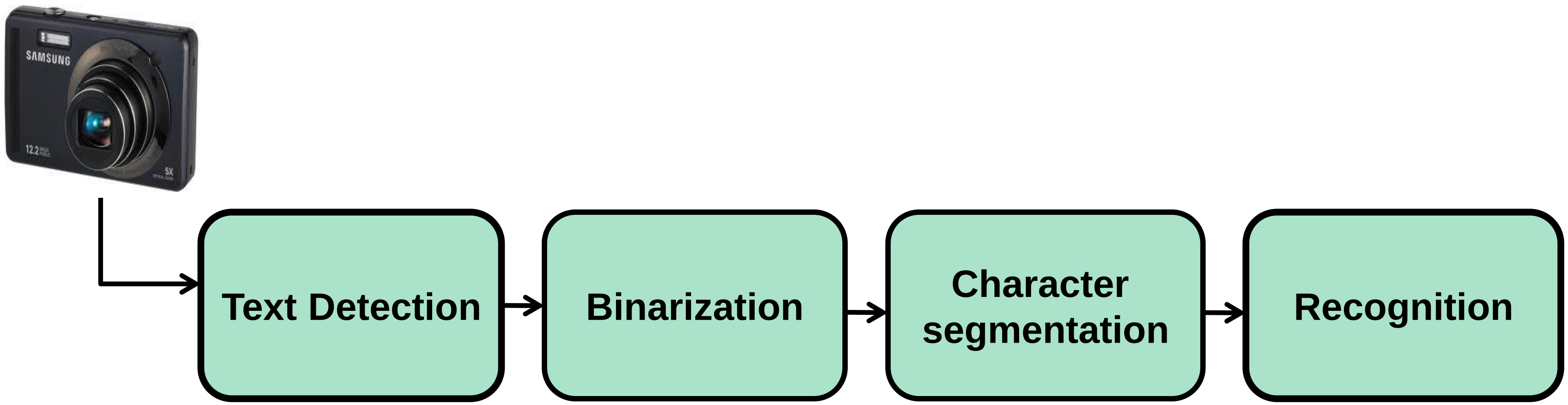
Many Applications

- Help for visually impaired

- Cross lingual access through cell phones

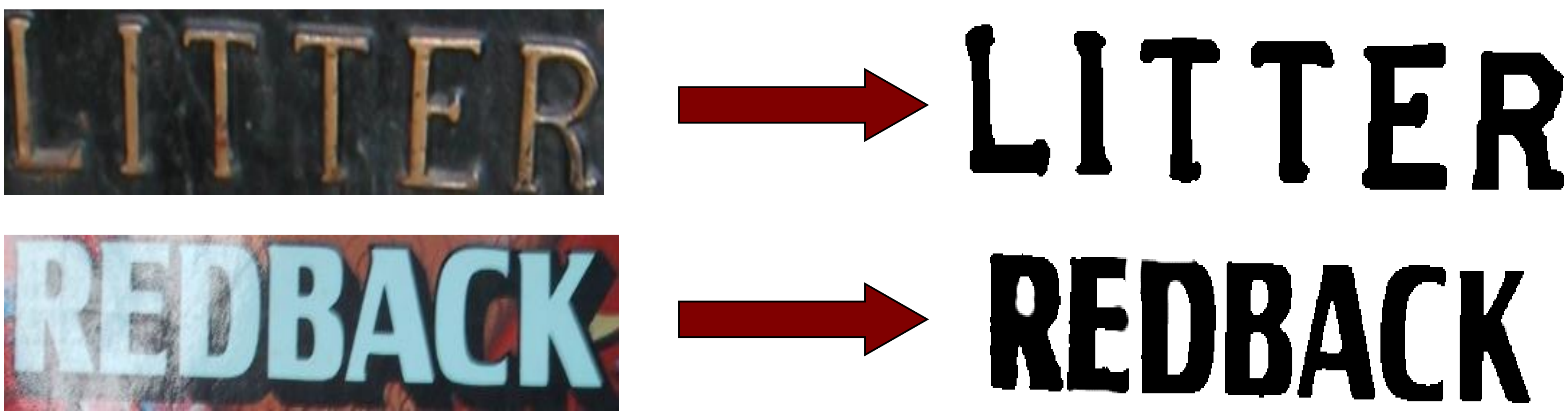
- Multimedia indexing
- Auto navigation

Natural Scene Text Recognition



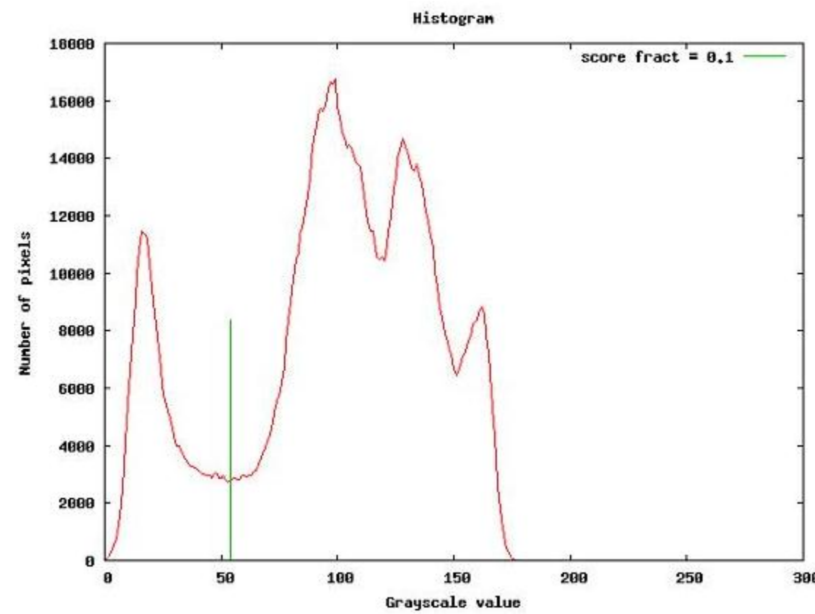
Natural Scene Text Recognition

- Text Detection → **Binarization** → Character segmentation → Recognition



Long History of Binarization

- Global Methods:
  - Otsu (1979)
  - Kittler (1985)
- Local Methods:
  - Niblack (1986),
  - Sauvola (2000)
- Uses local or global statistics
- Works satisfactorily well for scanned documents



Long History of Binarization

- **K-means and SVM based method**

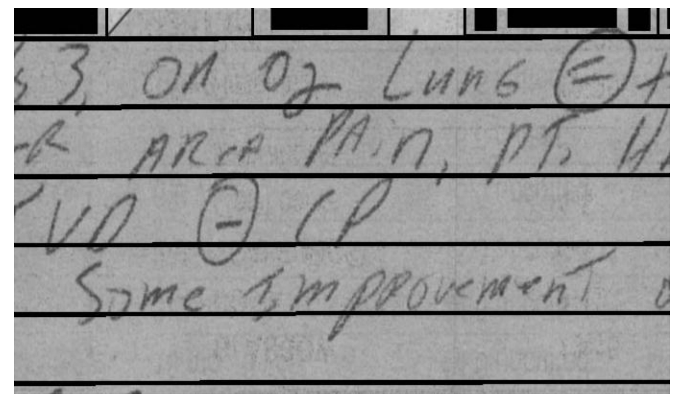
  \- Kita and Wakahara (ICPR 2010)

- **MRF model:**

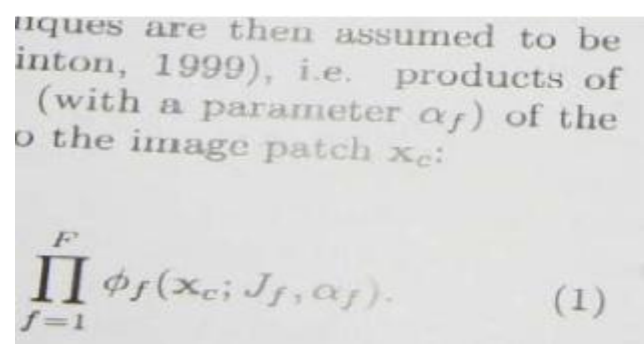
  \- Cao and Govindraju (CVPR 2007)

  \- Kuk and Cho (ICDAR 2009)

  \- Peng et al. (ICVGIP 2010)



- Many recent works: more suitable for scanned or handwritten documents



niques are then assumed to be
inton, 1999), i.e. products of
(with a parameter α_f) of the
o the image patch x_c:

$$\prod_{f=1}^{F} \phi_f(x_c; J_f, \alpha_f). \quad (1)$$

Natural Text Binarization: Challenges

- Similar text-background colours

Natural Text Binarization: Challenges

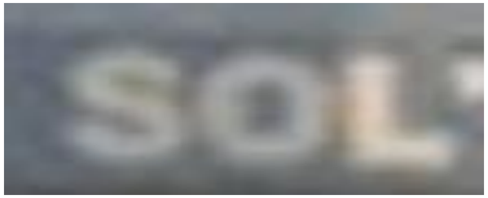
- Similar text-background colours
- Variable Illumination

Natural Text Binarization: Challenges

- Similar text-background colours
- Variable Illumination
- Noise

Natural Text Binarization: Challenges

- Similar text-background colours
- Variable Illumination
- Noise
- Low contrast



Natural Text Binarization: Challenges

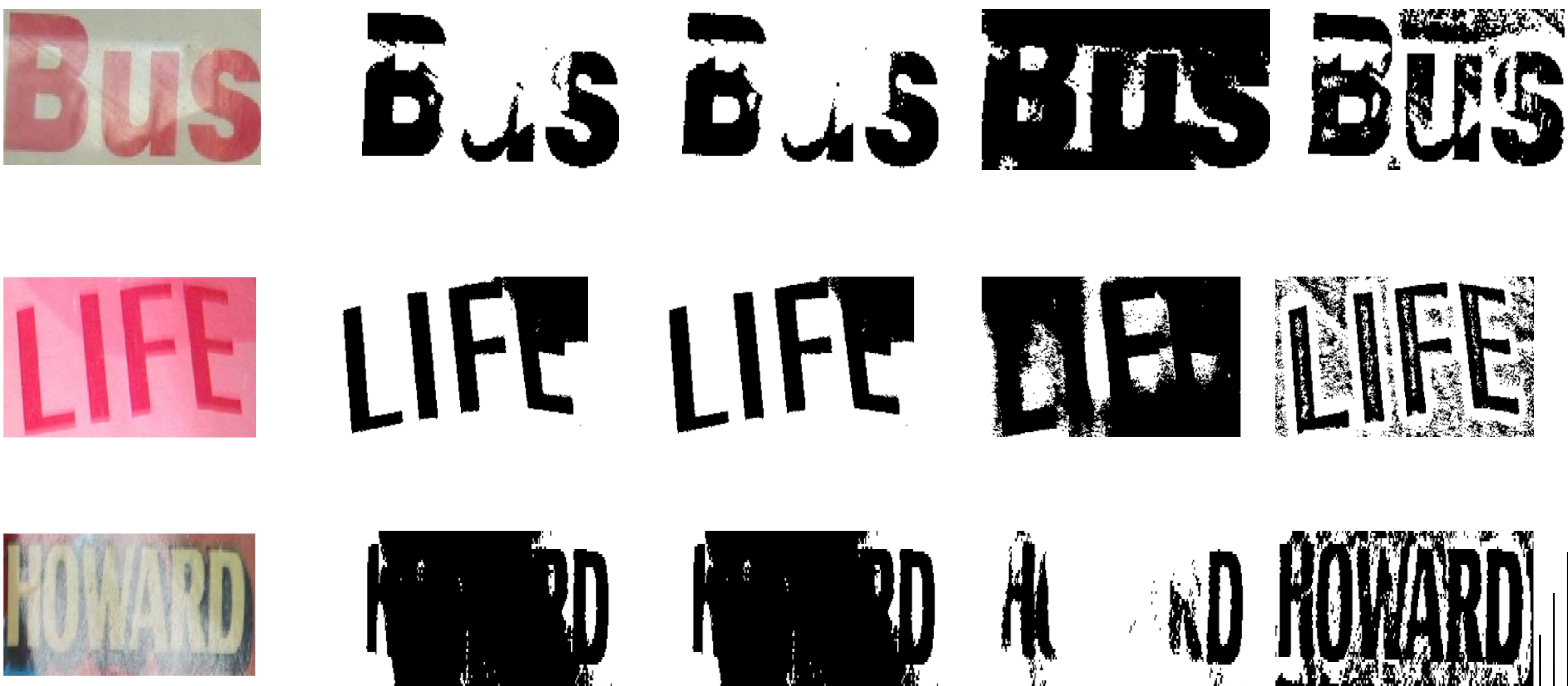
- Similar text-background colours
- Variable Illumination
- Noise
- Low contrast
- Non-uniform background

Natural Text Binarization: Challenges

- Similar text-background colours
- Variable Illumination
- Noise
- Low contrast
- Non-uniform background
- Imaging problems

Typical Failures

Otsu Kittler Sauvola Niblack



An MRF based Binarization

Assign a label to each pixel from L = {Text (0), Background(1)}

An MRF based Binarization

Assign a label to each pixel from L = {Text (0), Background(1)}

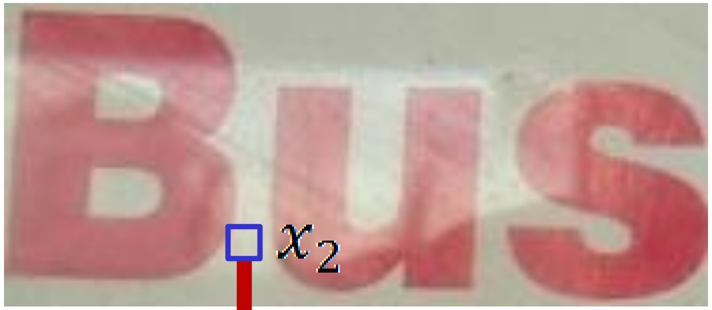
Many labelings possible, we are interested in
“the optimal” one

An MRF based Binarization

z_i: *Pixel colour at pixel position i*
fg: *foreground (text)*
bg: *background*

$p(fg|z_1)$ *is high*

An MRF based Binarization



z_i: *Pixel colour at pixel position i*
fg: *foreground (text)*
bg: *background*

$p(bg|z_2)$ *is high*

An MRF based Binarization

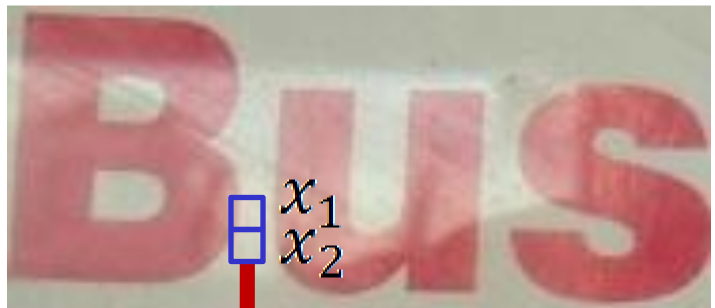
z_i: *Pixel colour at pixel position i*
fg: *foreground (text)*
bg: *background*

Minimize

$$E(x) = -\sum_i \log p(x_i|z_i)$$

Unary (data) Term

An MRF based Binarization



z_i: *Pixel colour at pixel position i*
fg: *foreground (text)*
bg: *background*

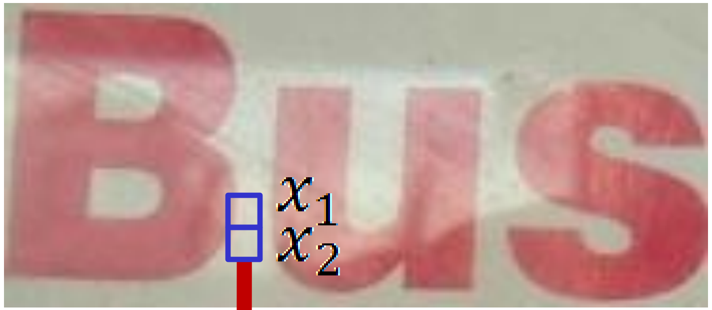
$p(bg, bg \mid z_1, z_2)$ *is high*

Minimize

$$E(x) = -\sum_i \log p(x_i|z_i)$$

Unary (data) Term

An MRF based Binarization



z_i: Pixel colour at pixel position i
fg: foreground (text)
bg: background

$p(bg, bg | z_1, z_2)$ is high

Minimize

$$E(x) = -\sum_i \log p(x_i|z_i) + \lambda_1 \sum_{i,j \in N} \exp\left(-\beta \, ||z_i - z_j||^2\right)$$

Unary (data) Term

Pair wise (smoothness) Term

An MRF based Binarization

Gradient magnitude at pixel position i

Pair wise term $= \lambda_1 \sum_{i,j \in N} \exp\left(-\beta \, ||z_i - z_j||^2\right)$

$$+ \lambda_2 \sum_{i,j \in N} \exp\left(-\beta \, ||w_i - w_j||^2\right)$$

Edginess Term

An MRF based Binarization

The problem is to minimize following energy (MRF energy):

$$\boldsymbol{E(x) = Unary\ term + Pairwise\ term}$$

An MRF based Binarization

The problem is to minimize following energy (MRF energy):

$$E(x) = Unary\ term + Pairwise\ term$$

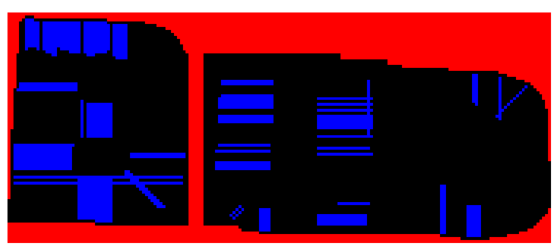
Two questions:

1) **How to learn the probabilities $p(x_i|z_i)$ used to compute the unary term?**
2) **How to find the minima of above energy?**

Learning Probabilities

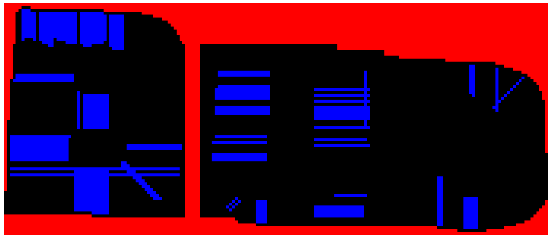
Canny Edge operator

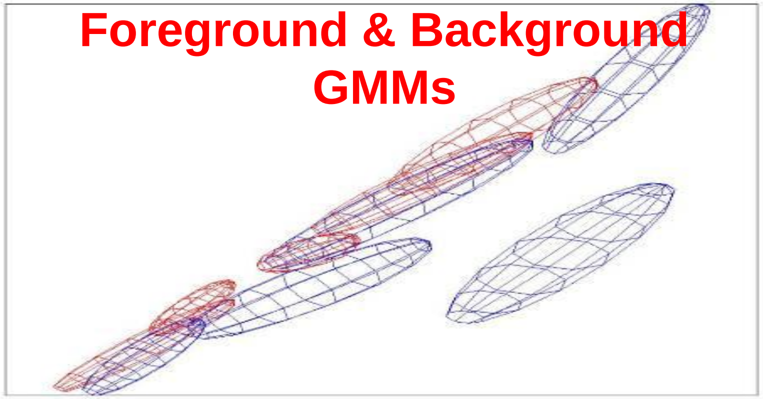
Find foreground - background seeds



Blue colour: Foreground
Red colour: Background

Colour modelling through GMMs





Unary term is calculated based on the probability of a pixel colour belonging to one of the GMM components

GrabCut (Carsten Rother et al., SIGGRAPH 2004) uses similar modelling for object segmentation problems

An MRF based Binarization

The problem is to minimize following energy (MRF energy):

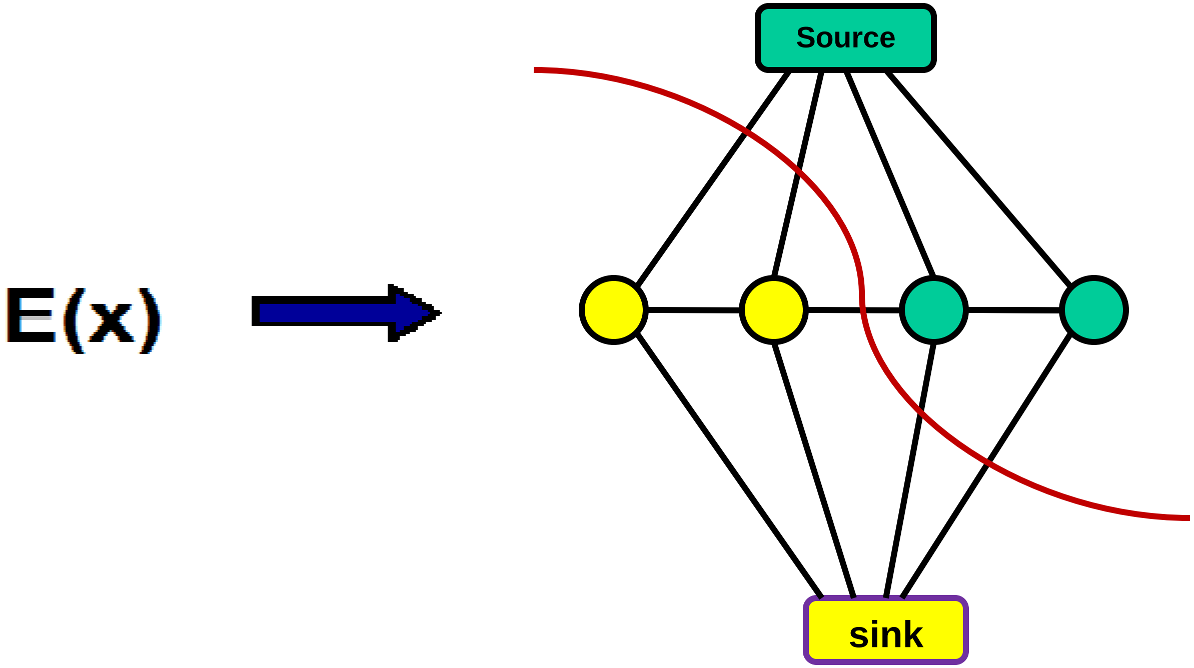
$$E(x) = Unary\,term + Pair\,wise\,term$$

Two questions:

1) **How to learn the probabilities $p(x_i|z_i)$ used to compute the unary term?**
2) **How to find the minima of above energy?**

Graph Cut

Minimum of MRF energy = min cut of graph



Efficient codes available to compute min cut of such graph

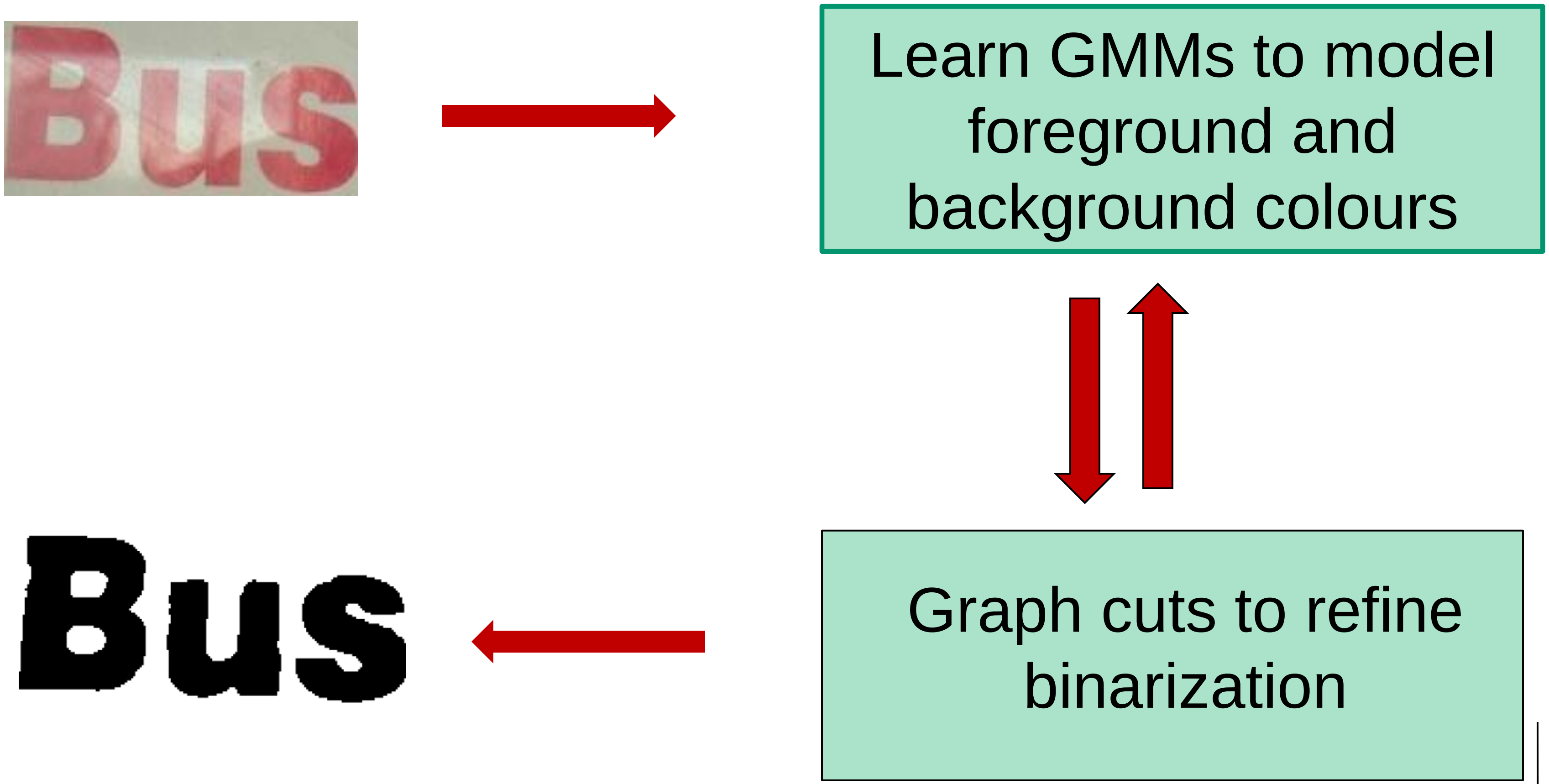
Vladimir Kolmogorov and Ramin Zabih, "What Energy Functions can be minimized via Graph Cut", PAMI 2004

An Iterative Graph Cut based Approach

Learn GMMs to model foreground and background colours

Graph cuts to refine binarization

An Iterative Graph Cut based Approach



Qualitative Results

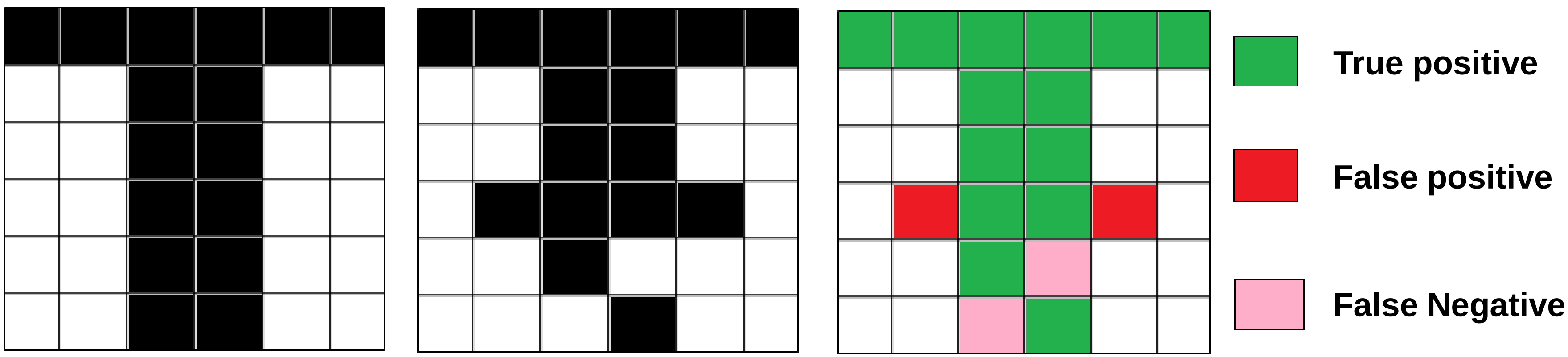
Qualitative Results

Quantitative Results

- OCR accuracy

Quantitative Results

- OCR accuracy
- Pixel level accuracy



Ground truth   Binarization result   Evaluation result

$$precision = \frac{Total\ number\ of\ green\ boxes}{Total\ number\ of\ green\ boxes + Total\ number\ of\ red\ boxes}$$

$$recall = \frac{Total\ number\ of\ green\ boxes}{Total\ number\ of\ green\ boxes + Total\ number\ of\ pink\ boxes}$$

$$f - score = \frac{2 \times precison \times recall}{precision + recall} \times 100$$

Results (ABBYY OCR Accuracy)

Method	Word Accuracy (%)	Character Accuracy (%)
Otsu	41.52	51.74
Sauvola	39.77	51.63
Niblack	39.18	42.31
Kittler	41.12	49.88
Otsu + CT	45.03	51.98
Proposed (without edginess diff.)	49.12	55.94
Proposed (with edginess diff.)	**52.04**	**60.14**

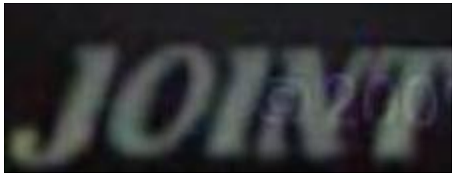
Results (Pixel Level Accuracy)

Method	f-score (%)
Otsu	79.32
Sauvola	73.87
Niblack	76.86
Kittler	72.89
Otsu + CT	78.12
Proposed (without edginess diff.)	87.84
Proposed (with edginess diff.)	**88.64**

More Results

Results based on Street View Text Dataset

Method	Word Recognition accuracy (%)
ABBYY	32.61%
Our Binarization + ABBYY	**42.81%**

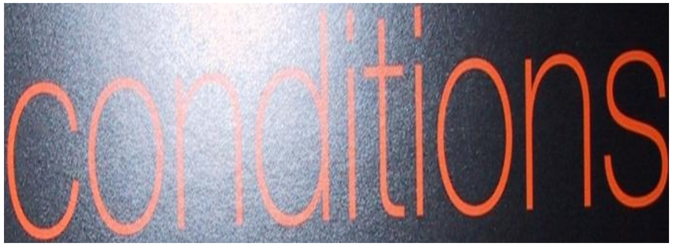


Kai Wang and Serge Belongie (ECCV 2010) have introduced a challenging Street View Text (SVT) dataset

Where we fail?

- Colour is not everything!!
  (At-least not always)

- Severe failure in learning text -background probabilities



Conclusions and Future Work

- A principled framework for challenging scene text Binarization
- Nearly 10 % improvement in accuracy
- **Future work:** Incorporating shape priors

Thank You

Supported by Microsoft Research India Travel Grant

Supplementary Slide

MRF based Methods in Literature

Method	Key points	Datasets
Cao and Govindraju (CVPR 2007)	1. Probability of character like patches are learnt 2. Does not handle intense illumination variation, complicated background, and blurring	Carbon copy handwritten images
Kuk *and Cho* (ICDAR 2009)	1. Text, Background and Near Text Regions are decided based on some local statistics 2. Graph cut is used for relabeling	Printed documents with uneven lighting
Peng *et al. (ICVGIP 2010)*	1. Graph cut is used to smooth initial binarization obtained by thresholding based methods 2. Along with intensity features, Stroke features are also used	Camera captured printed document