Methods for text segmentation from scene images

Deepak Kumar

Research Advisor: **Prof. A. G. Ramakrishnan**

Medical Intelligence and Language Engineering (MILE) Laboratory
Department of Electrical Engineering (EE)
Indian Institute of Science (IISc)
Bengaluru–560012, INDIA

Overview

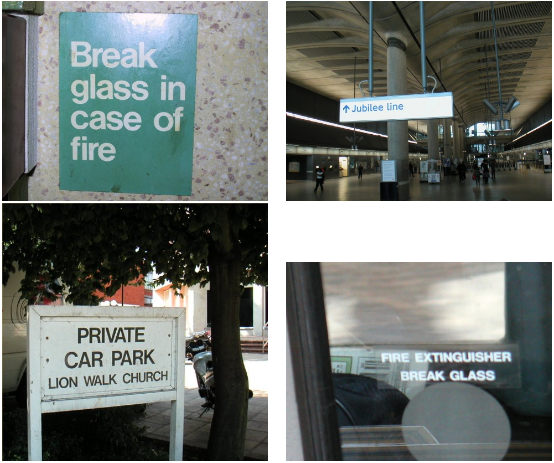
1. Scene text
2. Text segmentation (OTCYMIST)
3. Word segmentation (PLT, MAPS and NESP)
4. Character and word recognition (DCT and Block DCT)
5. Conclusion and Future work

Section I

Scene Text

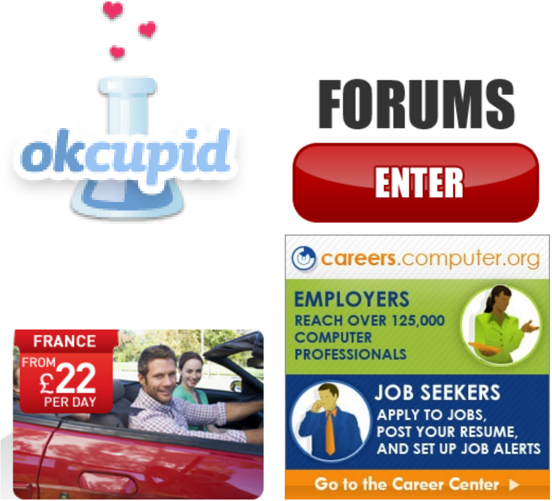
Scene text images

Figure: Samples from ICDAR 2003 competition data set with non-uniform illumination, perspective deformation and motion blur.


Break glass in case of fire
↑ Jubilee line
PRIVATE CAR PARK
LION WALK CHURCH
FIRE EXTINGUISHER
BREAK GLASS

Born-digital images

Figure: Samples from ICDAR 2011 competition data set.


okcupid
FORUMS
ENTER
careers.computer.org
EMPLOYERS
REACH OVER 125,000
COMPUTER
PROFESSIONALS
JOB SEEKERS
APPLY TO JOBS,
POST YOUR RESUME,
AND SET UP JOB ALERTS
Go to the Career Center
FRANCE
FROM
£22
PER DAY

Scene text analysis

Text detection, Text localization, Text segmentation/extraction, Word recognition, Character recognition

Text style complexities

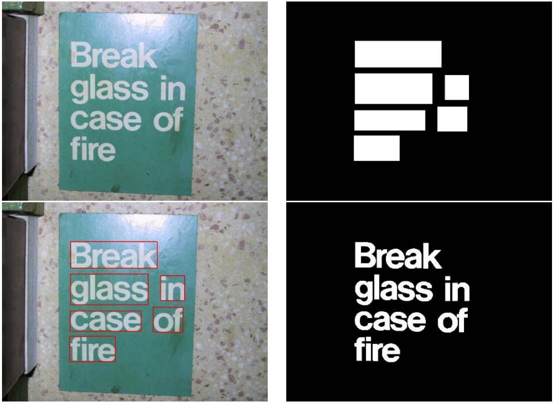
Randomly placed text, Resolution, Handwritten or Artistic font

Degradations

Skew, occlusion, motion blur, glossy (behind a glass frame) or non-uniform illumination

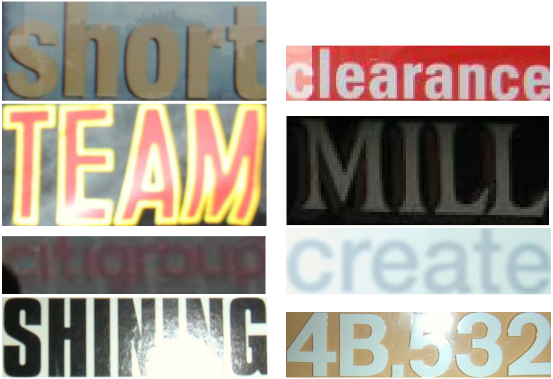
Text localization and segmentation

Figure: Sample image from ICDAR 2003 data set. (a) Original image (b) Text localization mask (c) Red coloured bounding box placed around each word (d) Text segmentation.


Break glass in case of fire
Break glass in case of fire
Break glass in case of fire

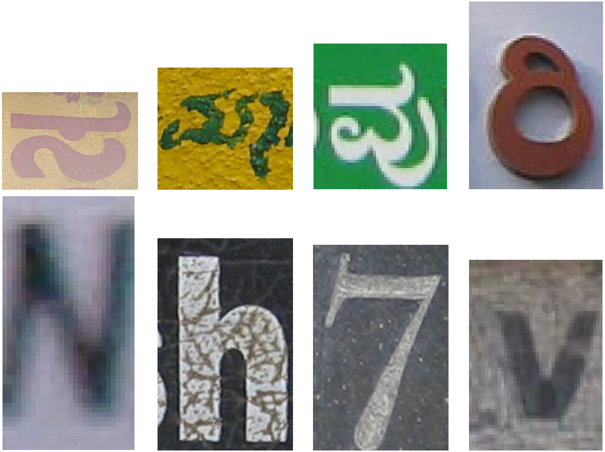
Word recognition

Figure: Cropped word image samples from ICDAR 2011 data set.


short
clearance
TEAM
MILL
citigroup
create
SHINING
4B.532

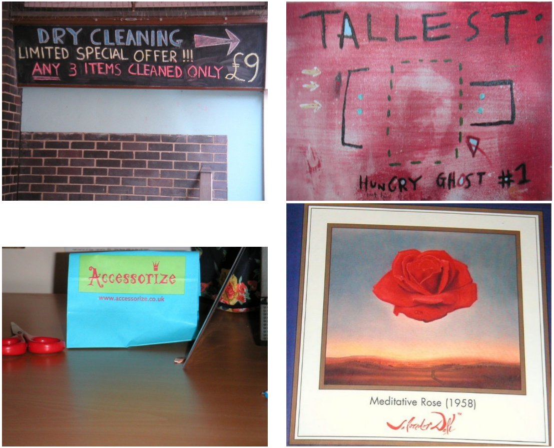
Character recognition

Figure: Kannada and English character image samples from Chars74k data set.



Handwritten or Artistic

Figure: Sample scene images with handwritten and artistic characters from ICDAR 2003 competition data set.


DRY CLEANING
LIMITED SPECIAL OFFER !!!
ANY 3 ITEMS CLEANED ONLY £9
TALLEST:
HUNGRY GHOST #1
Accessorize
www.accessorize.co.uk
Meditative Rose (1958)

Section II

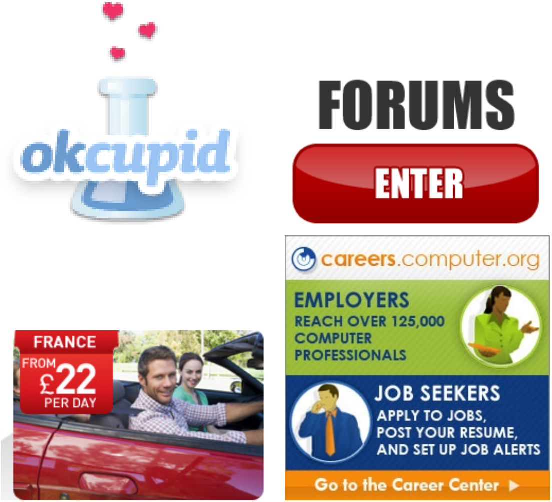
Text segmentation

Born-digital image

- An image with text superimposed by a software is known as born-digital image.
- The resolution of text present in the images and anti-aliasing of text with the background, form the major differences between scene and born-digital images.
- Otsu-Canny Minimum spanning tree (OTCYMIST) method is proposed for text segmentation task on born-digital image.

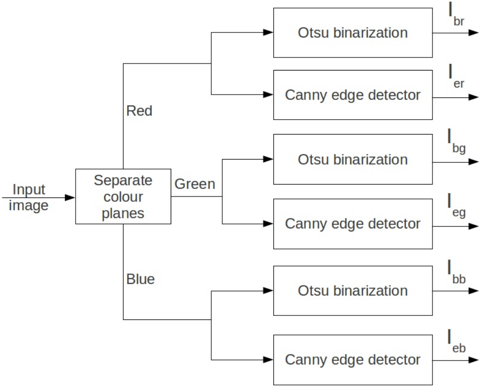
Born-digital images

Figure: Samples from ICDAR 2011 competition data set.


okcupid
FORUMS
ENTER
FRANCE
FROM £22 PER DAY
careers.computer.org
EMPLOYERS
REACH OVER 125,000 COMPUTER PROFESSIONALS
JOB SEEKERS
APPLY TO JOBS, POST YOUR RESUME, AND SET UP JOB ALERTS
Go to the Career Center ▶

Initial segmentation

Figure: The first part of the proposed OTCYMIST method. This comprises the binarization and edge map modules for the individual colour channels.



Input image
Separate colour planes
Red
Green
Blue
Otsu binarization
Canny edge detector
Otsu binarization
Canny edge detector
Otsu binarization
Canny edge detector
I_{br}
I_{er}
I_{bg}
I_{eg}
I_{bb}
I_{eb}

Discrimination calculation using Otsu method

$$\sigma_B^2(k*) = \max \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]}$$

where,

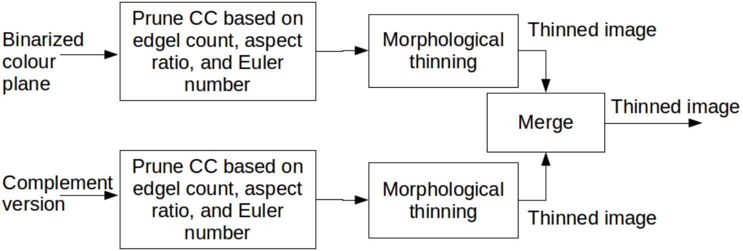
$$\omega(k) = \sum_{i=1}^{k} p_i$$

$$\mu(k) = \sum_{i=1}^{k} i p_i$$

$$\mu_T = \sum_{i=1}^{L} i p_i$$

Pruning connected components

Figure: Second section of OTCYMIST method separately processes each binarized colour plane (I_{br}, I_{bg} and I_{bb}) and its complement to form a thinned image.



Binarized colour plane

Prune CC based on edgel count, aspect ratio, and Euler number

Morphological thinning

Thinned image

Merge

Thinned image

Complement version

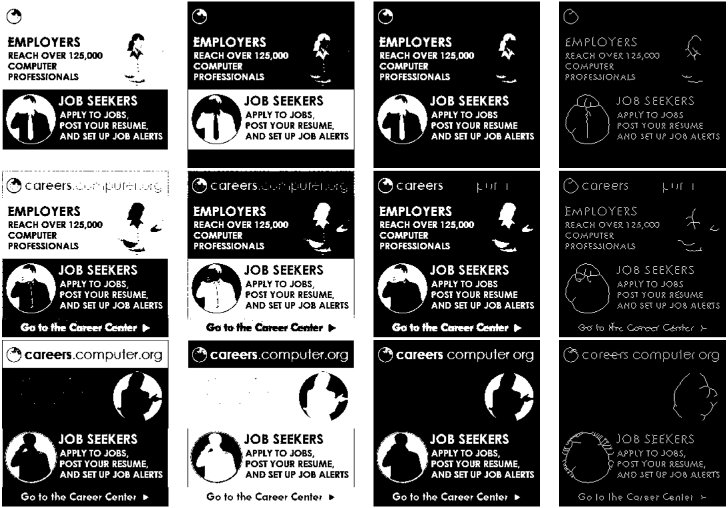
Prune CC based on edgel count, aspect ratio, and Euler number

Morphological thinning

Thinned image

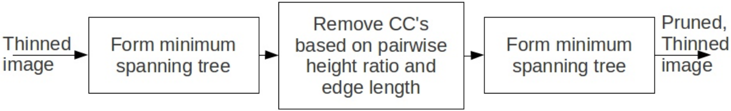
Segmentation and thinning results

Figure: Pruning and thinning of connected components in each colour plane.


careers.computer.org
EMPLOYERS
REACH OVER 125,000
COMPUTER
PROFESSIONALS
JOB SEEKERS
APPLY TO JOBS,
POST YOUR RESUME,
AND SET UP JOB ALERTS
Go to the Career Center ▸

Minimum spanning tree

Figure: Minimum spanning tree and height ratio check in the proposed OTCYMIST method. This section operates on each of the three thinned planes (I_{tr}, I_{tg} and I_{tb}) obtained from the previous section.



Thinned image

Form minimum spanning tree

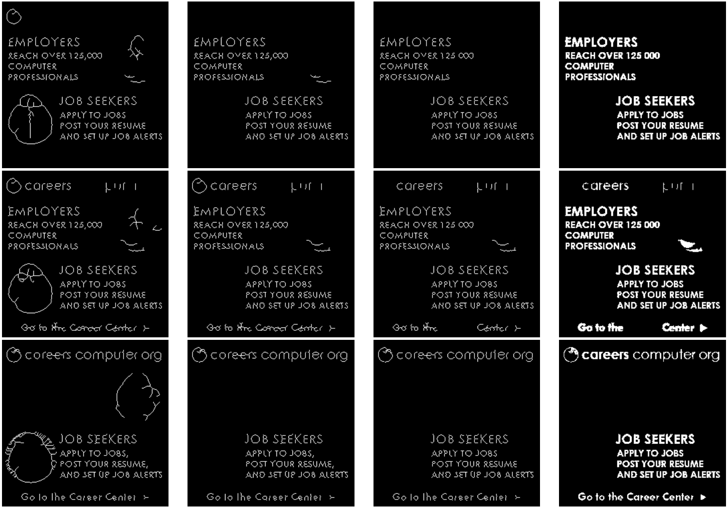
Remove CC's based on pairwise height ratio and edge length

Form minimum spanning tree

Pruned, Thinned image

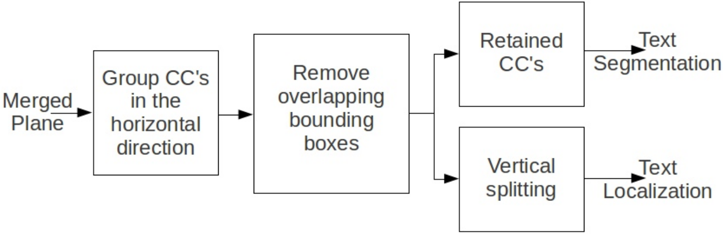
Minimum spanning tree results

Figure: Analysis of minimum spanning tree on thinned colour planes.


EMPLOYERS
REACH OVER 125 000
COMPUTER
PROFESSIONALS
JOB SEEKERS
APPLY TO JOBS
POST YOUR RESUME
AND SET UP JOB ALERTS
careers
Go to the Center
careers computer org
Go to the Career Center

Horizontal grouping

Figure: The final section of the OTCYMIST method which segments and localizes the text. Vertical splitting block locates the individual words from a group of words.


Merged Plane
Group CC's in the horizontal direction
Remove overlapping bounding boxes
Retained CC's
Vertical splitting
Text Segmentation
Text Localization

An example of horizontal grouping

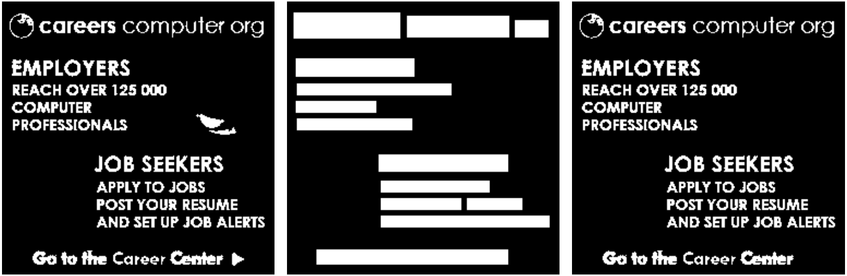
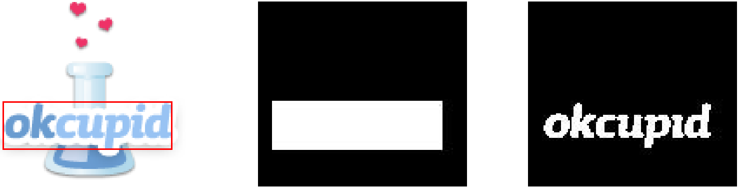

careers computer org
EMPLOYERS
REACH OVER 125 000
COMPUTER
PROFESSIONALS
JOB SEEKERS
APPLY TO JOBS
POST YOUR RESUME
AND SET UP JOB ALERTS
Go to the Career Center

Figure: CC's before grouping. After grouping, each group is replaced by a white mask; thirteen distinct groups can be seen. Most of the non-text components are removed due to the overlap, which does not happen for the text components.

Text localization and segmentation results

Figure: Result of the proposed OTCYMIST method for the born-digital image. (a) Localized text. (b) Text Localization mask. (c) Segmented text.



Text localization results - ICDAR 2011

Table: Text localization results (%) of ICDAR 2011 Robust Reading Competition: Challenge-1 on BDI evaluated using Wolf and Jolion method.

Method	Recall	Precision	H-mean
TDM IACAS	69.70	85.83	76.93
TH-TextLoc	73.06	80.39	76.55
Textorter	69.08	85.54	76.43
Baseline Method	69.94	83.92	76.30
OTCYMIST	75.65	63.85	69.25
SASA	64.91	67.38	66.12
TextHunter	58.43	75.52	65.88

Text segmentation results - ICDAR 2011

Table: Text segmentation results (%) of ICDAR 2011 Robust Reading Competition: Challenge-1 on BDI evaluated using Clavelli's method.

Method	Component Level			Pixel Level		
	Well Segmented	Merged	Lost	Recall	Precision	H-mean
OTCYMIST	64.14	15.69	20.15	80.62	72.06	76.10
Textorter	58.12	9.50	32.37	65.23	63.63	64.42
SASA	41.58	10.97	47.43	71.68	55.44	62.52

Text localization results - ICDAR 2013

Table: Text localization results (%) of ICDAR 2013 Robust Reading Competition: Challenge-1 (on BDI) evaluated using Wolf and Jolion method.

Method	Recall	Precision	Hmean
USTB_TexStar	82.38	93.83	87.74
TH-TextLoc	75.85	86.82	80.96
I2R_NUS_FAR	71.42	84.17	77.27
Baseline	69.21	84.94	76.27
Text Detection	73.18	78.62	75.81
I2R_NUS	67.52	85.19	75.34
BDTD_CASIA	67.05	78.98	72.53
OTCYMIST	74.85	67.69	71.09
Inkam	52.21	58.12	55.00

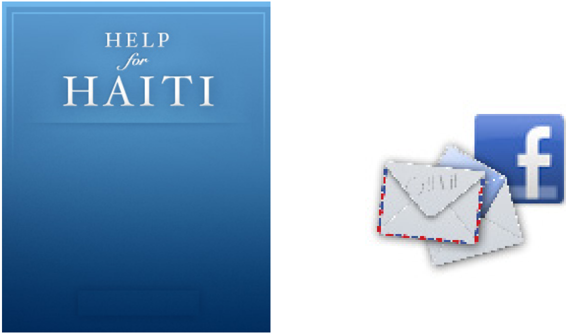
Text segmentation results - ICDAR 2013

Table: Text segmentation results (%) of ICDAR 2013 Robust Reading Competition: Challenge-1 (on BDI) evaluated using pixels and atoms.

Method	Pixel Level			Atom Level		
	Recall	Precision	F-score	Recall	Precision	F-score
USTB_FuStar	87.21	79.98	83.44	80.01	86.20	82.99
I2R_NUS	87.95	74.40	80.61	64.57	73.44	68.72
OTCYMIST	81.82	71.00	76.03	65.75	71.65	68.57
I2R_NUS_FAR	82.56	74.31	78.22	59.05	80.04	67.96
Text Detection	78.68	68.63	73.32	49.64	69.46	57.90

Failure cases

Figure: Images from the born-digtial competition data set, where OTCYMIST method fails. The first image has varying illumination. The second image has a single character, which gets eliminated in the proposed method.



Conclusion - Text segmentation

- Recall is better than the other methods, since the individual channels are used for segmentation by OTCYMIST.
- The performance of OTCYMIST method on bipolarity text was good.
- Minimum spanning tree is used in OTCYMIST method.

Video demonstration

Text segmentation video.

Section III

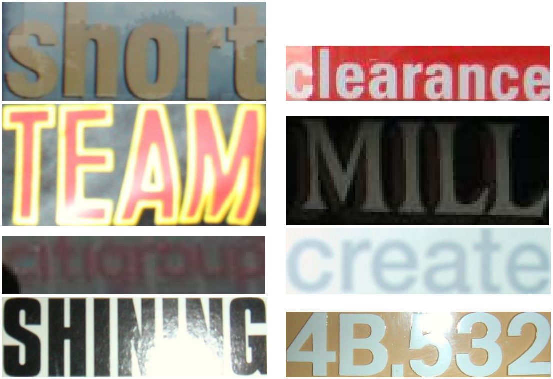
Word segmentation

Introduction

- The task of word recognition is considered as a part of object recognition in the field of computer vision.
- Recognition of a word obtained by manually cropping the scene or born-digital image, is not trivial.
- If the cropped word is first segmented, it can be more easily recognized by traditional OCR engines.
- It has two advantages: 1. avoids building character classifier from scratch and 2. reduces word recognition task to word segmentation task.
- We propose two bottom-up approaches for the task of word segmentation. These approaches choose statistically different features at the initial stage of segmentation.

Sample cropped images

Figure: Sample cropped word images from ICDAR 2011 data set.


short
clearance
TEAM
MILL
citigroup
create
SHINING
4B.532

Competition on word recognition

- First competition was organized in ICDAR 2003. But,no entries were received.
- 52% word recognition rate was shown on sample data set.
- 62.7% word recognition rate with a custom lexicon.
- 41.2% word recognition rate was reported in the competition on ICDAR 2011 data set.

Methods

Table: Distinction between the three approaches proposed for segmentation of word images. MAPS method begins by binarizing the middle line pixels, whereas PLT and NESP methods operate on the gray and colour values of the pixels, respectively.

Method	Initial input to the method	Segmentation approach
PLT	Gray values of an image	Non-linear enhancement before applying Otsu's threshold
NESP	Colour values of an image	Independent non-linear enhancement of each plane before applying Fisher discriminant and Otsu's threshold
MAPS	Gray values of only the middle row of an image	Foreground and background segmentation using Min-cut/Max-flow algorithm

Power-law transform

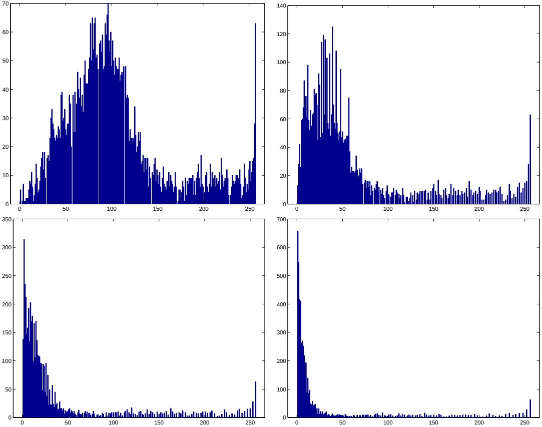
- In images with poor contrast, which is a kind of degradation, the adjacent characters merge during segmentation.
- The pixel values are modified to avoid the merge between adjacent characters.

$$f_{out}(x, y) \;=\; C\, f_{in}{}^{\gamma}(x, y)$$

- A variety of devices used for image capture, printing, and display respond according to a power-law.

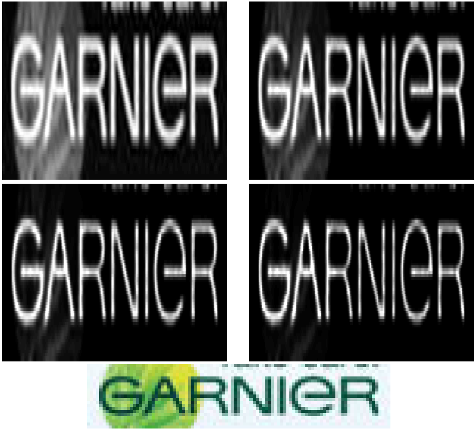
Histogram changes due to PLT

Figure: Histogram plots obtained after power-law transformation with $\gamma = 1, 2, 3$ and 4, respectively. The modified pixel values change the appearance of histogram.



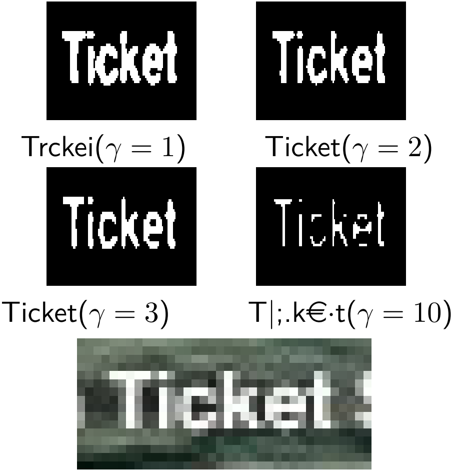
Contrast enhancement

Figure: Contrast enhancement after power-law transformation with $\gamma = 1, 2, 3$ and 4, respectively. The gray patch behind the letters ‘G’ and ‘A’ gradually decreases and becomes invisible for $\gamma = 4$.


GARNIER
GARNIER
GARNIER
GARNIER
GARNIER

Word recognition

Figure: Segmented outputs and their respective OCR results for different values of γ. To begin with, increasing the value of γ yields proper output and further increase in the γ value deteriorates the character stroke width.



Trckei($\gamma = 1$)

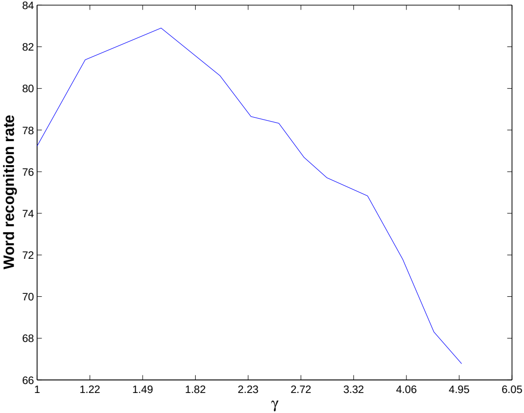
Ticket($\gamma = 2$)

Ticket($\gamma = 3$)

T|;.k∈·t($\gamma = 10$)

Recognition performance Vs. γ

Figure: A plot of word recognition rate against $\log \gamma$ on ICDAR 2011 born-digital word data set. Gamma value is fixed in each run to estimate the recognition rate on the entire test data set.



Non-linear enhancement and selection of plane

A word image is split into Red, Green, Blue, Gray and Lightness (CIE Lab) components, which we refer to as planes.

$$d_c^2(k*) = \max_k \frac{[\mu_{ct}\omega_c(k) - \mu_c(k)]^2}{\omega_c(k)[1-\omega_c(k)]}, \quad c \in [R,\ G,\ B,\ I,\ L]$$

where,

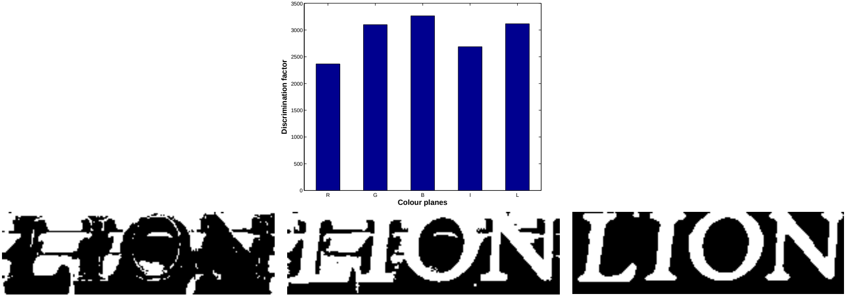
$$\omega_c(k) = \sum_{i=1}^{k} p_{ci}$$

$$\mu_c(k) = \sum_{i=1}^{k} ip_{ci}$$

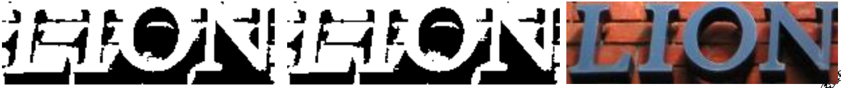
$$\mu_{ct} = \sum_{i=1}^{N} ip_{ci}$$

Selection of plane

Figure: Illustration of the effectiveness of plane selection in the NESP approach.



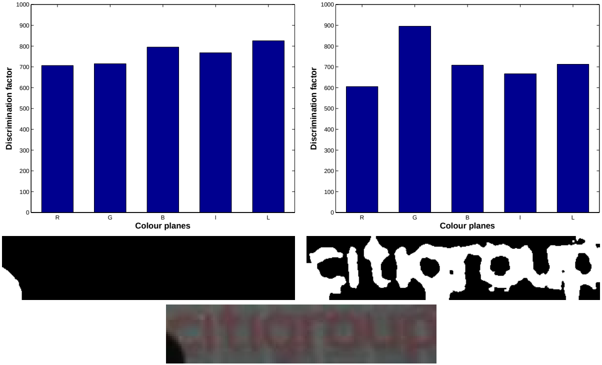
(a) Red (b) Green (c) Blue



(d) Intensity (e) Lightness (f) Original

Effect of Power-law

Figure: Top row: Discrimination factor computed by our method for different colour planes without and with power-law transformation. Second row: Results of segmenting the selected plane. Bottom row: Original image.



Middle line analysis and propagation of segmentation

- The gray values of an image are processed in this method.
- This method starts by segmenting the middle row pixels.
- Then for segmentation of entire image, the label information from the middle row pixels is passed on to the other pixels.

Segmentation of middle line - Niblack

Niblack's method

$$T_i = \mu_i + \ k_n \ * \ \sigma_i$$

$$\mu_i = \frac{1}{N_w} \sum_{j \in N_w} f(x_{i+j}, y)$$

$$\sigma_i^2 = \frac{1}{N_w} \sum_{j \in N_w} (f(x_{i+j}, y) \ - \ \mu_i)^2$$

Segmentation of middle line - Min–Max

Min-Max method

$$f_L(x_i, y) = [f(x_{i-N_m+1}, y), ..., f(x_i, y)]$$

$$f_R(x_i, y) = [f(x_i, y), ..., f(x_{i+N_m-1}, y)]$$

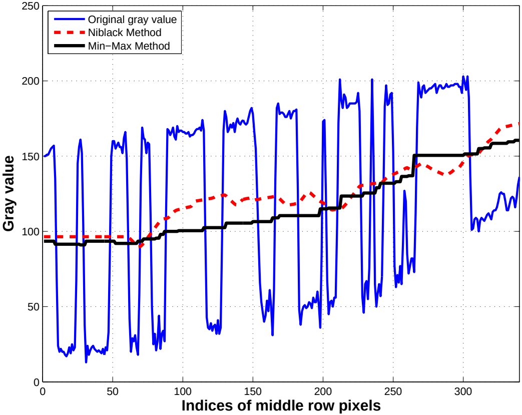
$$T_{max} = \min(\max_{N_m} f_L(x_i, y), \max_{N_m} f_R(x_i, y))$$

$$T_{min} = \max(\min_{N_m} f_L(x_i, y), \min_{N_m} f_R(x_i, y))$$

$$T = (T_{min} + T_{max})/2$$

Plot of middle line segmentation

Figure: A plot of gray values of the middle row of an image with the binarization thresholds given by Niblack and Min-Max Methods.



Classification of other pixels

$$\mu_0 = \frac{1}{N_0} \sum_{i \in C_0} f(x_i, y)$$

$$\sigma_0^2 = \frac{1}{N_0} \sum_{i \in C_0} (f(x_i, y) - \mu_0)^2$$

$$\mu_1 = \frac{1}{N_1} \sum_{i \in C_1} f(x_i, y)$$

$$\sigma_1^2 = \frac{1}{N_1} \sum_{i \in C_1} (f(x_i, y) - \mu_1)^2$$

Bayesian classification

$$p(C_0|f(x,y)) = \frac{p(f(x,y)|C_0)p(C_0)}{p(f(x,y)|C_0)p(C_0) + p(f(x,y)|C_1)p(C_1)}$$

$$p(C_0) = N_0/N$$

$$h(f(x,y)) : \begin{cases} p(C_0|f(x,y)) \geq p(C_1|f(x,y)), & f(x,y) \in C_0 \\ p(C_0|f(x,y)) < p(C_1|f(x,y)), & f(x,y) \in C_1 \end{cases}$$

$$p(f(x,y)|C_0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\{-\frac{(f(x,y) - \mu_0)^2}{2\sigma_0^2}\}$$

Min-cut/Max-flow algorithm

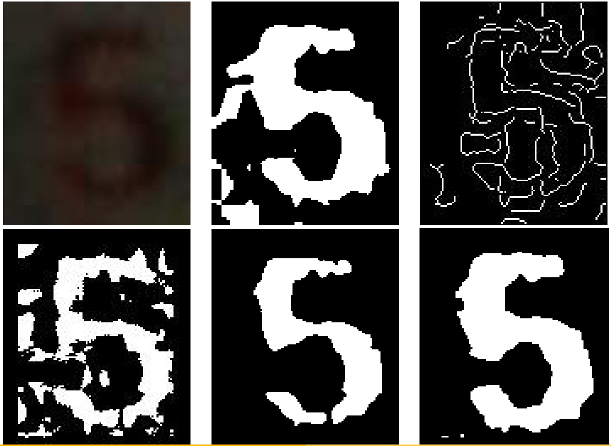
Energy minimization using Pott's model.

$$E(L) = \sum_{i \in \mathcal{I}} D(f_i, L_i) + \sum_{i \in \mathcal{I},\ j \in \mathcal{N}(i)} V(f_i, f_{(i+j)}, L_i, L_{(i+j)})$$

where $L = L_i | i \in \mathcal{I}$ is a labeling of the image $\mathcal{I}$, $D(.)$ is a data penalty function, $V(.)$ is an interaction potential, and $\mathcal{N}(i)$ denotes the neighborhood of i.

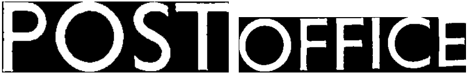
Middle line segmentation results

Figure: Segmentation of a sample word image from ICDAR 2003 data set. (a) Original image. (b) Otsu's method. (c) Canny's method. (d) Niblack's method. (e) MAPS technique (f) NESP technique ($\gamma = 1$ and Green plane).



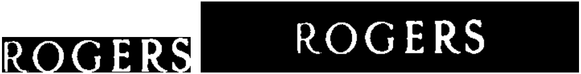
Post-processing of word images

Figure: The background is broken, where the text connected components touch the boundary of the word image. This creates ambiguity in determining the text polarity.



al OFFI

Figure: Post-processing the segmented image by padding background pixels to eliminate foreground-background ambiguity. (i) The text connected components touch the boundary. (ii) Foreground and background are clearly separated, after background padding.



IttentitIO ROGERS

(i) (ii)

Datasets

- ICDAR 2003: 1110 images
- Sign Evaluation (PAMI) 2009: 215 images
- Street view text (SVT) 2010: 647 images
- ICDAR Born-digital (BDI) 2011: 918 images
- ICDAR Scene 2011: 716 images

Benchmark

What is the maximum achievable recognition rate for properly segmented images?

Four months for creation of manually segmented images by three people.

Word recognition rates without lexicon

Table: Comparison of word recognition rates (WRR) of images segmented by manual (benchmark), MAPS and NESP method with the best results in the literature and baseline - for the seven publicly available word image data sets. (* Custom lexicon is used)

Data sets	ICDAR 2003	PAMI 2009	SVT 2010	BDI 2011	ICDAR 2011	BDI 2013	ICDAR 2013
Benchmark	83.9	89.3	79.6	88.5	86.7	—	—
NESP	**66.2**	80.9	35.2	79.4	**72.8**	81.7	65.9
MAPS	64.5	80.0	39.6	**82.8**	71.7	**83.8**	66.0
Best result in literature	62.7*	86.1*	73.6*	61.5	56.4*	82.2	82.8
Baseline (Omnipage)	41.0	50.2	27.7	63.0	31.4	61.0	45.3

Lexicon

- Searching an exact word from a vocabulary, covering up to a million words, is time consuming. If constraints are applied on the vocabulary, then the search list is reduced drastically.
- A custom lexicon is used to improve word recognition rates.
- Applications of word recognition with customized lexicon (only specific cases) are locating road or business names on highway or within a city, detecting the room or house number in a department or an apartment and the number on license plate of a vehicle in a state or country.

Word recognition rates with lexicon

Table: Comparison of word recognition rates (WRR) with and without lexicon for images segmented by MAPS and NESP methods for the seven publicly available word image data sets. A synthetic custom lexicon is used for SVT data set and a limited lexicon for other data sets, both derived from the ground-truths of the respective test sets.

Data sets	ICDAR 2003	PAMI 2009	SVT 2010	BDI 2011	ICDAR 2011	BDI 2013	ICDAR 2013
NESP + Lex	74.9	92.1	56.4	88.2	80.5	89.1	76.6
NESP	66.2	80.9	35.2	79.4	72.8	81.7	65.9
MAPS + Lex	74.2	88.8	63.5	89.9	79.2	90.0	76.7
MAPS	64.5	80.0	39.6	82.8	71.7	83.8	66.0
Best result in literature	62.7*	86.1*	73.6*	61.5	56.4*	82.2	82.8

Conclusion - Word segmentation

- We have proposed methods for segmentation and recognition of words from camera-captured and born-digital word image data sets.
- NESP method picks the right plane and the right value for gamma to process a word image.
- MAPS method performs initial operation on middle line pixels.

Video demonstration

English word recognition video.

Section IV

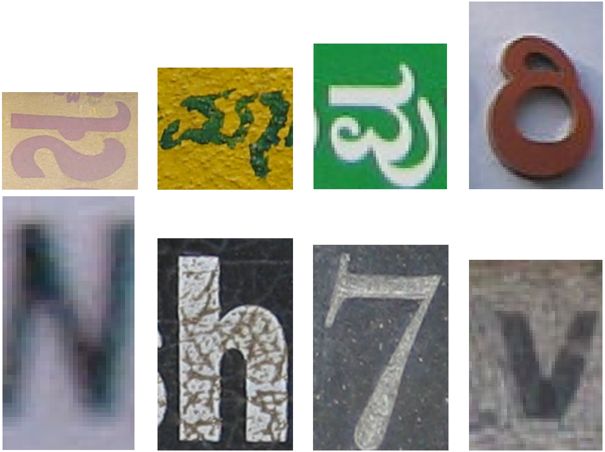
Character and word recognition

Features

- A method is proposed for feature extraction and classification of manually isolated characters from scene images.
- Chars74k dataset for English and Kannada scripts.
- Discrete Cosine Transform (DCT) is used to extract the features.
- MILE laboratory Kannada OCR samples are used for training the classifier and build Kannada word recognizer.

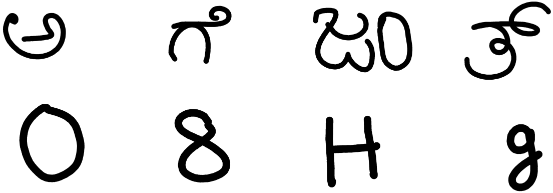
Character images

Figure: Kannada and English character image samples from Chars74k data set.



Handwritten samples

Figure: Kannada and English handwritten samples from Chars74k dataset.



Descriptors

Point based descriptors

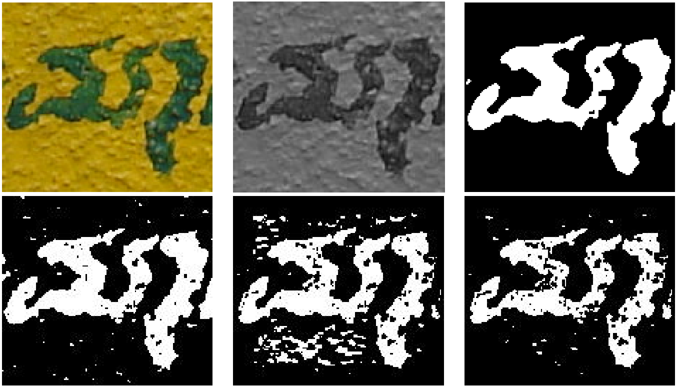
- Shape context
- SIFT
- Spin images
- Geometric blur
- Filter response (MR-8)/ Edge response
- HOG

Region based descriptors

- Patch descriptors
- Discrete Cosine Transform (DCT)

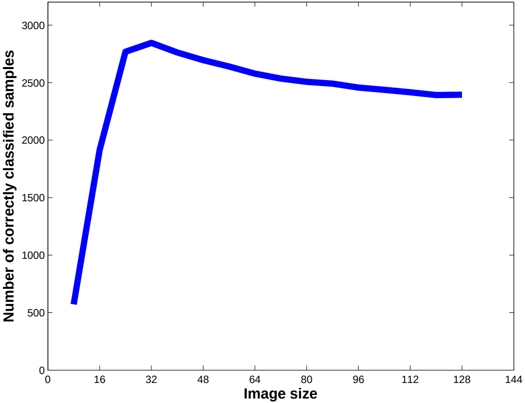
Binarization results using different methods

Figure: Comparison of outputs of different binarization techniques. (a) A sample image. (b) Gray scale image. (c) Classification of pixels using energy minimization function. (d) Otsu's global thresholding. (e) Niblack's local thresholding. (f) Sauvola's and Pietäikinen's local thresholding.



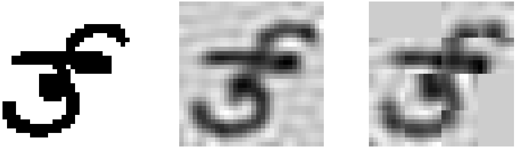
Determining optimal size for normalization

Figure: Plot of number of correctly classified Kannada character samples as a function of normalization size of the image.



Approximation by Global and Block DCT

Figure: Original image of a Kannada character ‘th’ and images reconstructed using Global DCT and Block DCT for 15% of the total number of coefficients.



Char74k dataset

- English *Fnt* : 1016 samples for each of 62 classes.
- English *Hnd* : 55 samples for each of 62 classes.
- English *Img* : 15 samples for each of 62 classes.
- Kannada *Hnd* : 25 samples for each of 657 classes.
- Kannada *Img* : Total 5135 samples for 990 classes.

MILE Kannada OCR samples

Machine printed: 347 base classes + 40 ottu classes (387 classes)

Results using *Img* dataset for training

Table: The classification results (%) on English *Img* dataset, using *Img* for training. Nearest neighbor classifier is used since the number of training samples available per class is limited. (5 for Chars74k-5 and 15 for Chars74k-15)

Feature vector	Chars74k-5	Chars74k-15
Block DCT	**47.71 ± .43**	**58.28**
Global DCT	**47.67 ± .65**	57.85
HOG	45.33 ± .99	57.5
MKL	—	55.26
Shape Context	26.1 ± 1.6	34.41
Geometric Blur	36.9 ± 1.0	47.09
Patches	13.7 ± 1.4	21.40
MR8	6.9 ± 0.7	10.43

Results using *Fnt* dataset for training

Table: The classification results on English *Img* dataset, using *Fnt* for training. The number of classes is 62 and there are 1016 training samples per class.

Feature vector	Recognition rate (%)
Block DCT	**66.4**
Global DCT	66.3
Sobel edges	71.6
Shape Context	44.8
Geometric Blur	54.3
SIFT	11.1
Patches	7.8

Results using *Hnd* dataset for training

Table: The classification results on English *Img* dataset, using *Hnd* for training. The number of classes is 62 and there are 55 training samples per class.

Feature vector	Recognition rate (%)
Global DCT	**46.4**
Block DCT	44.6
Shape Context	31.1
Geometric Blur	24.6
SIFT	3.1
Patches	1.7

Kannada *Hnd* dataset

Table: The cross validation results with different features on Kannada *Hnd* dataset consisting of 657 classes and 25 samples per class.

Feature vector	Recognition rate (%)
Global DCT	**33.3**
Block DCT	33.1
Shape Context	29.9
Geometric Blur	17.7
SIFT	7.6
Patches	23.0

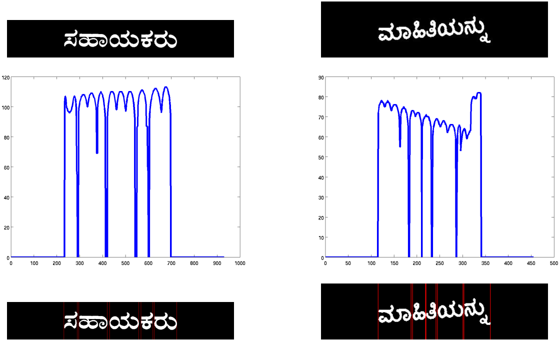
Kannada *Img* dataset

Table: The classification results on 5135 test samples Kannada *Img*, using Kannada *Hnd* dataset for training. The number of classes is 657 and there are 25 training samples per class.

Feature vector	Recognition rate (%)
Global DCT	**11.4**
Block DCT	11.1
Shape Context	3.5
Geometric Blur	2.8
SIFT	0.3
Patches	0.1

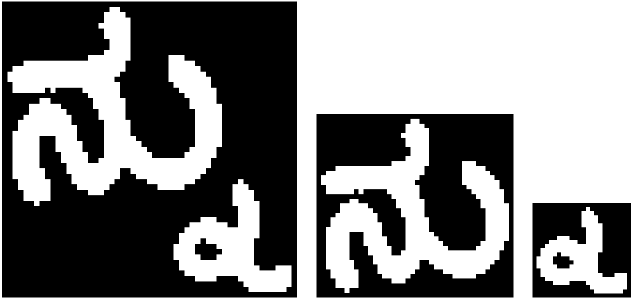
Segmentation of binarized word into components

Figure: Top row: Two manually segmented word images. Middle row: The plot of bottom text pixel from top row. Minima show the property of baseline. Bottom row: The word images are split into components using a threshold. Segmentation locations are shown by red lines.

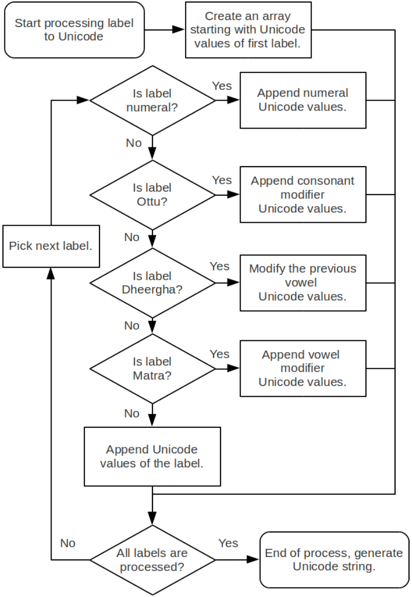


Splitting Ottu component

Figure: The segmented image has a base and a Ottu component. The Ottu is segmented based on its low overlap with the base component. The base component is shown in the middle and the Ottu component in the last column.



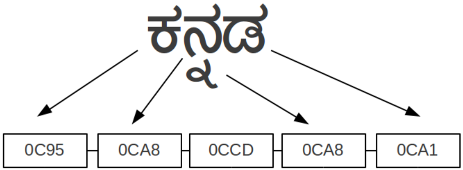
Kannada label to Unicode generation


Start processing label to Unicode
Create an array starting with Unicode values of first label.
Is label numeral?
Yes
Append numeral Unicode values.
No
Is label Ottu?
Yes
Append consonant modifier Unicode values.
Pick next label.
No
Is label Dheergha?
Yes
Modify the previous vowel Unicode values.
No
Is label Matra?
Yes
Append vowel modifier Unicode values.
No
Append Unicode values of the label.
No
All labels are processed?
Yes
End of process, generate Unicode string.

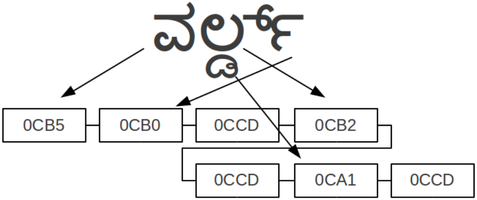
The flowchart of Kannada labels to Unicode conversion. The four important sections in the flowchart check the possible combination of previous and present label that modify the generated Unicode.

Mapping of Kannada symbols to Unicode

Figure: Mapping of Kannada symbols to their respective Unicode during the generation of a Kannada word.


ಕನ್ನಡ
0C95
0CA8
0CCD
0CA8
0CA1

(a) Linear mapping.


ವರ್ಲ್ಡ್
0CB5
0CB0
0CCD
0CB2
0CCD
0CA1
0CCD

(b) Complex mapping.

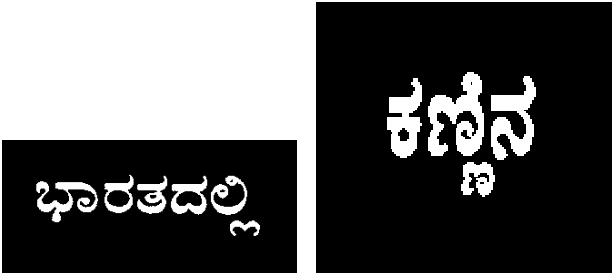
Kannada word recognition rate

Table: The word recognition rate of Kannada test samples using Tesseract OCR and Block DCT-NN. The number of words in the test set is 243.

Methods	Tesseract OCR		Block DCT-NN	
	Word recg.(%)	Edit distance	Word recg.(%)	Edit distance
Benchmark	11.11	178.68	12.35	141.09
NESP	5.76	212.26	6.58	200.41
PLT	5.35	210.61	7.82	194.29
MAPS	4.94	209.07	7.41	182.41

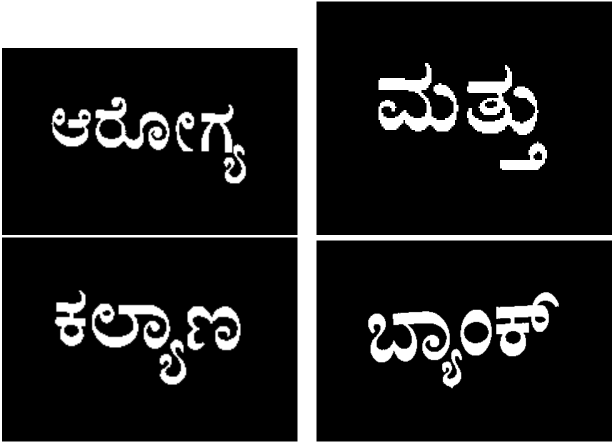
Kannada word recognized by both OCR

Figure: Common Kannada words, with Ottu, recognized by both Tesseract OCR and Block DCT from manually segmented word images.



Kannada word recognized by Block DCT

Figure: Kannada words, with Ottu, recognized by Block DCT but not by Tesseract OCR from manually segmented word images.



Kannada word recognition rate with lexicon

Table: The word recognition rate of Kannada test samples using Tesseract OCR and Block DCT. The number of words in the test set is 243 and used as a custom lexicon.

Methods	Tesseract OCR		Block DCT	
	Word recg. (%)	Edit distance	Word recg. (%)	Edit distance
Benchmark	32.10	155.35	43.21	110.96
NESP	14.81	201.58	25.10	164.90
PLT	15.64	197.57	27.98	158.59
MAPS	17.70	192.97	29.22	147.64

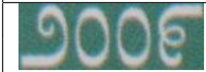
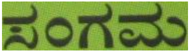
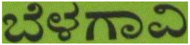
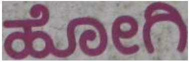
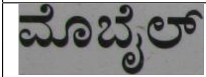
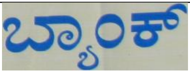
Different training samples

Table: Comparison of the classification accuracy(%) on Kannada test set from Chars74k data set using the *Hnd* training samples from Chars74k dataset and MILE Kannada OCR samples.

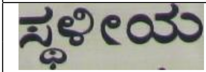
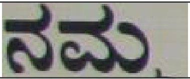
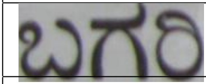
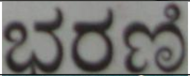
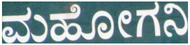
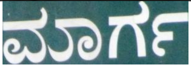
Methods	Chars74k samples	MILE Kannada OCR samples
Block DCT	11.1	36.8
Global DCT	11.4	34.3

Kannada word recognized

Table: Recognition result of MRRC training samples using MAPS binarization, Block DCT features and nearest neighbor classifier.

Word image	Recognized word (English transliteration)	Word image	Recognized word (English transliteration)
	೨೦೦೯ (2009)		ಸಂಗಮ (sangama)
	ಬೆಳಗಾವಿ (belagaavi)		ಹೋಗಿ (hoogi)
	ಮೊಬೈಲ್ (mobile)		ಬ್ಯಾಂಕ್ (bank)

Kannada word recognized

Word image	Recognized word (English transliteration)	Word image	Recognized word (English transliteration)
	ಸ್ಥಳೀಯ (sthaliya)		ನಮ್ಮ (namma)
	ಬಗರಿ (bagari)		ಭರಣಿ (bharani)
	ಮಹೋಗನಿ (mahogani)		ಮಾರ್ಗ (maarga)

Conclusion - Character and Word recognition

- The classification accuracy obtained indicates that using DCT as features from binarized image, results in higher recognition rates than the other methods.
- The word recognition rate for Kannada words is less, but with a custom lexicon, it is improved by 20%.
- A supervised classifier can replace the existing nearest neighbor classification approach.
- Training a classifier with partial characters may overcome degradations in scene images.

Video demonstration

Kannada word recognition video.

Section V

Conclusion and Future work

Conclusion

- Text segmentation, text localization, word recognition and character recognition problems in a scene image are addressed by analyzing camera-captured and born-digital images.
- Text localization and segmentation deal with a scene or born-digital image, whereas word and character recognition deal with cropped image obtained from a scene or born-digital image.

Contribution

- A system to segment born-digital images with minimum spanning tree modules to determine words and non-words in the image.
- Non-linear enhancement on gray scale to improve segmentation and further recognition on the data set.
- A middle line analysis for self-training on image itself. Thus, the segmentation is dependent on the middle line property in the image.
- Analysis on the number of class labels with different feature descriptors for the recognition of Kannada characters.

Future work

- Image segmentation is approached by splitting the colour channels of an image. If the pixels have values close to gray scale, then only one channel is sufficient for text segmentation.
- Images affected by slowly varying and strong illumination are not segmented properly. Illumination in the image has to be analyzed while segmenting the image.
- Word images have variation in the stroke width of characters. Incorporating uniformity in the character stroke width during segmentation may further improve the recognition rates.

Future work

- Multiple rows can be used as input to the sub-image segmentation stage, to avoid dependency on middle line or by including the colour information at classification stage in MAPS method.
- Only lexicon was provided, to increase word recognition rate, in the top-down approach. If a classifier is built based on partial character, then characters affected due to the degradations such as occlusion and strong illumination can be detected and processed for recognition.
- Skew, slant and curved words in an image can be aligned horizontally and passed to OCR for recognition. One-third of English word test set of Multi-script Robust Reading Competition has skew, slant and curved words.

Acknowledgements

I express my thanks to the members of MILE Laboratory.

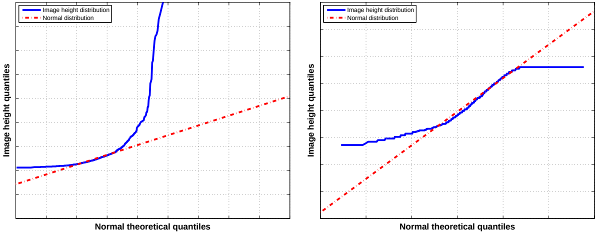
Thank you

Preprocessing of word images

- Rule 1: If the height of an image is less than 60 pixels, then it is rescaled by a factor of '3'.
- Rule 2: If the height of an image lies between 60 and 180 pixels, then it is not rescaled.
- Rule 3: If the height of an image exceeds 180 pixels, then it is scaled to a height of '180' pixels.

Q-Q plots of image heights for ICDAR 2003 data set

Figure: The actual image height quantiles is plotted against the normal theoretical quantiles. (i) Before height normalization of images. (ii) After height normalization of images.



(i) (ii)

Detection of text polarity

- Is the ratio of number of white pixels along the boundary of segmented word image to the length of boundary greater than 0.5?
- Is the ratio of number of white pixels on the vertical sides of the segmented word image to the total length of the side walls greater than 0.5?
- Is the ratio of maximum widths of ‘white‘ to ‘black’ connected components in the segmented word image greater than 1?

Discrete Cosine Transform (DCT)

$$C(u, v) = \alpha(u)\alpha(v)$$
$$\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2M}\right]$$

for $0 \leq u \leq (N-1)$,$0 \leq v \leq (M-1)$ and

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{P}} & k = 0 \\ \sqrt{\frac{2}{P}} & otherwise \end{cases}$$

where, $k = u,\ v$ and $P = N \; or \; M$.