
BR Mahadeshwara

Technologies for Indian Languages

A G Ramakrishnan
Medical Intelligence & Language Engineering (MILE) Lab
Chairman, Department of Elec. Engg, IISc, Bangalore.

Research in MILE Lab –
Solving real problems of real people

Research is identifying the real challenges

- Challenge in medical image analysis is segmentation.
- The challenge in speech recognition is not recognition of words from a vocabulary.
- In optical character recognition and online handwriting recognition, the ***challenge is NOT recognition or machine learning***; it is segmentation.
- With noisy and old documents, we get cuts (breaks) & merges & non-character components (noise of various kinds). PR systems are not designed to handle unseen classes. No research on classifiers can solve this.
- ➔ We need creative approaches.

What we have worked on/ future

- Indic Language Reading Machines for People with Visual Disability (PWVD) - “Automated Book Reader”
- Any printed material in Indian languages becomes accessible ➔ document analysis & recognition.
- Text to speech (TTS) conversion.
- Need to deal with bilingual & trilingual text ➔ script recognition at the word level.
- Posters, road signs, menu card, notice boards ➔ Camera based document analysis & recognition.
- Coloured text printed on complex background.
- Online Handwriting Recognition (OHWR)
- Machine Listening

OCR for Tamil & Kannada

Mozhi Vallaan & Lipi Gnani:

Recognition of Printed Text

– a quick overview.

Features of our OCR

Supports multiple image formats.
Outputs text in international METS/ALTO and MILE XML formats.
Handles fonts of different shape, size, style.
Accuracy over **94% on 5000 pages printed from 1950 to 2002.**
Braille and RTF output can be obtained. Soon to come: DAISY books (PrintToBraille tool or ShriVeRa GUI).
Pages of any layout can be digitized using block selection feature.
Integrated Unicode string search option.
Graphical Interface to edit the text output, with auto save feature.
Select and remove images before OCR.
Special features to OCR whole books, including multicolumn ones.
Covers pre-1960 Tamil Letters.

Developed by many students and staff, mostly non-Tamil !

Layout Analysis Results – contd.

20th CRS

The 20th Component Repair Squadron provides intermediate repair support for the avionics and propulsion systems installed in the wing's aircraft. The Accessories Branch performs a variety of on and off aircraft maintenance support ranging from hydraulics, electrical systems, and aircraft mechanical systems, to structural repair, welding, machine shop support, and a non-destructive inspection (NDI) laboratory.

The CRS Precision Measurement Equipment Laboratory (PMEL) provides repair and calibration support for all US forces in the southern half of the United Kingdom, and some RAF installations. CRS is also responsible for management of the F-111 Aircrew Training Device or flight simulator. This device provides for more rapid upgrading and proficiency training of aircrews in navigation, instrument flight, and emergency procedures for the F-111.

20th Supply Sq.

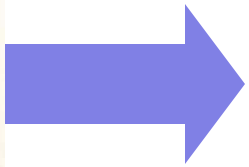
The 20th Supply Squadron provides supplies, equipment, and fuel support for the 20th Tactical Fighter Squadron and assigned tenant units. The squadron also supports RAF Croughton, Barford St. John, Greenham Common, Welford High Wycombe Air Station, Navy London and the Department of Defense Schools. Because of the day/night flying commitment of the F-111E, supply must meet the needs for continuous maintenance of the aircraft and the major communication links between Europe, the states and the U.K. They are a test unit for new computer programs, testing them before release to the rest of the Air Force. They operate under the standard base supply system using dual-system Univac 1050-II system.

20th CSG

The 20th Combat Support Group operates RAF Upper Heyford, Croughton and Barford St. John in direct support of the 20th TFW and its assigned and tenant units. The combat support group commander supervises the functions of Personnel; Chaplain; Administration; Morale, Welfare and Recreation; Civil Engineering; Security Police; Dependent Schools; Operations and Training; Disaster Preparedness; Services; Family Support Center and Base Headquarters Squadron.

F-111s on ramp at Aviano AS, Italy during a deployment from the 20th Tactical Fighter Wing.

UH-8



20th CRS

The 20th Component Repair Squadron provides intermediate repair support for the avionics and propulsion systems installed in the wing's aircraft. The Accessories Branch performs a variety of on and off aircraft maintenance support ranging from hydraulics, electrical systems, and aircraft mechanical systems, to structural repair, welding, machine shop support, and a non-destructive inspection (NDI) laboratory.

The CRS Precision Measurement Equipment Laboratory (PMEL) provides repair and calibration support for all US forces in the southern half of the United Kingdom, and some RAF installations. CRS is also responsible for management of the F-111 Aircrew Training Device or flight simulator. This device provides for more rapid upgrading and proficiency training of aircrews in navigation, instrument flight, and emergency procedures for the F-111.

20th Supply Sq.

The 20th Supply Squadron provides supplies, equipment, and fuel support for the 20th Tactical Fighter Squadron and assigned tenant units. The squadron also supports RAF Croughton, Barford St. John, Greenham Common, Welford High Wycombe Air Station, Navy London and the Department of Defense Schools. Because of the day/night flying commitment of the F-111E, supply must meet the needs for continuous maintenance of the aircraft and the major communication links between Europe, the states and the U.K. They are a test unit for new computer programs, testing them before release to the rest of the Air Force. They operate under the standard base supply system using dual-system Univac 1050-II system.

20th CSG

The 20th Combat Support Group operates RAF Upper Heyford, Croughton and Barford St. John in direct support of the 20th TFW and its assigned and tenant units. The combat support group commander supervises the functions of Personnel; Chaplain; Administration; Morale, Welfare and Recreation; Civil Engineering; Security Police; Dependent Schools; Operations and Training; Disaster Preparedness; Services; Family Support Center and Base Headquarters Squadron.

UH-8

F-111s on ramp at Aviano AS, Italy during a deployment from the 20th Tactical Fighter Wing.

Good segmentation is the key

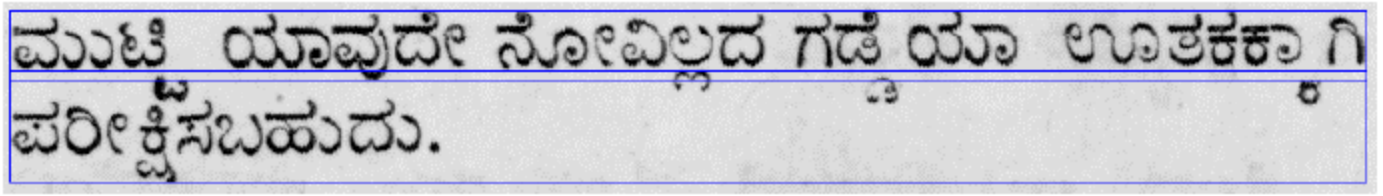
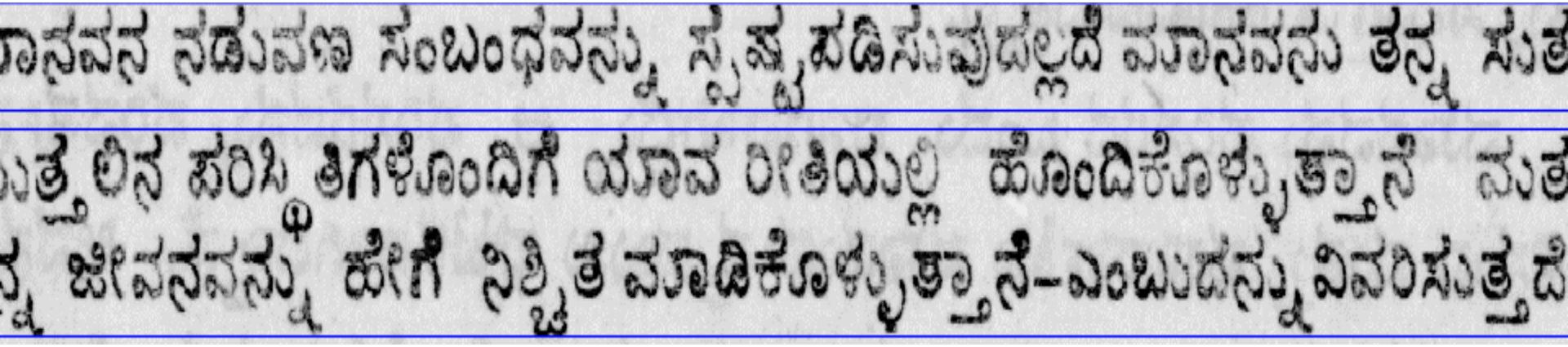
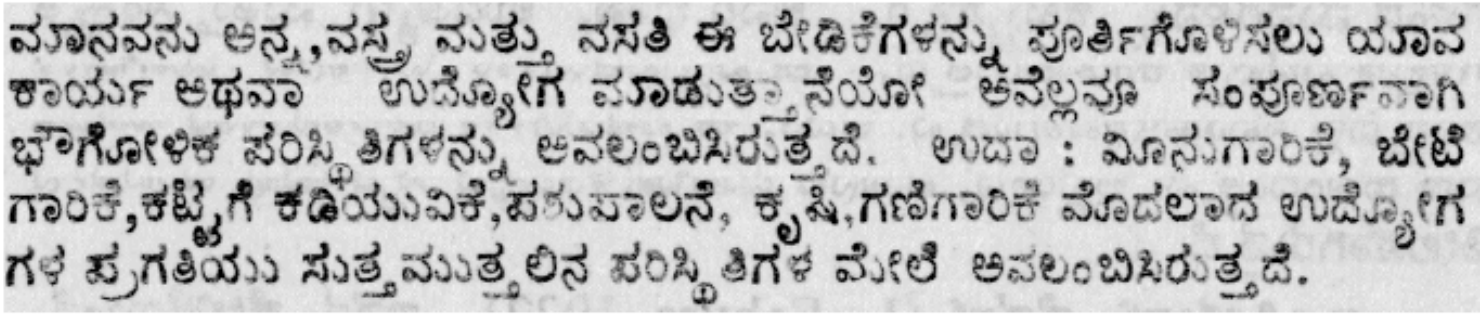


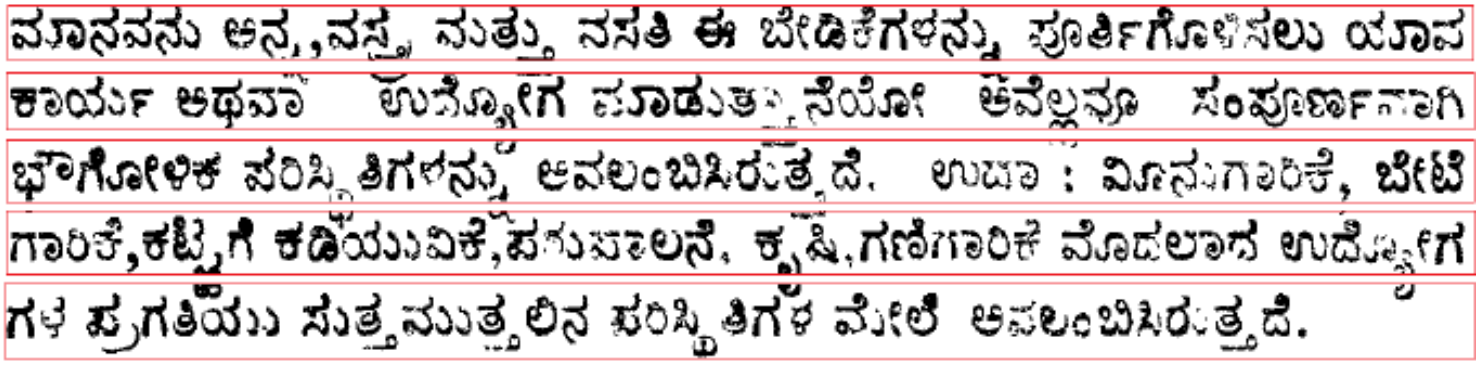
Fig. 6: Input image to illustrate the need for ordering of CCs.



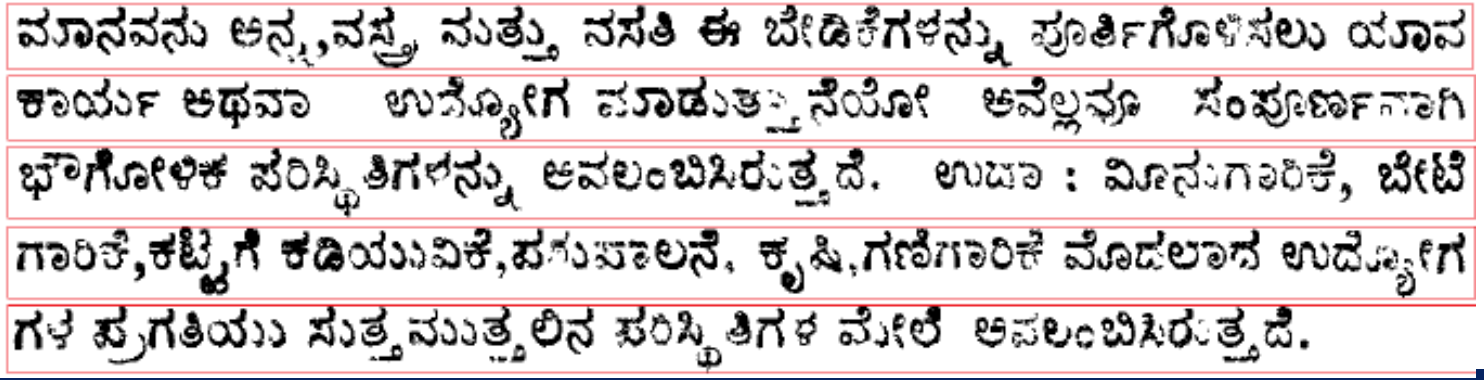
Interval tree based robust segmentation



(a) A section of a Kannada printed page, with overlapping text lines.



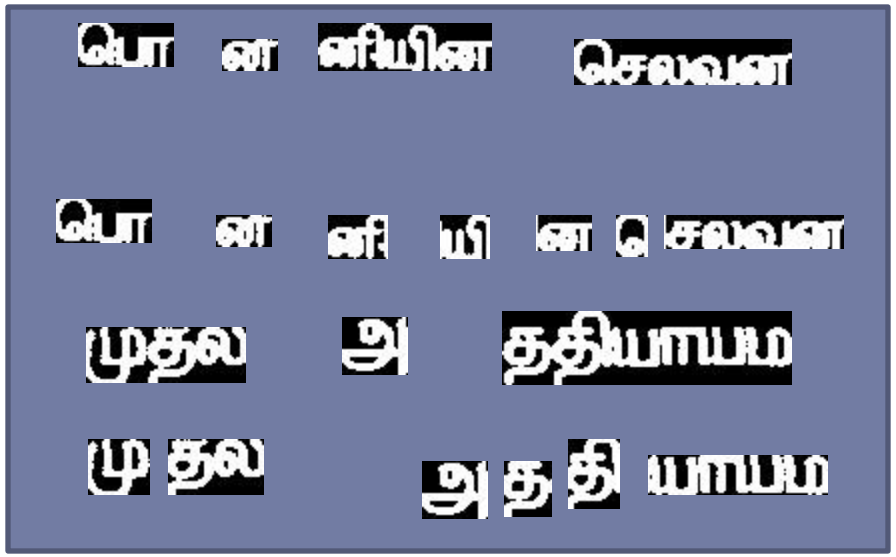
(b) Best segmentation result using Gaussian smoothed horizontal projection profile.

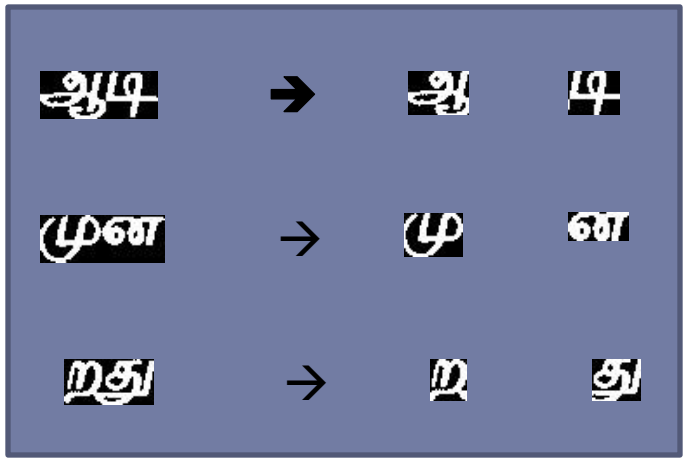


Principal Challenge in OCR

- **Challenge** is dealing with ***broken and merged characters*** in very old books (1950 – 2000).
- Good segmentation of recognizable components is the important step that can boost performance.
- In most image processing problems, ***segmentation is the key issue.***
- ***Same issues in online handwriting, where different letters may overlap or be written with many strokes.***
- ***We have proposed "attention-feedback" strategies for obtaining high performance segmentation (neuro inspired).***

Detection and Segmentation of merged components





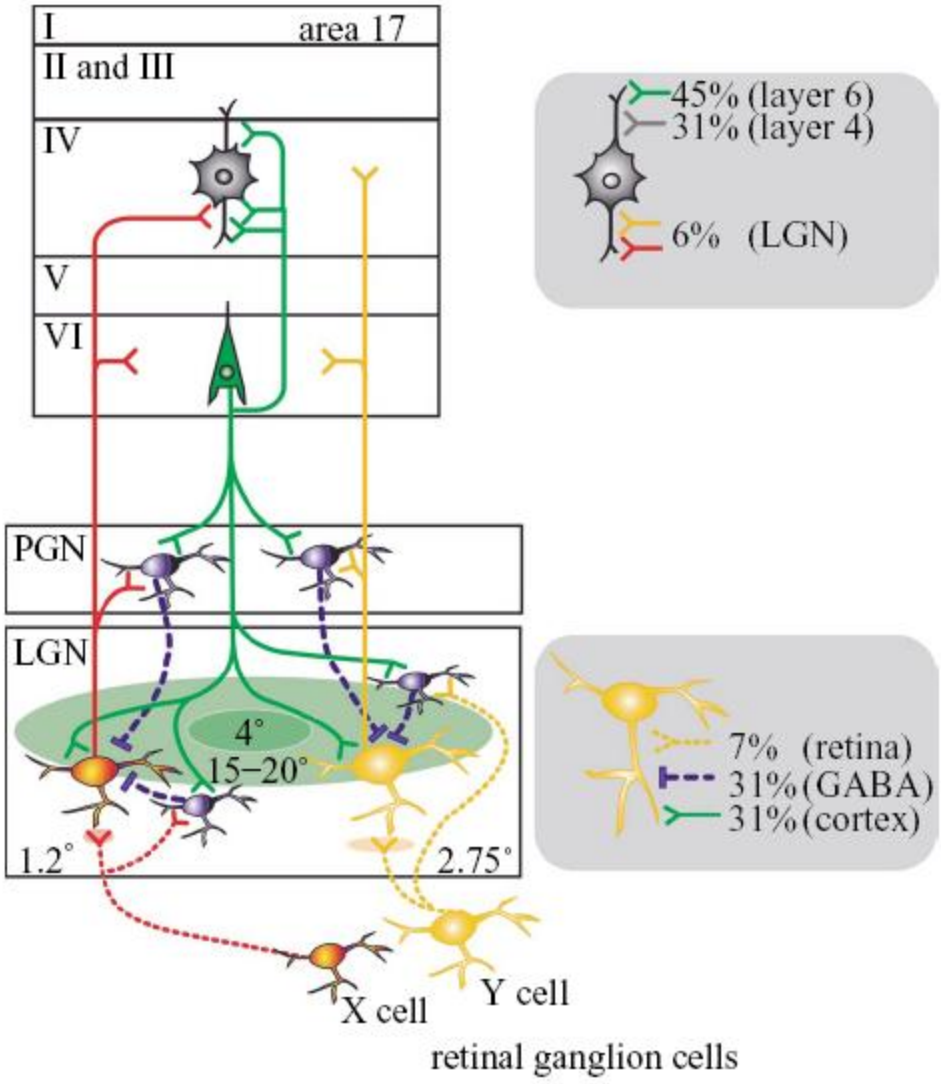
In old books, many characters get merged. In some cases, multiple number of characters or a whole word gets merged.

Attention on script specific information ***(eg. aspect ratio)*** and ***feedback*** from the classifier ***(eg. Confidence level)*** are used to correctly segment these merged components.

Shiva Kumar HR

Patent filed #

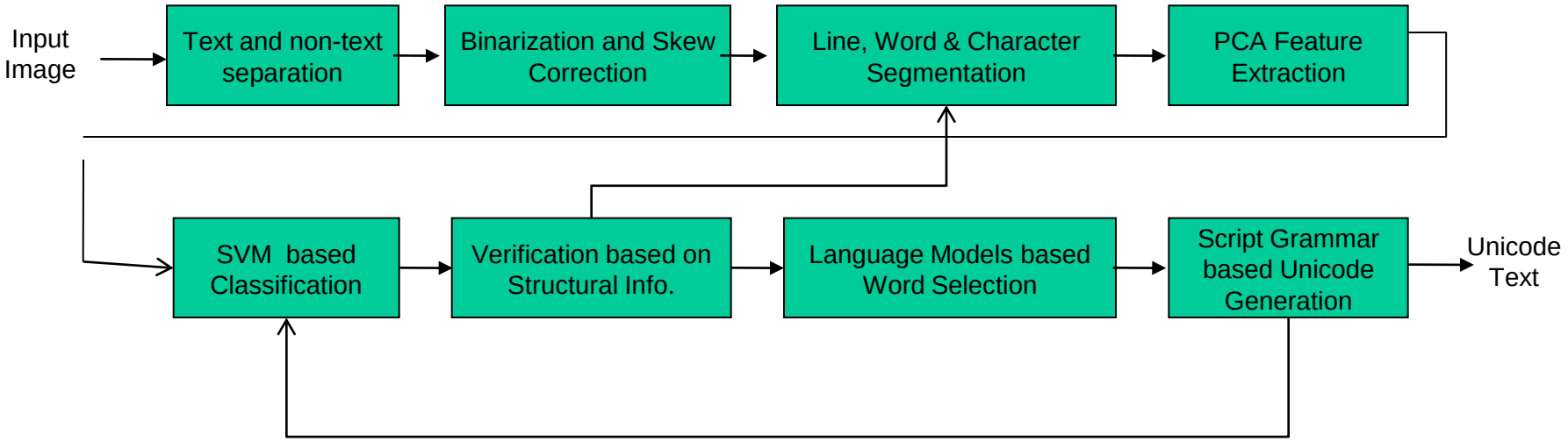
Rich Feedback in the ascending visual pathway



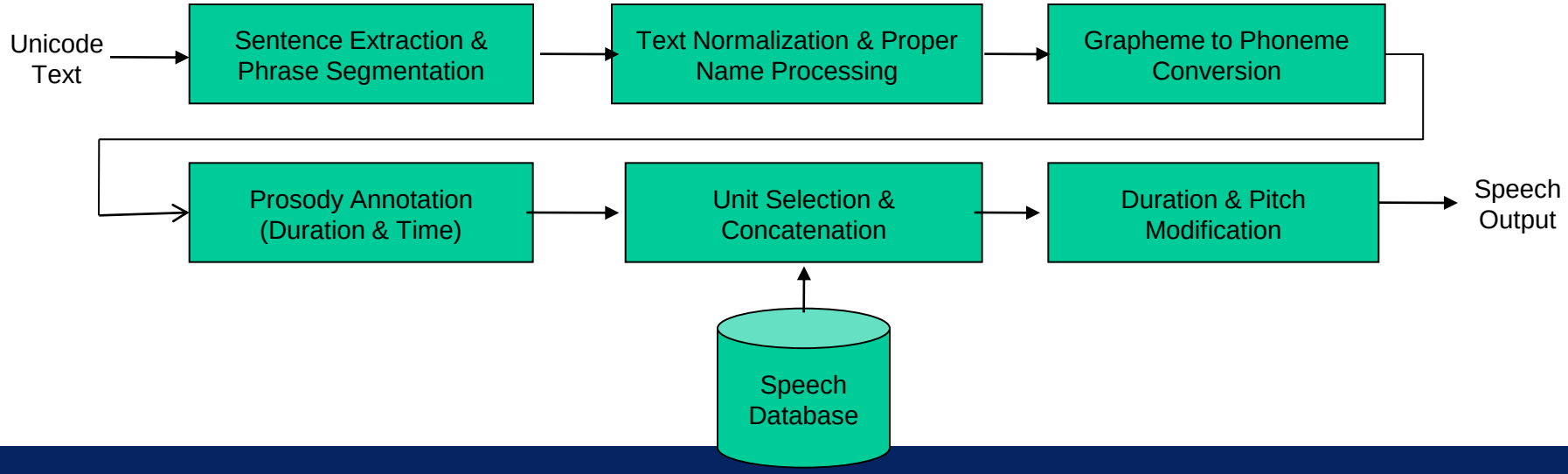
- Both excitatory and inhibitory feedback connections exist, and thus have a profound influence on the response mode of LGN relay cells.

Ref: Adam L Sillito and Helen E Jones, "Corticothalamic interactions in the transfer of visual information," Phil. Trans of the Royal Society – Biology, 2002, 357, 1739-1752.

Optical Character Recognizer (OCR) :



Text To Speech (TTS) Synthesis:



Segmentation of merged characters

Test results on a Tamil book with 258 pages

Number of merged images	Segmentation path selection algorithm	Character segmentation accuracy	Character recognition accuracy	Overall character recognition accuracy of the book
3348	None	NA	0%	73.4%
	VPP based selection	87.6%	82.6%	84.8%
	Valley matching based selection	92.3%	90.1%	86.1%

Segmentation of merged characters

Test results on a Kannada book with 80 pages

Number of merged images	Segmentation path selection algorithm	Character segmentation accuracy	Character recognition accuracy	Overall character recognition accuracy of the book
3348	None	NA	0%	82.2%
	VPP based selection	84.3%	83.4%	83.9%
	Valley matching based selection	89.6%	91.2%	84.6%

OCR Performance – CDAC Noida
Test Report - 2014

Script	Tested Pages from material printed from 1950 – 2002	Substitution Error Rate	Overall Accuracy
Tamil	4902	3.6%	93.9%

Recognition accuracy is about 7 to 8% higher than Google's Tesseract OCR – Tamil & Kannada.

Deployment of Mozhi Vallaan

- Using our Tamil OCR, Mozhi Vallaan, RCMCT Worth Trust, Chennai has already digitized over 600 Tamil books (~ 50,000 pages) and the Braille books are already being used by hundreds of PWVD.
- Pondicherry University library, Tamil Virtual Academy, Ramakrishna Math, Mitrajyothi and Canara Bank Braille Transcription Centres, Samskriti Foundation Mysore, Parankushachar Institute of Vedic Studies, Bangalore and individual blind students are other users of our OCR.

South East Asia & Asia Pacific

mile
MEDICAL INTELLIGENCE AND
LANGUAGE ENGINEERING LAB

GYTI Award 2015

Handwriting Recognition

Recognition of Online Handwritten text – a quick overview.

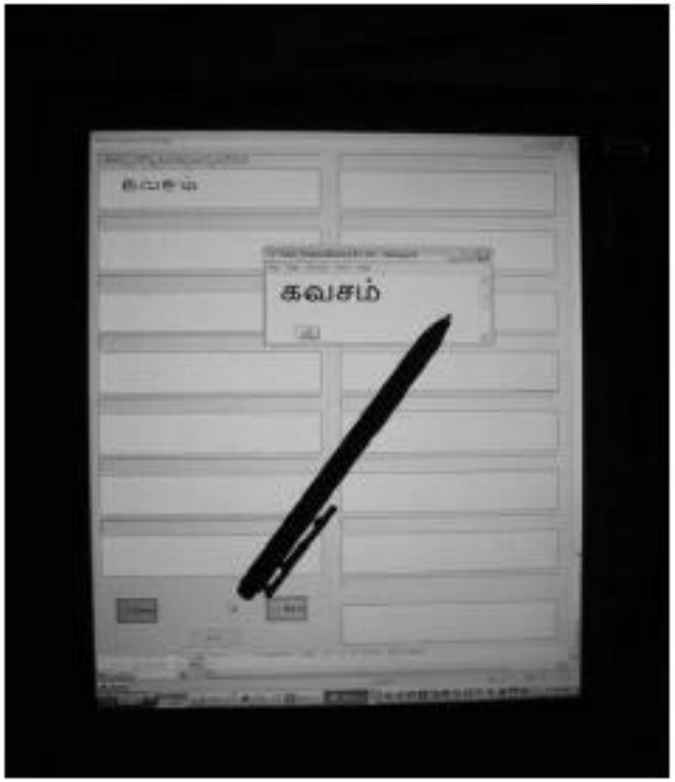
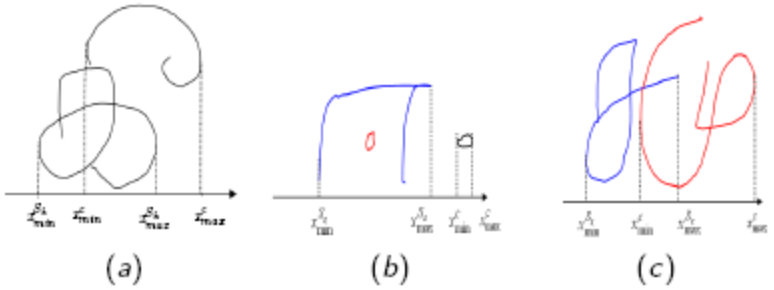


Fig: Picture of the tablet PC with the stylus used to record the handwritten data

Segmentation Issues



(a) (b) (c)

Fig: Parameters employed for computing the overlap in the DOCS scheme.

Detection of under-segmented stroke groups with feature attention

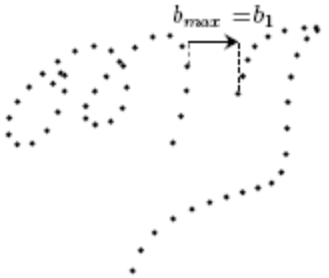


Fig: Distinct symbols wrongly merged to a stroke group by DOCS.

Stroke groups wrongly segmented with the DOCS

(a) (b)

AFS applied for patent

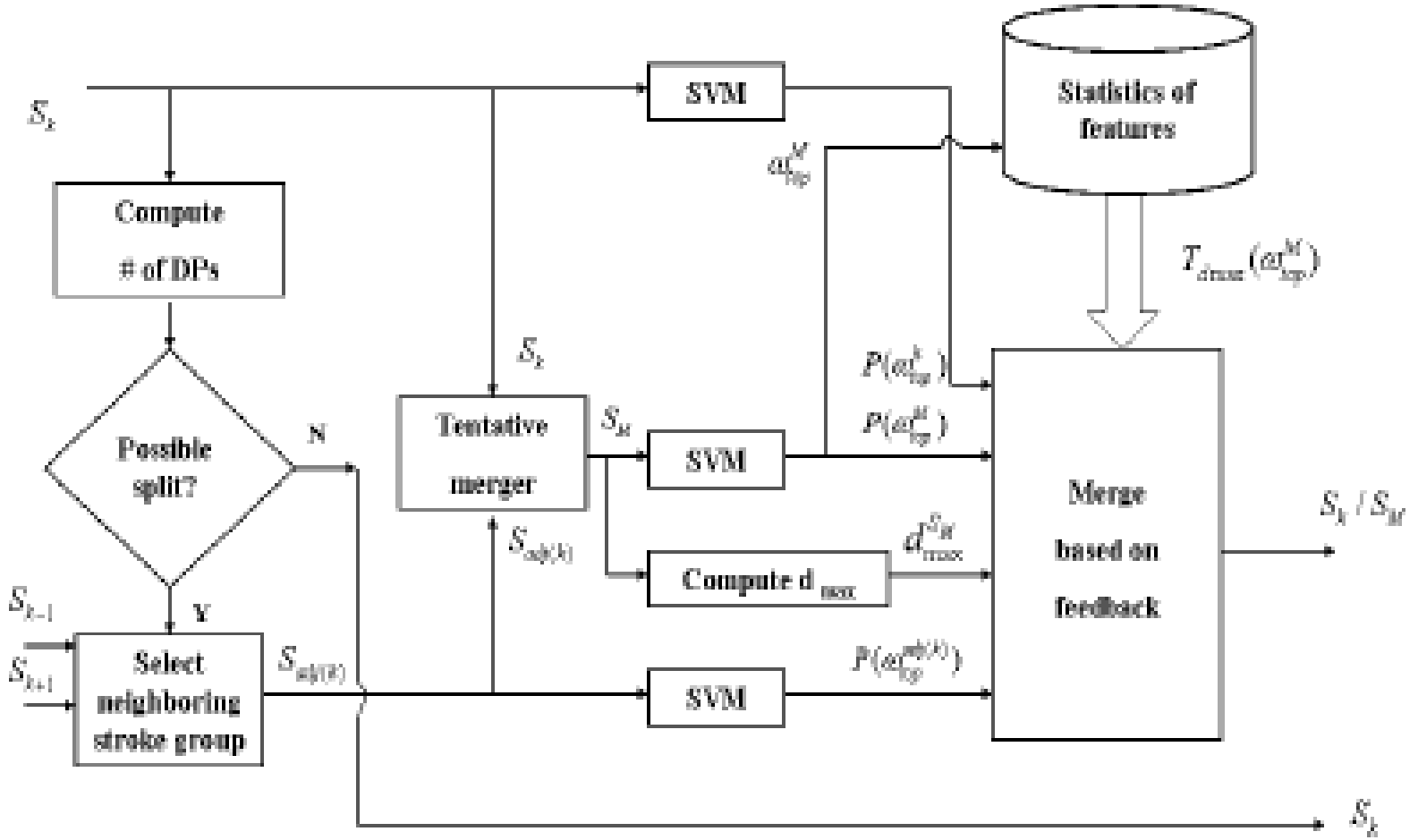
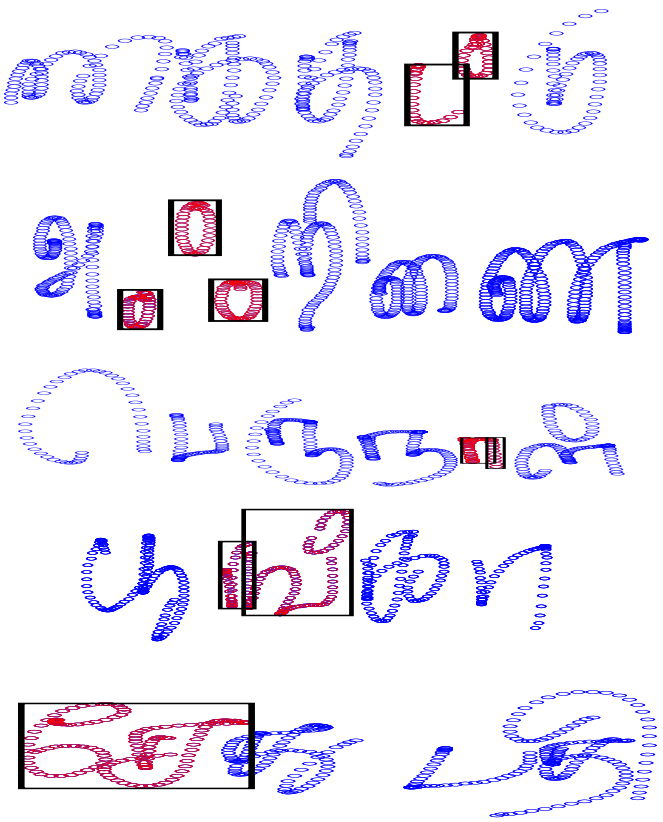


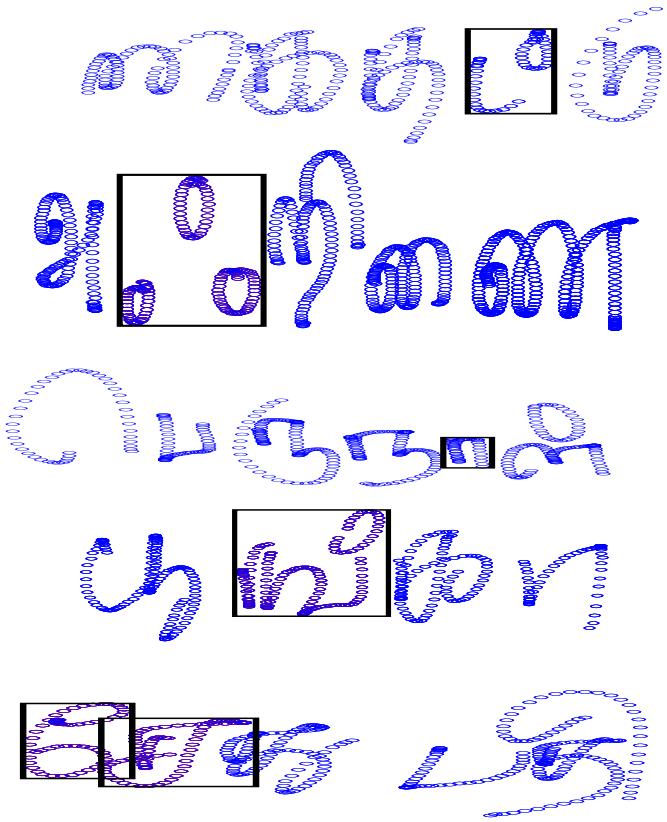
Fig: AFS module for resolving over-segmented stroke groups.

Results of Attention-Feedback segmentation

Segmentation by Overlap Criterion

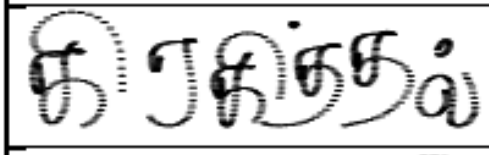
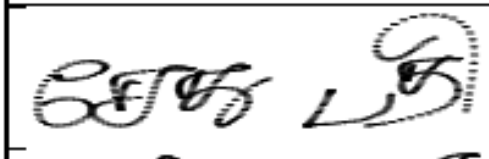
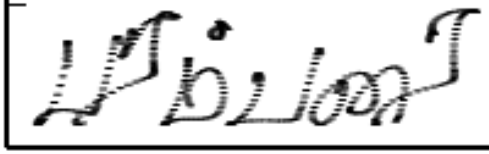


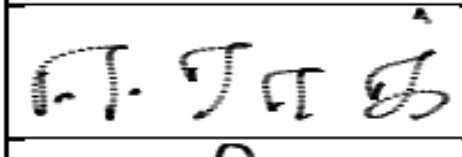
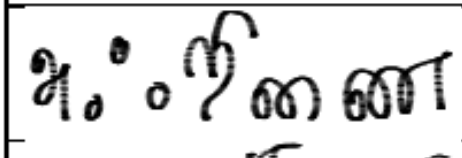
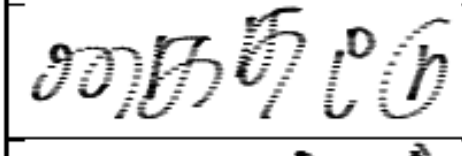
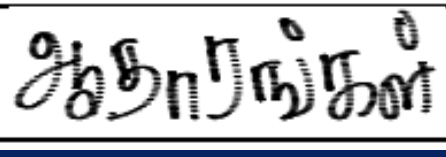
Word Segmented after Attention feedback



Feedback to segmenter (**LGN**) from feature extractor (intermediate junctions) and/or classifier (*visual cortex*)

Attention – Feedback Segmentation

Input Word	OCS o/p	AFS o/p
	கிரஓதல்	கிரகித்தல்
	சஷ்துபதி	சேதுபதி
	ஹுபங்	பரம்பரை

Input Word	OCS o/p	AFS o/p
	ஈஃஃராக்	ஈராக்
	அப்ய்ட்றிணை	அஃறிணை
	கைதடபடு	கைதட்டு
	ஆதபீநூரங்கள்	ஆதாரங்கள்

Effectiveness of AFS

Impact of the AFS scheme on the segmentation and recognition of symbols in the MILE Word database.

Total Words	Total No. of characters	Recognition of properly segmented letters	Overall correctly recognized letters
45, 405	2,53,095	95.9%	87.3%

Text to Speech Synthesis (TTS) – a quick overview: Madhura Vaachaka & Thirukkural

Uniqueness

- **Produces different instances of speech for the same text, to simulate naturalness**.
- Detects the language and switches the grapheme to phoneme conversion, as well as text normalization.
- SAPI interface plugs into any screen reader**.**
- ***Can handle Tamil, Kannada and English (external TTS)***

live mint
28 November 2015 | E-Paper

Last Modified: Fri, Nov 13 2015. 12 50 AM IST

Voice for the blind

Text to speech engine Madhura can read out e-text in two regional languages—Kannada and Tamil

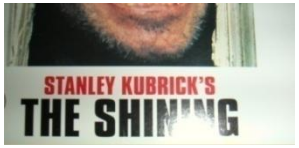
Arundhati Ramanathan

Mobile reading of unknown languages

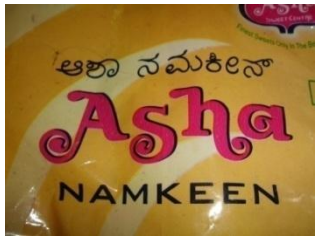
- Camera based document analysis and recognition
- Text extraction from scene images
- Segmentation of coloured scene word images
- Recognition of the segmented words
- Translation of the words into the target language
- Text to speech conversion of the words
- ICDAR 2013 Competition Results
- Computationally very efficient algorithms

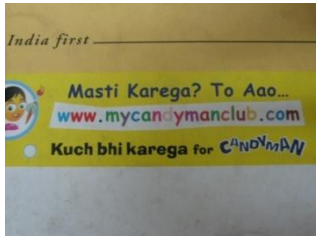
Challenges of CCDIA

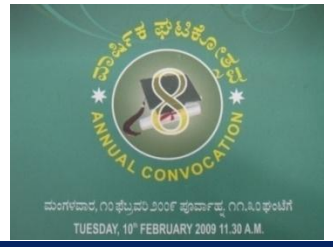
Text in camera-based images originating from text on real world objects

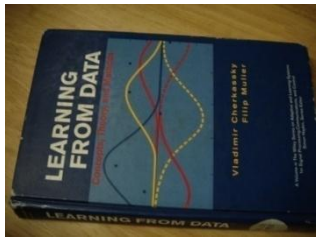


It may also have arbitrary text layout and multi-script content









Results on scene text extraction

$(p = 0.98, r = 0.96)$

$(p = 0.79, r = 0.94)$

$(p = 0.97, r = 0.91)$

$(p = 0.69, r = 0.91)$

$(p = 0.95, r = 0.93)$

$(p = 0.86, r = 0.94)$

$(p = 0.91, r = 0.84)$

$(p = 0.98, r = 0.95)$

$(p = 0.81, r = 0.93)$

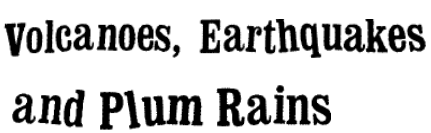
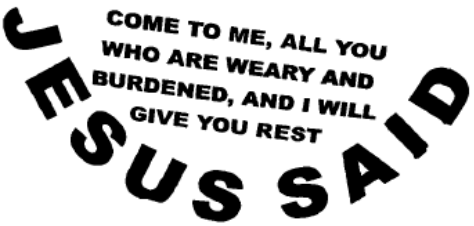
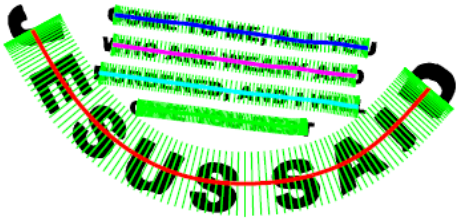
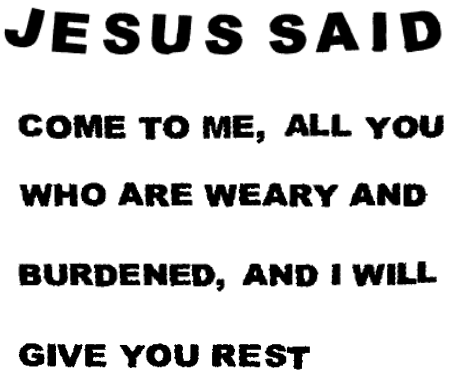
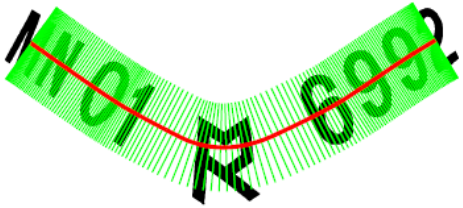
$(p = 0.97, r = 0.91)$

$(p = 0.53, r = 0.9)$

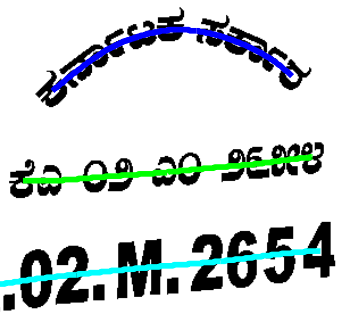
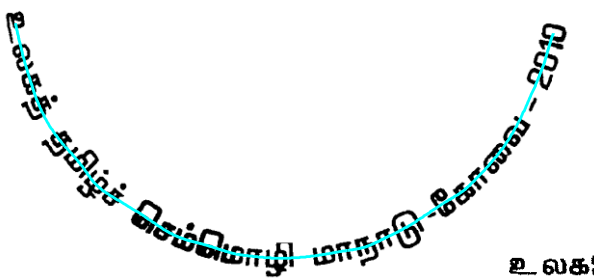
$(p = 0.91, r = 0.95)$

Precision	Recall	f
0.8	0.86	0.83

Recognition by Nuance Omnipage 16

Input image	Identified text strings and estimated normal vectors	Rectified images of individual text string	Corresponding OCR outputs
			Volcanoes, Earthquakes and Plum Rains
			JESUS SAID COME TO ME, ALL YOU WHO ARE WEARY AND BURDENED, AND I WILL GIVE YOU REST
			MN01 i.I 6992

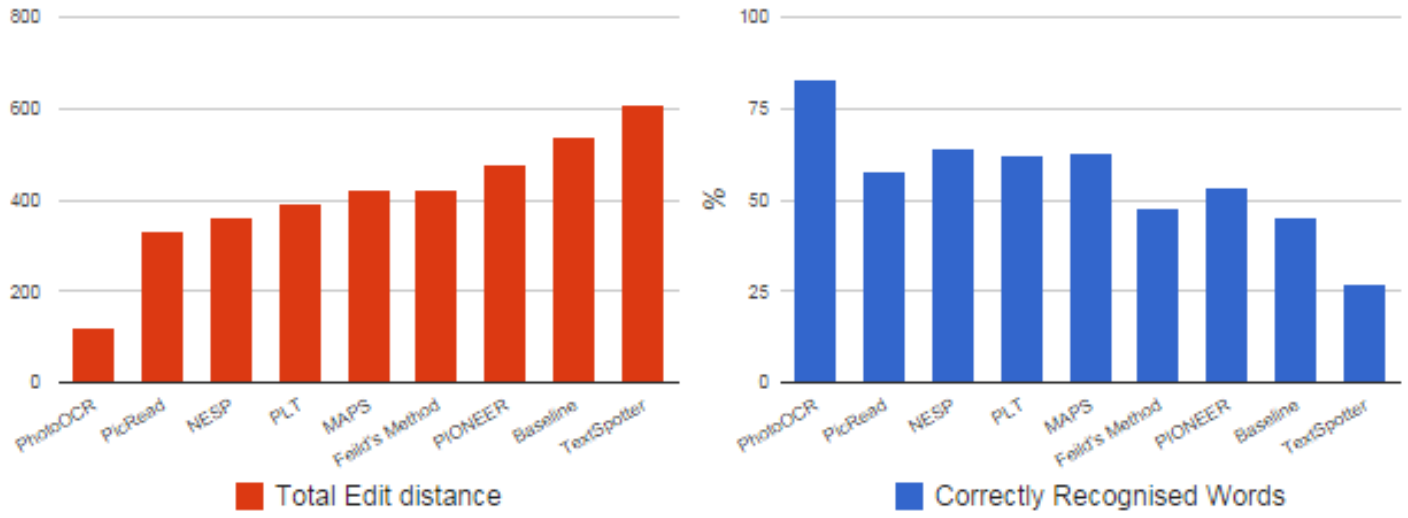
Applicable to Kannada and Tamil curved text

Input image	Extracted text string/s	Rectified text string/s
		
		

ICDAR 2011 – RRC – BDI Task

Ranking for Task 3 - Word Recognition

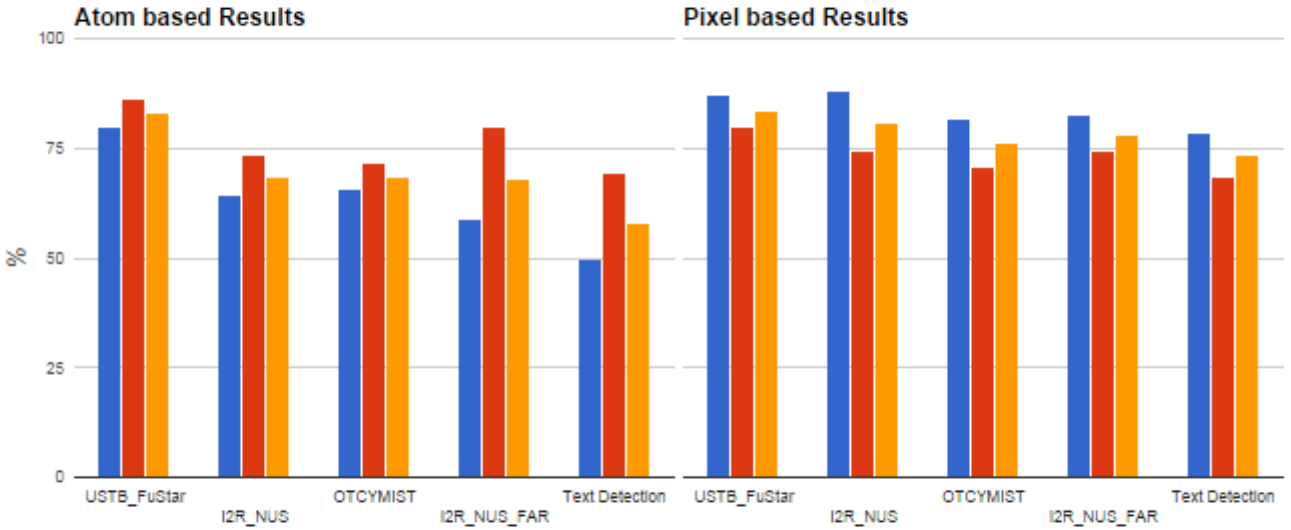
Method	Total Edit distance	Correctly Recognised Words	T.E.D. (upper)	C.R.W. (upper)
PhotoOCR	122.7	82.83 %	109.9	85.3 %
PicRead	332.4	57.99 %	290.8	61.92 %
NESP	360.1	64.2 %	345.2	64.84 %
PLT	392.1	62.37 %	375.3	63.11 %
MAPS	421.8	62.74 %	406	63.29 %
Feild's Method	422.1	47.95 %	390.6	52.33 %
PIONEER	479.8	53.7 %	426.8	55.71 %
Baseline	539	45.3 %	517.9	46.58 %
TextSpotter	606.3	26.85 %	597.3	28.13 %



ICDAR 2013 – RRC – BDI Task

Ranking for Task 2 - Text Segmentation

	Pixel Results			Atom based Results									
Method	Recall	Precison	F-Score	Well s.	Merged	Broken	Br.-Mer.	Lost	False p.	Detected	Recall	Precision	Fscore
USTB_FuStar	87.21 %	79.98 %	**83.44 %**	6258	920	56	1	587	370	7260	80.01 %	86.20 %	**82.99 %**
I2R_NUS	87.95 %	74.40 %	**80.61 %**	5051	1584	30	6	1151	685	6878	64.57 %	73.44 %	**68.72 %**
OTCYMIST	81.82 %	71.00 %	**76.03 %**	5143	1420	34	2	1223	1083	7178	65.75 %	71.65 %	**68.57 %**
I2R_NUS_FAR	82.56 %	74.31 %	**78.22 %**	4619	1474	12	1	1716	156	5771	59.05 %	80.04 %	**67.96 %**
Text Detection	78.68 %	68.63 %	**73.32 %**	3883	2716	36	0	1187	210	5590	49.64 %	69.46 %	**57.90 %**



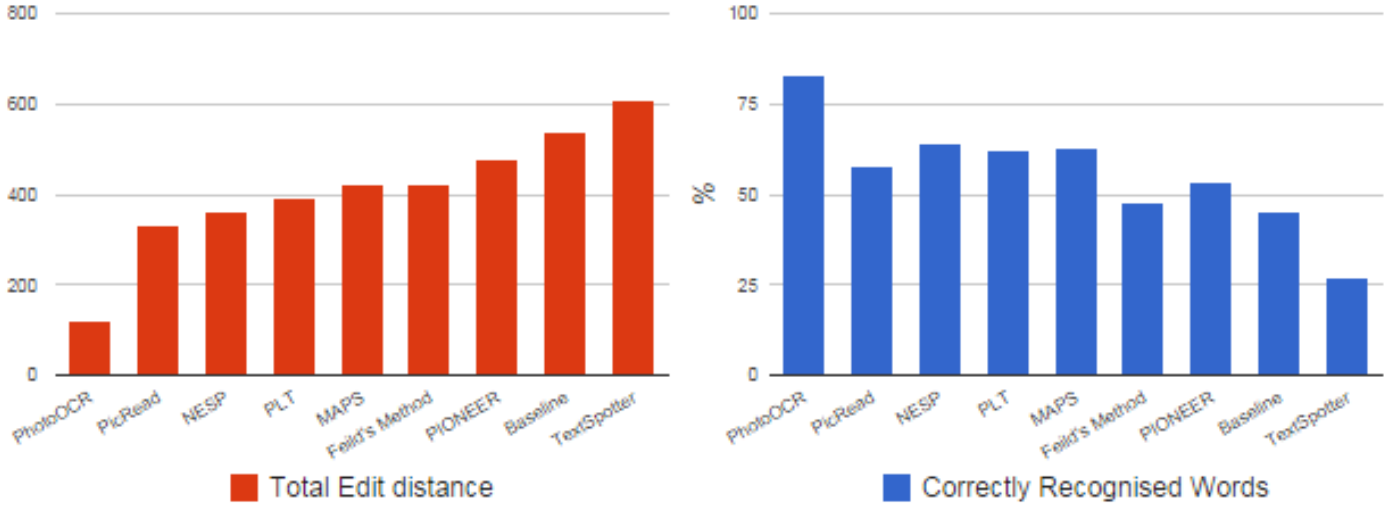
Ranking for Task 3 - Word Recognition

Method	Total Edit distance	Correctly Recognised Words	T.E.D. (upper)	C.R.W. (upper)
PhotoOCR	105.5	82.21 %	88.8	85.41 %
MAPS	196.2	80.4 %	186.4	81.51 %
PLT	200.4	80.26 %	190.9	81.38 %
NESP	214.5	79.29 %	198.2	80.75 %

ICDAR 2013 – RRC – Scene Word

Ranking for Task 3 - Word Recognition

Method	Total Edit distance	Correctly Recognised Words	T.E.D. (upper)	C.R.W. (upper)
PhotoOCR	122.7	82.83 %	109.9	85.3 %
PicRead	332.4	57.99 %	290.8	61.92 %
NESP	360.1	64.2 %	345.2	64.84 %
PLT	392.1	62.37 %	375.3	63.11 %
MAPS	421.8	62.74 %	406	63.29 %
Feild's Method	422.1	47.95 %	390.6	52.33 %
PIONEER	479.8	53.7 %	426.8	55.71 %
Baseline	539	45.3 %	517.9	46.58 %
TextSpotter	606.3	26.85 %	597.3	28.13 %



Machine Listening
– making sense of complex audio.

- **Today's best speech recognition (transcription) systems cannot deal with natural & man-made noise, multiple speakers/ languages or *even* laugh, cough or clearing of throat of the speaker!**

Acknowledgment

- Inspiring teachers and researchers
- Research Students
- ME Project Students
- Project Staff
- BE and ME Interns
- Collaborators
- Funding Agencies
- Science & Technology Policy Makers
- Enlightened Citizens, who pay taxes (Income Tax, Sales Tax, Excise Duty, Customs Duty, etc.)
- My parents, Prof. TMS, Swami Chinmayanada, J K, Werner Erhard, ..
- Each and everyone of you !