and allowing knowledge to be shared [2]. Invoice imaging applications are used in many businesses to keep track of financial records and prevent a backlog of payments from piling up [4]. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes. As the technology continues to develop, more and more applications are found for OCR technology, including increased use of handwriting recognition. Furthermore, other technologies related to OCR, such as barcode recognition, are used daily in retail and other industries.

1.3 Benefits of OCR

1.3.1 No More Retyping

If you lose or accidentally erase an important digital file, such as a proposal or invoice, but still have a hard copy, you can easily replace it in your digital filing system by using OCR software to scan the paper original or most recent draft.

1.3.2 Quick Digital Searches

OCR software converts scanned text into a word processing file, giving you the opportunity to search for specific documents using a keyword or phrase. For example, you could effortlessly search hundreds of invoices and locate a specific name or account in moments, without having to thumb through extensive files.

1.3.3 Save Space

Free up storage space by scanning paper documents and hauling the originals off to storage. You can easily turn a filing cabinet worth of information into editable digital files, and create a backup system consisting of a single CD.

1.3.4 Edit Text

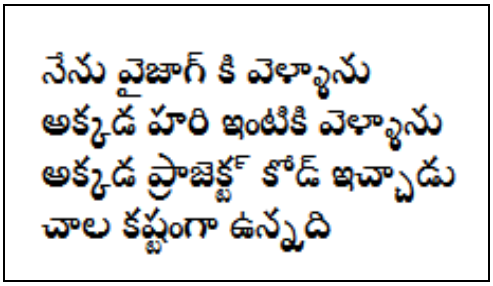
Once you've scanned your document using OCR, you have the option to edit the text within a word processing program of your choice. Scan items that may need to be updated in the future to help expedite the editing process:

- Typed family recipes
- Rental agreements
- Resumes
- Contracts

2. Segmentation Process

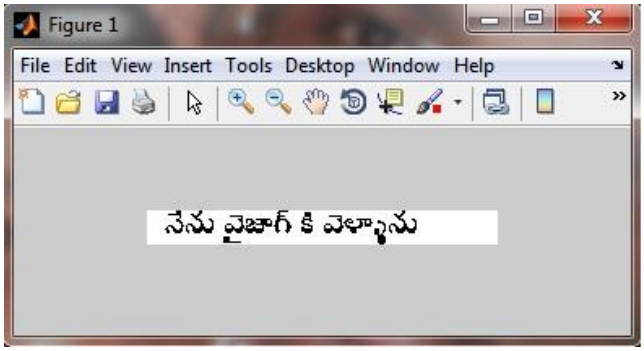
2.1 Line Segmentation

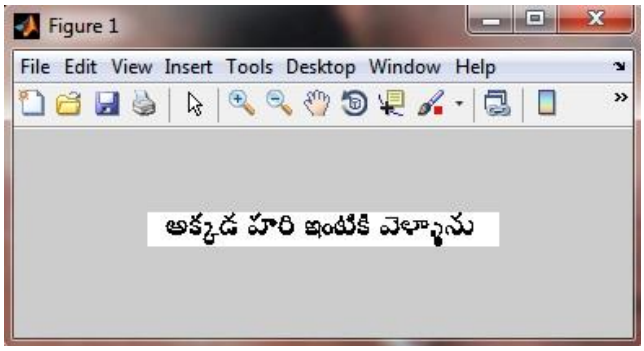
Second step of segmentation process is segmenting the text region into lines, also called as line segmentation. Generally, each text line is separated from the previous and following lines by white spaces. Therefore, the horizontal projection of a document image is the most commonly used technique to extract the lines from the document. If the lines are well separated and not tilted, the horizontal projection will have well separated peaks and valleys. These valleys can be detected easily and used to determine the locations of the line boundaries .shown in below fig2.1

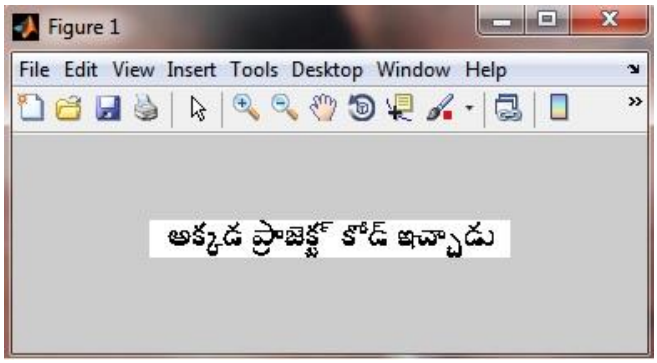


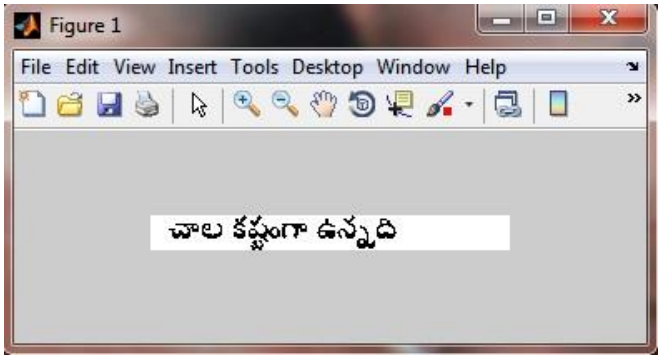
2.1 Input Image

Output Images Of Line Wise Segmentation:



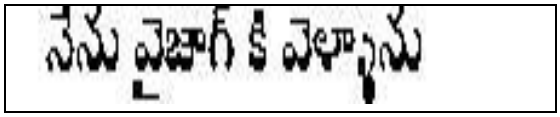






2.2 Word Segmentation

From the extracted text lines, words get separated. Usually, applying vertical projection profile (VPP) and detecting some specific threshold exceeding horizontal gaps, words are separated from a text line. An example is shown in Fig.2.2

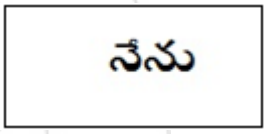


2.2 Input Image

Output Images of Word Wise Segmentation

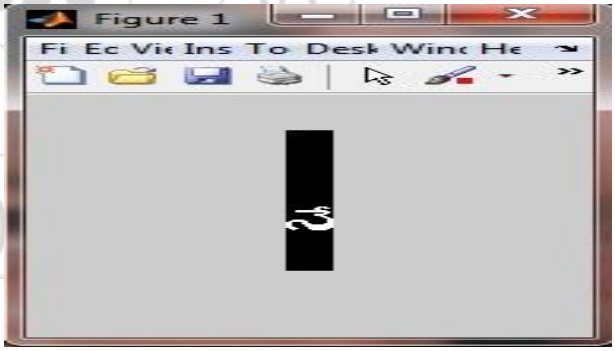
2.3 Character Segmentation

Segmentation of characters from the isolated words is the most challenging part of the script segmentation phase. Since, in computer composed scripts some characters in a container word may partially overlap with one another, it becomes very difficult to isolate those characters properly. Especially the modifiers (both vowels and consonants) most of the time coincide with the modifying characters as shown in Figure 2.3. These kinds of nontrivial combinations of characters make the whole process of character segmentation extremely challenging. Besides, some symbols, like Chandra-Bindu, often come between two consecutive characters in a word; then isolating those becomes a tough job. An example is shown in Figure 2.3.



2.3 Input Image

Output Images Of Character Wise Segmentation:



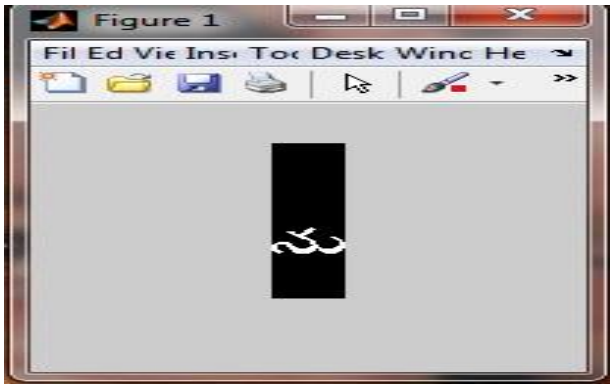


Table 1: Segmentation Table

Sr.No	Line wise	Word wise	Charter
1 (Big text image)	70%	80%	90%
2 (Small text image)	60%	70%	80%

3. Conclusions

In this project we have designed an OCR for the recognition of Latin Printed Document Images. We have conducted experiments to evaluate its performance in which we have got good results on reasonable diverse quality documents. However the performance of the OCR varies with the diversity of the font size and style. For Indian script documented image with sufficient large font the segmentation accuracy is more than 90% and for small font size, the segmentation accuracy declines and will be in the range of 80% to 9

References

[1] S. Mori, C. Y. Suen and K. Yamamoto, "Historical review of OCR research and development", *Proceedings of the IEEE*, Vol. 80(7), pp. 1029-1058, 1992.
[2] U. Pal and B. B. Chaudhuri, "Indian script character recognition: a survey", *Pattern Recognition*, Vol. 37(9), pp. 1887-1899, 2004
[3] R.G.Casey and E. Lecolinet, "A survey of methods and strategies inCharacter segmentation", *IEEE Transactions on PAMI*, Vol. 18(7), pp. 690-706, 1996
[4] C. E. Dunn and P. S. P. Wang, "Character segmentation techniques for handwritten text - a survey", in the Proceedings of 11th ICPR, Vol. 2, pp. 577-580, 1992

References

B. Harikumar completed his M.Tech in Electronics and Communications Engineering in Aurora's Scientific Tech & Research Academy, Hyderabad from 2014. Presently he is working Assistant professor (ECE) Wellfare Institute of science Technology & Management Pinagadi Village, Near Pendurthi, Andhra Pradesh He has 3 years Exp Teaching. His area of interests is ImageProcessing, Optical Character Recognition

N. Sateesh completed his M.tech in "Digital systems signal processing" from Gitam University,Vizag. Presently he is working Assistant professor (ECE) Wellfare Institute of science Technology & Management Pinagadi Village, Near Pendurthi, Andhra Pradesh He has 2 years Exp Teaching. His area of interests is Digital systems& VLSI