II. REVIEW OF RELATED WORK

In this section, we give a brief review of prior research

on the problem of table detection in scanned images that

contain at least some ruling lines. One of the earliest works

on identifying tabular regions in document images is the

method proposed by Watanabe et al. [1]. The method identifies

individual item blocks enclosed by vertical and horizontal line

segments on the basis of the interpretation of detected lines.

Firstly, line segments are detected and the corner points are

subsequently determined. The connective relationships among

the extracted corner points and hence the individual item

blocks are interpreted using global and local tree structures.

Laurentini and Viada [2] proposed a method to detect tables

where text and lines are horizontal or vertical. Text regions

are identified using a bottom-up approach and the detected

characters are grouped into words and subsequently phrases.

Based on threshold on the horizontal and vertical run length,

lines are obtained. The arrangement of these detected lines

is compared with that of the text blocks in the same area.

Further, using the horizontal and vertical projection profiles,

the algorithm attempts to add missing horizonal and vertical

lines in order to fully understand the table structure.

Green and Krishnamoorthy [3] proposed a model-based top-

down approach for table analysis by a hierarchical character-

ization of the physical cells. Horizontal lines, vertical lines,

horizontal space and vertical space are used as features to

extract the table region. Elementary cell characterization is

performed to label individual cells in such a way that the cells

belonging to an underlying nesting or overlapping of logical

units can be properly extracted. These raw labels are matched

to a table model such that the relational information in the

table can be extracted.

Cesarini et al. [4] present a system for locating table regions

by detecting parallel lines. They use a recursive analysis of the

modified X-Y tree of a page to identify regions surrounded

by horizontal (vertical) lines. The search is refined by looking

for further parallel lines that can be found in deeper levels of

the tree. The hypothesis that a region corresponds to a table is

verified by the presence of vertical (horizontal) lines or spaces

in the regions included between the two parallel lines. After

the complete tree analysis, sub-tables belonging to one table

are merged while tables smaller than a given threshold were

discarded. The method requires that at least two parallel lines

are present.

CONCLUSION AND FUTURE SCOPE

We have presented a system which can be used to scan a table from a magazine, a book or a document in seconds and can have digital copy of it with him forever and if you have a series of complicated tables as part of a larger image OTG will be extremely useful to be able to extract them with a scanner and edit them on screen. With further advancement it can be even more accurate than human eye. Future Scope for OTG is very vast. Our future work is to extend the method to handle tables without border lines too using additional cues from the image such as the arrangement of text and white spaces. We will also try to implement our system on a different language such as Hindi.