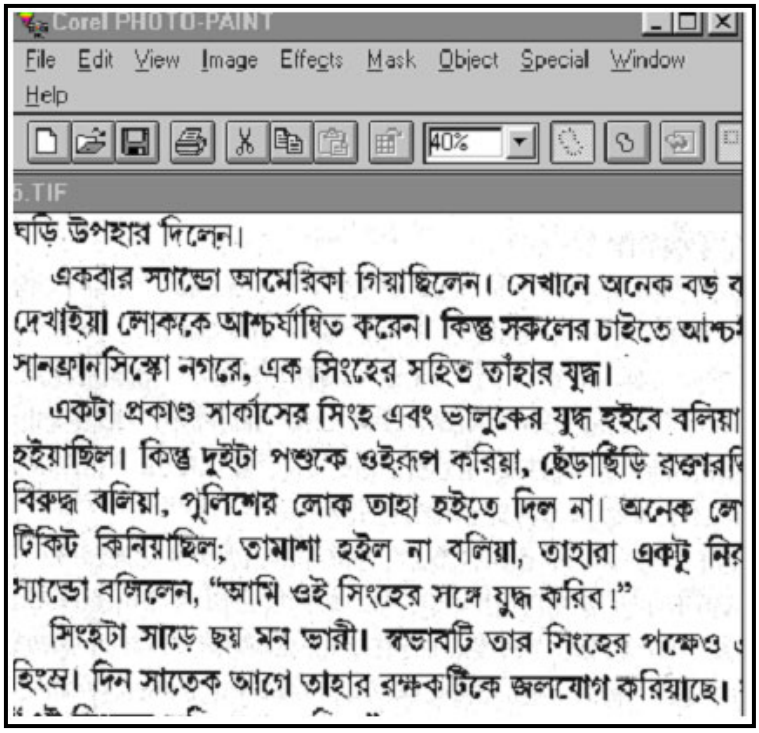
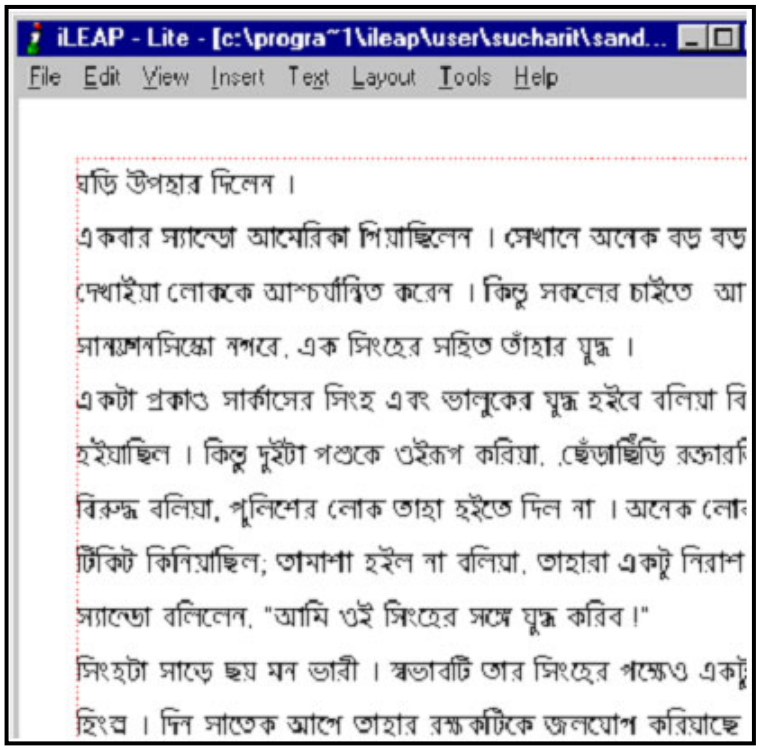
Technologies / Products Developed by ISI - Kolkata :

1. Bangla OCR

1.	Name of the Technology	Bangla Optical Character Recognition
2.	Nature of Technology	Printed Character Recognition Software for Bangla Language
3.	Level: (Product / Technology / Sub-system)	Technology
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	Bangla OCR System converts printed Bangla text into electronic version. It has many application potentials like reading aid for the blind (OCR and speech synthesis), automatic text entry into the computer, desktop publication, library cataloging and ledgering, automatic reading for sorting of postal mail, bank cheques and other documents. Also, document data compression is possible by converting the document image into ASCII format. **Performance Summary:** It deals with single column text image only. It can also handle small amount of skew in the range of –5 to + 5 degree. The system has been extensively tested on desktop and old published documents (more than thousand pages) of various font-types and sizes. The average character level accuracy is about 96% and word level accuracy is about 86%. **Supporting platform(s):** Linux, UNIX, DOS, MS-Windows. **Memory requirements:** 16MB RAM.

5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Bangla OCR: The Image:  The Corresponding Ground Truth: 

6.	Scalability / Portability / Expandability:	**Portability:** The system is implemented using ANSI C and hence portable to any platform.
7.	Readiness of Transfer of Technology (ToT) :	Yes
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology :	β- Testing done
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	Technology transferred to C-DAC, Pune for productization.
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

2. Devnagari OCR

1.	Name of the Technology	Devnagari Optical Character Recognition
2.	Nature of Technology	Printed Character Recognition Software for Devnagari Script
3.	Level: (Product / Technology / Sub-system)	Technology

4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	This Devnagari (Hindi) OCR System converts printed Devnagari text into electronic version. The software will be useful for automatic data entry, reading aid for blinds, postal sorting, data compression etc. **Performance Summary:** It deals with single column text image only. Five different popular fonts used by publication industries are recognized with high accuracy namely, 96% (approx) at the character level and 90% (approx.) at the word level. **Supporting platform(s):** DOS, MS-Windows. **Memory requirements:** 16MB RAM.
5.	Representative Snapshot / screenshot of the Technology / Product:	Similar to Bangla OCR:
6.	Scalability / Portability / Expandability:	**Portability:** The system is implemented using ANSI C and hence portable to any platform.
7.	Readiness of Transfer of Technology (ToT) :	Yes
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	Technology transferred to C-DAC, Pune, and ER&DCI, Noida for productization.
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

3. Assamese OCR

1.	Name of the Technology	Assamese OCR
2.	Nature of Technology	Adaptation of Bangla OCR to Assamese script
3.	Level: (Product / Technology / Sub-system)	Technology

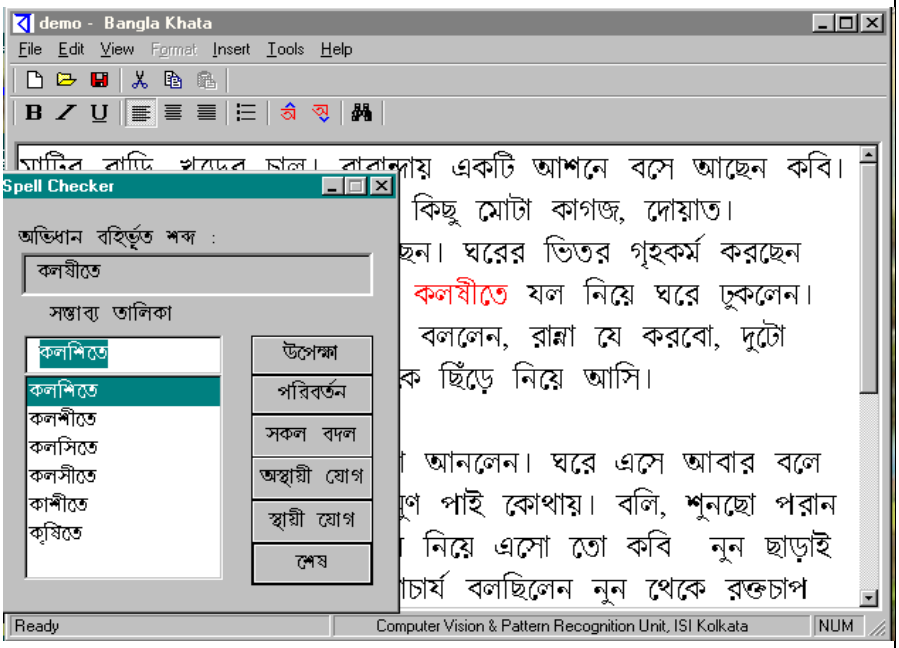
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	Since Assamese and Bangla share the same script with minor differences, the OCR system for Bangla can be successfully used for Assamese documents after some modifications. The modification is needed mainly in the post-processing stage when language specific OCR error correction is needed. The segmentation of a document image into lines, words, and characters, and the recognition of segmented characters is dependent on the script only. Thus, the modules of the existing OCR system for Bangla can be used for Assamese OCR. **Performance:** Test pages are selected from three Assamese books. Pages are scanned at 300 dpi. In total, 50 pages are used in testing phase. Analysis of test results shows a character-level accuracy of about 95%. Error analysis indicates that since the font used for printing Assamese materials are somewhat different from the fonts used in Bangla, generation of a new prototype library considering major Assamese fonts would improve the overall accuracy of the system.
5.	Representative Snapshot / screenshot of the Technology / Product:	Similar to Bangla OCR:
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in any UNIX and WINDOWS platform.
7.	Readiness of Transfer of Technology (ToT) :	Yes
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	**Technology Transfer:** The source code of the system along with the technical details has been transferred to IIT, Guwahati.
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road, Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

4. Oriya OCR

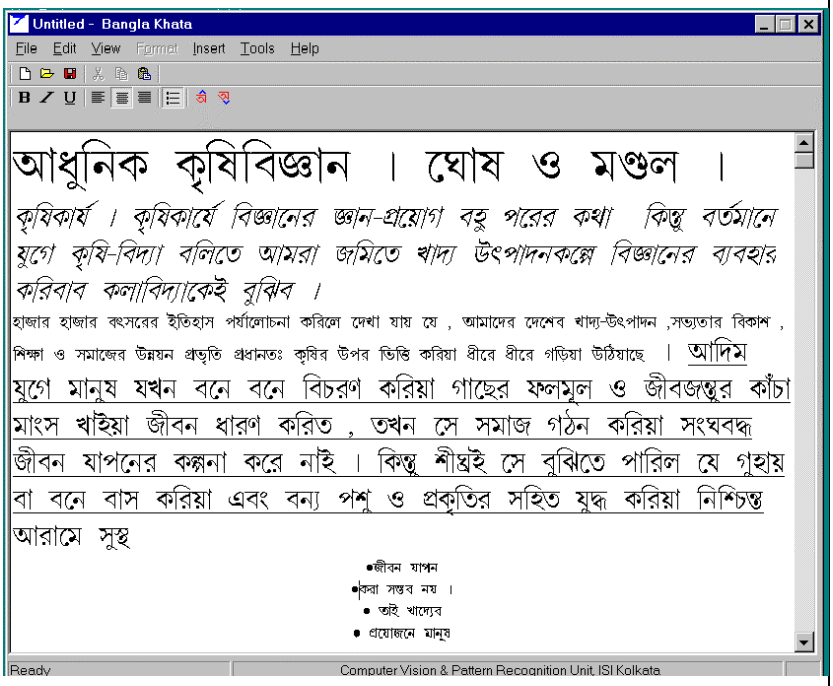
1.	Name of the Technology	Oriya OCR
2.	Nature of Technology	Printed Character Recognition Software for Oriya Script
3.	Level: (Product / Technology / Sub-system)	Technology
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	The purpose of this system is to recognize printed Oriya script automatically. In this recognition system, the document image is first captured using a flatbed scanner. The image is then passed through different preprocessing modules like skew correction, line segmentation, zone detection, word and character segmentation, etc. Next, individual characters are recognized using a combination of stroke and run-number based features, along with features obtained from the concept of water overflow from a reservoir. **Performance Summary:** On average, the system recognizes characters with an accuracy of about 96.3%, i.e. the overall error rate is 3.7%. **Supporting platform(s):** This system is developed in C language in UNIX platform. A WINDOWS based version is also available.
5.	Representative Snapshot / screenshot of the Technology / Product:	Similar to Bangla OCR:
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in any UNIX and WINDOWS platform.
7.	Readiness of Transfer of Technology (ToT) :	Yes
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	Technology transfer to OCAC, Bhubaneshwar, has been initiated.
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** (see above)

5. Bangla Spell-Checker

1.	Name of the Technology	Bangla Spell Checker (Banan Parikshak)
2.	Nature of Technology	Automatic Spell Checking and correcting Bangla words in Bangla editor
3.	Level: (Product / Technology / Sub-system)	Product Software ready for transfer for commercialization
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	Bangla spell checker can act in both offline and online modes. It has a Graphics User Interface via an editor (described below). Whenever the user types Bangla text, it checks for wrong spelling and gives suitable suggestions. On the other hand, one can run this software on previously typed material such as an OCR document. All misspelled words are captured by the system. For a single-error word, intended word is found within top four words in the suggestion list. Words having more than one error are so captured and for most of them intended words are in the upper half of the suggestion list. However, suggestions cannot be given on some inflected words. It has facility to add new words in the dictionary against which spellings are checked. The spell-checker is a tool to detect errors in Bangla words and correcting them by providing a set of correct alternatives which includes the intended word. An erroneous word can belong to one of two distinct categories, namely, non-word error and real-word error. Let a string of characters separated by spaces or punctuation marks be called a candidate string. A candidate string is a valid word if it carries a meaning. A meaningless string is a non-word. By real word error we mean a valid but not the intended word in the sentence, thus making the sentence syntactically or semantically ill-formed or incorrect. In both cases, the problem is to detect the erroneous word and either suggest correct alternatives or automatically replace it by the appropriate word. In this spell-checker, we consider only the non-word error. **Supporting platform(s):** Windows 95/98/2000/NT **Memory requirements:** 2MB or above

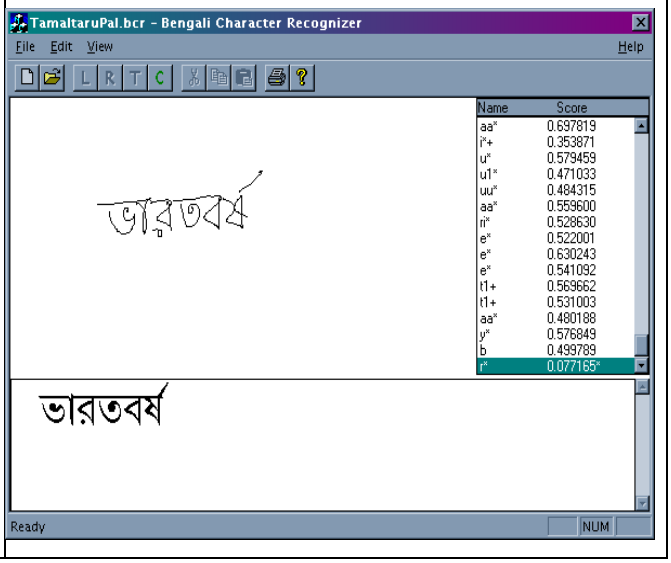
5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Application of Spell Checker in Bangla Editor 
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in WIN9X platform .
7.	Readiness of Transfer of Technology (ToT) :	Yes
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	None
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

6. Bangla Editor

1.	Name of the Technology	Bangla Editor
2.	Nature of Technology	Bangla Word processor
3.	Level: (Product / Technology / Sub-system)	Product
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	A Bangla text editor with the following feature has been developed where text document in Bangla can be created, saved (both in font code and ISCII code) formatted (bold, italic, size variation, underline etc.,) with different alignments (left, right, center). The online spellchecker mentioned above has been attached with it. **Supporting platform(s):** Windows 95/98/2000/NT **Memory requirements:** 2MB or above
5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Bangla Editor: 
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in WIN9X platform .

Items 7- 13: as for Bangla Spell-checker

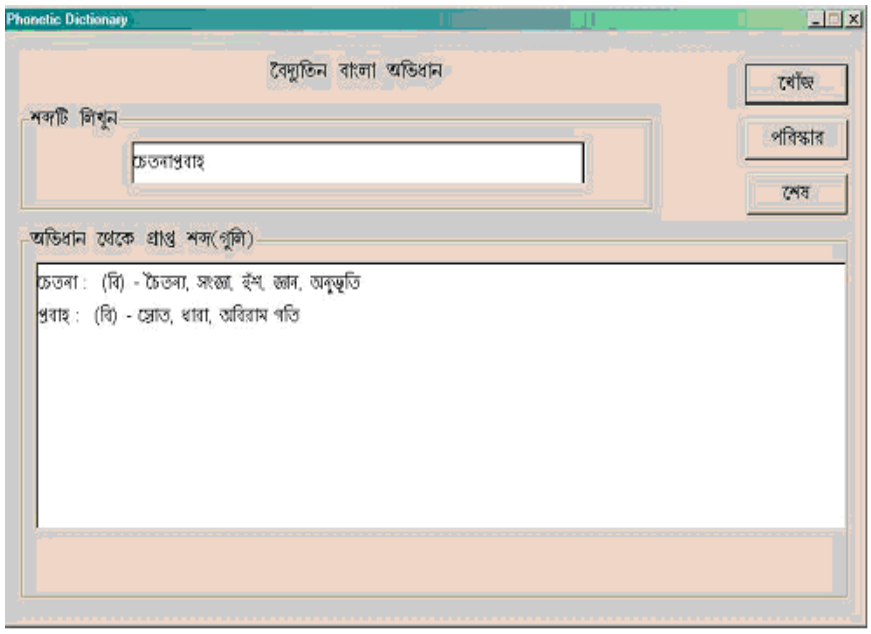
7. On-line handwritten Numeral and Character Recognizer:

1.	Name of the Technology	On-line handwritten Numeral and Character Recognizer
2.	Nature of Technology	On line Handwritten Character Recognition Software for Bangla Script
3.	Level: (Product / Technology / Sub-system)	Technology
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	This system is designed to recognize isolated handwritten characters as well as numerals in Bangla. This is one example of pen-based computing. A writer can write characters on a tablet using light-pen while the software recognizes them and stores the corresponding ISCII code in a file. The recognized characters are displayed using an Editor that supports Bangla font. **Performance Summary:** At present, system can recognize a character set that includes all basic characters, modifiers and numerals. It has almost cent percent accuracy and recognition is writer independent. On the other hand, the character level accuracy is about 95% which is writer dependent. **Supporting platform(s): DOS, MS-Windows.** Add-on H/W requirements: Tablet/Stylus. Memory requirements: 64MB RAM.
5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of On line Bangla OCR 

6.	Scalability / Portability / Expandability:	**Portability:** The system can run in WIN9X platform .
7.	Readiness of Transfer of Technology (ToT) :	No
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	Nil
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road, Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

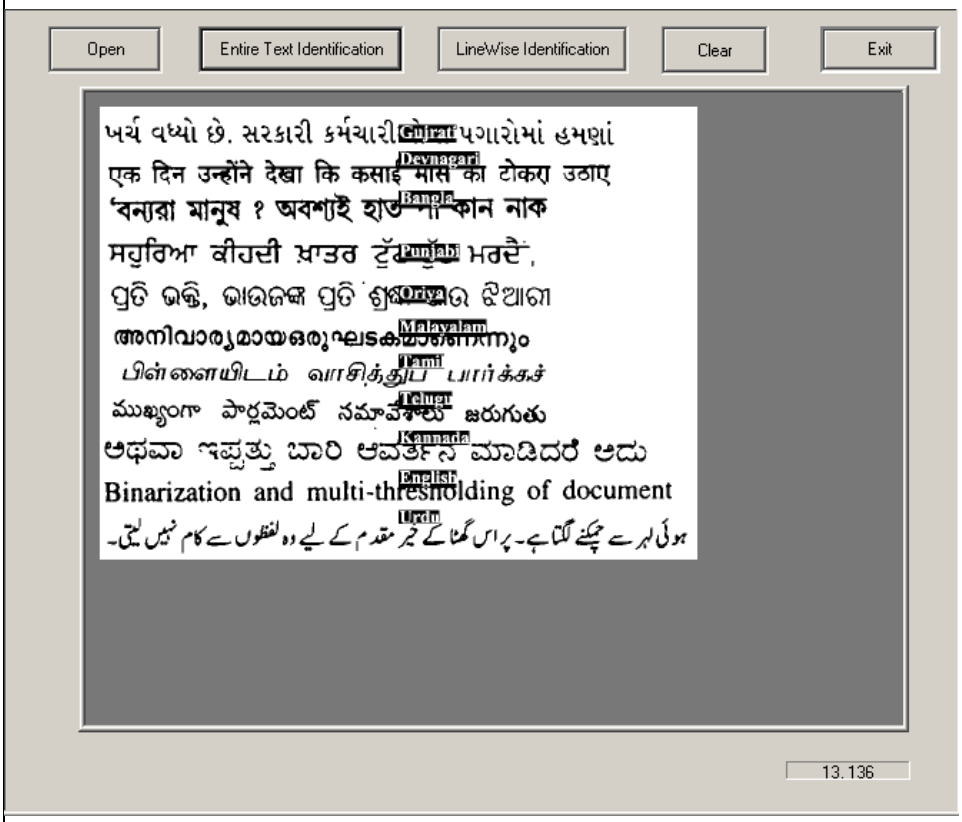
8. Digital Dictionary (Bangla-Bangla)

1.	Name of the Technology	Digital Dictionary (Bangla-Bangla)
2.	Nature of Technology	Digital Dictionary for Bangla to Bangla synonyms.
3.	Level: (Product / Technology / Sub-system)	Product
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	At present, the Dictionary contains about 65,000 words with parts of speech and meaning. It shows instant meaning of word when spelling is known. Automatic searching and finding the meaning for unknown spelling is also possible. This software also supports meaning for euphonic and assimilated words (not available in the dictionary). Automatic search of words with particular character-combination/with particular parts of speech and other attributes is also possible. It generates a simplified Thesaurus using dictionary. It has the facility to add new words with meaning and parts of speech in the (personal) dictionary. This software is a useful tool to the researchers and common people. **Supporting platform(s):** Windows 95/98/2000/NT **Memory requirements:** 32MB or above

5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Digital Dictionary: 
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in WIN9X platform .
7.	Readiness of Transfer of Technology (ToT) :	NA
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10.	IPR / Open-source:	Not open source
11.	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12.	User–agency tie-up:	Nil
13.	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road, Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

9. Script Identification and Separation From Indian Multi-Script Documents

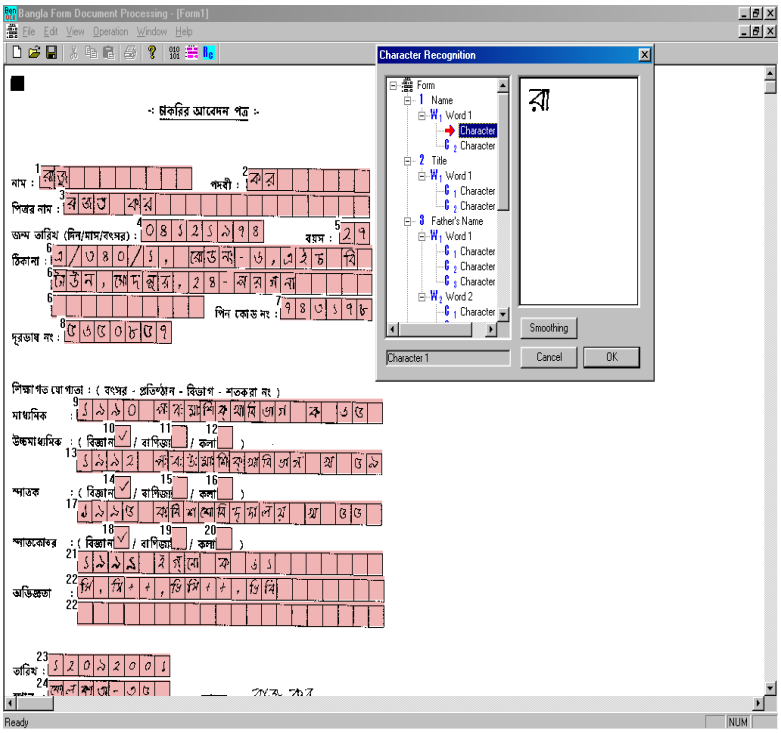
1.	Name of the Technology	Script Identification and Separation From Indian Multi-Script Documents
2.	Nature of Technology	Script separation software from Indian Multi-Script Documents
3.	Level: (Product / Technology / Sub-system)	Sub-system
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	India is a multi-lingual multi-script country, where a single document page (e.g. a passport application form, an examination question paper, a money order form, bank account opening application form etc.) may contain lines in two or more language scripts. So, for this type of document page there is a need for separating different scripts before feeding them to their individual OCR system for such a country. The purpose of this system is to identify different script regions of the document and hence separate them. The system works in two stages. In the first stage it separates each line in the scanned document page. Line segmentation is based on Horizontal Profile based technique. Secondly it identifies the script for each line in which it is actually written based on the distinguishing features between different Indian scripts. . **Performance:** The system is tested based on data taken from various sources like Journals, Newspaper, Synthetic Document etc. Currently the system has an overall accuracy of 97.52%. **Supporting platform(s):** This system is developed in C language in UNIX platform. A user interface for this system has been built in VC++ 6.0 in WINDOWS2000 platform.

5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Script Separation Software 
6.	Scalability / Portability / Expandability:	**Portability:** Both the C version and the VC++6.0 version of this system can run in any UNIX and WINDOWS platform respectively.
7.	Readiness of Transfer of Technology (ToT) :	NA
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10	IPR / Open-source:	Not open source

11	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12	User–agency tie-up:	Nil
13	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road, Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***

10. Form Document Processing

1.	Name of the Technology	Form Document Processing
2.	Nature of Technology	Automatic Processing of Hand-printed Table-Form Documents
3.	Level: (Product / Technology / Sub-system)	Technology
4.	Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:	In office environment, thousands of documents containing tables may be handled while processing application forms. In Indian context, in most of the cases, these table-form documents contain hand-printed text (like customer's name, date, item details/quantity/price etc.) mixed with printed text (invoice no., challan no., etc.). Once these forms are collected from the stockiest, customers, or other business centers, the computer operators manually enter the data into the computer to maintain an electronic version of the same. This manual approach makes this processing time consuming, boring and inefficient. Hence, an automatic approach is called for. Moreover, the work may be the basis of handwriting recognition of Indian languages. This takes the scanned gray-level image of a form document as an input to the software. It locates various parts of the document containing actual information. Different pieces of information are tagged. Each such piece of information is scanned again and different isolated characters forming the information are extracted and stored in the respective fields. It may be applied to automatic processing of various types of forms like application forms, examination papers, business forms, tax forms etc **Performance:** Currently, the system is trained using 50 forms for each of three different layouts. It is tested on distinct sets of another 500 forms for each type. In almost 99% cases, it could extract the information correctly on the test set.

5.	Representative Snapshot / screenshot of the Technology / Product:	Screenshot of Form Document Processing 
6.	Scalability / Portability / Expandability:	**Portability:** The system can run in WIN9X platform .
7.	Readiness of Transfer of Technology (ToT) :	No
8.	Availability of documentation:	Yes
9.	Testing of the Product / Technology:	Locally tested
10	IPR / Open-source:	Not open source
11	Potential Beneficiaries:	Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
12	User–agency tie-up:	Nil
13	Name and address of the Resource Person:	**B. B. Chaudhuri,** Professor & Head, Computer Vision & Pattern Recognition Unit Indian Statistical Institute 203 Barrackpore Trunk Road, Calcutta 700 108, INDIA **E-mail** : ***bbc@isical.ac.in***