International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 4, Issue 10, October 2015)

Secure Transmission of Hard Copy to Soft Copy Using OCR and Data Hiding Technology.

Riya. P. Ahuja[1], Mohini. R. Pasalkar[2], Ameya. J. Jadhav[3], Asst. Prof. Swati Shirke[4]
Department of Computer Engineering, STES's NBN Sinhgad School of Engineering, Ambegaon (BK) Pune 411041

***Abstract*— Most of the times you have to face this situation that you need the matter which you have in hard copy into a soft copy, like any official paper, notes etc. That you have in hard copy but you want them into a soft copy. So that you can use them through system or you can use them as your security purpose. What will you do to face this kind of a problem? Will you type all those papers or matter to get them into soft copy? In this article we will see how to convert your hard copy into a soft copy by using OCR and also for secure transmission by using data hiding techniques because Information hiding has importance in information security. With the help of a scanner or MFD you get software which is called OCR. Which converts your hard copy into soft copy, but when this copy is to be secure that is information is to be hidden. The soft copy which you get is encrypted into an image a Steganography is used for information hiding. Where Edge adaptive image steganography based on LSB is a popular type of steganographic algorithms used for hiding image in secured manner.**

***Keywords*—Opticalcharacterrecognition, Multi-function device, encrypt, Steganography, EDGE, List-significant bit.**

I. INTRODUCTION

Since traditional way of sending hard copy into soft copy by post ,or by any android apps like whastsapp, messenger, hike etc. and also by social networking sites(facebook,mail,etc.) are not safe or securely transmitted and if you want to share the data securely by post but there can be data lost while posting. The data will not be able to reach the second person and the time guaranty will not be fixed. If you do go by second way that is social networking sites or by android apps it may take less time to share the data but it won't be safe because the data which we are sending can be stored in some database and can be easily hacked or misused. To avoid this situation we have got a new approach that is converting hard copy into soft copy using OCR where we get the soft copy in .txt file. This file can be used in a normal manner to get hard copy in soft copy. But when we want to transmit the data securely .steganography based algorithm is used. Where the data will be encrypted into an image using edge adaptive image steganography based LSB .From this new approach we can send our confidential data in secure manner.

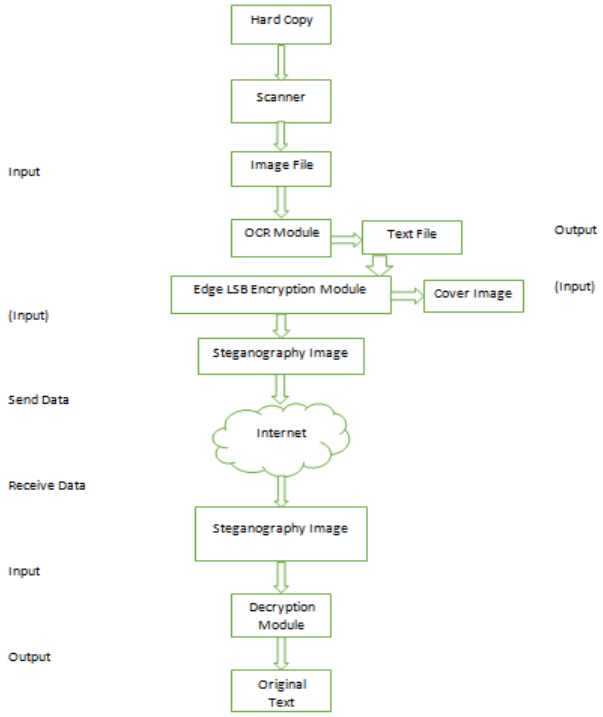

Figure 1: Architecture

A.OCR (Optical Character Recognition):

OCR is the acronym for Optical Character Recognition. This technology allows to automatically recognizing Characters through an optical mechanism. In case of human Beings, our eyes are optical mechanism. The image seen by Eyes is input for brain. The ability to understand these inputs Varies in each person according to many factors [2]. OCR is a Technology that functions like human ability of reading. Although OCR is not able to compete with human reading Capabilities.

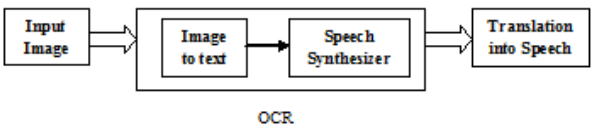

Figure 2: Block Diagram of OCR

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 4, Issue 10, October 2015)

OCR is a technique in which it enables you to convert different Types of documents as scanned paper documents, PDF Files or images captured in a digital camera into editable and Searchable data. Images captured on a mobile camera differ from scanned documents or image. They often have defects Such as noisy data at the edges and dim light, makes it Difficult for most OCR applications, to appropriately recognize them Text.

B. STEGANOGRAPHY: Comes from the Greek Words: STEGANOS – "Covered", GRAPHIE – "Writing".

Generally the sender writes an innocuous message and then conceals a secret message on the same piece of paper. The main goal of steganography is to communicate securely in a completely undetectable manner and to avoid drawing suspicion to the transmission of a hidden data. It is not only keeping others from knowing the hidden information, but also is to keep others from thinking that the information even exists. The advantages of Least-Significant-Bit (LSB) steganographic data embedding are that it is simple to understand, easy to implement, and it produces stage-image that is almost same as to cover image and its visual faithlessness cannot be demister by unaided eyes. Several steganography methods based on LSB have been proposed and implemented.

The previous algorithms LSB and Random LSB concentrate on hiding data in the least significant bit position of all or some selected pixels. They are not particularly concentrating on special pixels. To overcome these problems we proposed a novel image steganography algorithm based on least significant bit embedding algorithm (LSB) for hiding secret Messages in the edges of the image.

Figure 3: Original Image

Figure 4: Stego Image

II. Related Work

The task of character recognition through natural scenes is related to the problems considering in a camera based document analysis. Most of the work in this field is based on location and rectifying the text areas followed by the application of OCR techniques. Such approached are limited to scenarios where OCR works well. Furthermore, even the rectification step is directly applicable to our problem, as it is based in detecting of printed document edges or assumes the image is dominated in text. Methods for off-line recognition of hand printing characters have successfully cracked the problem of intra-class variation due to different writing styles. However such approached typically consider only limited number of appearance classes not dealing with variations in foreground, background color and texture. For natural scenes, some researches have designed systems that integrate text detection, segmentation and recognition in a single framework to accommodate contextual relationships. For instance used insights from natural languages processing and present a Markov chain framework for parsing images. Introduced composition machines for constructing probabilistic hierarchical image models which accommodate contextual relationships. This approach allows re-usability of parts among multiple entities and non-Markovian distributions. Proposed a method that fuses image features and language information (such as bi-grams and letter case) in a single model and integrates dissimilarity information between character images. Simpler recognition pipelines based on classifying raw images have been widely explored for digits recognition and other works on the MNIST and USPS datasets. Another approach is based on modeling this as a shape matching problem several shape descriptors are detected and extracted and point-by-point matching is computed between pairs of images.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 4, Issue 10, October 2015)

The data can be visible in basic formats like: Audio, Video, Text and Images etc. These forms of data are detectable by human hiding, and the ultimate solution was Steganography. The various types of steganography include:

A) Image Steganography: The image steganography is the process in which we hide the data within an image so that there will not be any perceived visible change in the original image. The conventional image steganography algorithm is LSB embedding algorithm.

B) Audio Steganography: Steganography can be applied to audio files i.e., we can hide information in an audio file, it can be called Audio Steganography. The audio file should be undetectable.

c) Video Steganography: Steganography can be applied to video files i.e., if we hide information in a video file, it can be called Video Steganography. The video file should be undetectable by attacker.

D) Text files Steganography: Steganography can be applied to text files i.e., if we hide information in a text file, it is called Text Steganography.

The general process of steganography i.e., preparing a stage object that will contain no change with that of original object is prepared but using text as a source. The basic image steganography algorithm is Least Significant Bit embedding. Least Significant Bit embedding (LSB) in a gray scale image each pixel is represented in 8 bits. The last bit in a pixel is called as Least Significant bit as its value will affect the pixel value only by "1". So, this property is used to hide the data in the image. Here we have considered last two bits as LSB bits as they will affect the pixel value only by "3". This helps in storing extra data. Least Significant Bit embedding

III. FUNCTIONAL AND NON FUNCTIONAL REQUIREMENTS

Any projects requirements need to be well thought out, balanced and clearly understood by all include, but perhaps of most important is that they are not compromised halfway through the project.

However, what exactly is the difference between „functional" and „non functional" requirements? It's not that complex, and once you understand the difference, the definition will be clear. The official definition of „a functional requirement" is that it actually specifies something the system should do.

Typically, functional needs will specify a behavior or function, for example: "Display the name, total size, available space and format of a flash drive connected to the USB port." Other examples are "add customer" and "print invoice".

Some of the more typical functional requirements include:

- Legal or Regulatory Requirements
- Transaction corrections, adjustments and cancellations
- Administrative functions
- Authentication
- Authorization levels
- Audit Tracking
- External Interfaces
- Certification Requirements
- Reporting Needs
- Historical Data
- Rules of Business

So what about Non-Functional Requirements? What are those, and how are they different?

Simply put, the difference is that non-functional needs describe how the system works, while functional needs describe what the system should do. The definition for a non-functional requirement is that it basically specifies how the system should work and that it is a constraint upon the systems behavior. One could also think of non-functional requirements as quality attributes for of a system.

Some typical non-functional requirements are:

- Performance
- Capacity
- Availability
- Maintainability
- Serviceability
- Security
- Regulatory
- Manageability
- Data Integrity
- Usability
- Interoperability
- Reliability
- Scalability
- Environmental
- Recoverability

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 4, Issue 10, October 2015)

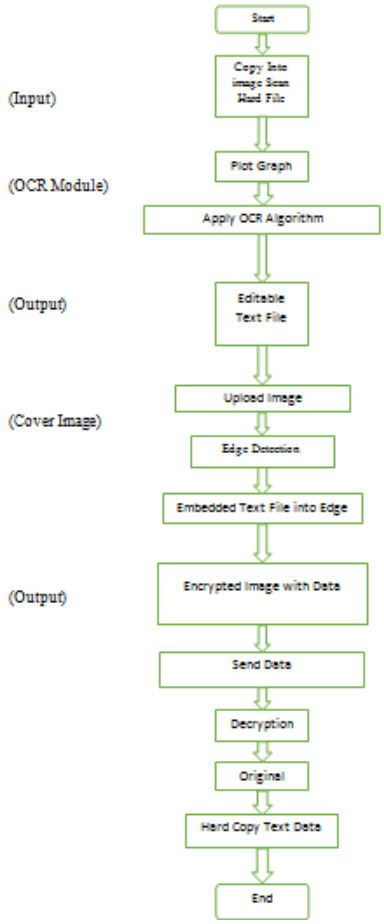

Figure 5: Flow Diagram

IV. CHALLENGES

1. Challenges to OCR of Camera Acquired Images:-

1. Uneven Lighting Conditions
2. Skewness
3. Tilting
4. Blur
5. Warping

2. Challenges to OCR due to inherent limitations of Mobile Devices:-

1. Limited Storage
2. Limited Computing Power

V. CONCLUSION AND FUTURE DIRECTIONS

Our next works with OCR Mobile Application will include the improvement of the results by the use of table boundaries detection techniques and the use of text post-processing techniques to detect the noise and to correct bad-recognized words. OCR application will also display the signatures and the other symbols as it is in the document. It will also update its features including the translation of one language to another. So that it will helpful for people from other countries who can't understand the local language.

REFERENCES

[1] K. Naveen BrahmaTeja1, Dr. G. L. Madhumati 2, K. Rama Koteswara Rao. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 11, November 2012).Data Hiding Using EDGE Based Steganography.

[2] Weiqi Luo, Member, IEEE, Fangjun Huang, Member, IEEE, and Jiwu Huang, Senior Member, IEEE. Edge Adaptive Image Steganography Based on LSB.Matching Revisited. IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 2, JUNE 2010

[3] Nadeem Akhtar, Shahbaaz Khan, Pragati Johri. An Improved Inverted LSB Image Steganography. 978-1-4799-2900-9/14/$31.00 ©2014 IEEE

[4] Mehdi Hussain, M. Hussain. Information Hiding Using Edge Boundaries of Objects**.** International Journal of Security and Its Applications Vol. 5 No. 3, July, 2011

[5] Susan Wiedenbeck, Jean-Camille Birget, Alex Brodskiy. Optical Character Recognition. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 46, NO. 1, MARCH 2014.