SIMULTANEOUS GAP FILLING AND TEXT REMOVAL USING IMAGE INPAINTING AND OPTICAL CHARACTER RECOGNITION

Gayatri Shahane[*1,]
Smita Shikare[2,]
Jyoti Karpe[3],
Vidya Jagtap[4]
[*123]U.G Student, Department of Computer Engineering, G.R.Raisoni COEM, Chas, Ahmednagar, India
[4]Associate Professor, Department of Computer Engineering, G.R.Raisoni COEM, Chas, Ahmednagar, India

ABSTRACT
Inpainting is a technique referred as a restoration of image or regeneration of image .In this paper we are combining two concepts namely Image inpainting and OCR i.e. Optical character recognition. The main problem is to identify the missing region or the region we want to Inpaint, to remove the text from the same and store that text using OCR. This paper gives the overall explanation about the algorithm which is Exemplar based inpainting and recreation of a new system. The main task is to identify the text written on the image, next to that is to remove that text and store that text. After that to fill the generated gaps using image inpainting. Anyhow, the main aim of any inpainting technique is restore or reconstruct the damaged area of an image

Keywords
Exemplar based image inpainting, texture synthesis, structure synthesis, OCR and inpainting

INTRODUCTION

The purpose of image inpainting is to fill in the missing areas or damaged area of an image. Here we are using Exemplar based image inpainting which is also known as Patched based image inpainting. This type of image inpainting is used to fill in solid parts or holes in an image by simply copying a similar part of the image from the known part of the image i.e. *known region* and paste it onto the damaged part or an *unknown region.* In proposed system the task is to be done in 3 steps: Using OCR, Save the text written on the image and then forward to next step which include to fill the holes from the image which further include 2 parts. First include larger and solid parts and second is for smaller parts. The working of Exemplar based image inpainting is given in figure below by [1].

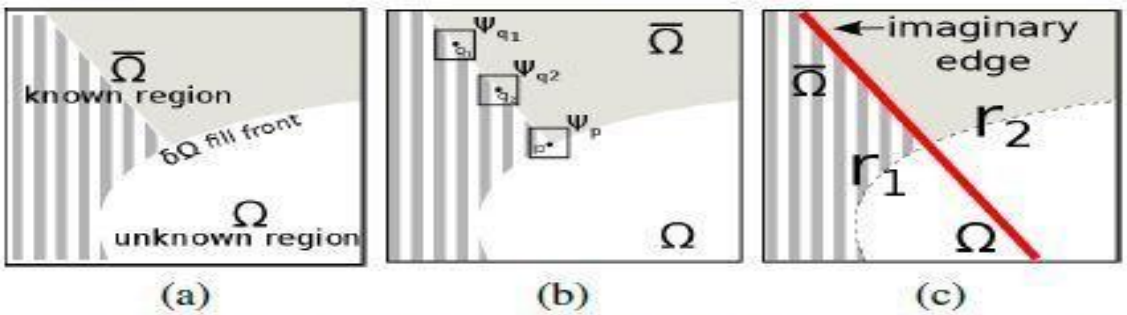

Fig. 1. Exemplar based inpainting. (a) Original image I with the unknown region Ω, its contour $\delta\Omega$, and the known region $\overline{\Omega}$. (b) K patches Ψ_{q_i} are found to synthesize the unknown area of the patch Ψ_p, with $p \in \delta\Omega$. (c) Example of the imaginary edge that helps the pdf computation for patches classified as structures.

Above figure shows the basic working Exemplar based inpainting.

LITERATURE SURVEY

Paper name: Exemplar-based image inpainting: fast priority and coherent nearest neighbor search.
Author name: R. Martınez-Noriega, A. Roumy, G. Blanchard
Review: Here they have introduced a better and efficient way to acquire the final result. They proposed an algorithm which will not only produce better completeness of an image but also they have reduced the error propogation problem. They have stated 2 ideas to improve the performance of Exemplar based inpainting.First, a modified priority for the filling order has been introduced so that it will reproduce the edges in the missing part more effectively and second is, in order to reduce the propagation of error in the inpainted image they stated a novel way to select the candidate patch.

Paper name: Region Filling and Object Removal by Exemplar-Based Image Inpainting.
Author name: A. Criminisi, P. Perez and K. Toyama
Review: In this paper they have proposed a new algorithm which will work on large objects and it will produce good effective background after removal of an object .They have introduced a result for simultaneous structure and texture synthesis using single algorithm only.
In given algorithm, each pixel hold its *color value* (it will be "empty" if pixel is unfilled) and *confidence value*, which will reflect the confidence in pixel value. The value will consider as *frozen* when it will get filled successfully.

Paper name: Digital inpainting on the basis of exemplar based Method.
Author name: Sanket Khedikar, P.N. Chatur
Review: Here author has given a simple but effective algorithm based on exemplar based image inpainting .Even though Proposed algorithm successfully removes the patches and damaged area successfully but the application of the same is very time consuming. Author also states the basic implementation of region filling algorithm.

Paper name: Image Inpainting on Satellite Image Using Texture Synthesis & Region Filling Algorithm.
Author name: Geeta K.Sarpate, Shanti K.Guru
Review: In this paper, algorithm is proposed to synthesize the texture and structure of an image and to fill the holes that left behind in undetected form. Proposed system is designed to compute the actual color values using region filling algorithm in addition with Exemplar based texture synthesis. They have shown result for both synthetic image and full color images including complex textures. But the patch size or the unknown region or the region to be remove must keep greater than the largest texture or the thickest structure(example: edges) in the source region.

Paper name: A complete optical character recognition methodology for historical documents
Author name: G.Vamvakas, B.Gatos, N. Stamatopoulos, S. J. Perantonis
Review: Author has presented a work for handwritten as well as printed documents. The presented algorithm is work in basic 3 steps. This system proposed a offline recognition system .The basic steps include pre-processing, top-down segmentation technique, creation of database and last is recognition stage.

SYSTEM MODEL

In this section we formalize system and algorithms.

System Architecture:
The core of the system is to complete the damaged digital image. In general, our system work in two different parts. First is to remove the important text from the image, store it in different and the second is to fill the gap or to remove the objects from the image

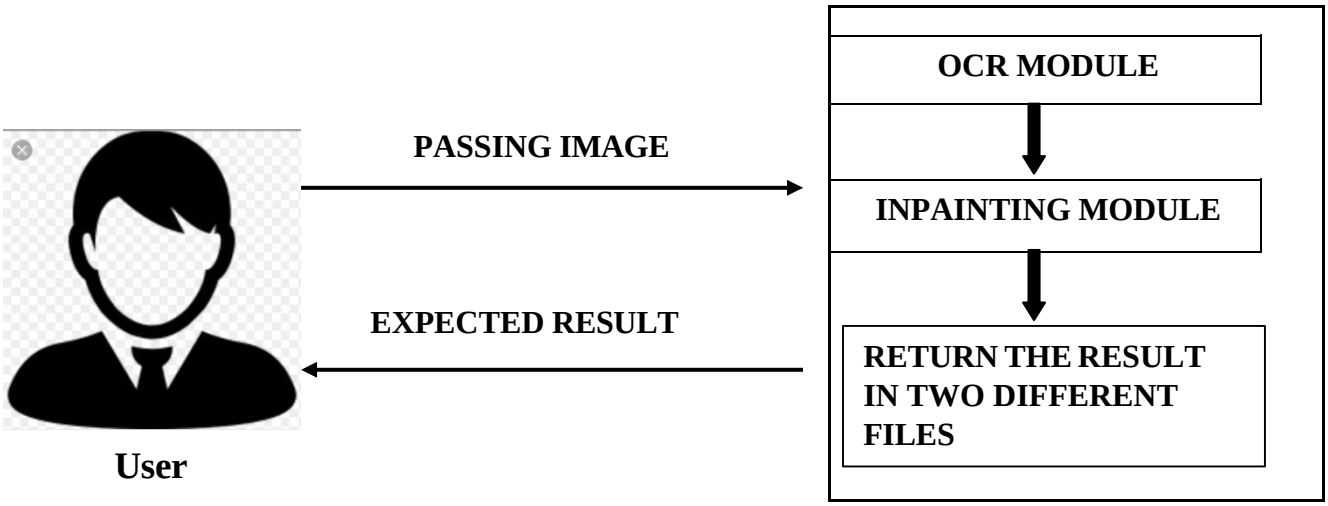

Figure 2: System architecture

As shown in figure, user need to pass the image to the system. An image has ample amount of data and colors in it. Our first task is to remove the text which is written on the image .After removal of text, that image is send to the inpainting module which further Inpaint the missing or damaged area of an image. We are using Exemplar based image inpainting for large , solid objects and texture and structure for smaller objects.

ALGORITHM

Steps for OCR:

As stated in [5], OCR implementation has several steps which are given below:

1. **Image binarization :** It is a very first step in which a grey scale image is converted into binary image.
2. **Segmentation:** After image binarization , a top down segmentation approach is applied .The first line from the page is detected. After that the words are detected from the document, they are extracted and finally segmented in characters.
3. **Character database**: At this step, we choose a set of images for processing and from these images characters are extracted following the segmentation procedure described in Section segmentation. This section consist of further more sub steps. Which are:
   - Feature extraction: As given in [5,6], they have presented two types of features. First feature is zone based and second one is given from area that is formed from the projection of lower and upper as well as of the lest and right character profile is calculated.
   - Clustering: Here we are using KNN for clustering.
     Algorithm is ,For the given object O, select top K dataset objects which are nearest to O by selecting distance measure. The Euclidean distance between these datasets is calculated by using following formula:
     Let Y1={y_{11}, y_{12}, y_{13},…., y_{1n}} and Y2={ y_{21}, y_{22}, y_{23},….., y_{2n}}
     There for,
     Distance(Y1,Y2)=$\sqrt{\sum_{i=0}^{n}(y1n - y2n)^2}$
4. **Recognition :** This is the last stage where the problem set or the processed document is converted into text file.

Steps for Image inpainting:

Here ,as explained in [1,2] throughout the paper, image is denoted by *I*. Let pixel denotes the pixel position of the image, whereas pixel value refers to as the value taken by the image at this pixel. More precisely, a pixel is mapped to a scalar (d = 1) for a gray level image and a vector (d = 3) for a color image:

The image *I* contains missing region or damage region whose information is unknown. As it is **Exemplar based inpainting** which was first introduced in [1,2] for **texture synthesis**. The patch Ψ_i is a square sub region. In

general, the size of Ψ_i is decided to be slightly larger than the largest distinguishable texture element. Here, the size of the patches is 9×9 pixels.

Aim of our algorithm is to complete the image *I* by filling Ω (i.e. unknown region) with information taken from Ω'(i.e. known region). Algorithm first selects the patch Ψ_p , with P ∈ δΩ, whose priority is maximum for filling among the patches of the fill front. Note that Ψ_p may contain the region who is already filled i.e. known region and the region with no information i.e. unknown region. Then we need to find *K*-most similar patch Ψ_{qi} to Ψ_p are found, where ⍰ =1,2,…,*K* and as shown in figure 1(b). So at last, Ψ_{qi} are linearly combined to form the unknown region Ψ_p . These steps are repeated until the unknown region Ω is filled completely.

This two step approach has been originally proposed in [1,2] i.e. filling order and information propagation. Originally, this algorithm is sequential. At each iteration n, the missing part of the current processed patch is repaired. Let P be the matrix that selects the known pixel from the current patch Ψ_p . Let Ω_n' and Ω_n denote the known and unknown region at iteration n. It is mathematically denoted by,

$f: \{0,\ldots,255\}^{d\times|\Psi_n \cap \Omega_n'|} \rightarrow \{0,\ldots,255\}^{d\times|\Psi_p|}$

$x = P\Psi_p \mapsto y$

CONCLUSION

In this paper, we have combined two different systems namely OCR i.e. optical character recognition and image inpainting. We have proposed a system which will give an efficient and fast algorithm for image restoration and optical character recognition. Moreover, we will try to improve the dots occur after text removal. Instead of only removing the damaged portion our algorithm also store the text. We are trying to improve our algorithms in order to better completeness of an image and to obtain more realistic image by means of texture and structure.

REFERENCES

[1] R. Martı'nez-Noriega, A. Roumy, G. Blanchard, "Exemplar-based image inpainting: fast priority and coherent nearest neighbor search", 2012 IEEE International workshop on machine learning for signal processing, Sept 2012

[2] A.Criminisi, P. Perez, K. Toyama, "Region Filling and Object Removal by Exemplar-Based Image Inpainting", IEEE Transactions on Image Processing, vol. 13, no. 9, sept 2004

[3] Sanket Khedikar, P.N. Chatur, "Digital Inpainting on the Basis of Exemplar Based Method", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , pp. 3943-3945, 2014

[4] Geeta K.Sarpate,Shanti K.Guru, "Image Inpainting on Satellite Image Using Texture Synthesis & Region Filling Algorithm", 2014 International Conference on Advances in Communication and Computing Technologies, 2014

[5] G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis, "A Complete Optical Character Recognition Methodology for Historical Documents", The Eighth IAPR Workshop on Document Analysis Systems, 2008

[6] G. Vamvakas, B. Gatos, S. Petridis and N. Stamatopoulos, "An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition", *9th International Conference on Document Analysis and Recognition (ICDAR'07)*, pp. 1073-1077, Curitiba, Brazil, September 2007.