International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6S, August 2019

# Noise Reduction Technique in Scanned Document using Cuckoo Optimization Algorithm

**S. S. Thakare, S. N. Kale**

***Abstract — While scanning, digitization and transmission, scanned documents can be contaminated with noise. Noise can be categorized by identifying its characteristics. Noise observed for similar pattern scanned document is the source for selecting appropriate noise removal techniques. Any image processing method can have few phases like (i) Pre-processing, (ii) Segmentation, (iii) Recognition and (iv) Post processing. This pre-processing stage is an essential stage, which primarily deals with noise removal. This paper involves the use of the Cuckoo Optimization Algorithm (COA) to remove noise in preprocessing. COA has shown its superior capabilities in fast convergence and better optimal global performance. It finds the most likely pixel value to restore noisy pixels.***

***Keywords: Cuckoo Optimization Algorithm (COA), Document, Scanned document, Image, Scanned image, Noise, Noisy pixel, Noise Reduction***

## I. INTRODUCTION

With the increased use of computers in a day to day life, the practice of digitizing all the important documents is increasing. Many types of software are available to recognize English text from the scanned images which stores data using ASCII codes. The characters are recognized and stored in the format of ASCII.

Optical character recognition normally abbreviated to OCR and digital character recognition (utilizing scanners and computer algorithms) are the electronic or mechanical interpretation of images of typewritten, printed or handwritten content (typically caught by a scanner) in machine-editable text. These are originally measured as separate fields (1, 2). This paper considers the offline document for example typewritten, manually written, or printed records caught by scanner (3).Documents scanning is the way to convert printed documents into digital format. But common problem is 'noise' generation during scanning. Noise generated in a image due to paper quality, the typewriting machine utilized, or it may be made by scanners while scanning procedure. In pre-processing, noise removal is a unique step. Noise can be produced before scanning or during scanning in an image (4).

The some noises usually observed in scanned images of documents are as below:

**Manuscript published on 30 August 2019.**
* Correspondence Author (s)
**Prof. S. S. Thakare**, Assistant professor, GCOE, Amravati, Maharastra, India(email: shubh_diwan@rediffmail.com)
**Prof. Dr. S. N. Kale**, Assistant professor, SGBAU, Amravati, Maharastra, India (email: Sujata.kale@gmail.com)



The lines present on ruled page are source of noise that interferes with text present. The dark patches present outside the page margin are the largest components of noise present in scanned document that may or may not have textual contents overlapped. A few types of background noise, for example uneven effects, contrast, interfering strokes, the background spots because of tilted documentation while scanning and hole in document record etc.(5)

This paper utilizes Cuckoo Optimization Algorithm (COA) propelled by the behavior of a bird called Cuckoo, for this noise removal. Actually the search algorithm for cuckoo worked as a optimization algorithm is a met heuristic algorithm developed by Suash Deb and Xin-She Yang in 2009. The breeding manners of the cuckoo bird are the main inspiration in the process of Cuckoo Optimization. These birds, for their reproduction, select their home randomly among the home of some other birds. Normally, this bird lays their eggs in chosen nest of the host bird and drops the host bird's egg. An outline of COA applications in different categories is to tackle the issues with optimization.

## II. NOISE REDUCTION

For any image processing application, the image produced by the scanner requires to go through the pre-processing step which incorporates the procedures such as shifting, binarisation, diminishing, smoothing and so on. The goal of the pre-processing phase is to reduce the level of the noise in scanned text document without changing the information in the document. (5)

Noise observed due to change in a pixel value or addition of bit pattern that are unwanted and has no meaning in the output. It may be introduced during its acquisition procedure due to image reproduction and transmission. Its like creating a gaps in the document lines, filled loops or a disconnected segments. The noise is usually found in documents of poor quality (7). This noise appears either as pixels that are isolated or as a off region pixels. One can remove this noise by filtering. It decreases bit patterns which are unwanted and is introduced by the scanner or device of data acquisition through writing surface or poor quality (4). It as well removes the background a bit textured or colored and refines the image. The purpose of structure of noise reduction process is to remove noise while holding all the applicable data (5). The filters are of two types, linear and nonlinear. The linear filter is utilized for noise reduction mainly mean filters and wiener filters, yet they have certain drawbacks like blurring of edges and its fine details, damage of lines.

So the nonlinear filters like median filters are used which overcomes the limitations of linear filters to some degree.
But it has its own disadvantages like it might cause the loss of corners and threads, blurring of text content in the document record.(5) Now, these limitations can be overcome by Using a novel technique i.e.Cuckoo Optimization Algorithm.

COA begins with primary cuckoo population. This cuckoo birds lay few eggs in the host bird's nests that precisely look like the host bird's eggs. Those eggs who look same, get the opportunity to live and develop. Others eggs will be spotted by the bird host and it may be wrecked by them. Thus the full-grown eggs reveal how the nests in a zone are good and profitable. If many eggs can endure, more profit will be devoted to that zone in an environment. So cuckoo is always in search of the best nests to lay eggs, thus maximum eggs can stay alive (10).

The Cuckoo Search simply follows the three rules that are very important and idealized

A cuckoo lays 1 egg at a time in a randomly chosen home.

The best home with the best egg quality will be transported to become next generation.

Fixed quantities of host nests are obtainable and the cuckoo egg is found by the bird host with a possibility of pa μ (0, 1).

As indicated by these three standard rules, the bird host can either discard the egg or leave nest and build entirely new home. For simplifying, the fraction pa of the home nests can be estimated by new nest homes (with new arbitrary solutions).

The arbitrary walks can then likewise be related to the similarity between the egg of a cuckoo bird and the egg of host bird that can be hard to execute. In certain conditions, which are very colossal, the new solution will be excessively a long way from the old solution (or even can jump out of the limits). It is doubtful that such a move will be acknowledged. In the event that is excessively little, the change is too little to be substantial, and this search along this line is not effective. Thus, an efficient search is imperative to take suitable steps (7).

A novel method for cancelling noise is developed using an evolutionary algorithm to remove the noise from the images. This algorithm works in a way that the details of images are conserved but on other hand, the noise is nearly eliminated.

Perhaps the first point that comes to the mind is about the time of restoration. But because this method as only applies to the noisy pixels, the overall filtering time will not be so much.

Rest of paper structure is: In Section III, a brief explanation of COA is presented. Noise removal algorithm, Implementations and simulation are represented in section IV. Lastly, the Conclusions are provided in the section V.

## III. INITIAL CUCKOO HABITAT GENEATION

The "habitat" is nothing but the optimization vector in Cuckoo Optimization Algorithm, which is denoted by a dimentional vector variable Nvar. A habitat refers to current living zone or a colony of a cuckoo bird under consideration and is represented as below:

habitat = [x1, x2, x3…. xNvar]

The Worth of a habitat is calculated from evaluating the following function:

Worth = fw(habitat)= fw(x1,x2,x3…xNvar)

Now, considering the population of COA habitat is of a size Npop × Nvar. Randomly this matrix is initialized and then, few eggs are distributed for all initial habitats. Generally 5-20 Eggs are laid by the birds, which are considered as lower and higher limits of number of eggs devotion in the subsequent optimization iterations. Additional important cuckoos habit is to lay eggs in a utmost span from initial positions.

This span is named as ELR (Egg Laying Radius), and defined as [7, 8].

ELR= ß X [No. of eggs of cuckoo bird (under consideration) /Total No. of eggs] X [var(hi) - var(low)]

where, ß is an integer, var(hi) is the higher search range and var(low) is the lower search range for the current optimization issue .

### *B. Egg Laying Technique*

Each one cuckoo lays eggs in randomly selected home nests of host birds but within the ELR span.

The basic incentive for this was Cuckoo's special lifestyle and its physiognomies in egg breeding and laying.

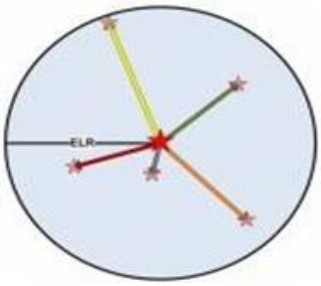


**Figure 1. Egg laying pattern of cuckoo within the ELR span[8]**

Ones the egg laying process is completed, if few eggs are distinguished as less like the eggs of the host bird then those are excluded. On an average 10 percent of total eggs will be eliminated in COA, with lower Worth and the remaining eggs will have an opportunity to survive.

*Process of immigration*

In this way, after new cuckoos develop into a mature ones, they make their own community to live in. Cuckoos begin to immigrate, when the season of egg laying comes. Always trying to locate better habitat where its eggs are alike to the eggs of host birds. Then they form groups at various zones to maximize the habitat worth. Other cuckoos also immigrate towards it and it becomes difficult to decide that a particular cuckoo is fit for which group while all cuckoos live in a same search space.

During immigration, cuckoos never travel all the way to the new target habitat. They follow an immigration pattern. Figure 2 shows this pattern.

As per Figure 2, any one particular cuckoo, moves some percent towards the target habitat and have a deviancy.

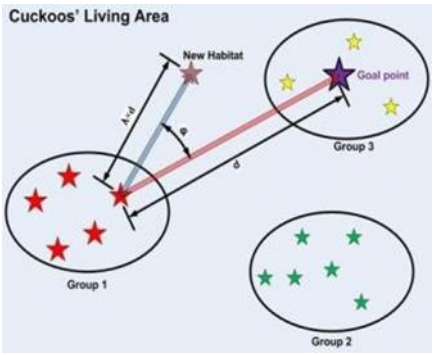


**Figure 2. Immigration scheme of Cuckoo[8]**

*Elimination of Worst Cuckoo*

Nature also keeps its balance on the Cuckoos populace size. Hence, the upper limit of alive cuckoos is governed by Nmax in each iteration, as the food is restricted and existence of predators. In standard COA, Nmax cuckoos that have more Worth will endure and the others will not [7].

Rather than utilizing filters as it might cause the loss of corners and threads, blurring of text content in the document (11), novel method is proposed for the elimination of noise from the scanned documents based on the algorithm of COA.

1: Set the initial values of host nest whose size n, probability pa ϵ [0,1] and maximum no. of iterations Max itr

2: Set t:=0, {counter initialization} , for image size (M,N)

3:For (i=1: i <=n) do

4: Generate primary population of n host x (t) ho

5: Evaluate the fitness function f(x) (t). ; Blob essentially be greater than 30 pixel & it must be 30 pixel

6: end for ; far from word

7: Repeat

8: Randomly create a new solution (Cuckoo) x (t+1) by Levy flight

9: Calculate the fitness function of a solution x (t=1) f(x (t+1))

10: Randomly select a nest xj amongst n solutions

11. If (f(x (t+1)) > f(x (t)) then

12: Replace the solution xj by the solution x (t+1) ; remove the blob (scanned noise)

13: end if

14: Abandon a fraction pa of worst nests.

15: Construct new nests at new places using Levy flight a fraction pa of worst nest

16: Keep the top solution (nest having quality solution) 17: Rank the solution and find the present best solution

18: Set t=t+1

19: Until (t>=Maxitr). {termination criteria are satisfied)

20: Produce the outstanding solution.

The proposed algorithm is executed and results are examined in the next section.

## IV.EXPERIMENTAL RESULT AND DISCUSSION

Algorithm is best suited to noise removal in scanned text with handwritten document.

The generated code can be briefed as:

Objective function:

Primary population of n host nests are generated

While (t<Max Generation) or (stop criterion)

Randomly created a new solution (Cuckoo) x (t+1)

and replaced its solution by performing Lévy flights;

Evaluated its quality/fitness

f(x (t+1)) > f(x (t))

Select Randomly a nest xj amongst n solutions

if (f(x (t+1)) > f(x (t)) ,

then replaced the solution xj by the new solution x (t+1) ; thus removed the blob (scanned noise)

end if

Abandoned a fraction pa of worst nests and new ones are built;

Keep the best solutions/nests;

Rank the solutions/nests and find the current top best solution;

Pass the present top best solutions to the next generation;

end while

Thus simplicity is the best advantage of this algorithm. Actually, associating with other population or agent-based meta heuristic algorithms, there is basically just a one single parameter pa in Cuckoo search (aside from the populace size). So, it is very easy to execute.

The image obtained from the scanner is shown in Figure 3.

The scanned document which contain printed Marathi text with handwritten text. It can be observed that the noise is there in scanned text document and hence, it requires pre-processing.

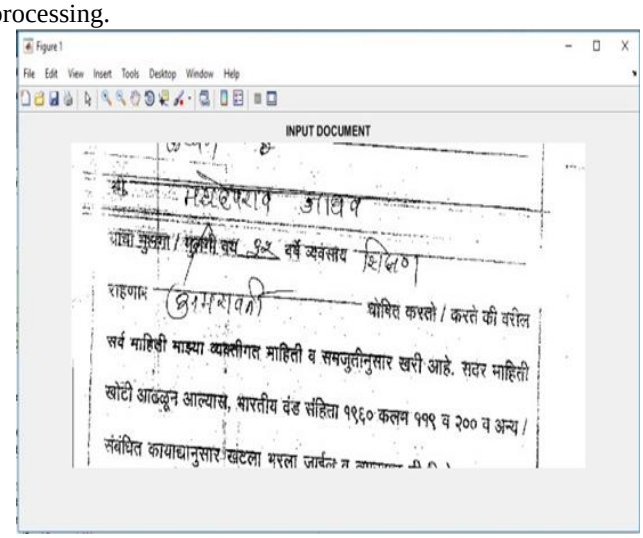


**Figure 3. Image obtained through Scanner**

Following Figure shows the all noisy pixels present in the document. After application of COA algorithm to this noisy document, as shown below, all noisy pixels will be removed.

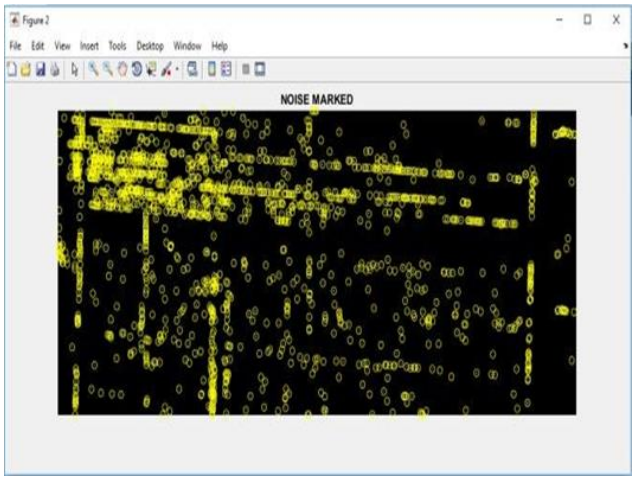


**Figure 4.Total Noisy pixels present in the document**

Thresholding is important task in noise removal. Here after separation of foreground and background image apply thresholds and As the document is handwritten Marathi document, the full stop, anuswar etc. are looks like a noise in the document so, even though here first apply thresholding, there remains noise near the words so for complete noise removal apply the COA algorithm and see the result as shown in Figure5.

After thresholding, noise will be detected and will be appear as shown below:

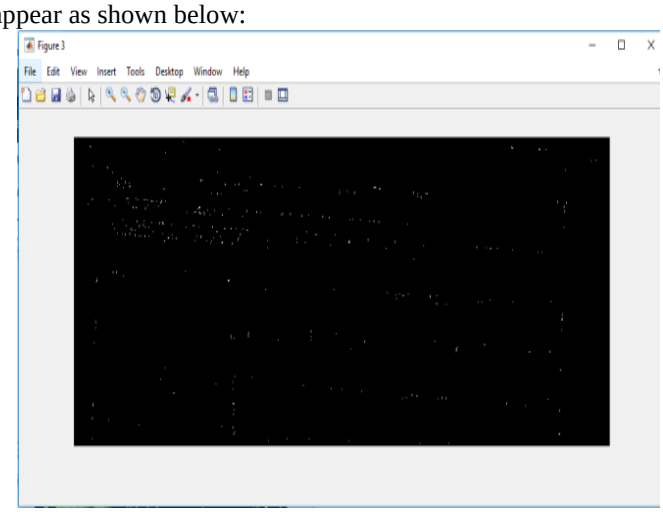


**Figure 4. Noise detected Image after thresholding**

Finally we get the output enhanced image which will be ready for any type of application

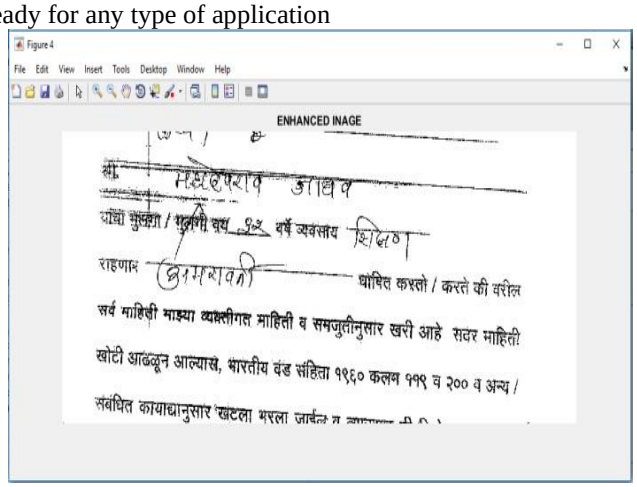


**Figure 5. Output Enhanced Image**

This novel technique finds a unique threshold for the picture, thus it expels background patterns effectively without removing a few details of an image. Then again, different methods preserve picture details yet some part of noise is classified as text content. In this paper, the results of other filters like anisotropic filter and wiener filter are also taken and compared it with proposed method.

The following Table displays the PSNR, SSIM and MSE values acquired for these methods.

In this paper, performance is checked on basis of some performance measures utilized for image compression such as MSE, PSNR and SSIM. Here three scan images of state government documents are taken, Scan image1,Scan image2 and Scan image3 and apply the above compression techniques by using Anisotropic Filtering, Wiener filtering and proposed filtering for noise removal. As per the following tables, Proposed filtering i.e. COA out performs in comparison to other methods in both MSE, PSNR and SSIM points of view.

Considering the MSE, PSNR and SSIM values with the results in the table, it is easily seen that novel algorithm is able to restore the noise corrupted pixels better than classic methods. The result shows that COA performs better than Anisotropic filters and Wiener filtering, that is generating less MSE and more PSNR, SSIM values.

| Scanned document Image | MSE | PSNR | SSIM | Method |
|---|---|---|---|---|
| Scan image1 | 1419.238984 | 28.6854438 | 0.703979624 | Proposed Filtering |
| | 3966.499964 | 24.2219245 | 0.011726187 | Anisotropic Filtering |
| | 4254.216496 | 23.9178032 | 0.034276527 | Wiener Filtering |
| Scan image2 | 3448.304562 | 24.8299429 | 0.027638482 | Proposed Filtering |
| | 4468.293429 | 23.7045822 | -0.054314368 | Anisotropic Filtering |
| | 3748.859977 | 24.4670069 | 0.058189216 | Wiener Filtering |
| Scan image3 | 3332.217443 | 24.9786658 | 0.09117171 | Proposed Filtering |
| | 4596.837861 | 23.5814072 | 0.01136018 | Anisotropic Filtering |
| | 3973.562024 | 24.2141991 | -0.011022462 | Wiener Filtering |

## V. CONCLUSIONS

Figure 1 indicates noisy image before enhancement, noisy image after enhancement and also removed noise image after enhancement. The dotted area in Figure 2 with yellow holes were marked. In the noisy image before up gradation, the noise can't be seen since the value of noise was low and seen as a dark like 0 value. After improving the noise value, the noise is seen clearly. Finally, the noise image that is removed, shows the noisy fragment in the past data is also removed.

This paper offers a novel method employed to remove noise from scanned document image. The performance of introduced method was compared with some classical methods, like simple wiener filter and anisotropic filter and also with conventional optimization algorithm, i.e. Cuckoo Search Optimization Algorithm. The results depicted that COA excels in comparison with other methods. It got a noisy scanned image and changed over it into a low noise binary picture. Our technique performed much better on pictures with identical font size and format. This algorithm is thus extremely useful for document record images of colleges, or official letters or schools.

## VI. ACKNOWLEDGMENT

I am thankful to many people who have supported and helped me to bring the results in this paper. I would like to gratefully acknowledge Sant Gadge Baba Amravati University, Amravati for permitting to use the resources of research lab and I would also like to thank my colleague for their continuous suggestions, which helped me in improving my work. I owe great thanks to my Guide Dr. S.N. Kale, Associate Professor, SGBAU, Amravati for her even willingness to give us valued advice and direction; I would too like to acknowledge the entire Electronic Department for being co-operative with me during this work. Above all, I am more thankful than I can express to my family for their moral support helping me through all times.

## REFERENCES

1. H. Deborah and A. M. Arymurthy, “Image Enhancement and Image Restoration for Old Document Image using Genetic Algorithm,” Proceedings of Second International Conference on Advances in Computing Control and Telecommunication
2. R. Rajabioun, “Cuckoo Optimization Algorithm”, Applied Soft Computing Journal, Vol. 11, pp. 5508-5518, 2011
3. M. Agrawal, D. S. Doermann: “Stroke-Like Pattern Noise Removal in Binary Document Images, ” ICDAR 2011: 17-21
4. Parul Agarwal,Shikha Mehta,“Nature-Inspired Algorithms: State- of-Art, Problems and Prospects” International Journal of Computer Applications (0975 – 8887) Volume 100 – No.14, August 2014
5. Dr. Ahmed Fouad Ali Suez Canal University, Dept. of Computer Science, Faculty of Computers and informatics Member of the Scientific Research Group in Egypt. Presentation slides on Cuckoo search algorithm
6. Rajesh Kumar Subudhi, 2Bibhuprasad Sahu, 3Pratyush Rn. Mohapatra, “A Novel Noise Reduction Method For OCR System”, IJCST Vol. 5, ISSue Spl - 2, Jan - March 2014 ISSN : 0976-8491 (Online)|ISSN : 2229-4333 (Print)Technologies (ACT 2010), p 108-12, 2010
7. Shi, Zhixin, Srirangaraj Setlur, and Venu Govindaraju, “Image enhancement for degraded binary document images,” Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011. Proceedings
8. A. Farahmand, A. Sarrafzadeh, and J. Shanbehzadeh, “Document image noises and removal methods,” IMECS, Newswood Limited, 436–440, 2013
9. F. Shafait and T. M. Breuel, “A Simple and Effective Approach for Border Noise Removal from Document Images,” Proc. 13th IEEE Int’l Multi-Topic Conf., Dec. 2009
10. M. Agarwal, D. Doermann, “Clutter noise removal in binary document images,” in [Proc. Intl. Conf. on Document Analysis and Recognition], 556–560 (2009)
11. Azizah Mohamad,Azlan Mohd Zain,Nor ErneNazira Bazin1,Amirmudin Udin, “Cuckoo Search Algorithm for Optimization Problems – A literature Review” Applied Mechanics and Materials Vol. 421 (2013) pp 502-506 © (2013)TransTechPublications,Switzerlanddoi:10.4028/www.scientific.net/AMM.42 1.502
12. R. Rajabioun, A. Mamizadeh “Impulse Noise Removal Using Cuckoo Optimization Algorithm”, 13th International Conference on “Technical and Physical Problems of Electrical Engineering” 21-23 September 2017, Yuzuncu Yil University Van, Turkey.