International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Improved LZW Compression Technique using Difference Method

Dibyendu Barman, Md. Bakash Ahamed

Abstract: ***This work attempts to give a best approach for selecting one of the popular image compression algorithm. The proposed method is designed to find the best performance approach amongst the several compression algorithms. In this work existing lossless compression technique LZW (Lempel-Ziv-Welch) is redesigned to achieve better compression ratio. LZW compression technique works on the basis of repetition of data. In a situation where all the values are distinct or repetition of data does not present the LZW can't work properly. To avoid this problem, difference method called difference matrix method which is actually calculate the difference between two consequence data and store it in a resultant matrix is used. In this case the matrix contains repetitive data which is more effective compared to LZW technique. Another problem of LZW is dictionary overflow, because of LZW works on ASCII character there is a limit of 256 dictionary length initially. In this work dynamic dictionary method is used without using the ASCII rather than this static method. As a result, this dictionary can contain the initial value anything in a range of -256 to 255. Here ASCII values are not used because the proposed method is applicable grayscale image, where the pixel values are between in range 0 to 255. Using these two changes the proposed improved LZW method becomes more powerful that can compress a non-repetitive set of data significantly. The proposed method is applied on many standard gray images found in the literature achieved 7% to 18% more compression the normal LZW keeping quality of the image same as existing.***

Keywords: **Image compression,** ***Lossless compression, Lossy compression, LZW, Difference method,***

I. INTRODUCTION

Image compression is a technique that used to reduce amount of storage medium required to store a digital image keeping quality of the image as close to the original image hence decrease the storage medium cost, communication medium cost [1-14]. In image compression technique the number of bits required to store an image is reduced based on the following data redundancy technique- a) Coding redundancy, b) Inter pixel redundancy and c) Psychovisual redundancy[13][16]. There are two types of image compression technique found in the literature

1) Lossless compression:

In case of lossless compression quality of the decompressed image almost same as original but less amount compression ratio achieved compare to the lossy technique. LZW, Arithmetic coding, Run length encoding (RLE),

PiCture eXchange(PCX)[6] are some well-known lossless image compression techniques [2-4].

2) Lossy compression:

In case of lossy image compression huge amount data loss occur as a result blocking artifacts, ring artifacts may appear in the decompressed image hence visual quality of the image degraded drastically but the amount of compression ratio achieved is very high. Vector Quantization (VQ)[7-13], Color Image Quantization(CIQ)[17-23], JPEGare the most popular lossy image compression technique[2-4].

The proposed work is based on focusing dictionary based lossless compression technique LZW.. Due to lossless property it can recover 100% original data from the compressed data. The main objective of this work is to improve the performance of normal LZW compression technique to achieve better compression ratio keeping visual quality of the image as close to the original image.

There is various lossless image compression technique. LZW (**L**empel - **Z**iv - **W**elch) is one of the most popular techniques, which is used in mainly GIF, ZIP file. But in some cases this technique does not provide the satisfactory result. The main objective of the proposed work is to improve the LZW compression technique so that in most of the cases much improved result is achieved. In literature study it is observed that LZW compression suffer from two drawback. i) Normal LZW can't compress on non-repetitive data. ii) Initial dictionary length is always 256 because it uses ASCII. The proposed work is designed by overcoming these disadvantages of LZW.

In this work two difference method techniques are used.

Difference matrix is used to produce repetitive data on a set of non-repetitive set of input data.

Dynamic dictionary is used to improve the space complexity rather than using of static 256 length ASCII dictionary.

LZW compression technique is widely used in TIFF, GIF, PDF and many commercial & non-commercial field. To improve this world wide recognized compression technique, two steps are included as discussed above.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Dibyendu Barman*, Department of Computer Science & Engineering, Government College of Engineering & Textile Technology Berhampore, WB, India. Email: dibyendu.barman@gmail.com
Md Bakash Ahamed, Department of Computer Science & Engineering, Government College of Engineering & Textile Technology Berhampore, WB, India. Email: mdbakash@gmail.com

The article is organized as-In Section 2 existing LZW compression technique is discussed. In section 3 brief discussion of the proposed method. Section 4 gives an experimental result and its analysis. Section 5 concludes the article.

II. RELATED WORD

This section briefly discussed the existing lossless compression technique.

A. *LZW compression technique:*

The **LZW**(**L**empel-**Z**iv-**W**elch) algorithm is a very common compression technique. This algorithm is typically used in GIF and optionally in PDF and TIFF. Unix 'compress' command, among other uses. It is a lossless compression, means no data loss occur at the time of compression. The algorithm is simple to implement and has the potential for very high throughput in hardware implementations. It is the algorithm of the widely used Unix file compression utility compress, and is used in the GIF image format. The Idea relies on re-occurring patterns to save data space. LZW is the foremost technique for general purpose data compression due to its simplicity and versatility. It is the basis of many PC utilities that claim to "double the capacity of your hard drive".

Working Principle:

LZW compression works by reading a sequence of symbols, grouping the symbols into strings, and converting the strings into codes. Because the codes take up less space than the strings they replaces so compression characteristic features of LZW is achieved,

LZW compression uses a code table, with 4096 as a common choice for the number of table entries. Codes 0-255 in the code table are always assigned to represent single bytes from the input file.

When encoding begins the code table contains only the first 256 entries, with the remainder of the table being blanks. Compression is achieved by using codes 256 through 4095 to represent sequences of bytes. As the encoding continues, LZW identifies repeated sequences in the data, and adds them to the code table. Decoding is achieved by taking each code from the compressed file and translating it through the code table to find what character or characters it represents.

Example: ASCII code. Typically, every character is stored with 8 binary bits, allowing up to 256 unique symbols for the data. This algorithm tries to extend the library to 9 to 12 bits per character. The new unique symbols are made up of combinations of symbols that occurred previously in the string. It does not always compress well, especially with short, diverse strings. But is good for compressing redundant data, and does not have to save the new dictionary with the data: this method can both compress and uncompressed data.

III. PROPOSED METHOD

The proposed method is based on the lossless image compression technique **LZW** (**L**empel - **Z**iv - **W**elch). This is one of the most popular compression techniques now a day. This is widely used in PDF, GIF, RAR file. In a situation input where the input values are distinct or the repetition of the values doesn't exist, in this case the LZW compression technique can't compress the data proper way. The proposed work is designed by overcoming these drawbacks to achieve better compression ratio keeping quality of the image almost same as existing method.

Main target is to improve the compression ratio so that a large file can be compressed significantly, though the input values are distinct or repetition doesn't exist. To do that a small, significant step is included in the existing LZW compression technique. The difference method technique is applied in LZW so it can work properly and gives a significant compression ratio.

In the difference method we follow the steps:

Let R is an array

$temp$ is a temporary variable

$R_1 = temp = \text{first value of input}$

For n as each element from input

$$R_n = n - temp$$

$$temp = n$$

Return R

In the above steps the difference between two consecutive values are generated and store it for the next step of LZW compression named 'Difference method'. It is done to generate the repetitive values in input sequence. As the proposed method is applicable on gray scale image, for that reason the maximum difference value presence between 255 and the new matrix contains value in a range of -256 to 255. This values are converted into string datatype for the further process. As we have noticed that the values that are going to feed the LZW contains negative and the traditional LZW doesn't work on it, we have to modify the LZW, so that it can process the sequence. Normally LZW works on ASCII values as a result it can encode character but in a range of 0-255. We have omitted the ASCII concept, and as the values are only string datatypes of integer values we have generated the initial dictionary in a way using two major changes.

1) Dynamic range based on input data,
2) String values rather than ASCII.

The implementation of this project is one of the most important task. We are trying to improve the LZW image compression technique so that it will be more powerful to compress, because it will contain more repeated data. We just add a simple step to do it. We just use difference method which just create a matrix that contain the difference of two consequent values. As a result, the new values are more repetitive and more helpful to compress through LZW technique.

Difference matrix:

As the new matrix contains more similar, repetitive values the compression ratio getting higher than the normal method. The difference method uses the following steps to encode:

Let R is an array

$temp$ is a temporary variable

R_1 = $temp$ = first value from input

For n as each element from input

R_n = $n - temp$

$temp$ = n

Return R

After that the new matrix can compressed with the LZW and will give us a higher compression ratio. The new difference matrix gives us some more advantages over the traditional method, those are as follows:

i. Here the repetition of the value is more which helps to compression method,

ii. In a situation where the repetition of values in input doesn't exists this method gives us repetitive environment to perform compression,

iii. The maximum number of values in the matrix are small.

And the decoding technique is just the reverse. To decode the data after the LZW decoding we follow the steps:

Let R is an array
$temp$ is a temporary variable
$R_1 = temp$ = first value from input
For n as each element from input
$R_n = n + temp$
$temp = n$
Return R

The resultant matrix of difference method gives us the values in a range between -256 to 255. But the traditional LZW operates in 0 to 255 range values. We have noticed that the values that are going to feed the LZW contains negative and the traditional LZW doesn't work on it, we have to modify the LZW, so that it can process the sequence. Normally LZW works on ASCII values as a result it can encode character but in a range of 0-255.

The existing LZW technique is modified is the following ways.

i. Dynamic range based on input data,
ii. String values rather than ASCII.

It is noticed that the initial dictionary size of the LZW algorithm is always 256lengths, because it is using the ASCII character to be considered. But rather than using ASCII character encoding as the proposed work applicable on gray scale image data and the values of the input are always integer so it is not need to work on ASCII. Rather than that has changed the dictionary type to the integer string key value pair. As a result, it is going the out of the bound. The initial dictionary length lies between the minimum & maximum value of the input sequence, which are feed by the difference matrix. But as the min & max value of the difference matrix lies between the -256 to 255 the dictionary initial value never crosses the boundary of this values. The flowchart of the proposed method is shown below.

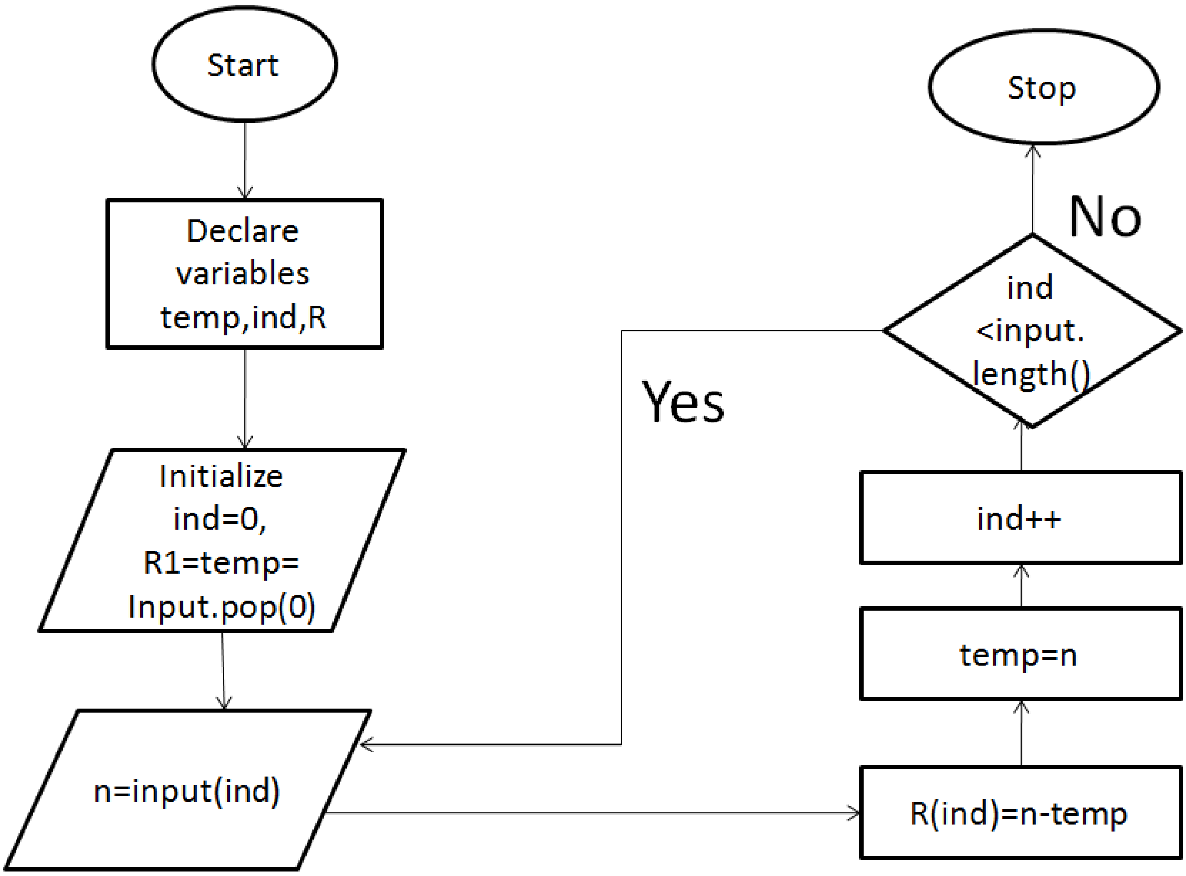

Fig. 1.Flow chart of the proposed difference method

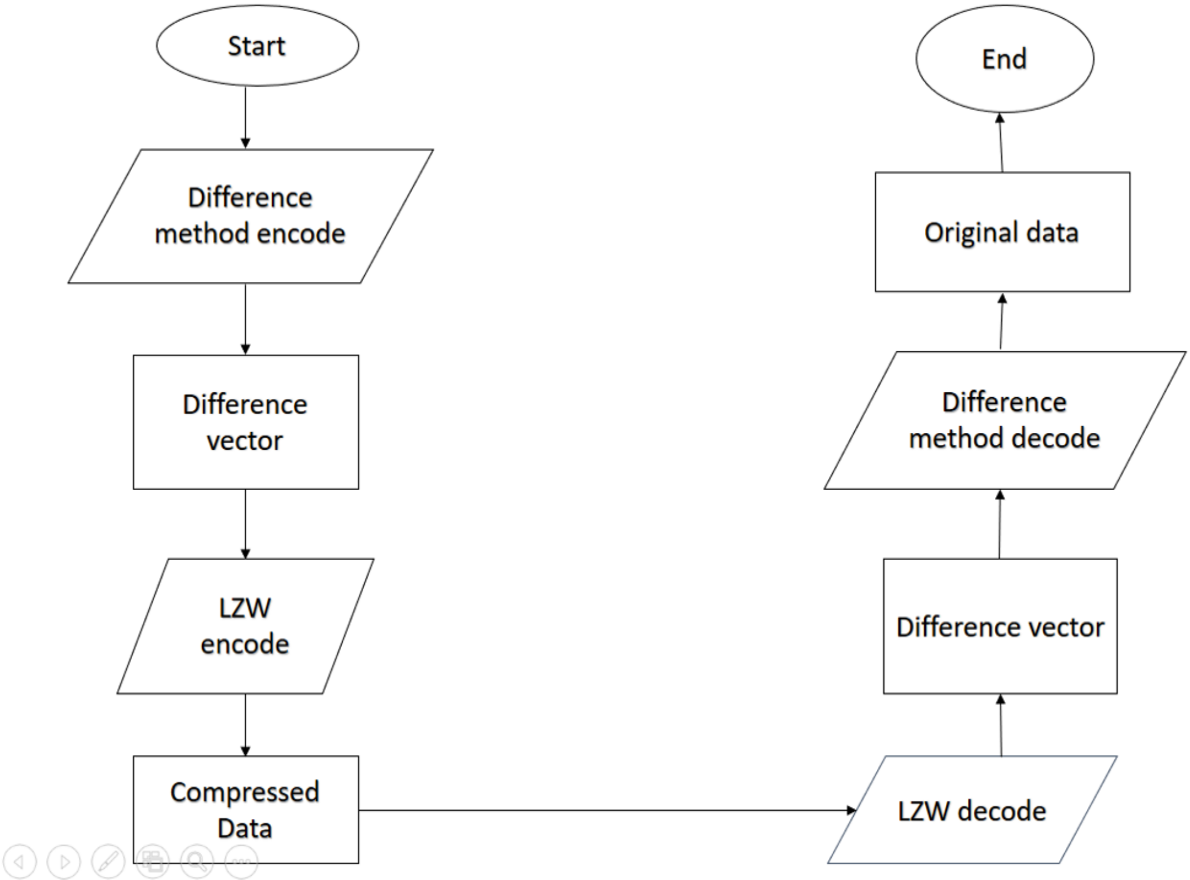

Fig. 2.Flow chart of the overall architecture of the proposed method

IV. RESULT AND DISCUSSION

The proposed method is implemented and coded in MATLAB2018 and applied on standard images found in the literature. The objective of this work is focused in retaining more image information during compression process to achieve better compression ratio than existing lossless LZW compression technique keeping quality of the image as close to the original image

TABLE I. EXPERIMENTAL RESULT USING PROPOSED METHOD

Image Name	Original Image size	LZW	Proposed Method	% compression using LZW	% compression using Proposed method
Elina512	258kb	136kb	103kb	47.74%	60.43%
Bird2	7.3kb	5.3kb	4.5kb	35.06%	46.24%
Lena	257kb	136kb	94kb	47.78%	64.46%
Terrace	257kb	140kb	121kb	46.06%	53.60%
Man	257kb	138kb	105kb	46.77%	59.42%
Peppers	257kb	135kb	86.6kb	47.52%	66.60%
Barbara	257kb	141kb	116kb	45.36%	55.35%
Couple	257kb	135kb	96.7kb	47.63%	65.25%

Figure 3 shows that average percentage of space reduction using normal LZW and proposed improved LZW compression technique. From the figure it is clearly observed that average space reduction using proposed method is far better than normal LZW.

Retrieval Number: E2216039520/2020©BEIESP
DOI: 10.35940/ijitee.E2216.039520
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

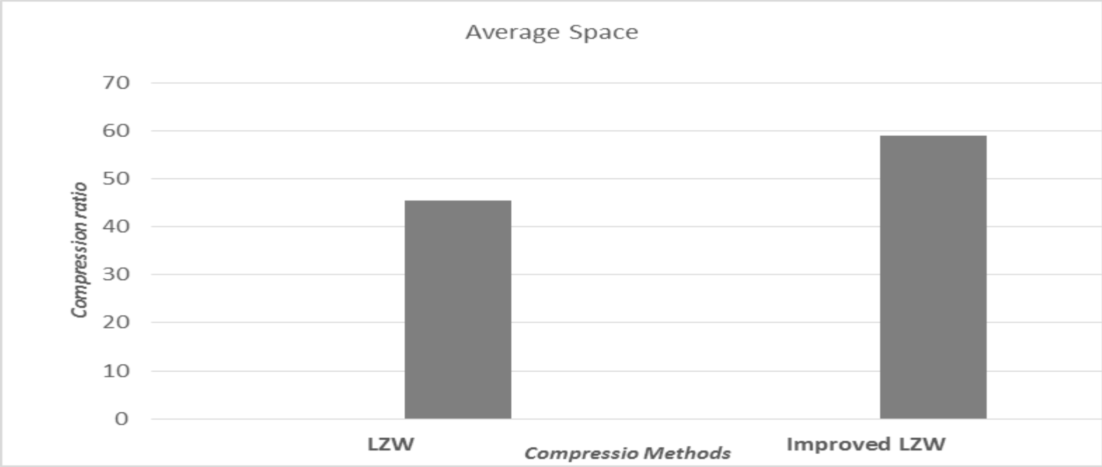

Fig 3. Average percentage of space reduction using proposed method and normal lzw

V. CONCLUSION

The proposed work is designed based on overcoming the two limitation of lossless LZW compression technique. i) LZW can't compress on non-repetitive data. ii) Initial dictionary length is always 256 because it uses ASCII. The proposed work is designed by overcoming these disadvantages of LZW. Here existing techniques is modified by adding two steps i) Difference matrix is used to produce repetitive data on a set of non-repetitive set of input data.ii)Dynamic dictionary is used to improve the space complexity rather than using of static 256 length ASCII dictionary. The proposed method is applied in many gray images found in the literature and achieved 7% to 18% more compression ratio than existing LZW keeping quality of the image same. Future work may be focused on application of the work in color images.

ACKNOWLEDGMENT

Author is thankful to Government College of Engineering and Textile Technology Berhampore, WB, India & Maulana Abul Kalam Azad University of Technolgy for permitting him to carry out research work.

REFERENCES

1. R. C. Gonzalez, R. E. Woods and S. L. Eddins, Digital Image processing using MATLB, Mc-Graw Hill, 2011.
2. G. Gan, C. Ma and J. Wu, "Data Clustering Theory, Algorithms and Applications", SIAM, 2007.
3. A. K. Jain and R. C. Dubes, "Algorithms for Clustering Data", Prentice-Hall, NJ, USA,2004.
4. D. H. Kil and F. B. Shin, "Reduced Dimension Image Compression And its Applications", Proc. of Int. .Conference Image Processing 3 1995, pp. 500-503.
5. C. K. Li and H.Yuen, "A High Performance Image Compression Technique For Multimedia Applications", IEEE Transactions on Consumer Electronics, 42(2) (1996), 239-243.
6. I. Avcibas, N. Memon and K. Sayood, "A Progressive Lossless / Near Lossless Image Compression Algorithm", IEEE Signal Processing Letters **9**(10) (2002), 312-314.
7. Y. Linde, A. Buzo and R. M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Transactions on Communicationsn COM 28(1) (1980), 84-195.
8. R. M. Gray, "Vector Quantization", IEEE ASSP Magazine **1**(2) (1984), 4-29.
9. L.Wenhua and E. Salari, "A fast Vector Quantization Encoding Method for Image Compression", IEEE Transactions on Circuits and Systems for Video Technology, **5**(2) (1995), 119-223.
10. S. D. Thepade, V. Mhaske and V. Kurhade, "New Clustering Algorithm for Vector Quantization using Slant Transform", ICETACS, St. Anthony's College, Shillong, India, 2013, doi:10.1109/ICETACS.2013.6691415.
11. D. K. Mahapatra and U. R. Jena, "Partitional K-Means Clustering based Hybrid DCT-Vector Quantization for Image Compression", IEEE Conference on ICT, Noorul Islam University Thuckalay, Tamil Nadu, India, 2013, doi: 10.1109/CICT.2013.6558278.
12. C. Karri and U, Jena, "Fast Vector Quantization Using a Bat Algorithm for Image Compression", Engineering Science and Technology an Int. Journal **9**(2) (2016), 769-781.
13. A. Hasnat, D. Barman, S. Halder and D. Bhattacharjee, "Modified Vector Quantization Algorithm to Overcome the Blocking Artefact Problem of Vector Quantization Algorithm", IOSP 32(5) (2017), 3711-3727.
14. S. Halder, A. Hasnat, D. Bhattacharjee and M. Nasipuri, "A Novel Low Space Image Storing and Reconstruction Method by K-Means Clustering Algorithm", Int.conference on Advances in Pure and Applied Mathematics (ICAPAM) 4(1) (2014). doi: 10.13140/2.1.4470.4648.
15. B. Freisleben and A. Schrader, "An Evolutionary Approach to Color Image Quantization", Proc. of IEEE Int.Conference on Evolutionary Computation 1997, pp. 459–464.
16. A. Hasnat, S. Halder, D. Bhattacharjee and M. Nasipuri, "A Novel Approach for Colorization of a Grayscale Image using Soft Computing Techniques", Int.Journal of Multimedia Data Engineering and Management (IJMDEM), IGI publication 8(4) (2017), 13-43..doi: 10.4018/IJMDEM.2017100102.
17. Z. Bing,S. Junyi and P. Qinke, "An Adjustable Algorithm for Color Quantization", Pattern Recognition Letters 25(16) (2004), 1787–1797.
18. M. G.Omran, A. P. Engelbrecht and A. Salman, "A Color Image Quantization Algorithm Based on Particle Swarm Optimization", Informatica 29 2005, pp. 261-269.
19. M.E. Celebi, Q. Wen and J. Chen. "Color Quantization using C-Means Clustering Algorithms", Proc. of 18th IEEE Int.Conference on Image Processing 2011, pp. 1729–1732.
20. S. C. Cheng and C. K. Yang, "A Fast and Novel Technique for Color Quantization using Reduction of Color Space Dimensionality", Pattern Recognition Letters 22(8) (2011), 845–856.
21. C. Ozturk, E. Hancer and D. Karaboga, "Color Image Quantization: A short Review and an Application with Artificial Bee Colony Algorithm", INFORMATICA 25(3) (2014), 485-503.
22. D. Barman, A. Hasnat, S. Sarkar and M. A. Rahaman, "Color Image Quantization using Gaussian Particle Swarm Optimization(CIQ-GPSO)", IEEE Int.Conference on Inventive Computation Technologies(ICICT), Coimbatore, TN, India, 2016. doi:10.1109/INVENTIVE.2016.7823295.
23. P. Scheunders, "A Genetic C-Means Clustering Algorithm Applied to Color Image Quantization", Pattern Recognition **30**(6) 1997, 859-866.

AUTHORS PROFILE

Dibyendu Barman, received his B.Texh degree fron University of North Bengal in the year of 2003 and M.Tech degree from Jadavpur University ,India, in 2006. Currently he is doing PhD at Maulana Abul kalam Azad University of Technology(MAKAUT) in the field of Image compression. In 2008 he joined in West Bengal General Service (WBGS). Now he is appointed as Assistant Professor in the department of Computer Science and Engineering, Government College of Engineering & Textile Technology, Berhampore, West Bengal, India His area of research interests is Image Processing.

Md. Bakash Ahamed received his M.Tech degrees from Jadavpur University,India, in 2011 and B.Tech in Computer Science and Engineering from University of Kalyani in 2004. In 2006 he joined in West Bengal General Service (WBGS-Gr A) under Department of Higher Education(Tech Branch), Govt of West Bengal. Now he is appointed as Assistant Professor in the department of Computer Science and Engineering, Government College of Engineering & Textile Technology, Berhampore, West Bengal, India His area of research interests is Multimedia Technology.