Design and Comparison of Error Correctors Using Clustering in Holographic Data Storage System

Sang-Hoon Kim*, Jang-Hyun Kim**, Hyun-Seok Yang*** and Young-Pil Park***

* Department of Mechanical Engineering, Yonsei University, Seoul, Korea
(Tel : +82-2-2123-4677; E-mail: hotkilsh@hotmail.com)
**Department of Electrical Engineering, Yonsei University, Seoul, Korea
(Tel : +82-2-2123-4677; E-mail: jhkim@control.yonsei.ac.kr)

Abstract: Data storage related with writing and retrieving requires high storage capacity, fast transfer rate and less access time in. Today any data storage system can not satisfy these conditions, but holographic data storage system can perform faster data transfer rate because it is a page oriented memory system using volume hologram in writing and retrieving data. System architecture without mechanical actuating part is possible, so fast data transfer rate and high storage capacity about 1Tb/cm3 can be realized. In this paper, to correct errors of binary data stored in holographic digital data storage system, find cluster centers using clustering algorithm and reduce intensities of pixels around centers. We archive the procedure by two algorithms of C-mean and subtractive clustering, and compare the results of the two algorithms. By using proper clustering algorithm, the intensity profile of data page will be uniform and the better data storage system can be realized.

Keywords: Holographic, Subtractive Clustering, C-mean Clustering, Data Storage

1. INTRODUCTION

Holographic Data Storage System[1], one of the next generation data storage principle, is a 2-dimensional page oriented memory using volume hologram in writing and retrieving process. In the HDS system, data management procedure is performed in parallel so fast data transfer rate can be realized. And the system store data in binary form (0 or 1), so that computers can use the digital data directly. In writing procedure, Laser of specific wavelength passes through a Spatial Light Modulator (SLM) to make 2-dimensional data page. A digital data 0 makes image of a black pixel (off-pixel) by blocking the light on SLM and digital data 1 is imaged as a white pixel (on-pixel) on CCD camera.

Diffraction, the nature of light, makes the laser which passes through an on-pixel surrounded by off-pixels on SLM to affect surrounding pixels by 2-dimensional Fourier Transform of a plane wave and when retrieving process the effects to surrounding pixels of an on-pixel are photographed by CCD camera and cause errors to binary data.

In this paper, we analyze the effect of an on-pixel (digital data 1) to surrounding off-pixels (digital data 0) and suggest a new method for reducing errors of binary data using two clustering methods. The error reduction algorithm is that ; find centers of on-pixel group and reduce the effect of on-pixels. To find centers of on-pixel groups, subtractive clustering and C-mean clustering algorithms are used. Figure 1 is the HDS system experiment booth.

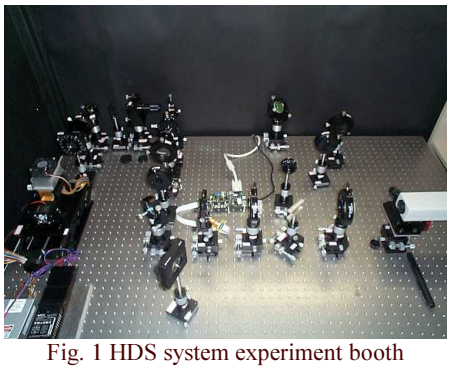


Fig. 1 HDS system experiment booth

2. IPI NOISE AND CLUSTERING ALGORITHM

2.1 IPI noise

By the Fraunhofer diffraction theory, the beam from the laser source emits as a plane wave. However after crossing the aperture – SLM, the beam propagates as a spherical wave from many point sources and diffracts. The on-pixels of SLM allow the beam to pass through, but the off-pixels do not. So an on-pixel acts like as an aperture, and the beam through the on-pixel of SLM will affect to surrounding pixels by two-dimensional Fourier Transform[2]. This effect acts as one of main noises, and is called “Inter Pixel Interference” noise.

Inter Pixel Interference (IPI) degrades the performance of the channel, and it tends to occur when an off-pixel is surrounded by on-pixels or vice versa. Where there are many on-pixels in comparison, errors from IPI noise will appear relatively high than other area of data page. Therefore if we find the cluster of on-pixels and reduce the intensities of the on-pixels, the possibility of error by IPI noise will decrease than before. And about the whole data page because the intensity of laser of brighter area will be decreased, uniform intensity profile will be obtained.

The IPI noise from two-dimensional Fourier transform is shown in figure 2.

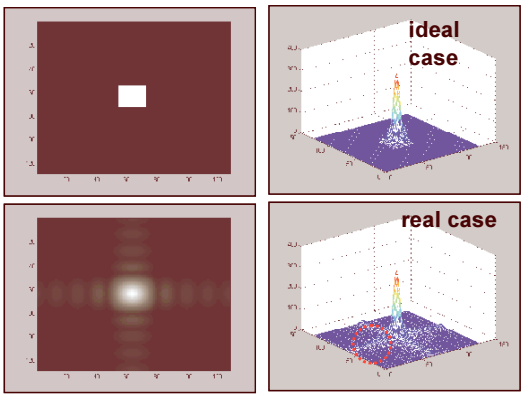


Fig. 2 Inter Pixel Interference

2.2 C-mean clustering

To find a region where there are many on-pixels in comparison, two algorithms are used and one of them is C-mean clustering.

Clustering is a method to divide distributed data into several groups. Data in the same group have similar characteristics and groups are divided by their average properties[3-4].

Clustering can be considered as connecting data set X with a matrix U, and expressed U_X. In C-mean clustering algorithm, to obtain optimal U_X the target function is summation of squared errors in generalized group like this ;

$$J_m(U,v)=\sum_{j=1}^{n}\sum_{k=1}^{n}(u_{kj})^m\left\|x_j-v_k\right\|^2, \qquad 1\le m\le\infty \tag{1}$$

x_j is d-dimensional measured data and v_k is d-dimensional k^{th} cluster center. One of the features of fuzzy clustering is that one can control the ambiguity of the equation by setting the value of m freely larger than 1. The parameters u_{kj} and v_k for local minimum of the equation (1) are set to $\hat{u}_{kj}$ and $\hat{v}_k$. Here, for $m>1$, the next equation is satisfied.

$$\hat{u}_{kj}=\left(\sum_{r=1}^{c}\frac{\left\|x_j-\hat{v}_k\right\|}{\left\|x_j-\hat{v}_r\right\|}^{\frac{2}{m-1}}\right)^{-1}, \quad \forall j,k \tag{2}$$

$$\hat{v}_k=\frac{\sum_{j=1}^{n}(\hat{u}_{kj})^m x_j}{\sum_{j=1}^{n}(\hat{u}_{kj})^m}, \quad \forall k \tag{3}$$

Repeating the next process, $\hat{u}_{kj}$ making equation (1) to be local minimum can be obtained.

Step 1 : Assume m and c (number of clusters) and define the norm by equation (2). Set $U^{(0)}$, the initial value of U, properly.($U^{(0)}$ can be chosen randomly without conscious of u_{kj})

Step 2 : Calculate cluster center $v_k^{(0)}$ using $U^{(0)}$ and equation (3).

Step 3 : Calculate $U^{(1)}$ using $v_k^{(0)}$ and equation (2).

Step 4 : Define critical value ε and iterate above processes until $\left\|U^{(p)}-U^{(p-1)}\right\|\le\varepsilon$.

2.3 Subtractive clustering

Subtractive clustering algorithm[4-5] defines the data as cluster center with the maximum of potential values, calculated by the function inversely proportional with the distance between each data with another. By the definition, the data with many neighborhoods has large potential value. Thus determine the data with the largest potential value as the cluster center, and the next cluster center should be determined without the effect of previous cluster center. This process will iterate until the potential value becomes smaller than pre-defined value.

The step of data clustering with subtractive clustering algorithm is indicated here. The input is n-dimensional vector $x_1, x_2, \cdots x_n$, and the output is one-dimensional vector $z_1, z_2, \cdots, z_n$, than the input-output data can be indicated by the input-output vector with N-numbers $X_1, X_2, \cdots, X_N$ in $(n+1)$-dimensional input-output space.

Step 1 : Normalize all data from 0 to 1

Step 2 : Calculate the potential of each data

$$P_i=\sum_{i=1}^{N}\exp\left(-\alpha\parallel X_i-X_j\parallel^2\right), \quad i=1,2,\cdots,N \tag{4}$$

Step 3 : Determine the largest potential value of P_i as P_i^*, and determine the data related with P_i^* as the first cluster center X_1^*

Step 4 : Calculate potential value P_i except for effect of the first cluster center

$$P_i^{'}=P_i-P_1^*\exp\left(-\beta\parallel X_i-X_1^*\parallel^2\right) \tag{5}$$

where $\beta=4/\gamma_b^2$ and $\gamma_b>\gamma_a$ so that the next cluster center would not appear near the current cluster center

Step 5 : Determine the largest value of $P_i^{'}$ as P_2^*, and determine the data related with P_2^* as the second cluster center X_2^*. Generally calculate potential value $P_i^{'}$ after removing the k th cluster center X_k^* and determine the data for the largest potential value as cluster center X_{k+1}^*.

$$P_i^{'}=P_i-P_k^*\exp\left(-\beta\parallel X_i-X_k^*\parallel^2\right) \tag{6}$$

Step 6 : Stop iterating if $P_k^*/P_1^*\le\underline{\varepsilon}$. If $P_k^*/P_1^*\ge\bar{\varepsilon}$, or $P_k^*/P_1^*>\underline{\varepsilon}$ and $\frac{d_{\min}}{\gamma_a}+\frac{P_k^*}{P_1^*}\ge 1$ than determine that as cluster center and repeat Step 5. If $P_k^*/P_1^*>\underline{\varepsilon}$ and $\frac{d_{\min}}{\gamma_a}+\frac{P_k^*}{P_1^*}<1$ than set the X_k^* as 0 and after determining the second largest potential value, determine that data as the new cluster center if $\frac{d_{\min}}{\gamma_a}+\frac{P_k^*}{P_1^*}\ge 1$ and repeat Step 5.

Where, $a=4/\gamma_a^2$ and γ_a is small positive real number less than 1 so that data out of the radius γ_a can effect to potential value little. Step 6 is the process for determination about selection of the calculated cluster center. Parameter

$d_{\min}$ expresses the shortest distance among determined cluster centers $X_1^*, X_2^*, \cdots, X_k^*$. Under the process for determination of cluster center, it makes data to be determined as a cluster center with large $d_{\min}$ though it has small potential value to establish upper bound $\bar{\varepsilon}$ and lower bound $\underline{\varepsilon}$.

3. SIMULATIONS AND RESULTS

The picture wanted to be stored is the Lenna picture in figure 3. Size of the picture is 120x90 pixels and 256 level gray-scaled.



Fig.3 The picture wanted to be stored

When converted to digital data the picture becomes 90x960 pixels and by 6:8 balanced coding method, it expanded to 90x1280 digital pixels array. To displayed on SLM the digital data array divided into eleven digital data pages with ten data page with 90x120 pixels and one data page with 90x80 pixels. The error correction process will be applied to all data pages, but the processes are equal about 11 data pages so it will be discussed about the first data page. The first digital data page of the Lenna picture is figure 4. And on figure 5, there are cluster centers found by the two clustering algorithms about the first data page – subtractive clustering and fuzzy C-maen clustering. The blue x markers are cluster centers found by subtractive clustering algorithm and red o markers are found by C-mean clustering algorithm.

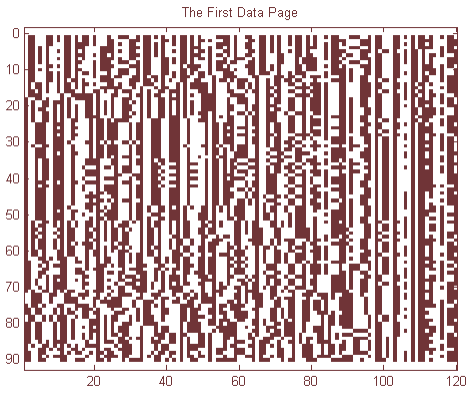


Fig.4 The first data page

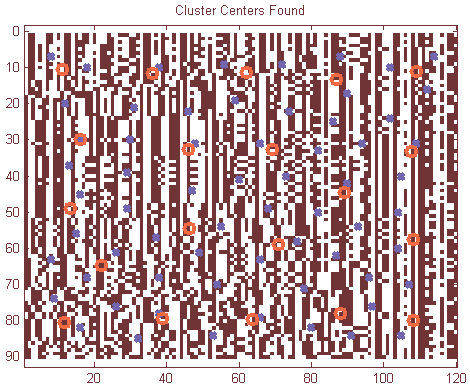


Fig. 5 Cluster centers found by the two clustering algorithms

The main objects are small errors of data page and short process time taken by each clustering algorithm. For these objects parameter values are selected and simulated. About the cluster centers found, intensities of pixels by a specific range are decreased by 0.2. The ranges are different about clustering algorithms, 4 about subtractive clustering and 9 about C-mean clustering. Figure 6 is the first data page with artificial IPI noise made by 2-dimensional Fourier Transform.

Figure 7 is the first data page modified with subtractive clustering algorithm and figure 8 is modified with C-mean clustering algorithm.

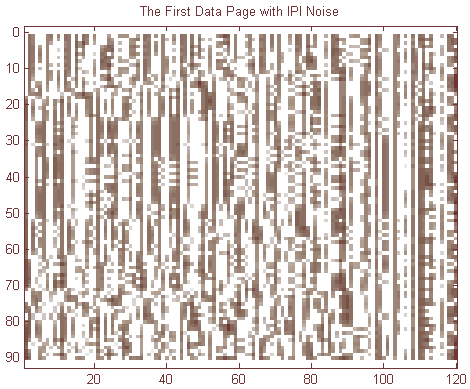


Fig.6 The first data page with artificial IPI noise

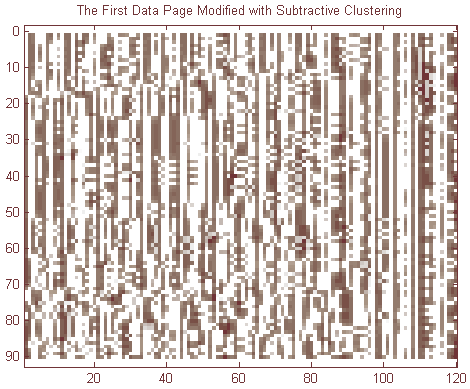


Fig.7 Data page modified with subclustering

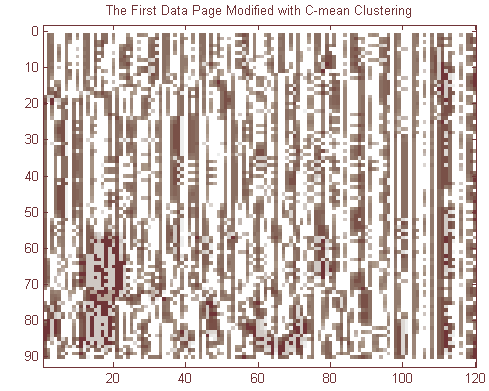


Fig.8 Data page modified with C-mean clustering

Apply the error reduction process to all data pages and reconstruct the picture. Figure 9 is reconstructed picture modified with subtractive clustering and figure 10 is with C-mean clustering.



Fig.9 Reconstructed picture using subtractive clustering

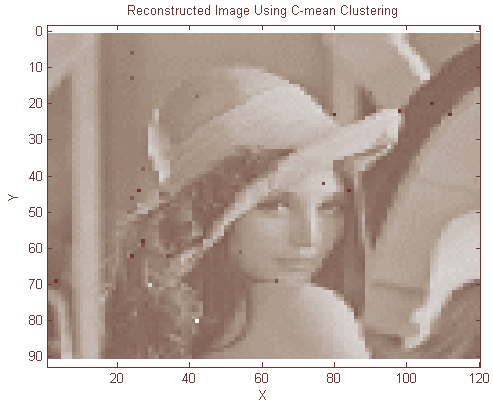


Fig.10 Reconstructed picture using C-mean clustering

Number of error pixels of figure 9 is 67 and it takes 123.31 time in arbitrary unit. About figure 10 number of error pixels is 75 and 100.5 time consumed.

When using subtractive clustering algorithm, number of error pixels is less than the case of using C-mean clustering algorithm but it takes more time about 1.25 times.

4. CONCLUSIONS

Inter Pixel Interference is one of a main noise sources in holographic data storage system. And it is required that making the intensity profile of data page to be as uniform as possible. By using clustering algorithm, clusters of on-pixels were found and intensities were decreased by a specific value. It was verified that number of error pixel was decreased through this procedure from simulation results.

It takes a time to proceed that, the algorithm which consumes less time is better. Therefore two clustering algorithms were simulated, subtractive clustering and C-mean clustering algorithm. Number of error pixels was smaller in subtractive clustering than C-mean clustering case but process time was longer. However the data page to be stored should be changed and each data page has its own characteristics, so subtractive clustering algorithm is better than C-mean clustering algorithm.

ACKNOWLEDGMENTS

This research was supported by the MOCIE (Ministry of Commerce, Industry and Energy) of Korea through the program for the Next Generation Ultra-High Density Storage (00008145) and the Korea Science and Engineering Foundation (KOSEF) through the Center for Information Storage Device (CISD) Grant No. R11-1997-042-11002-0.

REFERENCES

[1] Hans J.Coufal, Demetri Psaltis and Glen T. Sincerbox, *Holographic Data Storage*, Springer, 2000.

[2] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, New Jersey, 2002.

[3] ChangYong Yoon, JungHo Park and Mignon Park, "Face recognition Wavelets and Fuzzy C-Means clustering," *Proceedings of the IEEK Conference*, pp. 583-586, 1999.

[4] Muriel Bowie, "Fuzzy Clustering, Feature Delection, and Membership Function Optimization", Department of Informatics University of Fribourg, Switzerland, Seminar Paper 2004.

[5] Min-Soo Kim and Seong-Gon Kong, "Fuzzy Control System Design by Data Clustering in the Input-Output Subspaces" *Journal of the Korean Institute of Telematics and Electronics*, Von. 34S, No. 012, pp.30-40, 1997.