



CLASSIFICATION OF BIOMEDICAL IMAGES USING CONTENT BASED IMAGE RETRIEVAL SYSTEMS

Yinghui Zhang¹, Fengyuan Zhang², Yantong Cui³, Ruoci Ning⁴

¹ University of California, Berkeley, USA

² Northeastern University, China

³ Anshan No.1 High School, China

⁴ Marian High School, USA

Abstract:

Because of the numerous application of Content-based image retrieval (CBIR) system in various areas, it has always remained a topic of keen interest by the researchers. Fetching of the most similar image from the complete repository by comparing it to the input image in the minimum span of time is the main task of the CBIR. The purpose of the CBIR can vary from different types of requirements like a diagnosis of the illness by the physician, crime investigation, product recommendation by the e-commerce companies, etc. In the present work, CBIR is used for finding the similar patients having Breast cancer. Gray-Level Co-Occurrence Matrix along with histogram and correlation coefficient is used for creating CBIR system. Comparing the images of the area of interest of a present patient with the complete series of the image of a past patient can help in early diagnosis of the disease. CBIR is so much effective that even when the symptoms are not shown by the body the disease can be diagnosed from the sample images.

Keywords: *Biomedical Images; Content-Based Image Retrieval (CBIR); Gray-Level Co-Occurrence Matrix.*

Cite This Article: Yinghui Zhang, Fengyuan Zhang, Yantong Cui, and Ruoci Ning. (2018). "CLASSIFICATION OF BIOMEDICAL IMAGES USING CONTENT BASED IMAGE RETRIEVAL SYSTEMS." *International Journal of Engineering Technologies and Management Research*, 5(2), 181-189. DOI: 10.5281/zenodo.1186565.

1. Introduction

1.1. Content-Based Image Retrieval

It is a process in which an image is given to a system and searching is performed on the complete dataset of images to find the similar images based on color, size, texture. Histogram Equalization, Standard Deviation, Discrete wavelet transform, HSL (Hue, Saturation, Lightness) and HSV (Hue, Saturation, Value) are some of the parameters that can be used for comparison of various feature matching techniques [1].

1.2. Histogram

Histograms are a type of bar plot for numeric data that group the data into bins. In Matlab histogram of an image is calculated using the following syntax [2]:

- ***imhist(I)***

Calculates the histogram for the intensity image I and displays a plot of the histogram. The number of bins in the histogram is determined by the image type
[https://in.mathworks.com/help/images/ref/imhist.html#buo3qek-2_1]

1.3. Gray-Level Co-Occurrence Matrix (GLCM)

Also known as the gray-level spatial dependence matrix is a statistical method for examining the texture using the spatial relationship of the pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixels with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix[3].

In Matlab gray-level co-occurrence matrix from the image is created using the following syntax:

- ***glcms = graycomatrix(I)***

This will create a gray-level co-occurrence matrix (GLCM) from the image I.

- ***glcms = graycomatrix(I,Name,Value,...)***

This statement will return one or more gray-level co-occurrence matrices, depending upon the values of NAME/Values.

graycomatrix creates the GLCM by calculating how often a pixel with gray-level (grayscale intensity) value i occurs horizontally adjacent to a pixel with the value j .

1.4. Correlation Coefficient

For matching the percentage comparison between the two images, correlation coefficient is calculated between them. The function for calculating the correlation coefficient in Matlab is [4]:

- ***r = corr2(A, B),***

Which returns the correlation coefficient r between A and B, where A and B are matrices
corr2 computes the correlation coefficient using the following Algorithm

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}}$$

where $\bar{A} = \text{mean2}(A)$, and $\bar{B} = \text{mean2}(B)$

2. Applications of CBIR

- 1) An information system can be developed related to the **Biodiversity**, that will contain information of various species in the form of images. In such a system an input in the form of an image can be given and similar images can be searched based upon various characteristics like size, texture, color, etc.
- 2) CBIR can help in the investigation of the **criminal cases**. Images related to a case can be uploaded to the database and then identification of a person or place can be searched accordingly.
- 3) **E-commerce** can use CBIR system that will recommend the relevant product to the customer based on feature extraction corresponding to color, shape texture as per the user's liking.
- 4) CBIR can be developed for **medical applications** such that treatment of the diseases, diagnosis of the disease at early stage etc. can be done in a more efficient manner [5].
- 5) **Remote sensing, weather forecasting, surveillance, historical research** are other research areas where content-based applications can be used for automation of the searching process and finding crucial information from vast data.

3. Literature Survey

Research Paper	Work Done
Rani et al. [6]	The main objective of their work is to use Support Vector Machine in a more efficient manner so that results can be produced in lesser execution time. In their proposed work, images are pre-processed before they are stored in the database. This process helps to enhance the quality of the images by removal of the noise. Then RGB model is used for clustering the images. Again, re-clustering using support vector machine algorithm (SVM).
Chaudhary et al. [7]	In their proposed work, integrated approach is used to extract color and texture feature from images. The help of the multi-featured extraction is taken so that efficient image retrieval process can be performed. A higher order of color moment is used for the extraction of color features. Texture extraction and face recognition are done using local binary pattern (LBP).
Trojcanec et al. [8]	In their proposed work, an improvement of the two-level CBIR architecture is performed. In the first level, the clinician's voting is included, and in the second level, inclusion to the previously computed agents' voting is performed. The main motive of these two levels is to increase the efficiency of the proposed work and accuracy of the retrieval process.
Antaniet al. [9]	Their research work is focused on developing techniques for hybrid text/image retrieval from the survey of text and image data. Their research also shows various challenges that come into existence, when a particular CBIR system is developed especially for the biomedical images.

Khan et al. [10]	<p>This research gives a comparison of three different approaches of CBIR based on image features and similarity measures taken for finding the similarity between two images. Main conclusions are as follows:</p> <ol style="list-style-type: none"> 1. <i>A Novel Fusion Approach</i> (NFA) incorporated both local and global level features and has fused them for retrieving images. So, it contained properties of both local image descriptors as well as global image descriptors. 2. <i>A Universal Model</i> (UM) method used only local features to construct a feature vector. 3. <i>A Genetic Programming Framework</i> (GPF) considered only global image features
Aloniet al. [11]	<p>In their proposed work a novel approach using classification is driven and learning based frame work is developed for the retrieval of the medical images having different modalities. A direct link between the classification and retrieval process is established for efficient working. Relevance feedback (RF) based technique is used for the updating of the feature weights based on feedback given by the user. The authorhave also used multiclass support vector machine (SVM) classifier for category prediction.</p>
Ramos et al. [12]	<p>In their proposed work in radiology, reports are done which supervise the CBIR. Inferring the relationship between the patients and subsequently applying them to supervise the metric learning algorithms are the main phases of the proposed work. Their research work calculates the text distances between exam reports in the exam image space to supervise a metric learning algorithm. Using this method, there is a consistent increase in the CBIR performance.</p>

4. Present Work

4.1. Dataset Characterization

In the present work, the data set from National Biomedical Imaging Archive (NBIA) is used. NBIA is a searchable repository of in vivo images that provides the biomedical research community, industry, and academia with access to image archives. Dataset corresponding to the cancer images for Breast having protocol name Bilat Breast Lesion Vib is used in the present work for observations [13].

4.2. Work flow of Proposed Work

- 1) Collection of the images from the repository of National Biomedical Imaging Archive (NBIA).
- 2) Select and image say X corresponding to which similar images are to be searched.
- 3) Draw Histogram of this image.
- 4) Input the percentage corresponding to which matching is being done.

- 5) Select the first image for the repository, calculate its GLCM and HISTOGRAM.
- 6) Calculate correlation coefficient of the these to images.
- 7) If the value of the correlation coefficient matrix is equal to or less than the image X, **then both the images are same.**
- 8) Load next image and repeat from Step No.5.

Diagrammatical representation of the workflow is shown in Fig.1.

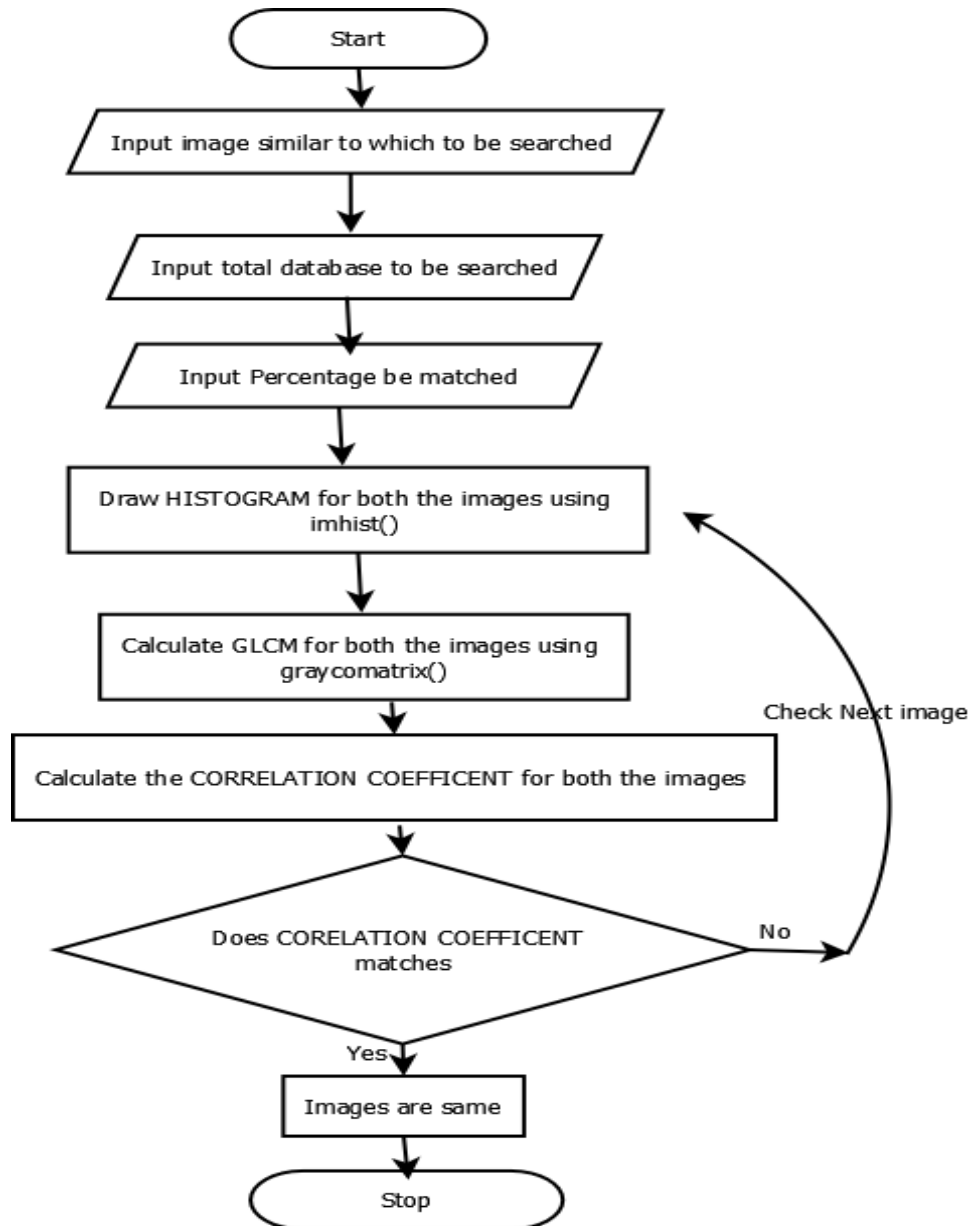


Figure 1: Workflow of the proposed work is as follows

Figure 2, 3, 4 represents images of a random image corresponding to Histogram, image features using GLCM. Table 2. Represents GLCM values corresponding to the input image

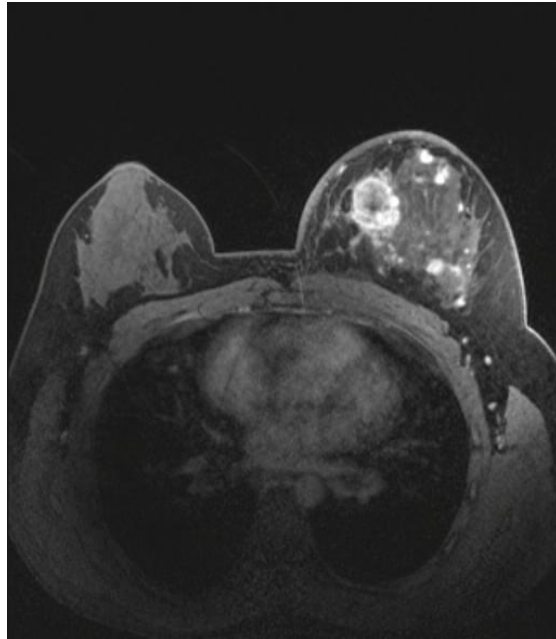


Figure 2: Representing sample image under observation

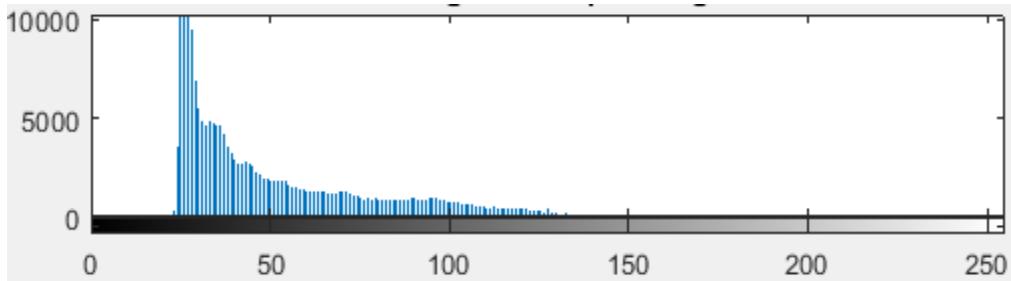


Figure 3: Representing Histogram corresponding to input image

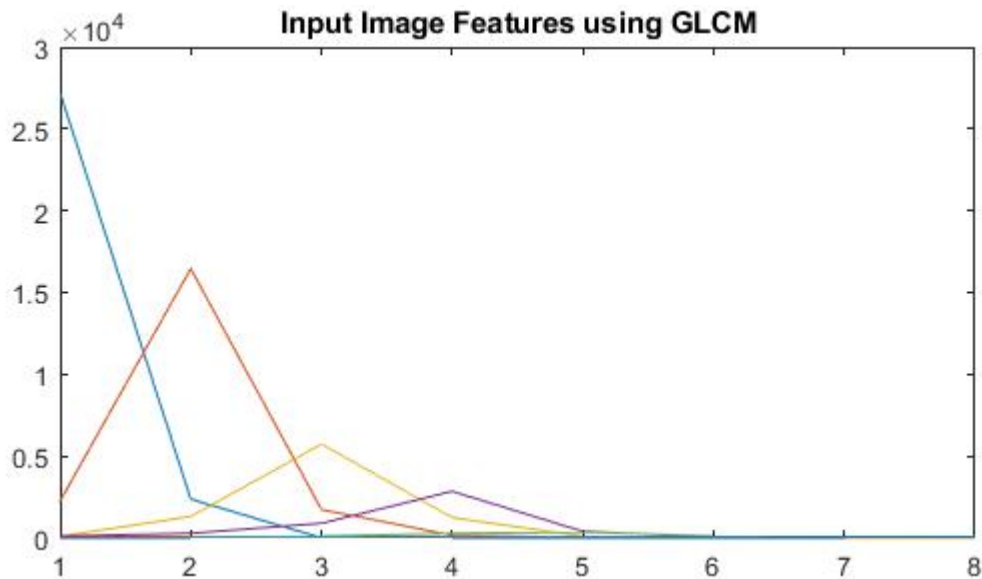


Figure 4: Representing features of the input image using GLCM

Table 2: Representing GLCM values corresponding to the input image

27218	2250	94	100	15	9	0	0
2363	16455	1311	278	94	18	3	0
2	1729	5733	891	106	28	3	4
1	163	1236	2846	272	55	21	3
0	21	97	394	360	73	34	17
0	3	19	63	109	85	43	32
0	2	6	13	28	53	70	44
0	1	0	12	12	33	42	57

4.3. Results

Fig.4, 5 represents proposed work in execution for input image and the corresponding output.

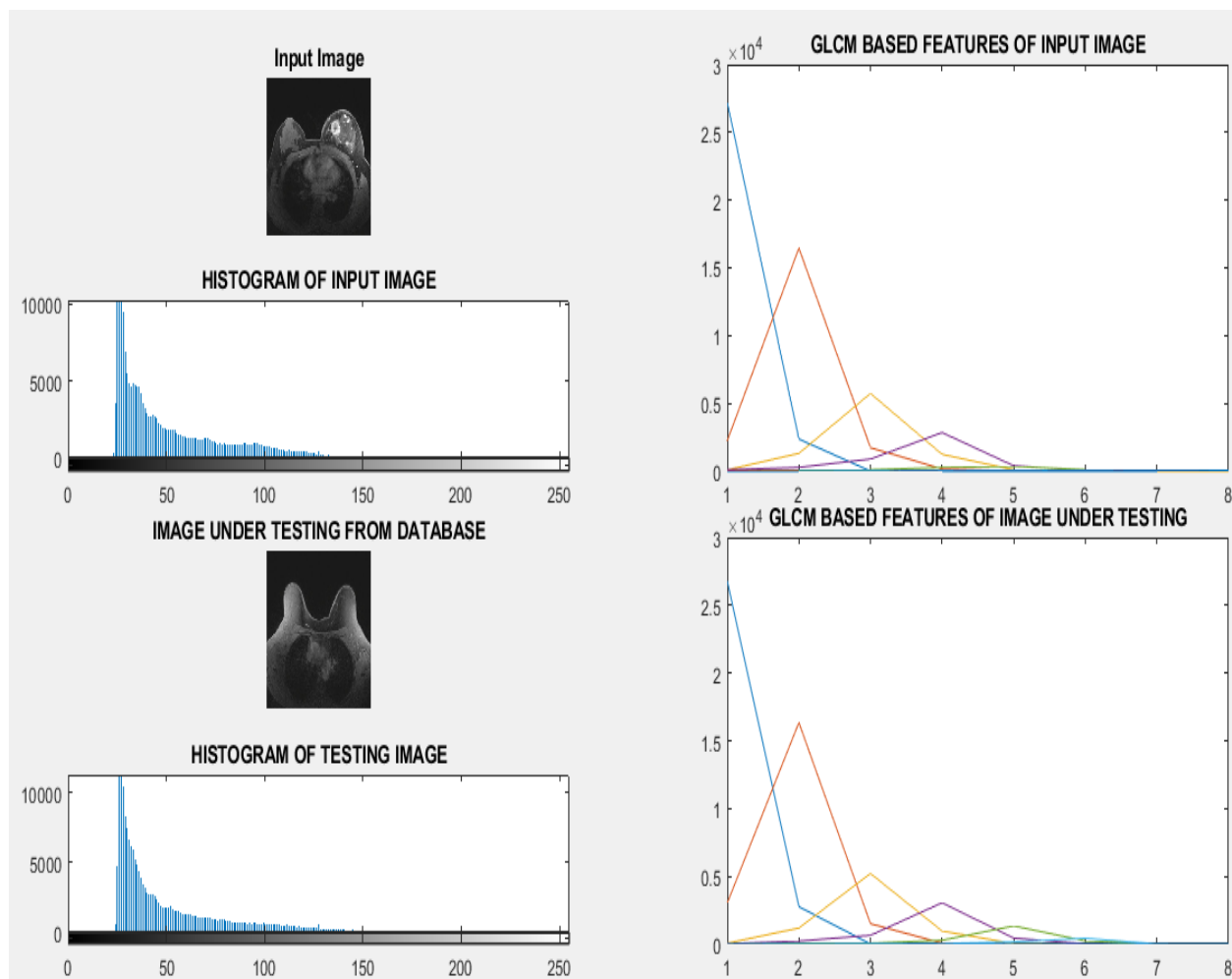


Figure 5: Represents proposed work in execution for input image



Figure 6: Represents output of the proposed corresponding to the input image

5. Conclusion and Future Scope

In the present work, CBIR is implemented using LCM and is used to search the similar images for various images of breast cancer. The proposed system can be very much helpful for the physicians. By matching the images of the present patient with the previous patient, diagnosis can be done in the early stage, and most efficient medicine can be prescribed accordingly. Present work implements sequential matching of each image one by one, and in future work, other programming languages having multithreading support [14] can be used so that parallel matching can be started at the time same time and execution time can be decreased. Use of Cache efficient algorithms can be done for faster execution [15] [16]. Hadoop based image processing techniques are also much effective to parallelize the tasks [17].

References

- [1] M. Madugunki, D. S. Bormane, S. Bhadoria and C. G. Deth, "Comparison of different CBIR techniques," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, 2011, pp. 372-375.].
- [2] "Histogram of image data," Math Works, [Online]. Available: [Accessed 18 February 2018] https://in.mathworks.com/help/images/ref/imhist.html#buo3qek-2_1
- [3] "Texture Analysis Using the Gray-Level Co-Occurrence Matrix," Math Works, [Online]. Available: [Accessed 18 February 2018] <https://in.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-gldm.html>.
- [4] "2-D correlation coefficient," [Online]. Available: [Accessed 18 February 2018] <https://in.mathworks.com/help/images/ref/corr2.html>.
- [5] K. Trojancanec, I. Dimitrovski and S. Loskovska, "Content based image retrieval in medical applications: an improvement of the two-level architecture," IEEE EUROCON 2009, St.-Petersburg, 2009, pp. 118-121.
- [6] Rani, Deepu, and Monica Goyal. "A Research Paper on Content Based Image Retrieval System using Improved SVM Technique." International Journal of Engineering and Computer Science 3.12 (2014).

- [7] Choudhary, Roshi, et al. "An integrated approach to content-based image retrieval." *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on.* IEEE, 2014.
- [8] Trojancanec, Katarina, Ivica Dimitrovski, and Suzana Loskovska. "Content based image retrieval in medical applications: an improvement of the two-level architecture." *EUROCON 2009, EUROCON'09.* IEEE. IEEE, 2009.
- [9] Antani, Sameer K., L. Rodney Long, and George R. Thoma. "Content-based image retrieval for large biomedical image archives." *Medinfo.* 2004.
- [10] Khan, Sumaira Muhammad Hayat, Ayyaz Hussain, and Imad Fakhri Taha Alshaikhli. "Comparative study on content-based image retrieval (CBIR)." *Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on.* IEEE, 2012.
- [11] Aloni, Sukhada. "Content Based Image Retrieval in Biomedical Images Using SVM Classification with Relevance Feedback." *International Journal of Scientific and Research Publications* 3.11 (2002): 1-7.
- [12] Ramos, José, et al. "Content-based image retrieval by metric learning from radiology reports: Application to interstitial lung diseases." *IEEE journal of biomedical and health informatics* 20.1 (2016): 281-292.
- [13] "National Biomedical Imaging Archive," [Online]. Available: [Accessed 18 February 2018] <https://imaging.nci.nih.gov/ncia/login.jsf>.
- [14] Bagga, Sachin, Akshay Girdhar, and Munesh Chandra Trivedi. "SPMD based time sharing intelligent approach for image denoising." *Journal of Intelligent & Fuzzy Systems* 32.5 (2017): 3561-3573.
- [15] Kaur, Manmeet, Akshay Girdhar, and Sachin Bagga. "Cluster Based Approach to Cache Oblivious Average Filter Using RMI." *The International Symposium on Intelligent Systems Technologies and Applications.* Springer International Publishing, 2016.
- [16] Bagga, Sachin, et al. "RMI Approach to Cluster Based Cache Oblivious Peano Curves." *Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on.* IEEE, 2016
- [17] Kaur, Gurbinder, Sachin Bagga, and Kulvinder Singh Mann. "Hadoop Approach to Cluster Based Cache Oblivious Peano Curves." *Advance Computing Conference (IACC), 2017 IEEE 7th International.* IEEE, 2017.