opción

Revista de Antropología, Ciencias de la Comunicación y de la Información, Filosofía, Lingüística y Semiótica, Problemas del Desarrollo, la Ciencia y la Tecnología

Año 35, 2019, Especial N° 20

Revista de Ciencias Humanas y Sociales

ISSN 1012-1587/ ISSNe: 2477-9385

Depósito Legal pp 198402ZU45

Universidad del Zulia

Facultad Experimental de Ciencias

Departamento de Ciencias Humanas

Maracaibo - Venezuela

Opcion, Año 35, Nº Especial 20 (2019):*2899-2921*
ISSN 1012-1587/ISSNe: 2477-9385

Memory Management Productivity in Big data Multimedia using Fireworks Algorithm

Jamal N. Hasoon[1], Assist. Prof. Dr. Rehab Hassan[2]

[1] Department of Computer Science, University of Almutansiryah, jamal.hasoon@uomustansiriyah.edu.iq
[2]Department of Computer Science, University of Technology, 110019@uotechnology.edu.iq

Abstract
The using of Big Data applications increased exponentially in the last few years. There are some knowledge extracted from huge volumes of data has been concerned by several enterprise. The working in Big Data faced challenges and should do hard. As a result, various types of distributions and technologies have been developed. This paper present a method for memory mapping of image depend on some aggregation function and optimization these function using Fireworks algorithm to reduce the size of memory required in searching process. Image features extraction used in aggregation function and used later for optimization. Three types of feature extraction used image descriptors techniques. The proposed method provides abstract domain that replaced total data domain and the performance increase the system throughput and reduce cost.

Keywords: Big data, Cloud Computing, Firework algorithm, feature extraction

Productividad de gestión de memoria en Big data Multimedia usando el algoritmo de Fireworks

Resumen
El uso de aplicaciones Big Data aumentó exponencialmente en los últimos años. Hay algunos conocimientos extraídos de grandes volúmenes de datos que han sido preocupados por varias empresas. El trabajo en Big Data enfrentó desafíos y debería ser difícil. Como resultado, se han desarrollado varios tipos de distribuciones y tecnologías. Este artículo presenta un método para el mapeo de memoria de la imagen que depende de alguna función de agregación y la optimización de estas funciones utilizando el algoritmo de Fireworks para reducir el tamaño de la memoria requerida en el proceso de búsqueda. La extracción de características de imagen se utiliza en la función de agregación y luego se utiliza para la optimización Tres tipos de extracción de características utilizan técnicas de descriptores de imágenes. El método propuesto proporciona un dominio abstracto que reemplazó el dominio de datos totales y el rendimiento aumenta el rendimiento del sistema y reduce el costo.

Palabras clave: Big data, Cloud Computing, algoritmo de Firework, extracción de características

1. Introduction

The era of Big data has caused a new methods aimed at efficiently using enormous amounts of data ,thus developing big data applications has become gradually significant in the latest years to support managing and knowledge detection events. The organization and investigation of this big data is a great job for old-fashioned data and data warehouse. The examination of big data continuously associated with managing data to come up with appropriate styles and arrangements that can be used to make decisions, enrich processes, and even drive a new pattern. The ideas of cloud computing relies on sharing the resource to achieve uniformity, but big data systems (Hadoop) are built on shared nothing principle, where every node is self-sufficient and independent. By integrating cloud computing technologies, big data, businesses and education institutes could have an improved path in the upcoming [1]. The ability in managing storing huge amounts of data in diverse structure with less processing time will result in data that can guide businesses and educational institutes in rapidly developing [2]. The moving to the cloud is a big issue regarding confidentiality and safety concerns that is the chief reasons, why companies and educational institutions is not moving to the cloud. In the following sections will have more about the characteristics, trends and challenges of big data. It explores the profits and the risks that may increase out of the integration between big data and cloud computing [3].

2. Perfect Combination from Cloud Computing and Big Data

Cloud Computing is an environment established on employing and offering services. The service-oriented systems can be clustered, therefore different categories will be there. One of the most used criteria to cluster these systems is the abstraction level that is offered to the system user. Three diverse levels are frequently distinguished: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) as shown in Figure 1 from the top to the bottom [4].

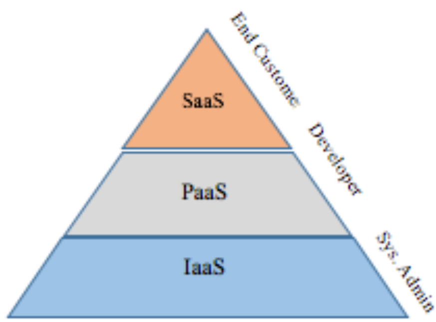

Figure 1: three levels services in

Some resolutions are proposed to deal with storing and retrieving big amounts of data that needed by Big Data and, some of them are presently used in Clouds. Internet-scale file systems such as the Google File System (GFS) attempts to afford the robustness, scalability, and dependability that assured Internet services need [5]. Further resolutions offer object-store abilities where files can be pretend across multiple geographical sites to enhance idleness, data availability, and scalability. Examples include Nirvanix Cloud Storage, Amazon Simple Storage Service (S3), Azure Binary Large Object (Blob) storage, and OpenStack Swift and Windows. Though those resolutions offer the redundancy and scalability that most of the Cloud applications want, they occasionally do not encounter the con- currency and performance needs of definite analytics applications. The data locality is the key aspect in providing performance for Big Data analytics applications. Because the volume of involved data, makes it excessive to transmit the data for handling [6]. This was the ideal option in usual high performance computing systems: in such systems, that usually concern execution CPU-intensive calculations over a reasonable to average volume of data, it is possible to transmit data to the computing units, since the fraction of data transfer time to processing time is minor. But, in the situation of Big Data, moving data to computation nodes would generate large ratio of data transfer time to processing time. Accordingly, a diverse method is chosen, where computation is done to the data located. The similar method of traveling data locality was discovered previously in scientific workflows and in Data Grids [7].

3. Image Features

The extraction of image featursse involves reducing the amount of resources required to describe a large set of data [8]. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power; also, it may cause a classification algorithm to over fit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy [9].

• Statistical Features

The features of an image can be extracted using some statistical method in first order and second order. These statistical features are mean, median, standard deviation, skewness, kurtosis, and co-efficient of variation. While in second order, the co-occurrence matrix used to find second moment features. The main of them are Contrast represent the intensity contrast between a pixel and its neighbor over the whole image. Correlation represent the correlation of a pixel to its neighbor over the whole image. Energy that represent the uniformity the sum of squared elements in the GLCM. Distribution of elements in the GLCM to the GLCM diagonal, Homogeneity is the closeness [10].

• Histogram Features

Histogram attribute extraction is considered as a global image feature extraction [11]. A color histogram is employed as the features classification, which gives an approximation of the distribution of the colors in the image. These methods will use to compare the images with other images. The number of the selected bins and the nearby values will be quantized to these bins.

• Local Binary Pattern Features

In visual descriptor, a Local binary patterns (LBP) is considered as a type for it[12]. It has since been found, it is a powerful feature for texture classification; it has further been determined that when LBP is combined with the histogram of oriented gradients (HOG) descriptor, it improves the detection performance considerably on some datasets. A useful extension to the original operator is the so-called uniform pattern, which is used to decrease the length measurement of the feature vector and apply a simple rotation invariant descriptor.

4. Fireworks Algorithm

Explosion of fireworks in night simulate in algorithm called FWA that produced sparks around each elimination [13] as shown in figure 1. It considered as Swarm intelligence algorithm, local search presented by each firework that spark around some positions in

range called amplitude. The new populations cooperated to find global search. FWA contains some characteristics that verify it from other algorithms such as simplicity, locality, distribute parallelism, diversity and extendibility [14]. The basic idea of FWA works as follow: first is initializing randomly N fireworks with their fitness value (quality), every one calculated to find the number of sparks and the explosion's range (explosion amplitude). The new generation is obtained by explosion the previous fireworks and everyone is considered as local search. The amplitude balance the global and local space and ensure diversity by generating large population in small range and small population in large range, when low fitness value, this technique give chance to escape from local minima. The mutation in FWA called Gaussian mutation or "Gaussian Spark" and the selection strategy may use subset of whole population in each iteration to balance the global search [15].

5. Explosion Operator

For each firework in fireworks need, explosion operator that select a number of sparks in explosion operator is found using equation 1 as follow:

$$S_i = m \times \frac{y_{max} - f(x_i) + \varepsilon}{\sum_{i=1}^{N}(y_{max} - f(x_i)) + \varepsilon} \ldots 1$$

Where S_i is number of spark in each firework, the total number of individual in firework is (m), the total number of spark is (N), y_{max} is the worst fitness value in population. $f(x_i)$ is the fitness value of xi, and ε is small value for avoiding division by zero. The amplitude of explosion is found using equation 2 as follow:

$$A_i = \hat{A} \times \frac{f(x_i) - y_{min} + \varepsilon}{\sum_{i=1}^{N}(f(x_i) - y_{min}) + \varepsilon} \ldots 2$$

Where A_i *is a* range (amplitude) of each one is, $\hat{A}$ is a fixed value specified (may be sum of all amplitude), and ε is used to same reason before.

These parameters used find distance on each firework using equation 3 as follow:

$$x_i^k = x_i^k + U(-A_i, A_i) \ldots 3$$

Where $U(-A_i, A_i)$ is random number in uniform order in the amplitude Ai interval.

The total steps for generation spark shown in algorithm 1.

6. Mutation Operator

The selected individual or the current individual are used to applied mutation operation on it as x_i^k, where i represent the interval from $(1 \ldots N)$ and *the* dimension of current state is represented by (k). The sparks of Gaussian explosion are calculated by using equation 4 as follow:

$$x_i^k = x_i^k \times RndGauss(1,1) \ldots 4$$

Where *RndGauss* is random number in Gaussian distribution. Gaussian mutation in FWA explain in algorithm 2.

7. Mapping Rule

The mapping rule is a process that keeps all the individuals in the accepted range of the population. Any individual result from FWA operation return inside range space if it lie out boundaries by applying modular arithmetic operation. The mapping rule utilizes a modular operation as stated as in the following equation (5):

$$x_i^k = X_{LBound,k} + x_i^k ArthMod(X_{UBound,k} - X_{LBound,k}) \ldots 5$$

Where x_i^k denotes sparks positions that lie out of bounds, while X_{UBound} and X_{LBound} are represent the upper and the lower boundaries of spark position. "ArthMod" represents modular arithmetic.

8. Selection Method

The selection method may need to find the distance between individuals for this purpose may use Euclidean distance as measurement to find the nearest one as explain in equation 6.

$$R(x_i) = \sum_{i=1}^{K} d(x_i, x_j) = \sum_{i=1}^{K} \|x_i - x_j\| \ldots 6$$

Where d represent distance (Euclidean distance for example) between any two individuals x_i and x_j. This distance may be combined sparks results by explosion operator or mutation operator.

Roulette Wheel one of methods may be used in FWA to select new generation, depend on $P(x_i)$, that calculate in equation 7 as follow:

$$P(x_i) = \frac{R(x_i)}{\sum_{j \in k} R(x_j)} \ldots 7$$

High distances have more chances to be in new generation and increased the diversity of the population.

9. Proposed method

The proposed method are used the feature extraction of images that considered as big data to reduce the amount of memory required for searching. These features used as abstract dataset replaced the original data. The Firework algorithm used for optimization the features vector values that useful when searching on similar images in all dataset. The proposed method steps are illustrated in figure (2).

• Image Feature Extraction

Reduction dimensionality of image useful in image matching and retrieval. Feature extraction used some characteristics of image for find the similarities or differences between images. Feature extraction, and matching used to solve computer vision problems in object detection and recognition, content-based image retrieval, face detection and recognition, and texture classification. Three kinds of features used in this work and all optimized using FWA as follow:

o Statistical Features extraction

Use twenty-eight features in total for color image and applied on three-band color. This method used mean, standard deviation, skewness, and variance of each column in image (RGB). For all result vector used mean and standard deviation. These all used as first order while in second order used the hue band of image after transformation from RGB color domain to HSV color domain and used as gray level co-occurrence matrix (GLCM) then find contrast, correlation, energy, homogeneity.

o Histogram Features Extraction

In Histogram feature used twenty-four bins and other values quantized to this bins values. These bin value will normalized by divide the results with the summation of total image pixels to find the features. Normalization step make the features in accept range and can used in measurement of the similarity and differences of images.

o LBP Features

LBP feature clue is enthused by the fact that some binary shapes arise more frequently in texture images than others. If the binary pattern contains at most two 0-1 or 1-0 transitions, so is called uniform. Such as, 00010000(2 transitions) is a uniform pattern, and 01010100(6 transitions) is not. In the computation of the LBP histogram, the histogram has a separate bin for each uniform pattern, and a single bin for all non-uniform patterns.

• Group by Features

Group by applied on feature vector and find the unique value of each one. Group by method applied on all features vector that extracted before. It groups a selected set of rows into a set of summary rows by the values of one or more columns or expressions. For each group, one row is returned. In the select clause list , the aggregate functions provide information about each group instead of individual rows.

• FWA Optimization

The main step in proposed method is optimization step that used FWA for optimization the output of previous step. The first insert image features that equal to k and each feature represent by the centroid of each cluster. The second step select the number of sparks. The next step is to select the number of groups for each k,and the number of features that represent the dimension of the algorithm. After these steps, the number of iteration is selected. For each spark, find the upper and lower bound and these and for each spark specify the dimension of it. Apply Gaussian spark for each group in FWA and find fitness function that have minimum values which represent the best value. Update the group values for each iteration and select the best solution that used minimum memory mapping at all and be useful in searching for image information.

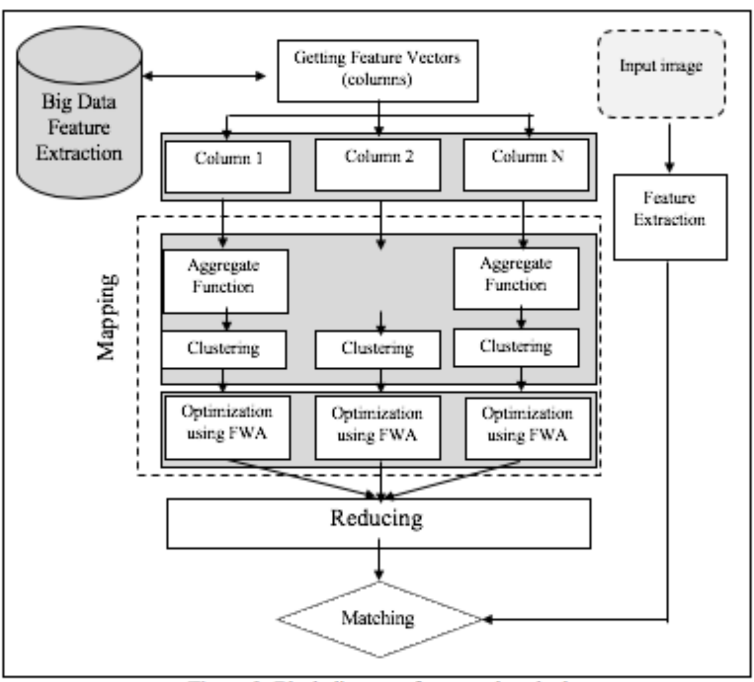

Figure 2: Block diagram of proposed method

10. Experimental results

In proposed method, three kinds of feature extraction applied to find the memory required of each kind of them. The first method of feature extraction was statistical feature, the second is histogram features, and the third method is local binary pattern features. The first method consist of twenty-eight features each column will first apply group by method as aggregation function as shown in figure (3). These will reduce required memory by used memory mapping. The FWA optimized required memory by

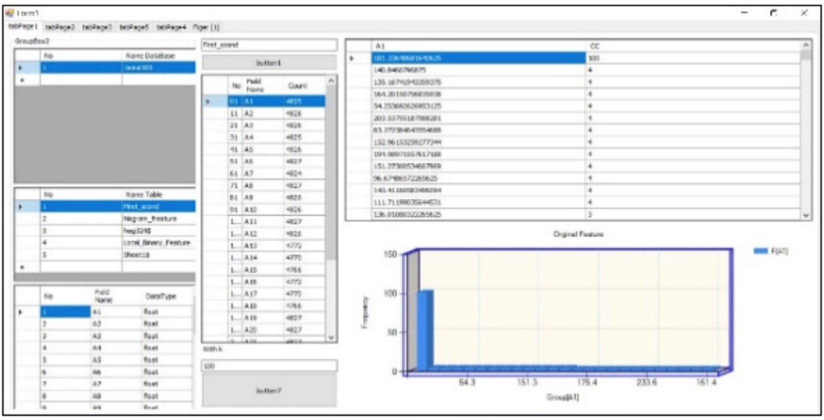

Figure (3): grouped by of Statistical Features

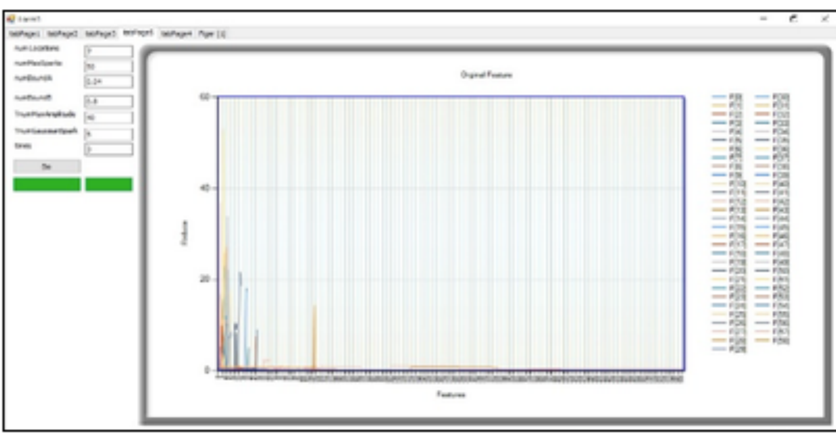

Figure (4): FWA Statistical features optimization

The second method consist of twenty-four features each column will first apply group by method as aggregation function. These will reduce required memory by used memory mapping. The FWA optimize the required memory by reducing the size of it.

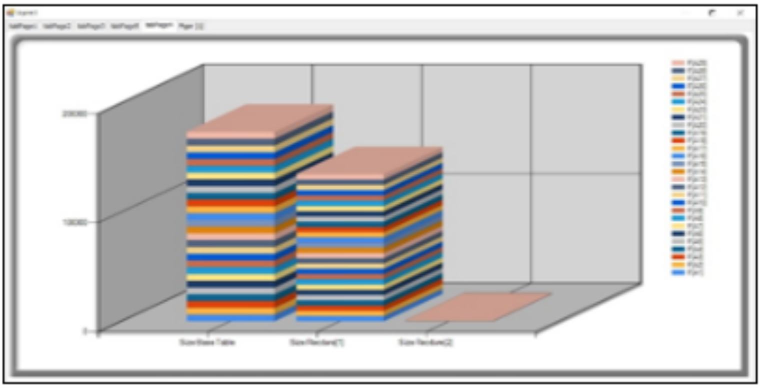

Figure (5): Reduction in total size of statistical features

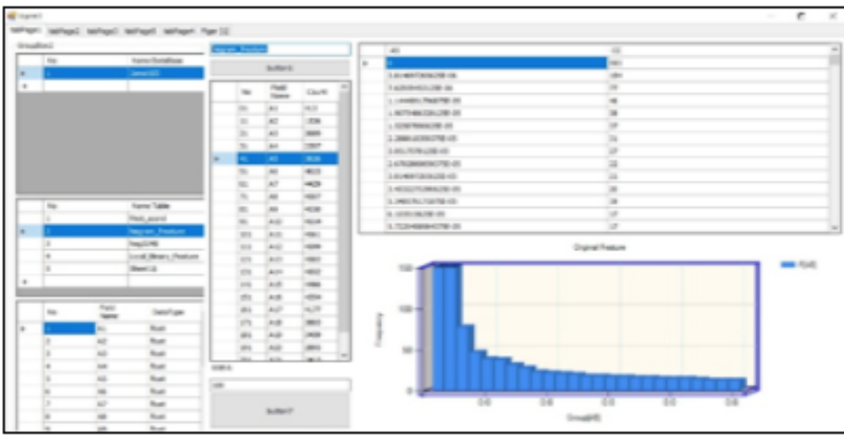

Figure (6): grouped by of histogram Features

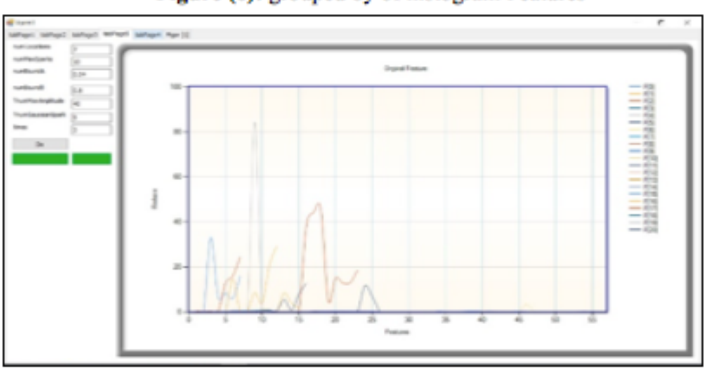

Figure (7): FWA histogram features optimization

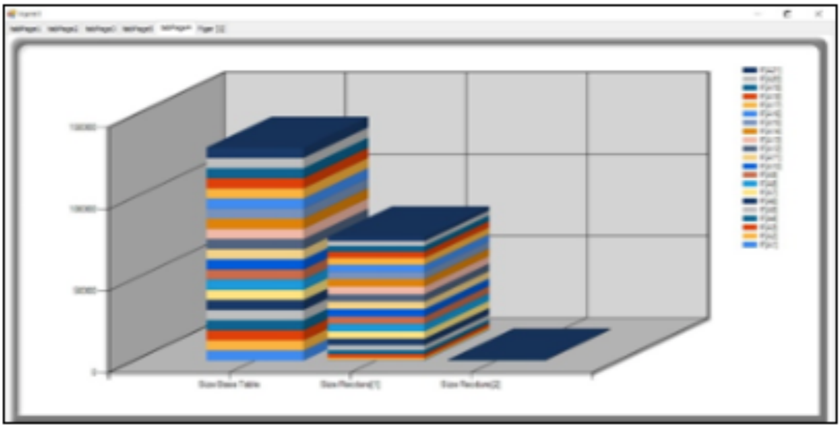

Figure (8): Reduction in total size of histogram features

The third method consist of sixty features each column will first apply group by method as aggregation function. These will reduce required memory by used memory mapping. The FWA optimize the required memory by reducing the size of it. The FWA optimize the required memory by reducing the size of it. The total steps of implementation of proposed method with LBP features explain in figure (9).

- Performance Measurement

The performance of FWA depend on several parameters such as the type of feature extraction used for description. Also the parameters used in FWA such number, of maximum sparks, the upper and lower bound, maximum amplitude, Gaussian mutation, number of iterations used to find the optimized value of each feature as aggregation function table (1) explain the result of features as a method of mapping.

Figure (9): Reduction in total size of LBP features

Feature extraction	No. of Records		Grouped by		FWA Optimization	
	F1	F2	F1	F2	F1	F2
Statistical Features	6200	6200	4825	4826	1357	1042
Histogram Features	6200	6200	4599	4561	987	871
LBP Features	6200	6200	4277	4827	843	822

Conclusion

The memory management productivity are present in this work depend on FWA as optimization method. It used for different kind of feature extraction. FWA implemented depending on the behaviour fireworks explosion and the amplitude of each spark. FWA present a better diversity solution through solving problem and high parallelism ability. The other characteristics is simplicity in mutation operator and in selection operator. FWA perform well and reduced required memory when compare the results with the using only aggregation function on the same image big data. To enhance convergence rate and the accuracy of FWA optimization problem may use different type of mutation method and may use dynamic evaluation on the parameters used in FWA iteration.

References

[1] Field Cady, “In-Memory Big Data Management and Processing: A Survey”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 7, JULY 2015.

[2] Yao Wang, LinBo Zhu, Jiwen Wang and Jianfeng Qiu, “The rise of “Big Data” on cloud computing: Review and open research issues”, Published February 3rd 2017 by Wiley.

[3] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan,” Big Data and Cloud Computing: Trends and Challenges”, this publication at: https://www.researchgate.net/publication/264624667, 2014.

[4] Samir A. El-Seoud, Hosam F. El-Sofany, Mohamed Abdelfattah and Reham Mohamed,” Big Data and Cloud Computing: Trends and Challenges”, International Journal of Interactive Mobile Technologies (IJIM) · April 2017.

[5] Dr.A.P.Mittal, Dr.Vanita Jain, and Tanuj Ahuja,“Google File System and Hadoop Distributed File System- An Analogy”, International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 4, Special Issue March 2015.

[6] Adriana Alexandru, Cristina Adriana Alexandru, Dora Coardos, and Eleonora Tudora,” Healthcare, Big Data and Cloud Computing”, WSEAS TRANSACTIONS on COMPUTER RESEARCH, 2017.

[7] Nader Gemayel, “Analyzing Google File System and Hadoop Distributed File System”, Research Journal of Information Technology, 2016.

[8] Atul Bansal, Ravinder Agarwal and R. K. Sharma “Statistical feature extraction based iris recognition system”, Indian Academy of Sciences, Vol. 41, No. 5, May 2016.

[9] Gaurav Kumar, and Pradeep Kumar Bhatia, “A Detailed Review

of Feature Extraction in Image Processing Systems", Fourth International Conference on Advanced Computing & Communication Technologies, 2014.

[10] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar," Image Texture Feature Extraction Using GLCM Approach", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.

[11] Hongming Zhang and Debin Zhao, "Spatial Histogram Features for Face Detection in Color Images", Springer-Verlag Berlin Heidelberg 2004

[12] Jae Young Choi, Konstantinos N. Plataniotis, and Yong Man Ro1, "Using Colour Local Binary Pattern Features for Face Recognition", IEEE 17th International Conference on Image Processing September 26-29, Hong Kong, 2010.

[13] Ke Ding, Shaoqiu Zheng, and Ying Tan, "A GPU-based Parallel Fireworks Algorithm for Optimization", GECCO'13, July 6-10, 2013.

[14] Shaoqiu Zheng, Andreas Janecek, Junzhi Li, and Ying Tan, "Dynamic Search in Fireworks Algorithm" the National Natural Science Foundation of China under grants number 61375119, 61170057 and 60875080, 2014.

[15] Xi-Guang Li, Shou-Fei Han, and Chang-Qing Gong, "Analysis and Improvement of Fireworks Algorithm", Academic Editors: Sergio Rajsbaum and George Karakostas, 2017

UNIVERSIDAD

DEL ZULIA

opción

Revista de Ciencias Humanas y Sociales

Año 35, N° 20, (2019)

Esta revista fue editada en formato digital por el personal de la Oficina de Publicaciones Científicas de la Facultad Experimental de Ciencias, Universidad del Zulia.
Maracaibo - Venezuela

www.luz.edu.ve

www.serbi.luz.edu.ve

produccioncientifica.luz.edu.ve