Hardware Accelerated Design on Bag of Words Classification Algorithm

[1]Chang-yong Lee, [2]Ji-yong Lee, [*3]Yong-hwan Lee
[1]*Kumoh National Institute of Technology, lcy42845220@gmail.com*
[2]*Siliconworks, jiyongend@naver.com*
[*3 Corresponding Author]*Kumoh National Institute of Technology, yhlee@kumoh.ac.kr*

Abstract

In this paper, we propose an image retrieval algorithm for real-time processing and design it as hardware. The proposed method is based on the classification of BoWs(Bag of Words) algorithm and proposes an image search algorithm using bit stream. K-fold cross validation is used for the verification of the algorithm. Data is classified into seven classes, each class has seven images and a total of 49 images are tested. The test has two kinds of accuracy measurement and speed measurement. The accuracy of the image classification was 86.2% for the BoWs algorithm and 83.7% the proposed hardware-accelerated software implementation algorithm, and the BoWs algorithm was 2.5% higher. The image retrieval processing speed of BoWs is 7.89s and our algorithm is 1.55s. Our algorithm is 5.09 times faster than BoWs algorithm. The algorithm is largely divided into software and hardware parts. In the software structure, C-language is used. The Scale Invariant Feature Transform algorithm is used to extract feature points that are invariant to size and rotation from the image. Bit streams are generated from the extracted feature point. In the hardware architecture, the proposed image retrieval algorithm is written in Verilog HDL and designed and verified by FPGA and Design Compiler. The generated bit streams are stored, the clustering step is performed, and a searcher image databases or an input image databases are generated and matched. Using the proposed algorithm, we can improve convenience and satisfaction of the user in terms of speed if we search using database matching method which represents each object.

Keywords: *Bag of Words, SIFT Algorithm, Bit Stream, Clustering, Image Classification, FPGA*

Ⅰ. Introduction

As the utilization of Internet search increases, the search methods are becoming more diverse. The most commonly used methods are keyword-based[1] and content-based[2]. The keyword-based method is a method of directly typing in a search window when a user wants to search on the Internet. The advantage of this method is that when the user knows the exact name or the word of the object to be searched, the user can obtain a desired result through a quick and accurate search. The disadvantage of this method is that it is difficult to get the desired search results if the user is not familiar with the name of the object or knows it abstractly. For example, if you want to search for items of interest through the visual media if you do not know your name or information of the products it is difficult to obtain the desired results. In addition, the exact word that you are looking for is entered into the search box, but not only the desired result, but also related results are searched. This is a feature created for the user, but it takes a long time to have the associated search results. The content-based image retrieval method complements the disadvantages of the keyword-based retrieval method. This method extracts data such as shape, texture, and color, which are unique information of an image, by dragging the image into the search window instead of inputting the keyword into the search window. Therefore, this method refers to a search method for determining the similarity between the unique information of

[*] Corresponding Author
Received: Dec. 12, 2018, Revised: Dec. 21, 2018, Accepted: Dec. 26, 2018
** 본 논문은 (사)ICT 플랫폼학회 2018 하계학술대회 우수 논문상 수상 논문임

the input data and the search database of the corresponding portal site. This method is advantageous in that the user can get the desired search results faster and more accurately than keyword-based method. The proposed method is a real-time image retrieval method based on the BoWs (Bag of Words) algorithm[3-4]. The BoWs method is a classifier for classifying documents, and it is necessary to grasp the distribution of words in a document, and then to identify the document. The proposed method uses the principle of BoWs to automatically classify images and use them as an image retrieval method. Fig. 1 shows a flow chart of the BoWs algorithm.

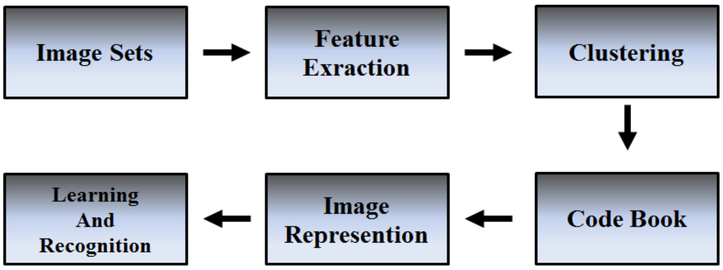

Figure 1. The flowchart of the BoWs algorithm

There is a problem with real-time processing when searching images with software using the BoWs algorithm. Thus, the feature points of each image are extracted to increase the search speed. Representative algorithms for extracting feature points include Scale Invariant Feature Transform (SIFT)[5], Fast Fourier Transform (FAST), and Speed Up Robust Features (SURF)[6] algorithms. The proposed algorithm used to extract feature points is SIFT algorithm. SIFT algorithm has the advantage of extracting feature points that are invariant to rotation and size. The feature points extracted using SIFT algorithm have a key point descriptor and a 128-dimensional real number value for each feature point. These real values are the main reason for slowing down real-time processing. In order to overcome these drawbacks, we propose a method of transforming a 128-dimensional real value having characteristic points, which are unchanged in magnitude and rotation, into a bit stream in order to process images at high speed through the SIFT algorithm. One feature point has one key point descriptor. In the key point descriptor, the bit stream is converted using the rotation and size information. These bit streams become new data representing the image, and when the bit stream is extracted for all the images, the clustering[7] and entropy are used to generate the final values representative of the respective images. In addition, deep learning can be applied to the proposed method, and more databases can be constructed to get desired results faster.

II. Image search algorithm

The image search algorithm uses two databases as the content-based retrieval method. One database is a database related to all the images of the searcher, and another one is a database of images inputted. Fig. 2 shows the block diagram of each database creation and algorithm based on the BoWs algorithm. The searcher image database extracts the feature points using the SIFT algorithm as shown in Fig. 2, and creates the database through the clustering step. The input image database creates the database in a manner similar to the searcher database, but does not use the clustering step.

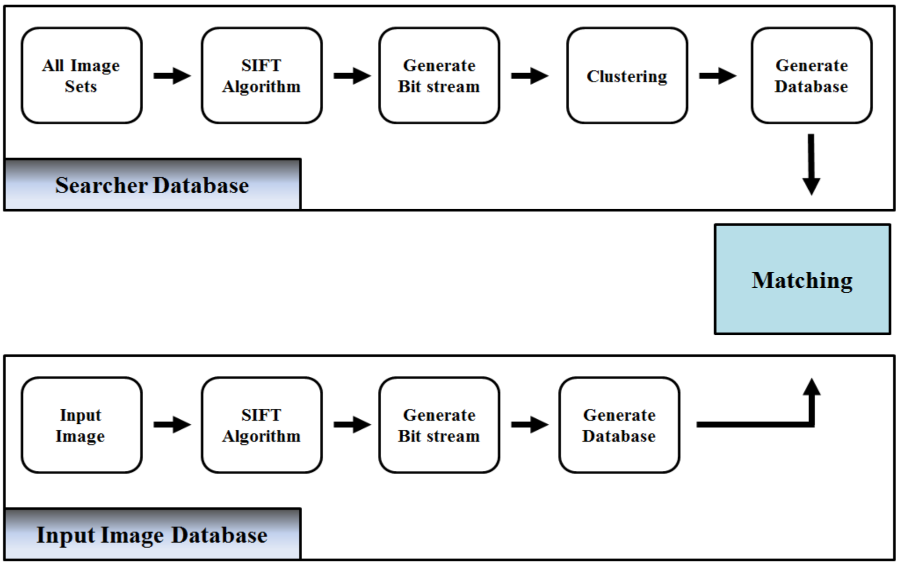

Figure 2. Each database block diagram

III. Algorithm verification

K-fold cross validation is used to test the performance of the image searcher algorithm. K-fold cross validation[8] is one of the typical methods for verifying the performance of datasets. In K-fold cross validation, the yellow area represents the database portion of the input image and the blue area represents the database portion of the searcher. Using the above method, perform the performance test excluding each yellow test image set from Round 1 to Round N. Finally, the average of each round is added to calculate the overall average. In this paper, seven classes are used, and each class has seven images. A total of 49 image data were tested for image performance and 2316 feature points of 49 images were extracted. The components of each class are ferry, camera, chair, piano, lamp, airplane, and watch.

Fig. 3 shows the performance test results using k-fold cross validation, and is the result of combining the bit-string length and the matching rate for the relationship of cluster. Each result is repeated 100 times with k-fold cross validation to show average matching rate.

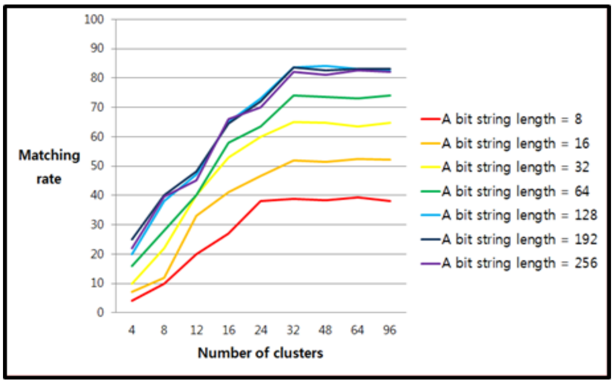

Figure 3. The result of performance test

IV. Hardware design

4.1. Hardware structure

Fig. 4 shows the design structure for the searcher database. The searcher database structure is largely divided into software, which is a preprocessing process, and hardware, which is a post-processing process. In the preprocessing process, each feature point from the image is detected by the SIFT algorithm using C-language and a bit string is generated. The generated data is stored in the single port ROM and the initial cluster values are stored in the single port RAM. Entropy step also uses a single port RAM. In the post-processing, clustering is performed and the clustering part that has the greatest effect on real-time processing is started. In the clustering process, clustering is performed for all bit strings. In the clustering process, the similarity of each bit string and cluster is calculated, and the final cluster value is obtained and a searcher database is generated.

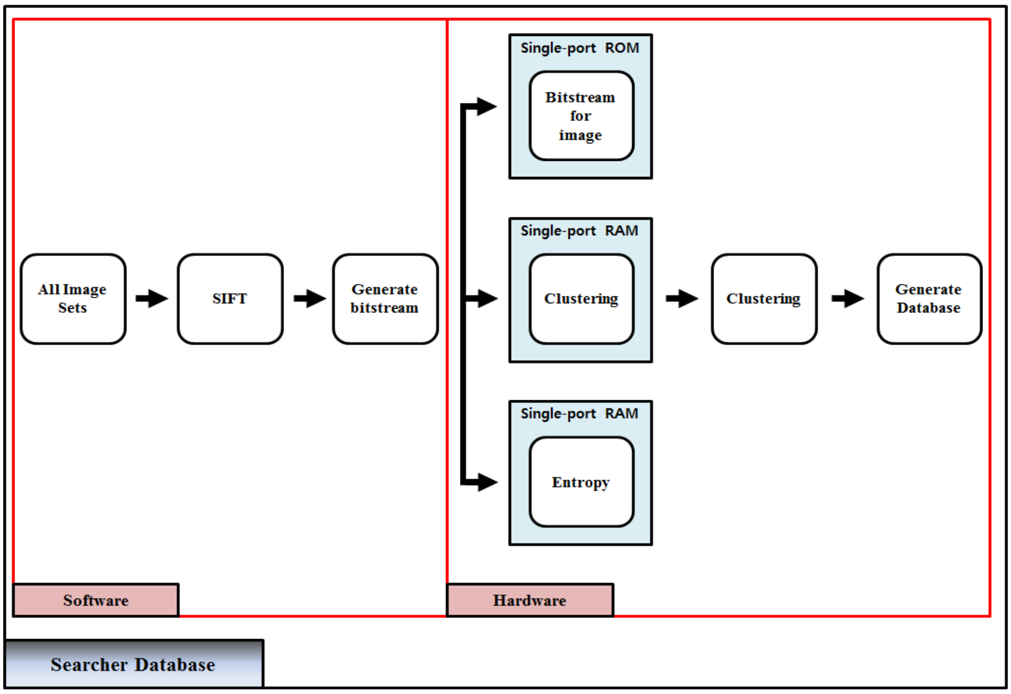

Figure 4. The design structure for the searcher database

4.2. Design Clustering

Fig. 5 shows the FSM(Finite State Machine) of hardware. One FSM is for clustering and one FSM is for database creation. The clustering FSM structure, which is the post-processing part for hardware structure design, is divided into 5 stages. stages are composed of 0, 1, 2, 3, and 4 stages. This process is terminated when the previous cluster value and the current cluster value are all equal. Otherwise, repeat the stage of step 5 to create a new cluster again. FSM structure for the database of images, when clustering is terminated and the searcher database and the input image database are generated by computing the final clustering value and all the images with the bit strings of the respective images. When a searcher database and an input image database are created, the searcher database and the input image database are matched using the Euclidian distance. The input image is searched for an image after matching. Clustering FSM Phase 5 is repeated until the value of the cluster is unchanged.

In Stage 0, clustering starts when the Start Enable signal is 1, and Stage 0 is maintained when the Start Enable signal is 0. In Stage 1, similarity calculation of each cluster and distance is performed for all bit strings in the bit string memory structure. Declaring a register as necessary for calculation of distance similarity can exponentially increase the circuit or affect the operating frequency. Therefore, to maximize the circuit size and operating frequency, the similarity distance value is calculated and the number of registers to temporarily store the calculated value is limited. When designing the comparator, the pipeline structure is designed to increase the operating frequency. Fig. 6 shows the structure of a four-stage pipeline comparator.

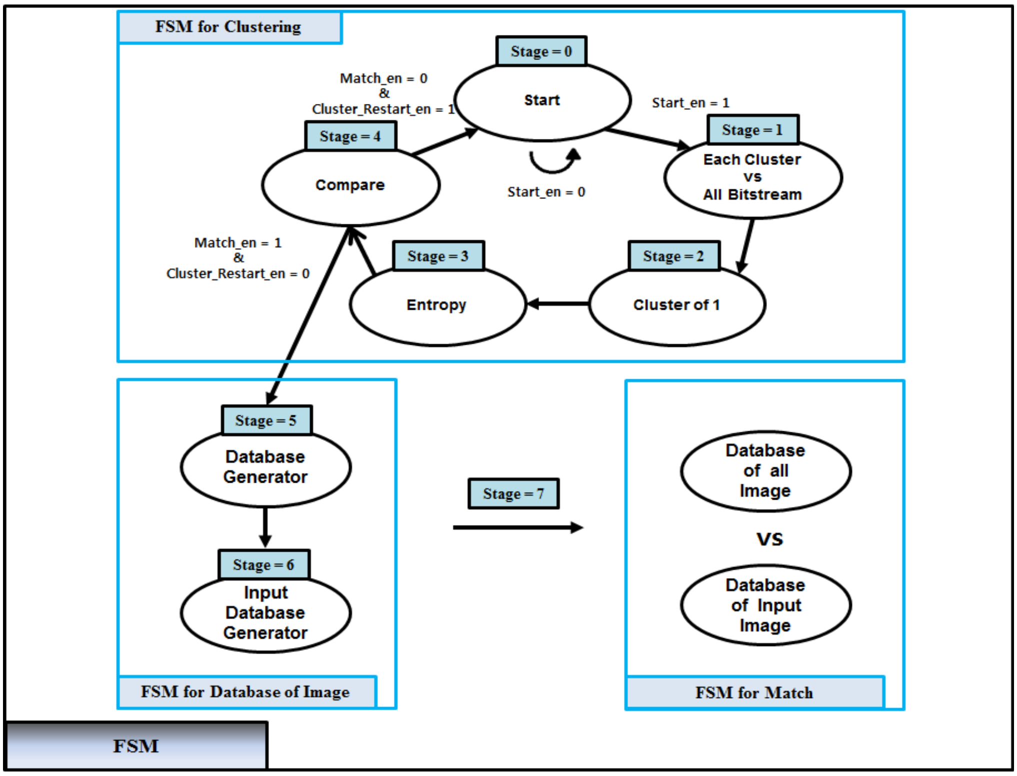

Figure 5. FSM of hardware

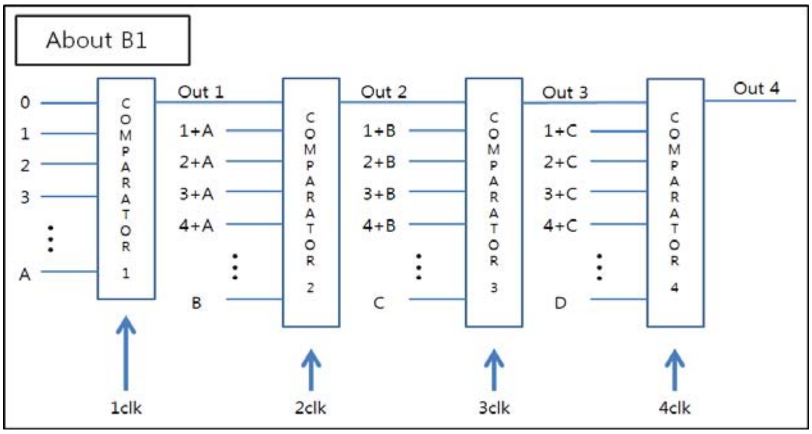

Figure 6. The structure of a four-stage pipeline comparator

In stage 2, the preprocessing process for Entropy counts the number of 1s for the bit sequences included in each cluster, and applies Entropy to the number of 1s. In stage3, new clusters are created with Entropy as 1, 0, and X(don't care) value as a process to create a new cluster. In the hardware structure, the multiplier and the divider are implemented using IP in order not to use the floating point, unlike the decimal point when calculating Entropy in the software structure. The cluster values after clustering are terminated are the primary data for creating a database of scanners in the image and a database of the input images. In stage 4, the clustering is terminated if the previous and current clusters are the same as the previous clusters and the current clusters. If not, the clustering process from stage 0 to stage 3 is repeated until they are the same. Fig. 7 shows the comparison process between past and present clusters and is determined as 1 when two values are equal. In stage 5, the database of the searcher is created. After clustering is terminated, a search database is created using the final cluster values and bit streams of each image in all images. In the Entropy part of the hardware structure, the divider and multiplier methods have been applied to the searcher database part to eliminate the decimal representation. In stage 6, the input image database is created. Fig. 8 shows a diagram of the input image database. As a preprocessing process for generating an input image database, feature points of an input image are extracted through a SIFT algorithm and bit strings are generated. Block ROM input image structure as shown in Fig. 8, and the input image database is generated with the bit streams stored in the block ROM input image.

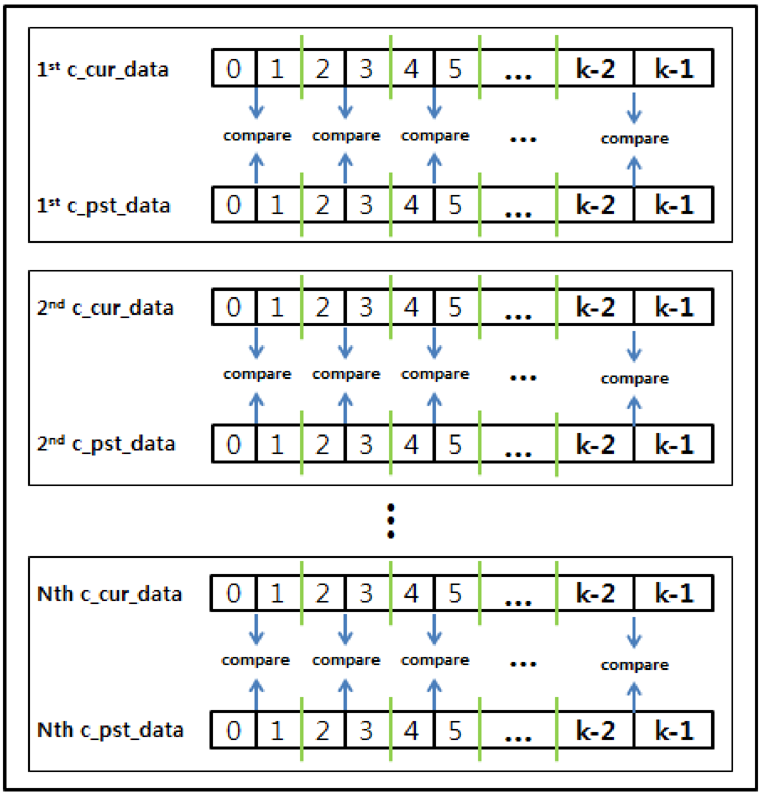

Figure 7. The comparison process between past and present clusters

Stage 7 is the matching step. Euclidean distance is used to measure the similarity between the searcher database and the input image database. The matching step is implemented by applying a three-stage pipeline. In the first clock, the difference between the two inputs a and b is found. In the second clock, the difference is squared. In the third clock, the values from the square are added up to n clocks. Finally, it puts the route and outputs the distance between the input image and one searcher database. When the distance calculation is performed for all the databases of the input image database and the searcher image by the above process, the output distance values are determined by using the comparator to determine which value has the smallest value, and if the class of the input image is in any class. The result is finally a value of 1. This value is the result that the input image is most similar to the first database.

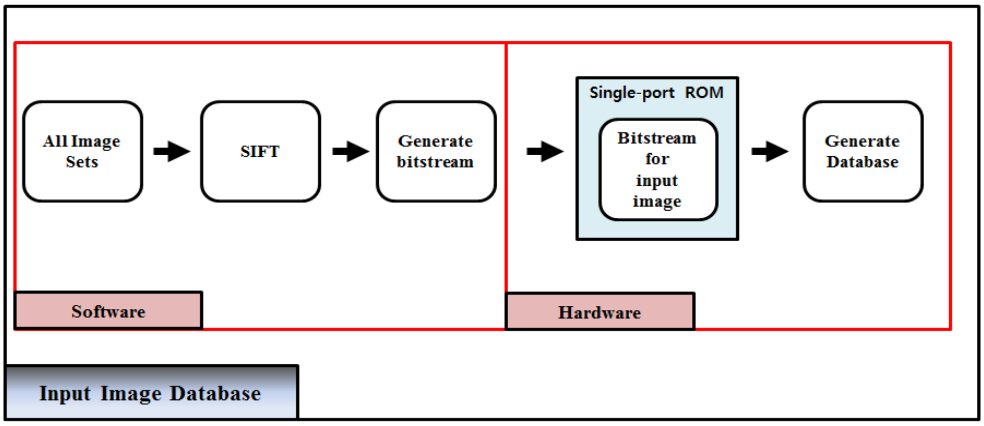

Figure 8. The diagram of the input image database

V. Conclusion

In this paper, we propose an image search algorithm for real-time processing and design it with hardware structure. We propose an image retrieval algorithm using bit streams based on BoWs, which is a good performance among existing image retrieval algorithms. In the accuracy measurement, 86.2% of the BoWs algorithm and 83.7% of the proposed hardware-accelerated software implementation algorithm were used, and the BoWs algorithm was 2.5% higher. In the speed measurement, the image retrieval processing speed of BoWs is measured as 7.89s and our algorithm is measured as 1.55s, which shows that our algorithm is 5.09 times faster than the BoWs algorithm. In the hardware structure, the algorithm code was written in Verilog HDL. The proposed image retrieval algorithm was designed and verified by FPGA and Design Compiler, which is a synopsis system tool.

For real-time processing, the bit streams of the main image data set are processed in parallel and four bit streams are arranged in the horizontal row of the block RAM so that four bit streams can be read and written at a time. In addition, the FSM is used to control signals in each state of the algorithm. In the computation process, the size of the circuit is reduced by parallel processing, pipeline application and memory reuse, and the desired operating frequency of 100 [MHz] is designed to operate.

In this paper, we propose a process to create an entire image database and a process of creating an input image database in order to retrieve one input image. If an actual user performs an image search in this way, as the entire image data exponentially grows, it will take a lot of time to create the database. In order to solve this problem, a database is created separately from the server that generates the database, and the database is created in a hardware structure and the created database is stored in a large memory.

VI. Acknowledgments

This paper was supported by Kumoh National Institute of Technology.

VII. References

[1] Chen Yi, Wang Wei, Liu Ziyang, "Keyword-based search and exploration on databases", Data Engineering (ICDE), 2011 IEEE 27th International Conference on 2011 Apr ,pp. 1380 - 1383 , 2011

[2] Marinovic Igor, Furstner Igor, "Content-based image retrieval", Intelligent Systems and

Informatics, 2008. SISY 2008. 6th International Symposium on 2008 Sept. ,pp. 1 - 6 , 2008
[3] Tsai, Chih-Fong, "Bag-of-Words Representation in Image Annotation: A Review", ISRN artificial intelligence v.2012 ,pp. 1 - 19 , 2012
[4] Feng Cheng, Wang Xiaohong, "Image retrieval system based on bag of view words model", Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on 2016 June ,pp. 1 - 4 , 2016
[5] Yue Huanjing , Sun Xiaoyan , Wu Feng , Yang, Jingyu, "SIFT-Based Image Compression", Multimedia and Expo (ICME), 2012 IEEE International Conference on 2012 July ,pp. 473 - 478 , 2012
[6] Bay, H., Tuytelaars, T., and Van Gool, L., "Surf: Speeded up robust features," in ECCV 2006
[7] Yan H, Zhang X, Li Y, "Kernel-based maximal-margin clustring algorithm", Journal of tsinghua University, v.42 no.1 ,pp. 132 - 134 , 2002
[8] Yadav, Sanjay, Shukla Sanyam, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification", Advanced Computing (IACC), 2016 IEEE 6th International Conference on 2016 Feb ,pp. 78 - 83 , 2016