

A Novel Quantization Approach for Approximate Nearest Neighbor Search to Minimize the Quantization Error

Uriti Archana¹, Urlam Sridhar²

P.G. Student, Department of Computer Science & Engineering, Sri Venkateswara College of Engineering & Technology, Etcherla, Srikakulam, Andhra Pradesh, India¹

Assistant Professor, Department of Computer Science & Engineering, Sri Venkateswara College of Engineering & Technology, Etcherla, Srikakulam, Andhra Pradesh, India²

ABSTRACT: As the data grow increases rapidly, the problem arises in fast retrieval of data from a large data set. In this paper, a novel quantization approach is proposed to minimize the quantization error. Generally the Nearest Neighbor (NN) Search is not benefit for large-scale datasets. To overcome the limitation of NN search, an iterative approach is defined to minimize the computational cost and also performs the state-of-the-art methods. The Approximate Nearest Neighbor (ANN) search performs the fast and efficient retrieval of data as the size of data grow increases rapidly. It explores the quantization centroids on multiple affine subspace. This approach is verified by using Net beans software.

KEYWORDS: Approximate Nearest Neighbor search, Binary codes, Large-scale retrieval, Subspace clustering, Cartesian product.

I. INTRODUCTION

The Nearest Neighbor (NN) Search is a linear search method in finding the exact nearest Neighbor. The NN search is widely used in pattern recognition, information retrieval and recommendation systems. In which, searching on large-scale datasets is not beneficial. The search on data set and distance calculation between sample pairs are computationally costly. To overcome the limitation of NN search, we propose an Approximate Nearest Neighbor (ANN) search [1] to calculate the pairwise distance between pairs of data points. It has proved to be a viable alternative and has so far achieved promising results. This paper focuses on vector quantization approach and focus on detailed review on hashing. The Vector quantization works by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Generally most of the algorithms are based on hashing. The Hashing method creates binary strings from sample vectors and compare those strings using the Hamming distance [2][3]. The Hamming distance measures the minimum number of substitutions required to change one string into the other. It minimizes the error occurrence. Generally the K-means is not directly applicable to large data set for very large number of centroids. In this paper, we develop an iterative approach to obtain the affine subspaces and codebook at the same time.

The idea of quantization is defined by Lloyd [4] which is closely related to K-means algorithm. The K-means algorithm is not directly applicable to large-scale data, for very large number of centroids. The improvement of Lloyd method is proposed by Jegou et al[5] for ANN. In their method called Product Quantization (PQ), the authors divide the sample vector into sub vectors and quantize each of them independently using subquantizers. This makes the quantization codebook a Cartesian product, where each centroid in this codebook is represented as a concatenation of the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

corresponding centroids from the sub codebooks. Therefore, for a small number of subquantizers, while each of them having a feasible number of centroids, obtaining the desired total number of centroids is made possible.

Another efficient coding on high dimensional vectors has been proposed by Jegou et al and later by Gordo et al. In which the data is decorrelated by applying the Principal Component Analysis (PCA)[6] and its dimension is reduced to a desired value. Gong and Lazebnik [7] propose a two-step method called Iterative Quantization (ITQ). In which dimensional reduction and orthogonal rotation on the data is applied iteratively. Brandt in proposes a method called Transform Coding (TC) to balance the variance corresponding to each code separately after PCA. In TC, each dimension is allocated a variable number of bits and a scalar quantization is performed on each principal component independently. Optimized Product Quantization [8] (OPQ) and Cartesian K-Means (CKM) [9] both produce an improvement over PQ by applying an iterative optimization process in order to balance the dimension variances. Many of the proposed methods so far transform or project the data into a new (sub)space, where vector dimensions are reduced, reordered or rotated using PCA. As these methods are equipped for large-scale datasets having a large number of dimensions, this example doesn't give a quantization performance comparison but it's advantageous to visualize the primary variations between your methods.

The proposed method [10] contributes the following:

- Vector Quantization is defined to minimize the Quantization error.
- An Iterative approach is proposed to minimize the Quantization error in order to create a novel quantization scheme.
- Achieves the State-of-art performance.
- Decreases the computational cost.

Paper is organized as follows. Section II describes System Architecture of Quantization method for ANN Search in reducing computational cost and quantization error and also performs the state-of-art algorithms. The architecture contains some of the modules that is given in Section III. Section IV presents experimental results showing results of dataset tested. Finally, Section V presents conclusion.

II. SYSTEM ARCHITECTURE

In this paper, quantization method is proposed to minimize the quantization error and also reduce the computational cost. The architecture itself represents the modules of the quantization approach by achieving the State-of-art method. Initially the query is provided to find the mean squared error and then applies the iterative approach to find the approximate nearest neighbor search. The architecture shows the flow of result obtaining from each module in the following fig.1. In the determination of subspaces approach, the affine subspaces is defined and then for each subspaces, the bits are distributed among dimensions are obtained. The number of dimensions of a subspace is determined according to the bit allocation strategy. The multiple affine subspaces are introduced in order to minimize the error generated by the projection onto a subspace. Since multiple PCAs are used to generate the projection matrices, the statistical dependencies between dimensions have been minimized i.e., quantization can be performed independently on each dimension and the code vector can be obtained as a Cartesian product of quantized values. The reduced number of dimensions for a subspace is the number of dimensions which has at least one allocated bit. Once the subspaces and their number of dimensions are established, the quantizers are obtained and each sample from the training set is assigned to its new cluster. Finally the clusters are updated. In order to calculate the quantization error for all subspaces, the sample should be projected onto each subspace and quantized by the corresponding subquantizer.

While [11] proves that the use of multiple PCAs improves the retrieval performance, optimizing the centroids together with affine subspace clustering as proposed in this paper has shown to outperform the state-of-the-art methods.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

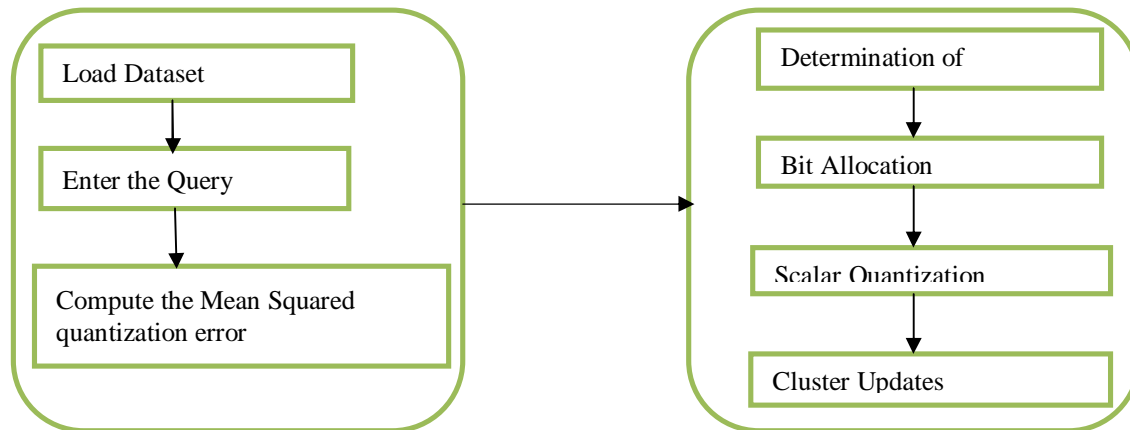


Fig.1. Block Diagram of System Architecture

III. MODULES

In the load dataset module, the admin loads the dataset and then user sends the query request for approximate nearest neighbor search. Now the admin will apply the iterative approach to minimize the quantization error. In the mean squared quantization module, the admin compute the mean squared quantization error by using the following formula. The admin also finds the number of rows in data.

$$MSE_Q = \frac{1}{N} \sum_i^N \|X_i - Q(X_i)\|^2$$

Similarly in the iterative approach module, it will define the determination of subspaces, bit allocation, scalar quantization and then provides the cluster updates to show all the approximate nearest neighbors. First it is aimed to define the affine subspaces and then for each subspaces, the bits are distributed among dimensions are obtained. The multiple affine subspaces are introduced in order to minimize the error generated by the projection onto a subspace. Since multiple PCAs are used to generate the projection matrices, the statistical dependencies between dimensions have been minimized i.e., quantization can be performed independently on each dimension and the code vector can be obtained as a Cartesian product of quantized values. The number of dimensions of a subspace is determined according to the bit allocation strategy. Once the subspaces and their number of dimensions are established, the quantizers are obtained and each sample from the training set is assigned to its new cluster. Finally the clusters are updated.

IV. EXPERIMENTAL RESULTS

Figures show the results of quantization method from the dataset and reduce the computational cost and also minimize the quantization error by giving the updated cluster. Fig. 2 shows the query entered from the given dataset taken and then finds the mean squared error value. Figs. 3, 4, 5, 6 shows the number of iterations taken place as determination of subspaces, bit allocation of data, scalar quantization and cluster updated.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

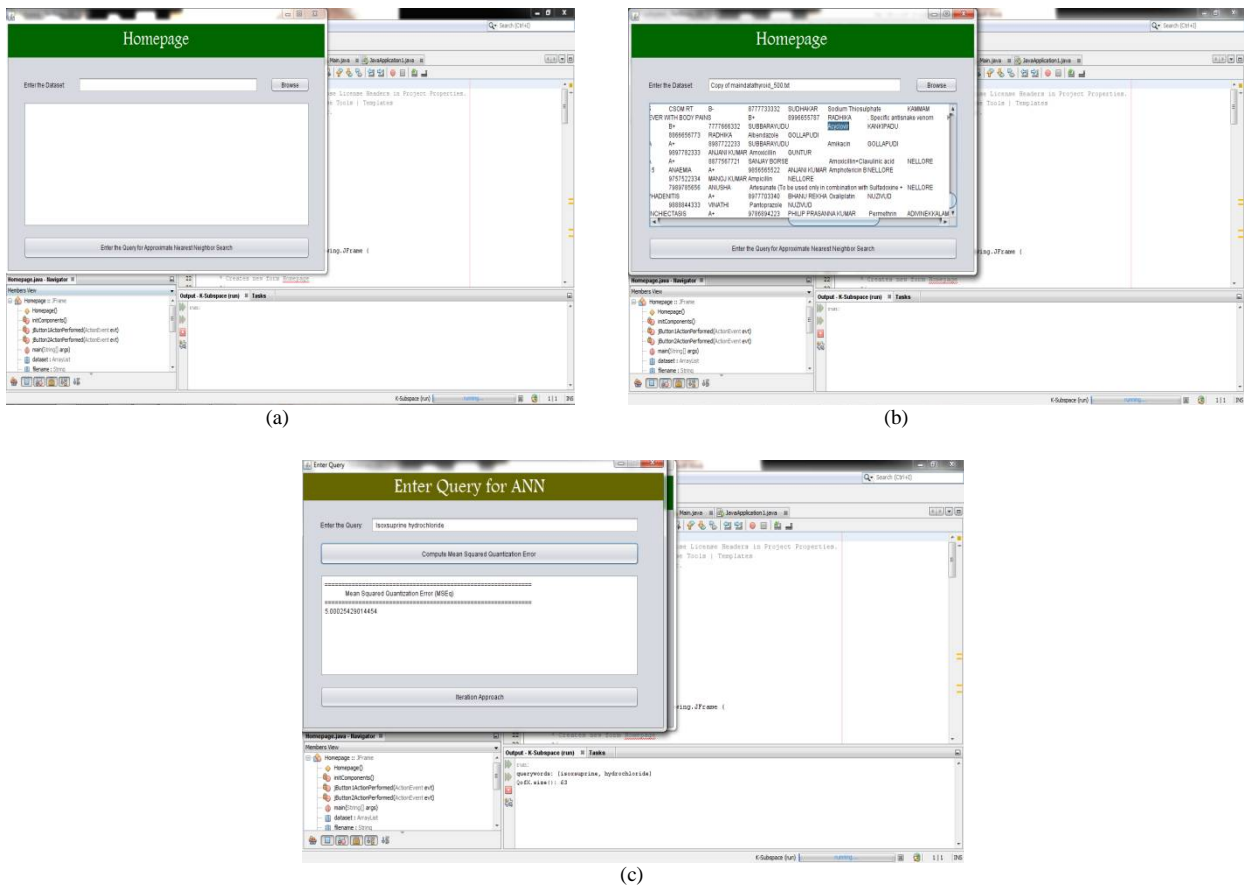


Fig.2.Results display the loading of data (a) Home page of data set entering. (b) Load the data from the dataset taken. (c) Enter the query and then compute the mean squared error to the given query.

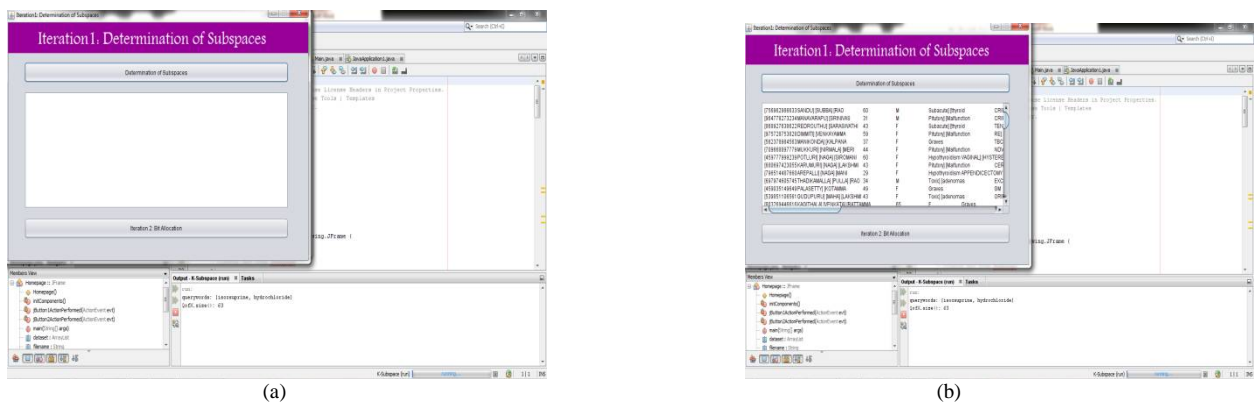


Fig.3.Iteration-1 Results: (a)Determination of subspaces page. (b) Shows the result of subspaces to the given dataset.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

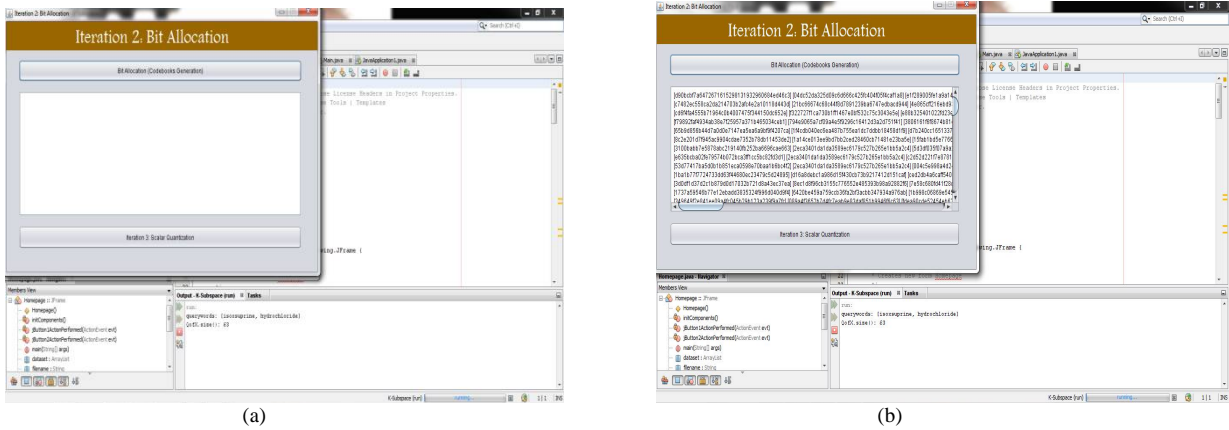


Fig.4.Iteration-2 Results: (a)Bit Allocation is done in second iteration. (b) Bit allocation of data set is represented.

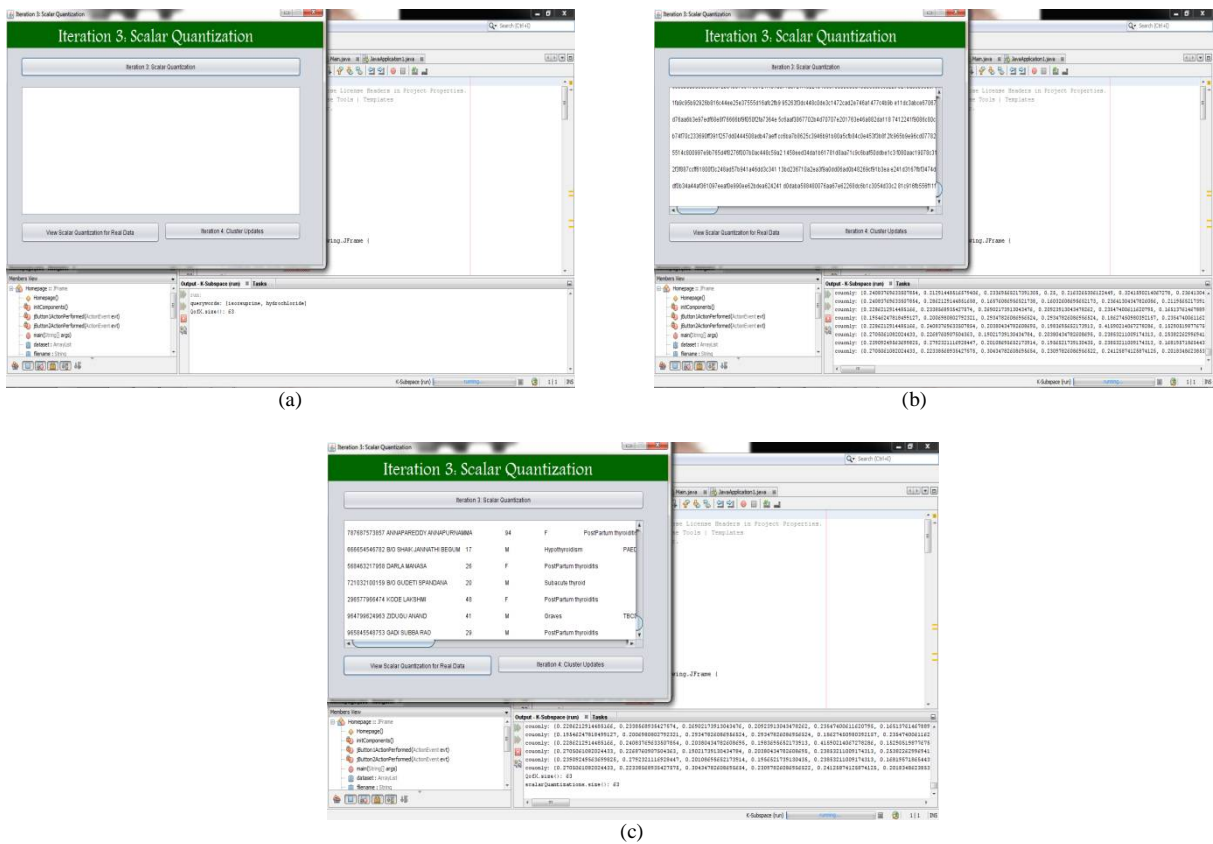


Fig.5Iteration-3 Results: (a)Scalar Quantization is done in iteration-3. (b) It results the scalar quantization of given bit data. (C) It also shows the scalar quantization for real data taken.

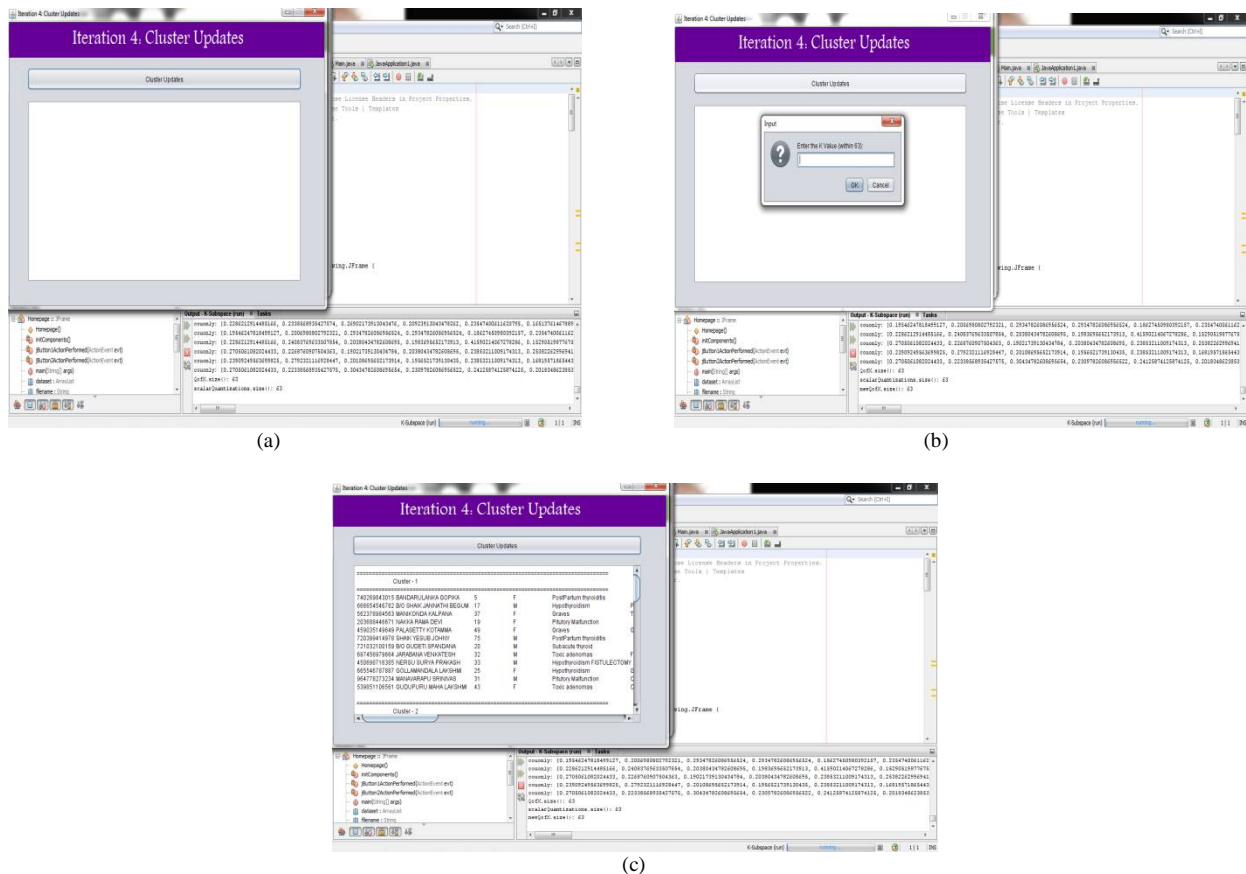


Fig.6.Iteration-6 Results: (a) Cluster updated page is shown. (b) Enter the k value as input to the resulted cluster. (c) Finally Cluster is updated based on the input value taken.

V. CONCLUSION

In this, a novel vector quantization algorithm is proposed for the approximate nearest neighbor search problem. The proposed method explores the quantization centers in affine subspaces through an iterative technique, which jointly attempts to minimize the quantization error of the training samples in the learnt subspaces, while minimizing the projection error of the samples to the corresponding subspaces. It is also shown that, dimension reduction is an important source of quantization error, and by exploiting subspace clustering techniques the quantization error can be reduced, leading to a better quantization performance. This method has proven to outperform the state-of-the-art methods, with comparable computational cost and additional storage.

REFERENCES

- [1] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," 2012, pp. 321-350.
- [2] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality sensitive hashing scheme based on P-stable distributions," in Proc. 20th Annu. Symp. Comput. Geom., 2004, pp. 253-262.
- [3] K. Terasawa and Y. Tanaka, "Spherical LSH for approximate nearest neighbor search on unit hypersphere," in Proc. 10th Int. Conf. Algorithms Data Struct., 2007, pp. 27-38.
- [4] Ezgi Can Ozan, Serkan Kiranyaz, "K-Subspaces Quantization for Approximate Nearest Neighbor Search", Senior member, IEEE and Moncef Gabbouj, Fellow, IEEE, 2016.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

- [5] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 1, pp. 117–128, Jan. 2011.
- [6] V. Gassenbauer, J. Kivaneh, K. Bouatouch, C. Bouville, and M. Ribardiere, "Improving performance and accuracy of local PCA," Comput. Graph. Forum, vol. 30, no. 7, pp. 1903–1910, Sep. 2011.
- [7] Y. Gong and S. Lazebnik, "Iterative quantization: A procrustean approach to learning binary codes," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2011, pp. 817–824.
- [8] T. Ge, K. He, Q. Ke, and J. Sun, "Optimized product quantization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 4, pp. 744–755, Dec. 2014.
- [9] M. Norouzi and D. J. Fleet, "Cartesian K-means," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 3017–3024.
- [10] Ezgi Can Ozan, Serkan Kiranyaz, "K-Subspaces Quantization for Approximate Nearest Neighbor Search", Senior member, IEEE and Moncef Gabbouj, Fellow, IEEE, 2016.
- [11] E. C. Ozan, S. Kiranyaz, and M. Gabbouj, "M-PCA binary embedding for approximate nearest neighbor search," in Proc. BigDataSE, 2015, pp. 1–5.