ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

IMAGE RETRIEVAL PROCESS BASED ON INPUT QUERY USING CLUSTERING TECHNIQUES

R.Dhivya

Research scholar, Department of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

Dr.M.Prabakaran

Research Supervisor, Asst. Prof. of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

ABSTRACT

A scalable approach for content-based image retrieval in peer-to-peer networks by employing the Bag-of-visual-words model. A peer-to-peer network regularly evolves dynamically, which makes a static codebook less effective for retrieval tasks. A dynamic codebook updating process by optimizing the communal information between the resultant codebook and relevance information, and the workload balance among nodes that manage different code words. The planned approach is scalable in developing and disseminated peer-to-peer networks, while achieving enhanced recovery accuracy.

KEYWORDS: Bag-of-visual-words, content-based image retrieval, peer-to-peer, information maximization, workload balance

INTRODUCTION

Peer-to-peer (p2p) networks, which are shaped by equally privileged nodes involving to each other in a self-organizing way, have been one of the most important architectures for data sharing. Popular P2P files-sharing networks such as eDonkey1 count millions of users and tens of millions of files. Unlike WebPages which mostly consist of textual documents such as news, blog article or meeting posts, multimedia files play a leading role in most P2P networks. The ever-growing amount of multimedia data and computational power on P2P networks expose both the need and potential for large scale multimedia retrieval applications such as content-based image sharing, and copyright infringement detection. While P2P networks are well known for their efficiency, scalability and robustness on files sharing, as long as extended search functionality such as content-based image retrieval (CBIR) faces the following challenges: 1) in contrast to centralized environments, data in P2P networks is distributed amongst diverse nodes, thus a CBIR algorithm needs to index and search for images in a distributed manner; 2) unlike distributed servers/clouds, nodes in P2P networks have incomplete network bandwidth and computational power, thus the algorithm should keep the network cost low and the workload between nodes balanced; and 3) as P2P networks are under unvarying churn, where nodes join/leave and files publish to/remove from the network the index needs to be updated dynamically to adapt to such changes. To support content indexing and avoid message flooding, structured overlay networks such as Distributed Hash Tables (DHTs) are often implemented on top of a physical network. By organizing the nodes in a structured way, messages can be efficiently routed between any pair of nodes, and the index integrity can be maintained during network churn. For the CBIR functionality, most of the existing systems adopt a global feature approach: an image is represented as a high-dimensional feature vector (e.g., color histogram), and the similarity between files is measured using the distance between two feature vectors. Usually, the feature vectors are indexed by a distributed high-dimensional index or Locality Sensitive Hashing (LSH) over the DHT overlay.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

RELATED WORKS

In [1] Robert Morris, David Karger, M. FransKaashoek ,HariBalakrishnan et al presents Peer-to-peer systems and applications are distributed systems lacking any centralized control or hierarchical organization, where the software management at each node is equivalent in functionality. A review of the features of topical peer-to-peer applications yields a long list: redundant storage, permanence, variety of close by servers, anonymity, search, authentication, and hierarchical naming. Despite this rich set of features, the core operation in most peer-to-peer systems is competent location of data items. The giving of this paper is a scalable protocol for lookup in a dynamic peer-to-peer system with frequent node arrivals and departures. The Chord protocol chains just one operation: given a key, it maps the key onto a node. Depending on the application using Chord, that node might be accountable for storing a value associated with the key. Chord uses a alternative of dependable hashing to dispense keys to Chord nodes.

In [2] Paul Francis Mark Handley Richard Karp1; Scott Shenker et al presents Hash tables which map "keys" onto "values" are an necessary building block in present software systems. We understand a similar functionality would be equally important to huge distributed systems. In this paper, we initiate the attention of a Content-Addressable Network (CAN) as a distributed transportation that provides hash table-like functionality on Internet-like scales. The CAN is scalable, fault-tolerant and finally self-organizing, and we disclose its scalability, heftiness and low-latency property during simulation. A hash table is a data structure that efficiently maps "keys" onto "values" and serves as a core structure block in the implementation of software systems. We inference that many large-scale distributed systems could likewise benefit from hash table functionality. We use the expression Content-Addressable Network (CAN) to depict such a distributed, Internet-scale, hash table.

In [3] Laura Ricci , Luca Ferrucci , Michele Albano , RanieriBaraglia et al presents The problem of defining a hold for multidimensional range queries on P2P overlays is currently an active field of investigate. Several approaches based on the addition of the basic functionalities offered by Distributed Hash Tables have been recently proposed. The main drawback of these approaches is that the district compulsory for the resolution of a range query cannot be certain by uniform hashing. On the other system, locality preserving hashing functions does not guarantee a good level of load balancing. This paper presents Hivory, a P2P overlie based on a Voronoi tessellation definite by the objects available by peers. Each object is map to a site of the Voronoi tessellation and the corresponding Delaunay Triangulation defines the P2P overlay. A hierarchy of Voronoi diagrams is defined by exploiting clusters of objects balancing with the same site of the Voronoi diagram. A new Voronoi diagram including the peers of the cluster is shaped so that the query resolution may be polished by a top down vacation of the Voronoi hierarchy.

In [4] Yuzhe Tang and Shuigeng Zhou et al presents DHT is a widely-used building block in P2P systems and complex queries are fast popularity in P2P applications. To sustain efficient query processing over DHTs, effectual indexing structures are essential. Recently, numerals of indexing schemes have been proposed. However, these schemes have focused on improving query efficiency, and as a trade-off, sacrificed preservation efficiency an important recital measure in the P2P context, where frequent data updating and high peer dynamism are classically incurred. In this paper, we propose LHT; a Low upholding Hash Tree, for efficient data indexing over DHTs. LHT employs a novel naming function and a tree summarization policy to elegantly distribute its index construction. It is flexible to any DHT substrates, and is easy to be implemented and deployed.

In [5] Herve J egou MatthijsDouzeCordeliaSchmid et al presents this article improves recent methods for large scale image search. We first evaluate the bag-of-features approach in the structure of approximate nearest neighbor search. This leads us to acquire a further accurate symbol based on Hamming embedding (HE) and weak geometric consistency constraints (WGC). HE provides binary signatures that refine the matching based on visual words. WGC filters matching descriptors that are not reliable in terms of angle and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

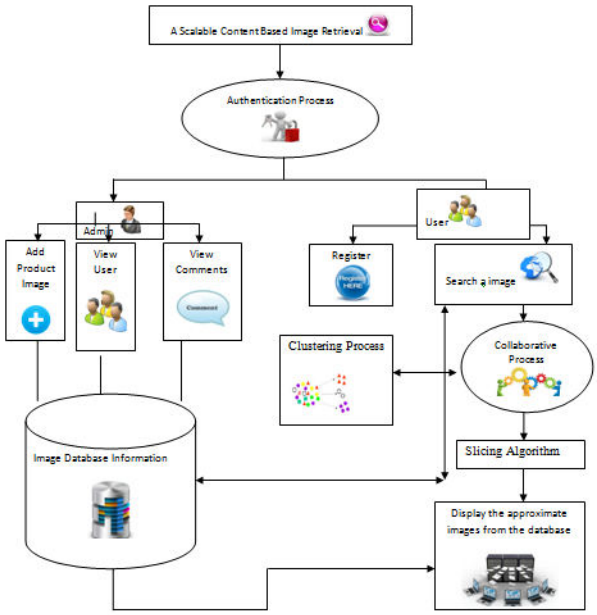
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

scale. HE and WGC are incorporated within an inverted file and are efficiently exploited for all images in the dataset. We then initiate a graph-structured quantizes which appreciably speeds up the assignment of the descriptors to visual words. A comparison with the state of the art shows the interest of our approach when high accuracy is required. Experiments performed on three reference datasets and a dataset of one million of images show a significant improvement due to the binary autograph and the feeble geometric constancy constraints, as well as their efficiency.

PROPOSED SYSTEM

In the measured technique Collaborative Filtering approach, this consists of two stages: clustering and collaborative filtering. Clustering is a preprocessing stair to separate massive data into expedient parts. Collaborative filtering methods have been purposeful to numerous dissimilar kinds of data including such as Camera, Laptop, Television By using slicing algorithm, since the number of services in a cluster is much less than the total number of services, the calculation occasion of CF algorithm can be muscular appreciably Since the ratings of similar services contained by a cluster are more related than that of dissimilar services, the suggestion accuracy based on user ratings may be improved. Finally, the updates are synchronized across the network at the end of each iteration. The discriminability and workload balance is optimized incessantly with the churn of the P2P network Collaborative filtering (CF) such as item and user-based methods are the leading techniques applied in RSs. The essential assumption of user-based CF is that people who agree in the past tend to concur again in the prospect. Different with user-based CF, the item-based CF algorithm recommend a user the substance that are parallel to what he/she has preferred before. Although traditional CF techniques are sound and have been effectively applied in many e-commerce RSs, they encounter two chief challenges for big data application

ARCHITECTURE DIAGRAM

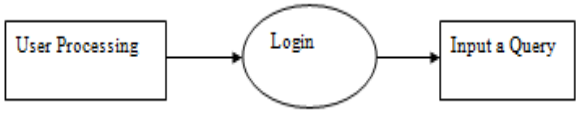


MODULES

- Input Image.
- Preprocessing method.
- Feature descriptor.
- FV/VLAD method.
- Sorting Algorithm.

INPUT IMAGE

Image Processing is processing of images using arithmetic operations by any sketch of signal processing for which the input is an image, a sequence of images, or a video, such as a photograph or video frame; the output of image processing may be also an image or a position of uniqueness or parameters related to the image. Most image-processing techniques absorb treating the image as a two-dimensional signal and apply archetypal signal-processing techniques to it. Images are also route as three-dimensional signals where the third-dimension being time or the z-axis. Get the input difficulty from the user and demonstrate the image in the threshold.

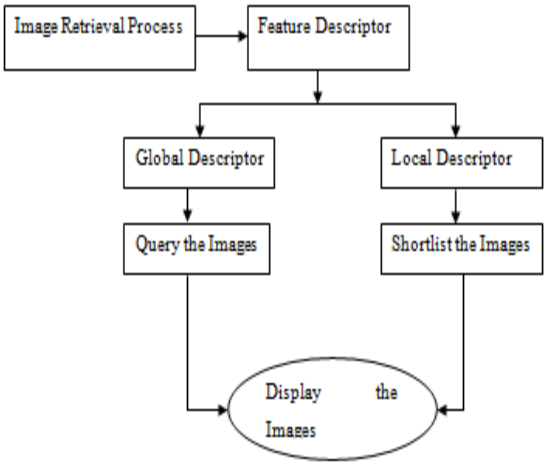


ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

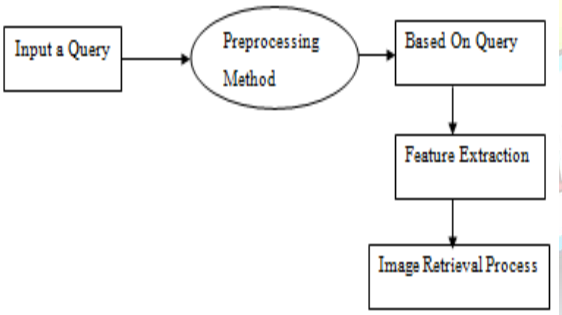
PREPROCESSING METHOD

These image preprocessing methods are accessible from the Analysis menu by selecting "Preprocessing" -> "X-block" and choosing "Custom "which presents the existing preprocessing methods. These image preprocessing methods are planned under "Image Filters". Preprocessing images repeatedly involve remove low-frequency backdrop noise, normalize the strength of the personality particles descriptions, remove reflections, and masking portions of images. Image preprocessing is the method of ornamental data images prior to computational processing.



FEATURE EXTRACTION

When the input data to an algorithm is too large to be process and it is supposed to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images accessible as pixels), then it can be misshapen into a condensed set of features (also named a features vector). This progression is called aspect selection.

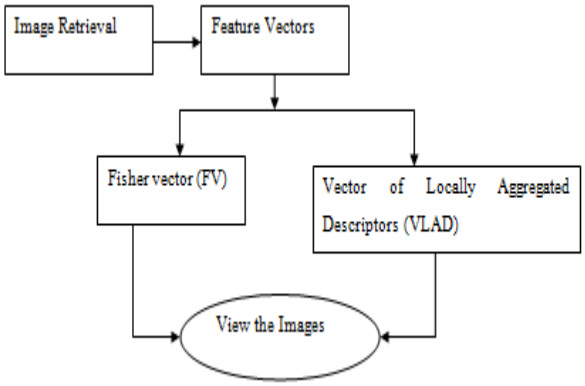


FEATURE DESCRIPTOR

At the first stage, we use the worldwide descriptor to calculate the similarity scores between query image and the dataset images, and get the shortlist from the top of the ranked list. At the second stage, the local descriptors of the query image are coordinated to the local descriptors of the shortlist images, and the geometric verification is applied for re-ranking the shortlist as a resultant criteria

FV/VLAD METHOD

Image retrieval tasks, feature vectors such as Fisher vector (FV), Vector of Locally Aggregated Descriptors (VLAD). In order to reduce its storage and CPU costs to a practical range but, requires more memory for retrieval phase. Vector of local descriptors (VLAD) and Fisher vector (FV). By combining them with a suitable coding technique, it is achievable to encode an image in a little dozen bytes while achieving excellent retrieval results.

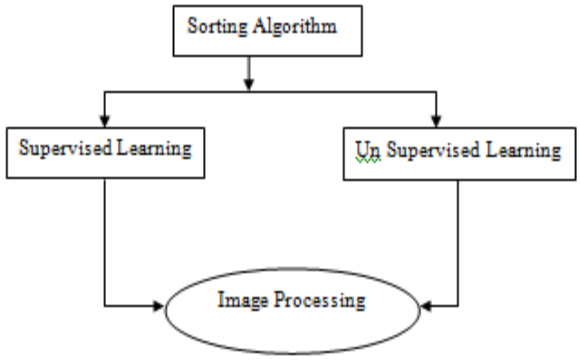


SORTING ALGORITHM

An efficient significance sorting algorithm considering both the supervised and the unsupervised cases, for image recognition and image retrieval. Supervised Cases Only from the Database Image retrieve. Unsupervised Cases unfamiliar image from the database

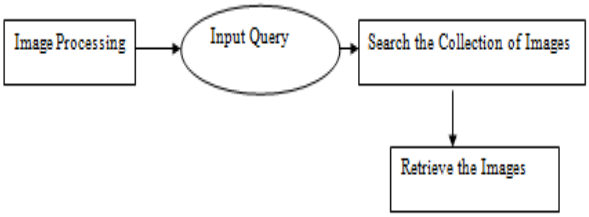
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018



SIMILARITY MEASURE

In similarity measure compare two data set at that time to create the similar images will be retrieved. An image retrieval scheme returns a set of images from a collection of images in the database to meet users' insist with similarity evaluations such as image satisfied similarity.



ALGORITHM

CLUSTERING ALGORITHM

Cluster analysis or clustering is the task of grouping a set of objects in such a system that objects in the same assemblage (called a cluster) are more analogous (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in lots of fields, including learning, pattern, image analysis, information retrieval, and bioinformatics. Cluster analysis itself is not one precise algorithm, but the universal chore to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular ideas of clusters embrace groups with modest distances with the cluster members, solid areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization trouble. The suitable clustering algorithm and limitation setting (including principles such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not a routine chore, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be needed to adjust data preprocessing and representation parameters until the product achieves the preferred properties.

COLLBORATIVE FILTERING

Collaborative filtering (CF) is a popular recommendation algorithm that bases its predictions and recommendations on the ratings or behavior of other users in the system. The fundamental assumption behind this method is that other users' opinions can be selected and aggregated in such a way as to provide a reasonable prediction of the active user's predilection intuitively; they suppose that, if users agree about the quality or relevance of some items, then they will likely agree about other items. The growth of the Internet has made it much more complicated to successfully extract useful information from all the presented online information. The overpowering quantity of data necessitates mechanisms for efficient information filtering. . In the case of static slicing, since the entire program unit is looked at irrespective of an exacting execution of the program, the affected statements in both blocks would be included in the slice. But, in the case of dynamic slicing we judge particular executions of the program, wherein the if block gets execute and the exaggerated statements in the else block do not get executed.

SLICING ALGORITHM

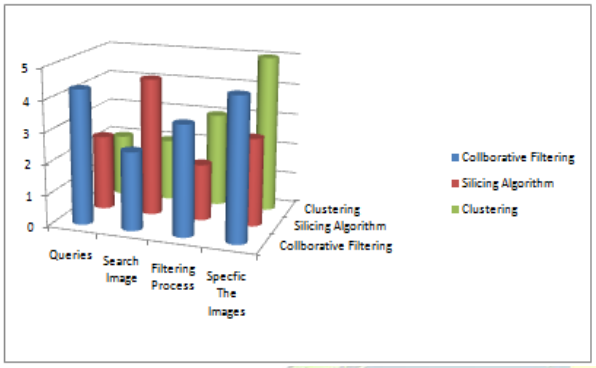
Program slicing is the calculation of the set of programs statements, the course slice, which may affect the values at a few point of interest, referred to as a slicing criterion. Program slicing can be worn in debugging to locate source of errors additional with no trouble. Other

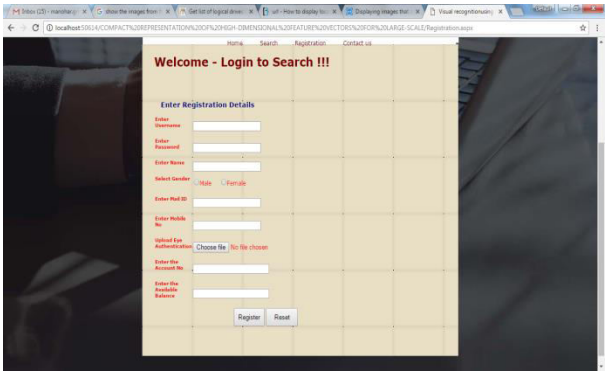
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

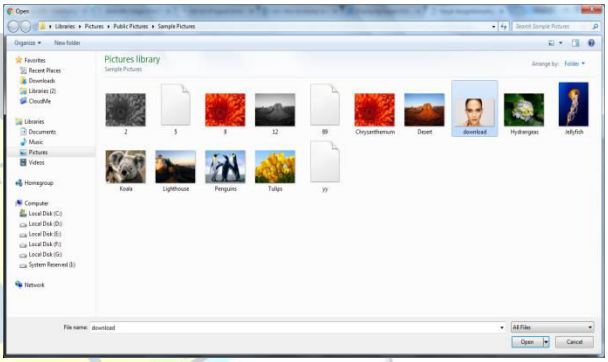
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

applications of slicing include software maintenance, optimization, and program analysis and information flow control.

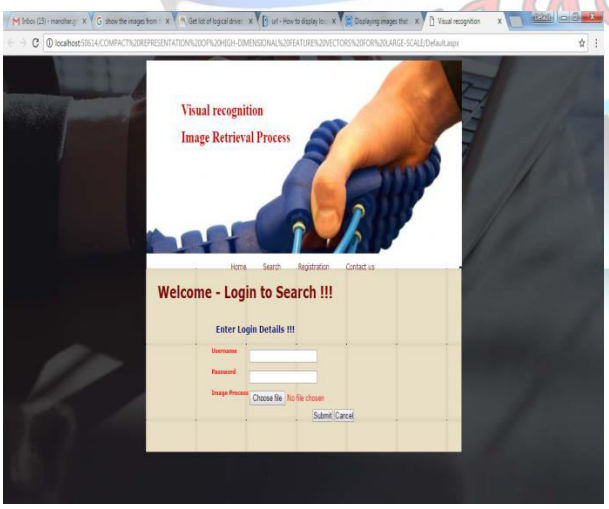
RESULT AND DISCUSSION

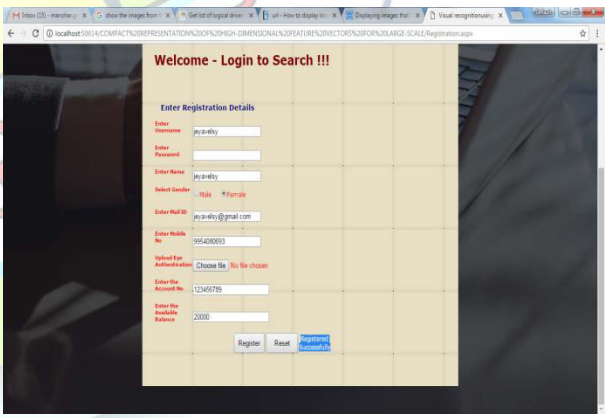


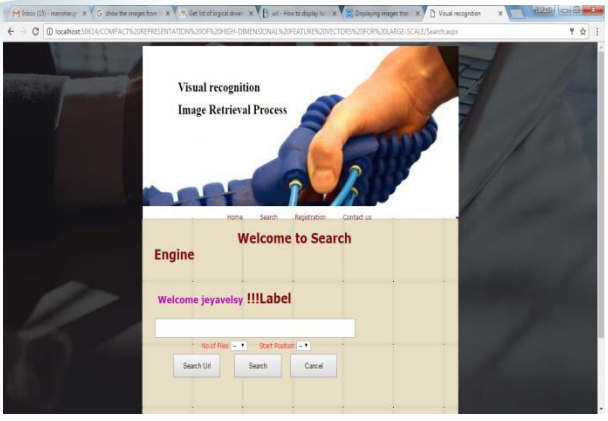




OUTPUT RESULT

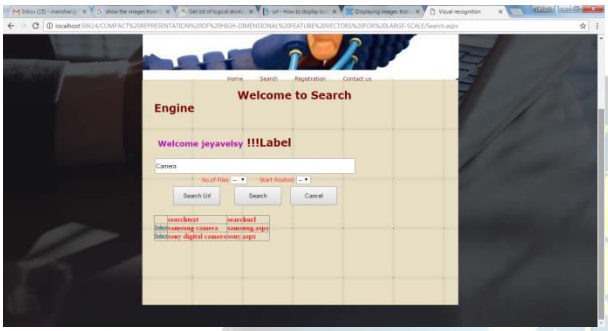


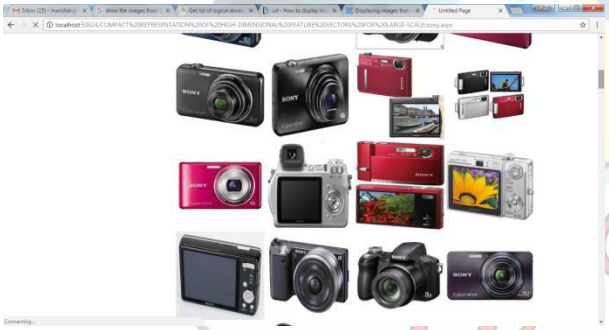


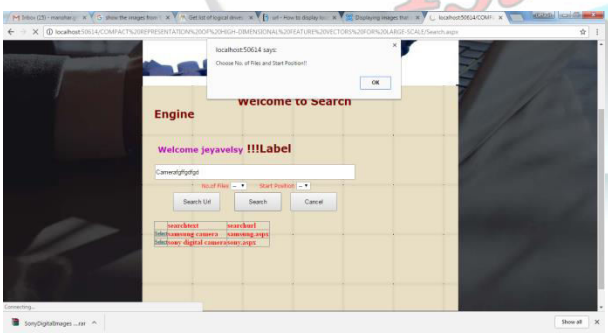


ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018







CONCLUSION

A bag-of-visual-words (BoVW) model based approach for content based image retrieval (CBIR) in peer-to-peer (P2P) networks. In order to conquer the complexity in generating and maintaining a global codebook when the BoVW model is deployed in P2P networks, we formulate the problem of updating an existing code book as optimizing the retrieval accuracy and workload balance. As a consequence, the planned approach is scalable to the number of images shared within a P2P network and the evolving nature of P2P networks. In order to additional get better the repossession performance of the proposed approach and lessen network cost, indexing pruning techniques are applied. We conduct comprehensive experiments to evaluate various aspects of the proposed approach while demonstrating its talented performance. In the future, we will investigate DHT specific optimizations for cost reduction, more advanced matching refinement and multi-modal fusion techniques in P2P networks, and extensions of this approach to other distributed architectures. In particular, for the CAN network we can embed the index into the CAN overlay. That is, we make the CAN tackle space equivalent to our feature space, and restore the CAN zones with codeword partitions. Such an embedding will eliminate the overhead of an extra DHT layer, as we can realize the SPLIT/MERGE operations as a CAN zone split/takeover, in its place of totaling and removing entries on DHT.

REFERENCE

[1] M. Steiner, T. En-Najjary, and E. W. Biersack, "Long term study of peer behavior in the KAD DHT," IEEE/ACM Transactions on Networking, vol. 17, no. 5, pp. 1371–1384, Oct. 2009.

[2]H.SchulzeandK.Mochalski,"Internetstudy2008/2009, "Internet Studies, ipoque, 2009.

[3] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," in ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, 2001, pp. 149–160.

[4] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in ACM Conference on Applications, Technologies,

Architectures, and Protocols for Computer Communications, 2001, pp. 161–172.

[5] M. Mordacchini, L. Ricci, L. Ferrucci, M. Albano, and R. Baraglia, "Hivory: Range queries on hierarchical voronoi overlays," in IEEE International Conference on Peer-to-Peer Computing, Aug. 2010, pp. 1–10.

[6] Y. Tang, S. Zhou, and J. Xu, "LIGHT: A query-efficient yet lowmaintenance indexing scheme over DHTs," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 1, pp. 59–75, Jan. 2010.

[7] L. Zhang, Z. Wang, and D. Feng, "Efficient high-dimensional retrieval in structured P2P networks," in IEEE International Conference on Multimedia and Expo Workshops, Jul. 2010, pp. 1439–1444.

[8] H. J´egou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search," International Journal of Computer Vision, vol. 87, pp. 316–336, 2010.

[9] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector,"inEuropeanConferenceonComputerVision,2002,pp.128– 142.