Recovering Images Based On Their Contents In A Distributed System

MADIPADHI RAMYA SREE
M.Tech Student, Dept of CSE, Malla Reddy College of Engineering and Technology, Hyderabad, T.S, India

G.MANOJ KUMAR
Associate Professor, Dept of CSE, Malla Reddy College of Engineering and Technology, Hyderabad, T.S, India

Dr. M.SAMBASIVUDU
Associate Professor, Dept of CSE, Malla Reddy College of Engineering and Technology, Hyderabad, T.S, India

Abstract: **The use of peer-to-peer networking as an alternative has become more common in recent years as a means of facilitating the scalable movement of multimedia data. The process of carrying out content-based retrieval in peer-to-peer networks, which are characterized by the distribution of huge quantities of visual data across several nodes, is an important yet challenging topic. In this study, we offer a scalable strategy for content-based picture retrieval in peer-to-peer networks utilizing the bag-of-visual-words paradigm. This is in contrast to most of the previous approaches, which were focused on indexing high-dimensional visual characteristics and had limitations on their scalability. When images are scattered over the whole of the peer-to-peer network, the key challenge lies in efficiently getting a global codebook. This is not a problem in centralized setups because it is easier to access the codebook. A static codebook is less helpful for retrieval tasks in a peer-to-peer network because of the dynamic nature of the growth of the network itself. In order to accomplish this, we present a method for dynamically updating the codebook. This method works by distributing the workload evenly across the nodes that are responsible for handling different code words and optimizing the mutual information that exists between the generated codebook and the relevance information. In order to speed up the retrieval process and cut down on network overhead, researchers are investigating several methods for index trimming. The comprehensive experimental data that we have collected indicates that the method that has been recommended is scalable in dynamic and scattered peer-to-peer networks, all while improving retrieval accuracy.**

Keywords: **Peer-To-Peer Computing; Feature Extraction; Indexing; Image Retrieval;**

I. INTRODUCTION:

One of the most important models for exchanging data has been the distributed (P2P) system, which is built around nodes with comparable strengths that communicate with each other in a self-organizing manner. In contrast to multimedia sites, which mostly consist of textual documents like news and blog entries, multimedia files play a significant role in most P2P networks. The demand and capacity for large-scale multimedia are made clear by the proliferation of multimedia data and computing resources on peer-to-peer networks [1]. Since most information in P2P networks is stored on several nodes, it is necessary for a CBIR algorithm to provide evidence and search for photos in a collaborative manner. P2P network nodes, unlike shared servers or clouds, have limited network bandwidth and computing capacity; hence, the algorithm must aim to minimise network costs while balancing the workload across all nodes. Structured overlay networks, such as Distributed Hash Tables (DHTs), are often used on top of a physical network to facilitate content indexing while avoiding heavy traffic. In a network, the order of nodes has an effect on the flexibility of the messages sent between them. Most current systems use a global feature approach to CBIR functionality, in which a picture is represented as a high-dimensional feature vector (such as a colour histogram), and the distance between two feature vectors is used to quantify the degree of similarity between two files [2][3]. Bag-of-visual-words (BoVW) models, on the other hand, have been effectively used for large-scale picture retrieval. The Bag-of-Visual-Words (BoVW) model treats each picture as a bag of local characteristics, much as the Bag-of-Words (BoW) model treats each document as a collection of random words. 2 There are three basic requirements for using the BoVW model. Initially, an image will be broken down into a series of local areas (using image segmentation or uniform image partitioning) or key points (using key point detection methods), with each local region or key point being represented by a unique high-dimensional descriptor. Several characteristics, including resolution, lighting fluctuations, and obscured objects, make the challenge of finding photographs according to their semantic contents very difficult. That's why it's important to have a representation in between, like a visual vocabulary. We've developed a P2P-CBIR system that makes use of the visual vocabulary for image retrieval and is based on the Bag of Words paradigm [4]. To begin using this method, a set of landmarks within a picture are chosen as points of focus. In order to extract the descriptors, it first constructs areas in the neighborhood of these sites

of interest, which have qualities like invariance to multiple picture transformations, illumination, etc. Images in a dataset are then given a numerical representation, called a "model vector," that is based on this visual lexicon. With the bag of words method, a collection of interesting locations is selected, and the features are extracted locally from the surrounding region in order to represent the image's contents (visual features). Gaussian scale-space localization is used to choose scale-invariant interest spots. The visual information of an image is represented by the extracted SURF (Speeded-Up Robust Features) features in the proposed P2P-CBIR. These details guarantee more consistency and individuality. When compared to the SIFT features, this method's extraction time is also much faster. The first step of this strategy is to identify key areas of focus. The Hessian matrix is first approximated quickly.

II. PROBLEM STATEMENT:

An image is represented as a high-dimensional feature vector (for example, a colour histogram), and the similarity between files is determined based on the distance between two feature vectors. This method is used by the systems that are now in use. Typically, the feature vectors are indexed using a distributed high-dimensional index, also known as locality-sensitive hashing (LSH), which is performed over the DHT overlay. P2P networks, in contrast to centralized systems, have data scattered across several nodes; hence, a CBIR algorithm has to index and search for pictures in a way that is distributed [5]. As peer-to-peer (P2P) networks are in a state of perpetual churn, in which nodes join or leave the network and files are added or removed, the index must be constantly updated in order to accommodate these kinds of changes. The techniques of dexing and locality-sensitive hashing When it comes to retrieval, the high-dimensional indexing-based methods save the feature vectors in a data structure that is often either a tree or a graph. This helps them accomplish successful search space trimming. The high-dimensional index in structured P2P networks is defined in a distributed manner across the P2P overlay, dexing, and locality-sensitive hashing. When it comes to retrieval, the high-dimensional indexing-based methods save the feature vectors in a data structure that is often either a tree or a graph. This helps them accomplish successful search space trimming [6]. The high-dimensional index is established in a distributed manner across the P2P overlay in structured P2P networks. The performance of high-dimensional indexing is known to suffer from the well-known "curse of dimensionality," even when it is performed in a centralized setting. Even if one is able to update the hash functions with the changing data, it is still a very difficult task to implement it across the DHTs. Because the data is stored among the nodes that correspond to the hash ID, even a one-bit change in the output of the hash function will cause a significant portion of the data, if not all of it, to be assigned to a different node, which will result in a significant increase in the amount of traffic that is generated across the network.

III. PROPOSED METHODOLOGIES:

We describe an innovative technique for dynamically generating and updating a global codebook that takes into consideration both discriminability and workload balance. When the queries are being processed, each node will gather data on the workload as well as the relevancy of the information. We optimize the information that the codebook provides about the retrieval results by using the relevance information, limiting the amount of information that is lost as a consequence of the quantization process. With the data from the workload, our goal is to distribute the burden evenly among the nodes, preventing any of them from being overloaded or under loaded. The partitioning of the codebook is frequently updated based on these two criteria. This is done by dividing and combining code words, which enables the codebook to expand or contract in line with the distribution of the data. The choice on whether or not a codeword should be split or merged is made on an individual basis by each management node in order to keep the expense of updating the codebook to a minimum. At the conclusion of each repetition, the changes are finally synchronized throughout the network. This occurs after each step. Because of this, the discriminability as well as the workload balance is continually adjusted with the churn of the P2P network. This is the first research of its kind to examine scalable CBIR using the BoVW model in P2P networks. It is recommended that an innovative objective function be devised for codebook optimization in a P2P setting, one that takes into account both the relevance information and the workload balance concurrently. It is recommended that an algorithm for updating remote codebooks be developed. This method should maximise the objective function while maintaining a low updating cost.

IV. ENHANCED SYSTEM:

Using a technique for feature extraction, the Feature Extractor module may extract features from photos that have been shared and produce thumbnails for those images. After that, the Image Indexer Module will index the picture collection by making use of the feature vectors that were extracted. In order to circumvent the bottleneck issue, it is helpful to distribute the workload of this time-consuming and computationally intensive activity among peers by giving them the ability to store and index their picture collection. Through the use of this module, the peer will join the P2P network. The Bootstrap Server is a programme that

keeps and answers a list of peers that are presently connected to the network. The Connection Management module sends a request to the Bootstrap Server. After acquiring the IP address, the peer next joins the P2P network by establishing connections with other peers that are currently online. The Feature Extractor module works on the query picture and puts together the Image Query message when a peer begins the retrieval process by issuing a query to look for other images that are similar to the one they are currently viewing. Similarly, when other peers receive image query messages, they carry out two actions: query message propagation and local index lookup. These processes are similar to the ones described before. After other pictures that are similar to the one being requested have been obtained, the peer will send the Query Hit back to the requester using the Packet Router module. Users will get a list of the locations of the photos that are matched to the query, as well as the size of those images, when an Image Query Hit message returns to the requester. The HTTP Agent will download the thumbnails that were created during the preprocessing step, or it will download the whole picture from the peer via the HTTP protocol. This is done in order to get the result of the query.

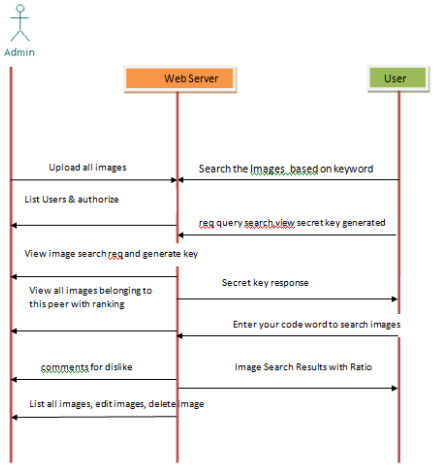

Fig 1: Sequence of System

V. CONCLUSION:

Within the context of peer-to-peer (P2P) networks, we describe a strategy for content-based image retrieval (CBIR) that is based on a bag-of-visual-words (BoVW) model. When the BoVW model is used in P2P networks, it might be challenging to generate and maintain a global codebook. In order to circumvent this challenge, we have recast the issue of updating an existing codebook as one of maximising retrieval accuracy and workload balance. As a result, the proposed method is adaptable to both the number of photos exchanged between peers in a P2P network and the dynamic nature of P2P networks. Techniques for indexing pruning are used in order to achieve the goals of further enhancing the retrieval performance of the proposed method and reducing the cost of the network. We execute exhaustive studies to examine the suggested method from a variety of angles while showing its potential for success.

REFERENCES:

[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. Schulze and K. Mochalski, “Internet study 2008/2009,” 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] J. Sivic and A. Zisserman, “Video Google: A text retrieval approachto object matching in videos,” in IEEE International Conference on Computer Vision, vol. 2, 2003, pp. 1470–1477.

[10] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, “Evaluating bag-of-visual-words representations in scene classification,” in ACM International Workshop on Multimedia Information Retrieval, 2007, pp. 197–206.

