A Scalable Approach for Content-Based Image Retrieval in Peer-to-Peer Networks

**ABSTRACT:**

Peer-to-peer networking offers a scalable solution for sharing multimedia data across the network. With a large amount of visual data distributed among different nodes, it is an important but challenging issue to perform content-based retrieval in peer-to-peer networks. While most of the existing methods focus on indexing high dimensional visual features and have limitations of scalability, in this paper we propose a scalable approach for content-based image retrieval in peer-to-peer networks by employing the bag-of-visual words model. Compared with centralized environments, the key challenge is to efficiently obtain a global codebook, as images are distributed across the whole peer-to-peer network. In addition, a peer-to-peer network often evolves dynamically, which makes a static codebook less effective for retrieval tasks. Therefore, we propose a dynamic codebook updating method by optimizing the mutual information between the resultant codebook and relevance information, and the workload balance among nodes that manage different codewords. In order to further improve retrieval performance and reduce network cost, indexing pruning techniques are developed. Our comprehensive experimental results indicate that the proposed approach is scalable in evolving and distributed peer-to-peer networks, while achieving improved retrieval accuracy.

**EXISTING SYSTEM:**

* 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 In contrast to centralized environments, data in P2P networks is distributed among different nodes, thus a CBIR algorithm needs to index and search for images in a distributed manner.
* P2P networks are under constant 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.
* Dexing and Locality-Sensitive Hashing. The high-dimensional indexing based approaches store the feature vectors in a data structure, usually a tree or a graph, to achieve effective search space pruning during retrieval. In structured P2P networks, the high-dimensional index is defined in a distributed way over the P2P overlay, dexing and Locality-Sensitive Hashing.
* The high-dimensional indexing based approaches store the feature vectors in a data structure, usually a tree or a graph, to achieve effective search space pruning during retrieval. In structured P2P networks, the high-dimensional index is defined in a distributed way over the P2P overlay.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Even in a centralized environment, the performance of high-dimensional indexing suffers from the well-known “curse of dimensionality”.
* Even when one can update the hash functions with changing data, implementing it over the DHTs is very challenging. As the data is stored among nodes of corresponding hash ID, a 1-bit change of the hash function output will result in large portion of (if not all) data being assigned to a different node, causing heavy network traffic.

**PROPOSED SYSTEM:**

* In this paper, we present a novel method to dynamically generate and update a global codebook, which considers both the discriminability and workload balance.
* While processing queries, each node collects the relevance information and workload data. With the relevance information, we maximize the information provided by the codebook about the retrieval results, thus minimizing the information loss incurred by quantization.
* With workload data, we aim to achieve a fair workload among nodes, thus avoiding overloading or under loading nodes. Based on these two criteria, the codebook partitioning is updated routinely by splitting/merging code words, thus allowing the codebook to grow/shrink in accordance to the data distribution.
* To minimize the cost of codebook updating, the decision whether a codeword should be split/merged is taken by its managing node individually. Finally, the updates are synchronized across the network at the end of each iteration.
* As a result, the discriminability and workload balance is optimized continuously with the churn of the P2P network.

**ADVANTAGES OF PROPOSED SYSTEM:**

* It is the first study to investigate scalable CBIR with the BoVW model in P2P networks.
* A novel objective function for codebook optimization in a P2P environment is proposed, which considers both the relevance information and the workload balance simultaneously.
* A distributed codebook updating algorithm based on splitting/merging of individual code words is proposed, which optimizes the objective function with low updating cost.

**SYSTEM SPECIFICATION**

**Hardware Requirements:**

* System : Pentium IV 3.5 GHz or Latest Version.
* Hard Disk : 40 GB.
* Monitor : 14’ Colour Monitor.
* Mouse : Optical Mouse.
* Ram : 1 GB.

**Software Requirements:**

* Operating system : Windows XP or Windows 7, Windows 8.
* Coding Language : Java / J2EE (Jsp,Servlet)
* Data Base : My Sql Server
* Documentation : MS Office
* IDE : Eclipse Galileo
* Development Kit : JDK 1.6
* Server : Tomcat 6.0