An Efficient Approach for Content-Based Image Retrieval Using Distributed networks

J. Srinija [1], Y.Rajesh[2] , Ravi Kumar Tenali[3]
[1]M.Tech, Dept. of CSE, Andhra Loyola institute of Engineering and technology, Vijayawada
[2,3]Asst.Professor, Dept. of CSE, Andhra Loyola institute of Engineering and technology, Vijayawada.

Abstract: *Peer-to-peer networking offers an adaptable answer for sharing media information over the system. With a lot of visual information circulated among various nodes, it is a vital yet difficult issue to perform content-based retrieval in peer-to-peer networks. While the majority of the current strategies concentrate on ordering high dimensional visual elements and have impediments of adaptability, in this paper we propose a versatile approach for content-based retrieval in peer-to-peer networks by utilizing the pack of-visual words display. Contrasted and concentrated conditions, the key test is to effectively get a worldwide codebook, as pictures are disseminated over the entire p2p system. What's more, a shared system frequently advances progressively, which makes a static codebook less compelling for recovery undertakings.*

Along these lines, we propose an element codebook refreshing strategy by upgrading the common data between the resultant codebook andpertinence data, and the workload adjust among hubs that oversee distinctive code words. So as to further enhance recovery execution and lessen arrange cost, ordering pruning procedures are created. Our farreaching test comes about show that the proposed approach is versatile in developing and distributed p2p systems, while accomplishing enhanced recovery exactness.

1. INTRODUCTION

Advancement of digital media coding and Internet technology have enabled, Peer-to-Peer (P2P) networks to share files, transferring real-time video streams, and performing ContentBased Image Retrieval (CBIR) in recent decades. In the P2P networks, each connected peer serves simultaneously as a server and a client, which can distribute computation and network traffics to peers to provide efficient streaming and CBIR services.

The CBIR has been developed over the past decade since international image/video coding standards, such as JPEG, MPEG-4, H.264, and HEVC, have been widely used and distributed over Internet. The CBIR engine can help to find user interested relevant multimedia contents, either through multiinstance query or relevance feedback control. Before the CBIR search engine being developed, text information isthe only precise data used to perform content similarity retrieval, such as filename, creator, and content descriptions.

However, the text-based CBIR requires human annotation and content categorization, such that large scale retrieval is not feasible. In addition, the categorization and annotations would be different through different human labeling, which would bias the retrieval results.

To perform server-client CBIR, the server has to record addresses and feature characteristics of all client peers. To respond a query, the server helps the query peer, P, to forward the query message, Q, to all peers with relevant contents, which would perform retrieval and transmit relevant images toward P. This centralized approach suffers heavy network traffics in that unnecessary retrieval and transmission are involved. To eliminate the centralized traffic loading, P2P CBIR is proposed.

Multi-instance query is widely used by the CBIR search engine to improve retrieval accuracy and reduce retrieval processing time. As one image would demonstrate various types of features, performing multi-instance query with multiple feature (MIMF) types can yield more accurate and robust retrieval results, which can be achieved by assigning weighting to different similarity ranks corresponding to different feature types. Authors in [1] proposed to perform MIMF on the P2P paradigm for progressive and scalable retrieval. In this method, the query search time was high because of the network delay and image matchingat each peer. We take this problem as our motivation and propose methods to optimize the query search time.

2. RELATEDWORK

Our work is identified with three noteworthy gatherings of research: substance based picture recovery, remove metric learning, and web based learning. In the accompanying, we quickly audit the firmly related agent works in each gathering.

A. Content-Based Image Retrieval:

With the rapid growth of digital cameras and photo sharing websites, image retrieval has become one of the most important research topics in the past decades, among which content-based image retrieval is one of key challenging problems [1], [2], [3]. The objective of CBIR is to search images by analyzing the actual contents of the image as opposed to analyzing metadata like keywords, title and author, such that extensive efforts have been done for investigating various low-level feature descriptors for image representation. For example, researchers have spent many years in studying various global features for image representation, such as color features, edge features, and texture features. Recent years also witness the surge of research on local feature based representation, such as the bag-of-words models using local feature descriptors. Conventional CBIR approaches usually choose rigid distance functions on some extracted lowlevel features for multimedia similarity search, such as the classical Euclideann distance or cosine similarity.

However, there exists one key limitation that the fixed rigid similarity/distance function may not be always optimal because of the complexity of visual image representation and the main challenge of the semantic gap between the low-level visual features extracted by computers and high-level human perception and interpretation. Hence, recent yearshave witnesses asurge of active research efforts in design of various distance/similarity measures on some low-level features by exploiting machine learning techniques, among which some works focus on learning to hash for compact codes and some others can be categorized into distance metric learning that will be introduced in the next section. Our work is also related to multimodal/multiview studies, which have been widely studied on image classification and object recognition fields.

However, it is usually hard to exploit these techniques directly on CBIR because (i) in general, image classes will not be given explicitly on CBIR tasks, (ii) even if classes are given, the number will be very large, (iii) image datasets tend to be much larger on CBIR than on classification tasks. We thus exclude the direct comparisons to such existing works in this paper. There are still some other open issues in CBIR studies, for example, the effectiveness and adaptability of the recovery procedure that frequently requires a powerful ordering plan, which are out of this current paper's degree.

When all is said in done, P2P IR frameworks canbe ordered into a few gatherings (cf. [6]) 1. Systems of the first group follow a semantic query routing approach based on peer summaries. Routing Indices [9] are among the first approaches presented in literature belonging to this group. Based on summary information of neighboring peers that is aggregated along multiple hops, a peer routes queries towards the direction of peers potentially containing relevant documents w.r.t. the query. In order to restrict the size of peer summaries, topics are indexed rather than individual terms. As opposed to Routing Indices, which follow a multi-hop semantic routingapproach, PlanetP [8] and its scalable extension Rumorama apply single-hop semantic routing. Therefore, summaries are sent to all peers in a subnetwork.

A third class of P2P IR systems is represented by distributed indexing structures with distributed hashtables (DHTs) as its most prominent class member. Minerva [2] has been designed for the administration of text documents, where term statistics are indexed in a DHT. Every peer is responsible for a certain set of terms. Novak et al. have presented a large-scale CBIR architecture [23] based on a DHT. Within DHTs, indexing data of a peer's content is transferred to remote peers with every peer being responsible for a certain range of the feature domain of an individual feature. Presumably, for example, correlations between geographic information and image content are difficult to exploit. If we e.g. assume an image from the Sahara Desert with shades of beige sand and blue sky, different peers might be responsible for indexing the geographic and the image content information. In this way, while circulating the ordering information of the Sahara picture, questioning for it, or expelling it from the system, two distinct companions (in any event) must be reached. Indeed, even with just a single component sort being filed, the frequent joining and leaving of peers leads to an increase in network traffic as term statistics are transferred to or removed from remote peers.

3. PROBLEM STATEMENT

The P2P-CBIR system diagram proposed in [1] is demonstrated in Fig. 1, which comprises off- and on-line stages.

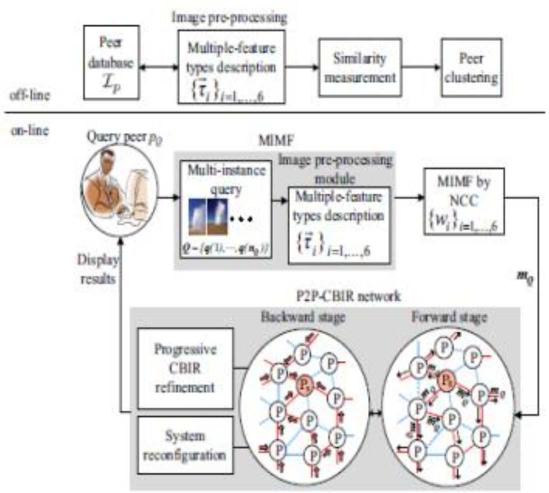

Fig. 1: The P2P-CBIR system diagram.

At the off-line stage, an image pre-processing module was integrated in each peer to extract descriptors of new images, such as color, texture, and shape. At the on-line stage, the query peer P that received the Q performs MIMF.

The problem in this approach is that the query feature is matched against each image in the peer during the online search stage. This results in bigger search time.

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. Ordinarily, the element vectors are recorded by a dispersed high-dimensional file or Locality Sensitive Hashing (LSH) over the DHT overlay asopposed to brought together conditions, information in P2P systems is appropriated amongvarious hubs, therefore a CBIR calculation needs to list and scan for pictures in a circulated way. P2P systems are under consistent beat, where hubs join/leave and documents distribute to/expel from the system, the record should be refreshed powerfully to adjust to such changes. Dexing and Locality-Sensitive Hashing. The high dimensional ordering based methodologies store the component vectors in an information structure, as a rule a tree or a chart, to accomplish successful hunt space pruning amid recovery. In organized P2P systems, the high-dimensional record is characterized distributedly over the P2P overlay, dexing and Locality Sensitive Hashing. The high-dimensional ordering based approaches store the element vectors in an information structure, as a rule a tree or a chart, to accomplish compelling inquiry space pruning amid recovery. In organized P2P systems, the high-dimensional record is characterized distributedly over the P2P overlay.

4. PROPOSED SOLUTION

We introduce two additional concepts in offline phase of P2PCBIR which will help to achieve reduced search time.

1. Clustering of Peers

2. Constructing search index for Peer cluster

In the online phase, the way of searching is modified to achieve fast search time.

4.1 Clustering of Peers

Peers must be grouped together based on any parameters like peers in a particular geographical area. Each peer will advertise the number of requests it has received till recent to all theneighbors in itsgeographical area with a distance of K hops. Once all the peers have exchanged the information, the peer node with maximum number of request till now will become the cluster head. All the other peers will come the part of the cluster. A peer node can also be part of two clusters. This occurs for the case of boarder peers. The cluster head peer will start a mobile agent. This mobile agent will visit all the peers in the cluster and visit the cluster head. The mobile agent will carry information of the image feature cluster information in each peer to the cluster head peer.

4.2 Constructing Search Index for Peer Cluster

The mobile agent will bring the feature vector of images in each peer to the cluster head peer. Based in this clustering is again done to group the image features. The mobile agent cannot bring the all image feature information from other peers at one shot. So the clustering algorithm which we will implement must be a agglomerative clustering algorithm. With this cluster information, a search index must be constructed. The search index will maintain the map of what features are available at the peer cluster as the whole and the whereabouts of each features in the peer cluster.

4.3 Modified Search

In the search phase, when the query is sent over P2P network, each peer will forward the query to its cluster head peer. Once the cluster head peer receives the query, it will search in the search index and look for the presence of the target feature presence or similarity with any features in its search index. If any matching entry is found the query is redirected to the peers in which feature is found. If the no matching entry is found, the cluster head will forward to neighbor cluster head.

5. CONCLUSIONS

We have detailed our proposed solution for reducing the search time for CBIR in P2P-CBIR system. In future we plan to analyze the performance of our solution in terms of search time for different cluster sizes and compare with existing solution. In addition we also plan to propose a representation of search index to enable faster search using spare coding representations.

REFERENCES

[1]. Chun-Rong Su, Jiann-Jone Chen, and Kai-Lin Chang "Content-Based Image Retrieval on Reconfigurable Peer-to-Peer Networks" 2013 International Symposium on Biometrics and Security Technologies.

[2]. Overview of the MPEG-7 Standard (version 10). ISO/IEC JTC1/SC29/WG11 N6828, Oct. 2004.

[3]. M. Flickner, et al., "Query by image and video content: the QBIC system," IEEE Computer, vol. 28, no. 9, pp. 23-32, 1995.

[4]. C. Carson, et al, "Blobworld: image segmentation using expectation maximizations and it application to image querying," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, pp. 1026V38,2002.

[5]. J. R. Smith and S. F. Chang, "Visualseek: a fully automated contentbased image query system," in Proc. ACM Multimedia, 1996.

[6]. T. Gevers and A. Smeulders, "Pictoseek: combining color and shape invariant features for image retrieval," IEEE Trans. Image Process, pp. 102-119, 2000.

[7]. P. Natsev, et al., "WALRUS: a similarity matching algorithm for image databases," Technical report, Bell Laboratories, Murray Hill, 1998.

[8]. R. Rahmani, et al., "Localized content based image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., 2008.

[9]. Tram Truong Huu& John Montagnat: Virtual Resource Allocations distribution on a cloud infrastructure (IEEE, 2010), pp.612-617.

[10]. Y. Chen, J. Bi, and J. Z. Wang, "MILES: multipleinstance learning via embedded instance selection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, pp. 1931-1947, Dec. 2006.

AUTHOR'S PROFILE:

Ms. J.Srinija Pursuing M.Tech

Department of Computer Science and technology, Andhra Loyola institute of engineering and technology, Vijayawada.

Mr. Y.RAJESH$_{\textbf{M.Tech.,Ph.D.,}}$

Asset.Professor, Dept. of Computer Science & Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada.