

# Improved Accuracy of Image Retrieval by Using K-CBIR

Snehal Yadav<sup>1</sup>, Sneha Varne<sup>2</sup>, Neha Jadhav<sup>3</sup>, Shamal Powar<sup>4</sup>, Priyanka Patil<sup>5</sup>

BE, Computer Science And Engineering, SGI, Maharashtra, India

\*\*\*

**Abstract** - The community networking sites like Facebook as well as Flickr allow users to upload the images. E-commerce web sites like flipkart and amazon etc. also provide number of products related images to user. Images which are present in social networking websites are accompanied by various tags, comments, annotations and other related information to form an image-rich information networks. Due to reality of different related information, retrieval of images from image rich information networks is very challenging and complex task. In this paper we used an integrated algorithm (S-Cbir) to retrieve images from image rich information networks. The algorithm consists of two different techniques such as link based image retrieval and content based image retrieval systems. In this paper K- Means algorithm is first applied to find out the node similarity in image rich information networks and then CBIR technique is applied to find out relevant images. Basic feature and texture extraction descriptors are used in CBIR technique. The search results improve the accuracy of search.

**Key Words:** K-Means, CBIR, Feature Extraction etc...

## 1. INTRODUCTION

The images which are present in social multimedia networking websites such as Facebook, Twitter etc. are accompanied by different tags, comments, annotations and other related information. Various types of images are also used in E-Commerce web sites like amazon, flipkart etc. They are accompanied by information such as links, producer and consumer related information, comments and annotations. Images in such type websites accompanied by all the links, annotations, comments and tags form an image rich information networks[8]. Information retrieval is the process of obtaining the relevant information for a given query from a collection of information resources. Information retrieval in image-rich information networks is very useful but challenging task, that is because existence of information such as text, image feature, user, group, and most importantly the network structure that is used by owner of that site.

Image search is a specialized data search used to retrieve relevant images from a dataset. To search images user need to enter a search query which can be keyword, image file/link or click on some image as a feedback after that system will return images similar to the search query.

After getting the image results one more problem is, how you can define the ranking of those images. Image Search Re-ranking (ISR) is defined as the improvement of search results by employ image visual information to alter the initial text-based search results. It comes from the observation that the noisy text-based search results still contain satisfactory images in top hundreds of search results. Current ISR approaches mainly focus on two important aspects: feature extraction and ranking function[7]. Based on that feature extraction currently we decide ranking of that images.

## 2. LITERATURE SURVEY

Z. Yang and C.-C. J. Kuo they extend the limited capabilities of proprietary solutions in identifying multimedia contents that exist today. a comprehensive survey of image database management techniques and systems is performed, and the status in the MPEG-7 standardization process is reported.

J. Laaksonen, E. Oja, M. Koskela, and S. Brandt they performed iRIN designed for performing image retrieval in image-rich information networks. We first introduce MoK-SimRank to significantly improve the speed of SimRank, one of the most popular algorithms for computing node similarity in information networks. Next, we propose an algorithm called SimLearn to extend MoK-SimRank to heterogeneous image-rich information network, and account for both link-based and content-based similarities by seamlessly integrating reinforcement learning.

X. Jin, J. Luo, J. Yu, G. Wang, D. Joshi, and J. Han, they designed link plus content based similarity approach is presented to perform image retrieval in heterogeneous image rich information network. Also personalized image search approach is presented in this paper which is based on annotations of the image. Presented image retrieval system will help user to get better results in the terms of relevance.

### 3. MOTIVATION OF PROJECT

The design of ranking function is key issue in ISR, which has attracted great research interests in recent years. ISR is categorized into 3 methods such as classification-based methods, graph-based methods, and learning-to-rank-based methods. These 3 methods are computationally expensive and require sufficient training data to train a learning model, which is not very practical in real image re-ranking applications.

Motivated by this we propose an integrated algorithm S-CBIR to address the limitations of ISR.

- Most of the image search engines use textual similarity to retrieve relevant images and then they use visual similarity to search visually relevant images from the dataset.
- Some integration based approaches discussed in this[5],[6]paper which is based on linear or nonlinear combination of textual and visual features of images.
- The K-means algorithm is used to retrieve images in image rich information networks and to find out the deficiencies and drawbacks of IWSL first we implement that algorithm.

### 3. PROPOSED WORK

We propose a framework for image re-ranking in image rich information network. Different techniques are used for image re-ranking. This project addresses the issue of performance of these techniques. By using these techniques we can retrieve the images from dataset. We examine the performance of these methods but there are some limitations of these methods such as it is computationally expensive and require sufficient training data to train a learning model, which is not very practical in real image re-ranking applications.

To overcome the limitations of the existing system integrated algorithm(S-CBIR) has been proposed.

### 4. SYSTEM ARCHITECTURE:

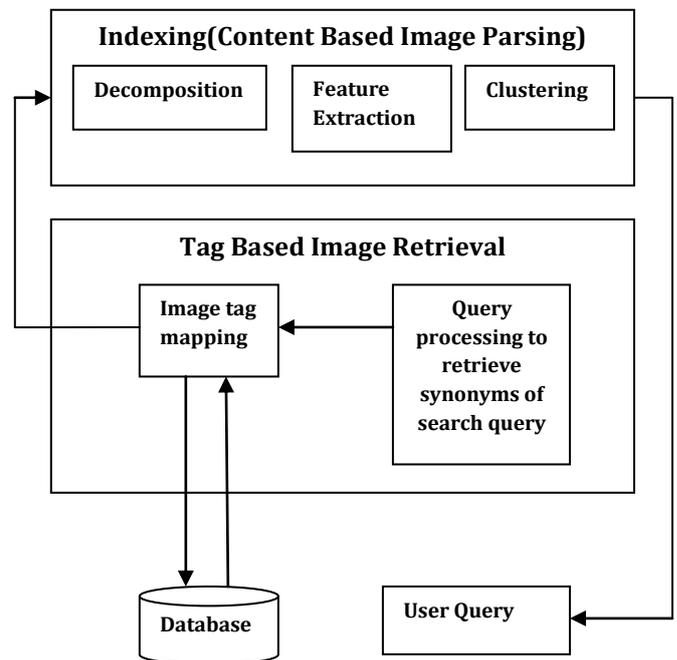


Fig. 1: System Architecture

### 5. PROPOSED INTEGRATED ALGORITHM

Main steps in proposed integrated algorithm are as follows:

#### 1. Preprocessing step:

Load all images from dataset.

- 1) Load image tag ids from image dataset.
- 2) Create indexing common tags from image dataset.
- 3) Compute K-means algorithm.
- 4) Extract and store features of all images in dataset.

#### 2. Integrated algorithm (S-CBIR)

1. Link based image retrieval technique:

**Input:** keyword query (tag)

- 1) Check whether entered keyword is present in dataset, If tag is present in dataset then get tag ids for that tag.
- 2) Find out image ids for retrieved tag ids.
- 3) Fetch K-Means for all the image ids retrieved in all other image nodes to get score.
- 4) Display images

**Output:** images

#### 2. Content based image retrieval:

**Input:** output images from link based image retrieval.

- 1) Selection of an image as input query.
- 2) Perform basic feature extraction.
- 3) Perform color extraction.
- 4) Perform texture extraction
- 5) Sort and display results.

**Output:** relevant images.

## 6. CONCLUSION

This paper presents an efficient way of searching relevant images by modeling a social image sharing website as image rich information network. We propose the S-CBIR algorithm which includes link based image retrieval and content based image retrieval. Both global and local features are considered while retrieving images by using S-CBIR algorithm. We have conducted the experiment on Flickr dataset and achieved results which shows proposed algorithm improves the accuracy of search results than the link-based images retrieval and content based image retrieval techniques. The proposed algorithm provides semantically and visually similar images as output.

## 7. FUTURE WORKS

Due to unavailability of amazon dataset we have created own dataset, we would like to conduct the experiment on heterogeneous datasets if we get another image dataset like amazon. The improvement in the CBIR is possible if SURF descriptor used instead of SIFT.

## REFERENCES

1. J. Huang, S. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on*, pp. 762– 768, Jun 1997.
2. Z. Yang and C.-C. J. Kuo, "Survey on image content analysis, indexing, and retrieval techniques and status report of mepg-7," *TAMKANG JOURNAL OF SCIENCE AND ENGINEERING*, vol. 2, no. 3, pp. 101– 118, 1999.
3. J. Laaksonen, E. Oja, M. Koskela, and S. Brandt, "Analyzing low-level visual features using content-based image retrieval," in *International Conference on Neural Information Processing (ICONIP)*, pp. 1333– 1338, 2000.
4. X. Jin, G. Wang, J. Luo, D. Joshi, J. Yu, and J. Han, "irin: Image retrieval in image-rich information networks", , ACM, April 26–30, 2010.
5. H. H. Wang, D. Mohamad, and N. A. Ismail, "Approaches, challenges and future direction of image retrieval," *CoRR*, vol. abs/1006.4568, 2010.
6. L. Duan, W. Li, I. Tsang, and D. Xu, "Improving web image search by bag-based reranking," *Image Processing, IEEE Transactions on*, vol. 20, pp. 3280–3290, Nov 2011.
7. M. S. Mehwish Rehman, Muhammad Iqbal and M. Raza, "Content based image retrival: Survey," *World Applied Sciences Journal*, 2012.
8. X. Jin, J. Luo, J. Yu, G. Wang, D. Joshi, and J. Han, "Reinforced similarity integration in image-rich

information networks," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, pp. 448–460, Feb 2013.

9. V. S. Kore, B. A. Tidke, and P. Chandre, "Article: Survey of image retrieval techniques and algorithms for image-rich information networks," *International Journal of Computer Applications*, vol. 112, pp. 39–42, February 2015.