

Image Retrieval using Perceptual Hashing

Harsh Kumar Sarohi¹, Farhat Ullah Khan²

¹(Department of computer science, Amity University, India)

²(Department of computer science, Amity University, India)

Abstract : This paper proposes a new perceptual hashing based approach for Image Retrieval. Information in form of images is increasing at a very fast rate therefore there is need for faster and effective retrieval of specific image from image repositories. Content Based Image Retrieval (CBIR) is considered most efficient to search large image collections. The proposed method uses perceptual hash function which calculates similar hash value for similar images. Finally, using an adequate distance or similarity function to compare two perceptual hash values, it can be decided whether two images are perceptually different or not. Perceptual image hash functions can be used e.g. for the identification, authentication or integrity verification of images. In this paper we also propose a hash algorithm that is robust and secure to non malicious manipulation and sensitive to the malicious tampering.

Keywords - Content based Image Retrieval (CBIR), Perceptual Hash algorithm, Image Retrieval using Perceptual Hashing, Query by Image Content (QBIC).

I. Introduction

In order to deal with searching in large image datasets two solutions are available

1. Image annotation

In this approach images are tagged and meta-data is attached to images [1], then keywords are used in search interface for getting specific image. However, tagging of images is time consuming, leads to inefficient data storage and since every image cannot be described in the form of keywords there is too much responsibility on the end user for forming the query. The solution to problem encountered is (Content Based Image Retrieval) CBIR.

2. Content Based Image Retrieval

In this technique images are retrieved based on content of an image. Two major steps in CBIR are

2.1 Feature Extraction Algorithm

This algorithm processes the image and compute feature vector. This feature vector [2] is used for the extraction of information from data set. Features can be human vision related or low-level features, objects related or middle-level features, high-level features or semantic related. Most commonly used features for retrieval are color, shape, texture.

2.2 Matching Algorithm

It performs comparison between extracted features of images to check whether they are similar or not and up to what extent they match. It uses distance functions like Euclidean distance, Mean Square Error for measurement of similarity.

II. Related Work

Existing CBIR systems are based on color, texture, shape. In color based approach color is extracted from the images to form feature vectors most. Color is represented using common color spaces models like sRGB, Adobe RGB and ProPhoto RGB, CMYK, HSV, HSL then similarity measure is computed using Euclidean distance. Texture describes statistical properties of an image; hence two images can be compared based on Texture. There are various region-based and boundary based techniques available for retrieval using shape. Swain and Ballard [3] proposed idea to use color histogram for retrieval. Lin et al. [4] proposed a scheme based on color-texture and color-histogram. Hiremath and Pujari [5] came up with a CBIR technique based on color, shape and texture. Rao et al. [6] proposed dominant color based retrieval system. CBIR has been used in many applications like recognition systems, tumor detection, detecting and blocking adult content.

III. Proposed Methodology

The proposed model is shown in Figure 1.

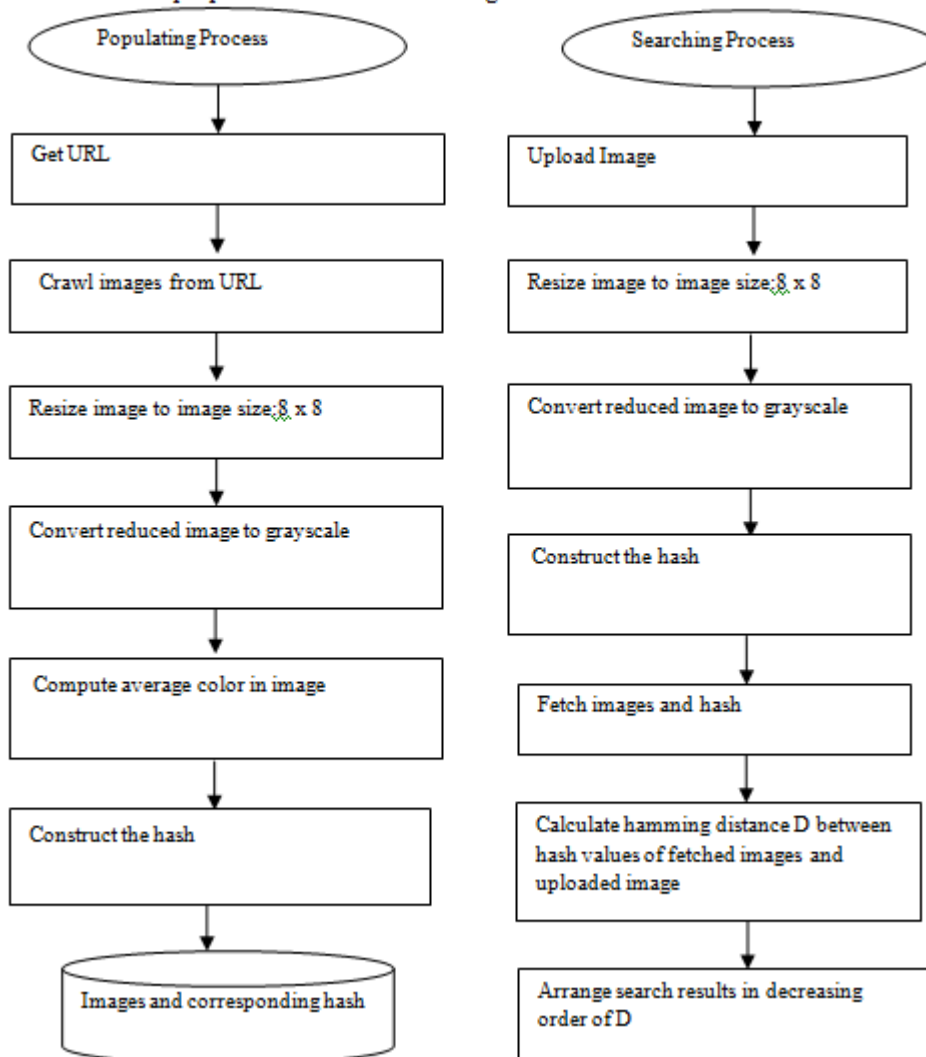


Fig 1: Flow chart of proposed system

1. Reduce size

Images are crawled resized to image size of 8 x 8 irrespective of all different image alterations so that we get same hash even if there is little variation. This results in no of pixels being reduced to 64

2. Reduce Color

Red, green and blue component [7] of images are extracted and image is converted to grayscale and RGB mean is computed

3. Hamming Distance

Images are compared using hamming distance [8] D to calculate how many bits are different in hash

4. Matching process

To compare two images, construct the hash from each image and count the number of bit positions that are different. A distance of zero indicates that it is likely a very similar picture. A distance of 5 means a few things may be different, but they are probably still close enough to be similar. A distance of 10 or more than that probably is a very different picture.

IV. Results

We have used a data set of 10,000 images of different colors. System is implemented on intel atom processor with 2GB RAM using PHP. We can see efficiency of using this method.



Fig 4: Query image



Fig 5: Search results

V. Conclusion

Color is main feature present in multimedia. This paper focuses only on efficient retrieval based on color. Proposed method uses color spaces and hamming distance to construct hash. Further, using perceptual hashing makes retrieval process very fast. CBIR based on color has been used in various applications like detecting skin lesions, retrieving tattoo images etc.

Results show that the efficiency the proposed algorithm used for perceptual hashing has low computation cost and satisfactory performance. Further, if method is used for large no of pixels it can produce great results with very high accuracy and preparing perceptual integrating other features like texture, shape will provide best results.

Acknowledgement

The authors would like to thank Amity School of Engineering and Technology (ASET) where the research was carried out.

References

- [1] T. Sumathi, C. Lakshmi Devasena, M. Hemlata, "An Overview of Automated Image Annotation Approaches", International Journal of Research and Reviews in Information Sciences, Vol. 1, March 2011.
- [2] A. Vadivel, A. K. Majumdar, S. Sural, "Characteristics of Weighted Feature Vector In Content-Based Image Retrieval Applications", Proceedings of Intelligent Sensing and Information Processing, 2004, ICISIP, 2004.
- [3] M. Swain, D. Ballard, "Color indexing", International Journal of Computer Vision, 1991.
- [4] C.H. Lin, R.T. Chen and Y.K. Chan, "A smart content-based image retrieval system based on color and texture feature", International Journal of Computer Vision, 1991.
- [5] P. S. Hiremath, J. Pujari, "Content Based Image Retrieval based on Color, Texture, Shape features using Image and its complement", 15th International Conference on Advance Computing and Communications. IEEE, 2007.
- [6] M.B. Rao, B.P Rao, A. Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features", International Journal of Computer Applications, Vol. 18, No.6, pp-0975-8887, 2011.
- [7] Manimala Singha, K. Hemchandran, "Performance analysis of Color Spaces Image Retrieval", Assam University Journal of Science & Technology: Physical Sciences and Technology, Vol. 4, 2011.
- [8] Mircea Ionescu, Anea Ralescu, "Fuzzy Hamming Distance in a Content-Based Image Retrieval System", International Journal of Information Technology and Knowledge Management, IEEE, 2004.