

Survey on Contextual Hashing for Searching and Retrieval of Images using Descriptors and Distance Measure

¹Subathra Muthuraman
(M.tech), Student,

Department of Information Science and Engineering
New Horizon College of Engineering
Bangalore, India.

²Mrs. Swathi Baswaraju
Assistant Professor,

Department of Information Science and Engineering
New Horizon College of Engineering
Bangalore, India

³Mrs. B. Mounica
Assistant Professor,

Department of Information Science and Engineering
New Horizon College of Engineering
Bangalore, India

Abstract— This paper covers increasing need for accurate image retrieval in the modern social media/web world and latest techniques that can be used to improve the accuracy of the image retrieval. “Content Based Image Retrieval” is one of the recently emerging methods, uses color, texture, and shape of the images for better matching. This paper explains about adding spatial context information along with local features which improves accuracy of image retrieval. Spatial Context information refers to the relation between the object of interest and surrounding objects. Spatial context is converted in to binary codes since the geometric verification can be done by efficient comparison of binary codes. Multimode Property of local features is explored. In the proposed method, Speeded up Robust Feature descriptor is used along with SIFT descriptor which detects edges and Hessian Affine detector which detects corners, in order to improve the performance of the process and reduce the error rate.

Keywords— Image retrieval, CBIR, spatial context modelling, hashing, geometric verification

I. INTRODUCTION

With explosive growth in social and multimedia, huge quantities of images are available in web. A Common man can use his digital cameras to capture the beautiful world in pictures and share with others. The images that are searched are available with different visual and semantic content. It is rapidly growing in size. Image retrieval refers to surfing, searching and retrieving images from a huge collection of digital images. Development of efficient image retrieval techniques would enable better utilization of such data.

Most of the current image retrieval is based on “Text descriptors” and a very little based on “Content Based Image Retrieval”, which searches and retrieves digital images from a huge database, where content, refers to the colour, texture and shape of the images. Content Based Image Retrieval is preferred as the images annotated by humans manually by entering keywords take more time.

Many Approaches depends on Bag of Visual Words model, where an image is represented by a set of visual words. This model defines a visual dictionary and quantizes the local features to the corresponding visual words. The Visual dictionary is built off-line by unsupervised clustering algorithms. In this way, an image is represented by a set of visual words. But since the spatial context information among the local features is not considered, the Bag of Visual Words model lacks accuracy. The proposed retrieval method improves the retrieval performance by including the spatial context information into the indexing structure. The storage and computation efficiency can be improved by binarizing the spatial context information that is converting the high dimensional data into binary code.

This article discusses the various methods for retrieving images in section II that are applicable to the field of image processing. The challenges in image retrieval during similarity matching are discussed in section III. The article is concluded in section IV.

II. DIFFERENT APPROACHES IN IMAGE RETRIEVAL

Semantics-Sensitive Approach

In [1], the author explains a Semantic-Sensitive approach to Content-Based Image Retrieval. In order to extract the feature correctly, a semantic categorization is followed. Following this categorization, similarity is computed based on the region. The important aspect of the proposed system is its retrieval

speed. For similarity matching purpose, the measure used here is integrated region matching which enables the faster retrieval. This approach categorizes the images as either textured-non textured graph-photograph. This approach was applied to a database with 200,000 images. Its retrieval of images was accurate and fast. It was robust to intensity variations, sharpness variations, scaling, rotation, and cropping. The limitation in this approach was that while classifying the image, it may fall into second semantic classes.

Blobworld Representation of Images

In [2], the author has explained the querying based on region using homogenous color texture segments known as blobs. It transforms the raw pixel data to a smaller set of image regions that are coherent in color and texture. This kind of image representation is known as Blobworld representation which is created by clustering points in a joint color-texture position feature space. Image to Image matching is not followed. If the blobs are identified by the user which is related to some concept such as "rose" then the users search will be looking for a rose within other images in several different backgrounds. This approach requires the involvement of the user. This algorithm has been run on 10,000 natural images. The proposed approach allows the user to view the internal representation of the submitted image and results.

Text retrieval approach to object matching in videos

In [3], an approach to object and scene retrieval which searches for and localizes all the occurrences of a user outlined object in a video is discussed by the author. As the object is represented by a set of viewpoint invariant region descriptor, it can be found out even when there is a change in view-point, illumination and partial occlusion. The paper's main aim is to retrieve the key frames and shots of a video containing a particular object. The author has used the approach of Text Retrieval. A text is retrieved by computing its vector of word frequencies and returning the documents with the closest vectors. The advantage of this approach is run-time object retrieval throughout a movie database without any delay by pre-computing the matches. The author has demonstrated the matching on two full length videos.

Content-Based Image Retrieval Using Shape and Depth features

In [4], the author describes an algorithm for retrieving images using the shape information in an image. It has also considered the 3D information of the image. The proposed linear approximation procedure captures the depth information based on the idea of shape from shading. The objects are retrieved using the similarity measure that combines both the shape and the depth information. This approach has been effective in retrieving engineering objects

Spatial-Bag-of-Features

In [5], the method of retrieving large scale images by creating a new bag-of-features that includes the information related to geometry of object in an image is described by the author. The Ordered bag-of-features are generated by projecting the local features of an image to different directions or points.

Depending on the ordered bag-of-features, distinct group of spatial-bag-of-features are designed to capture the invariance of object scaling, translation and rotation. Representative features are selected to generate the new bag-of-features. The author has designed the new bag-of-features for image retrieval goals such as the features of new bag and current bag are in the same format and the new bag determines the spatial information. The author summarizes that if the goals are met, then the information such as local features and spatial relationships are successfully organised in a single inverted index which makes the image searching fast and accurate. The proposed retrieval technique by the author leads to efficient image matching and indexing for large scale image retrieval and it also includes the spatial information of local features to improve the retrieval accuracy. The author first evaluated the spatial bag-of-features which was compared with bag-of-features, and then the bag of features was compared with RANSAC re-ranking. This experiment was carried on Oxford5K dataset. To test the effectiveness and scalability of spatial bag-of-features the dataset Panoramio1M was leveraged.

Hashing Shape Context Descriptors

In [6], a method that describes how to arrange and index the logo digital libraries is discussed. The Proposed retrieval system compares the query image with the logo images present in the database and retrieves them based on their similarity. These logos are described by a variant of the shape context descriptor. The Locality Sensitive Hashing indexing structure is used to arrange the descriptors to carry out the search process. Hashing techniques speeds up the indexing and retrieves logos based on similarity. The author has carried out the experiment on the tobacco-800 dataset to demonstrate the effectiveness and efficiency the proposed method. The author has validated the approach by using a repeated random sub-sampling validation scheme.

Bag of Visual Words

In [7], the approach of Bag of Visual Words to retrieve the relevant word images from a big database correctly is discussed. This approach is based on the principles of text retrieval system. The representation of word images are in the form of histogram of visual words. The histogram carries the information of the features in the image. Visual words are quantized to represent local features in an image. Since Bag of Visual words method does not explain the spatial relationship among visual words; the author has applied re-ranking method to the retrieved list of images in order to improve the performance. The author has validated this approach on four Indian languages and it is proved to be language independent and scalable. The author has demonstrated the utility of the proposed system across four Indian languages by using the dataset of 100K words. To demonstrate the scalability, the author has used large dataset of 1M words. The performance is measures by precision. The limitation of this process is the re-ranking step which is time consuming.

Wavelet Based Color Histogram Image Retrieval

In [8], the author has explained the Content Based Image Retrieval using the color and texture feature called Wavelet Based Color Histogram Image Retrieval. The texture features are extracted through wavelet transformation and color histogram are used to extract the color features. The combination of both texture and color feature is robust to scaling and translation of objects in an image. A clear understanding about color and representation of color in digital images is required to extract the color features from digital images. The texture descriptor provides measures for the properties such as smoothness, coarseness, and regularity. Distance function is used to find the similarity between the images. The author has demonstrated the proposed retrieval method on a WANG image database with 1000 general purpose color images and compared with the results of different authors. The result of the proposed method was shown to have better performance than others with the average retrieval time as 1 minute.

Principal Visual Word Discovery

In [9], the author has described a method to detect license plates in various observation angles, scale changes and illumination variation. It can also detect multiple plates in the image. Scale Invariant Feature Transform descriptors are used to deal with different angles, scale changes etc. The process is carried out with the help of Principal Visual Word discovering and Visual Word Matching. The author has demonstrated the proposed method on two different dataset such as LP dataset which is built by the author itself and another dataset named Caltech Cars which is downloaded from the internet. The detection rate of the proposed approach is lower than the other approaches such as HLPE, LPE, and ESM. This approach makes a false positive rate of only 1.0%. The author has investigated the time efficiency from two aspects such as feature extraction time and detection time after the feature extraction. It takes less time in average to process one image. The limitation of this approach is that, it fails to detect the license plate when the image quality is poor.

Comparison of SIFT and SURF

Two methods are discussed in [10] such as, Scale Invariant Feature Transform to extract the unique features from the images, where features in the images remain the same whatever may be the image scale and rotation and Speeded Up Robust Features, a scale and rotation invariant interest point detector and descriptor which makes use of integral images. The author has discussed about the process of Image registration which converts the different groups of data into a single coordinate system. Image registration is a difficult task in many applications. In Image registration, first the features are detected and matched, then a transformation function is derived according to the features in the image and finally the image is reconstructed based on the transformation function. The author has taken two images for the experiment. The features were detected in both the images by the descriptors. The author found out that SIFT descriptor detects more features when compared to SURF, but SURF was found to be fast.

Scalable Partial-Duplicate Mobile Search

In [11], the author has explained the large-scale partial duplicate image search on mobile platforms. Since the Scale Invariant Feature Transform descriptor, a histogram based descriptor is not the finest one for the search; the author has proposed the Edge-SIFT (Scale Invariant Feature Transform) descriptor. Edge-SIFT descriptor is built upon edge maps of local image patches and keeps both location and orientations of edges. The author has also proposed an inverted file based indexing framework in order to make use of Edge-SIFT descriptor in big scale partial duplicate mobile search. The author has used the Oxford Building dataset to test the effects of different parameters and to evaluate the validity of Edge-SIFT compression and compares with SIFT and ORB. Edge-SIFT perform better than the other two approaches in terms of retrieval accuracy, efficiency, and transmission cost and memory consumption. The limitation of Edge-SIFT descriptor is that it cannot be used like SIFT or SURF to perform other tasks such as recognition, classification etc.

Semantic-aware Co-indexing

In [12], the author explains an algorithm called Semantic-aware Co-indexing algorithm for vocabulary tree based image retrieval. It searches the images based on the conditions such as the similarities between the images are based on the local features and semantic attributes into the inverted indexes. After the search, the retrieval process considers not only the images having same local features but also allows consensus in their semantic similarities. The proposed algorithm leverages semantic attributes from advanced object recognition to update the inverted indexes of local features quantized by a large vocabulary tree. The methods used for the extraction of the BOW features from the images are HOG and LBP, and the SIFT descriptors are extracted to use the vocabulary tree. The online indexing of the features produced in this process is simple and it is less memory consuming. The author has demonstrated the proposed approach on three different datasets such as Holidays, UKbench, and Oxford. The proposed approach increases the overall discriminative capability of the inverted indexes which gives good retrieval results.

Object based Image Retrieval using Combined Features

In [13], the image retrieval system that retrieves the image based on the local and global features is discussed. The Bi-directional Empirical Mode Decomposition technique is used to detect the edges and Harris Corner detector is used to detect the corner points of an image. HSV color feature is used as the global feature. The author has applied the system on the ten categories of images each with seventy two different orientations from COIL-100 image database.

HSV-Color Histogram and GLCM

In [14], the Content Based Image Retrieval, which retrieves the images based on the similarity of color and texture features of image sub blocks is discussed. The image is segmented into sub-blocks of equal size. From each sub-block, the color is extracted by quantifying the HSV color space into non equal intervals and it is represented by cumulative histogram. Gray-level occurrence matrix is used

to get the texture of the sub-blocks. The Similarity measure used here is Euclidean distance. This method has better performance than the other system that uses only HSV color, only GLCM texture and combined HSV color and GLCM texture.

Deep learning for Content-Based Image Retrieval

In [15], the author has introduced a deep learning framework for the Content Based Image Retrieval by training large-scale deep convolution neural networks for learning effective feature representation of images. Deep learning is a class of machine learning. In this technique layers of information are exploited for pattern classification for feature classification. The author has conducted an extensive set of empirical studies for comprehensive evaluations of deep convolution neural networks with applications to learn feature representations for a variety of CBIR tasks under varied settings. The author has aimed to evaluate the performance of feature representation scheme on new CBIR tasks such as object retrieval using the Caltech256 dataset, landmark retrieval tasks using the Oxford and Paris dataset and facial image retrieval tasks using the Pubfig83LFW dataset.

III. CHALLENGES IN CONTENT BASED IMAGE RETRIEVAL

There are some challenging issues in content based image retrieval during the similarity matching such as distance functions, semantic gap, and goal of users. Many distance functions are available. Out of which the one that characterizes the underlying visual similarity between the images has to be chosen. The gap between the image semantics and the low-level features has to be bridged in image retrieval process. The goal of the user has to be targeted effectively.

IV. CONCLUSION

This review article illustrates the various methods used for the retrieval of images. An Image searching processing algorithm which finds the corresponding images with the help of SIFT descriptors and Hessian affine detector is proposed. SURF descriptor is used along with SIFT descriptor and Hessian affine detector in order to improve the performance of the retrieval rate. SURF descriptor will reduce the error rate.

REFERENCES

1. J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(9), 947–963, 2001.
2. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-maximization and Its Application to Image Querying", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.
3. J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec.2003, pp. 1470–1477.
4. Amit Jain, Ramanathan Muthuganapathy, and Karthik Ramani, "Content-Based Image Retrieval Using Shape and Depth from an Engineering Database", G. Bebis et al. (Eds.): *ISVC 2007, Part II, LNCS 4842*, pp. 255–264, 2007.
5. Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-of features", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3352–3359.
6. Marçal Rusinol, Josep Lladós, "Efficient Logo Retrieval through Hashing Shape Context Descriptors", in *Proc. ACM*, 2010.
7. Ravi Shekhar and C.V. Jawahar, "Word Image Retrieval using Bag of Visual Words", in *Proc. IEEE*, 2012.
8. Manimala Singha and K.Hemachandran, "Content Based Image Retrieval using Color and Texture", *Signal & Image Processing: An International Journal (SIPIJ)* Vol.3, No.1, February 2012.
9. W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection", *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4269–4279, Sep. 2012.
10. P M Panchal, S R Panchal, S K Shah, "A Comparison of SIFT and SURF", *IJIRCCE*, Vol.1.Issue 2, April 2003.
11. S. Zhang, Q. Tian, K. Lu, Q. Huang, and W. Gao, "Edge-sift: Discriminative binary descriptor for scalable partial-duplicate mobile search," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2889–2902, Jul. 2013.
12. S. Zhang, M. Yang, X. Wang, Y. Lin, and Q. Tian, "Semantic-aware co-indexing for image retrieval", in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Apr. 2013, pp. 1–8.
13. H.Kavitha, M.V.Sudhamani, "Object based Image Retrieval from Database using Combined Features", *International Journal of Computer Applications*, Vol.76, Issue No.8, August 2013.
14. Deepak John, S.T.Tharani, and SreeKumar.K, "Content Based Image Retrieval using HSV-Color Histogram and GLCM", *An International Journal (IJARCSMS)*, Vol.2, Issue No.1, January 2014.
15. Ji Wan, Dayong Wang, Steven C.H. Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, Jintao Li, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study", in *Proc. ACM Int. Conf. Multimedia*, 2014.