A Survey on Large-Scale Image Search Techniques

Usha.D
Department of Computer Science and Engineering
Sri Ramakrishna Institute of Technology, Coimbatore, Tamilnadu, India.

Jim Mathew Philip
Department of Computer Science and Engineering
Sri Ramakrishna Institute of Technology, Coimbatore, Tamilnadu, India.

Abstract - Rapid growth of technologies and memory storage techniques have resulted in the enormous growth of digital images. Searching an image from large database using text results in inefficient search. This leads to the emergence of Reverse image search technique. This technique is used to search images by using a query image. This paper provides the survey of various techniques used for the efficient search of similar images from a large database using a query image. This survey also provides memory size required and accuracy in searching images for various techniques.

Keywords - Reverse Image Search, Content Based Image Retrieval, SIFT descriptor, Mean average precision.

I. INTRODUCTION

With the explosive growth of technologies and memory storage techniques, number of digital images is increasing a lot. Searching an image from the large database using text query such as keywords, tags is an increasing problem. This method retrieves only the images containing the given text and it is very difficult. This leads to the emergence of searching image by image query technique such as Reverse image search. Reverse image search is a kind of Content Based Image Retrieval (CBIR) technique in which an image is given as a query to find the similar images based on their contents such as colors, shapes, textures from the large database. Similar images are retrieved such that the most similar image is ranked first in the search results.

Image searching technique initially performs feature detection process. Feature is referred as the significant structures of an image and it includes points, edges, regions and objects. Initially features can be detected by using feature detection methods such as Hessian-affine detector, Differences of Gaussian(DOG), Maximally stable extremal regions (MSER).Then the detected feature can be described by using the discriminative descriptor such as Scale Invariant Feature Transform(SIFT) [16] method, Speeded Up Robust Features(SURF) [15].Then the descriptor can be converted into vector representation by using methods such as Bag of visual words (BoVW), Fisher vector(FV),Vector of locally aggregated descriptors(VLAD) [13] and GIST representation. Finally indexing and similarity measurement methods are performed to find the similar images.

Searching an image from large database has two main issues. Increased memory size required for indexing and reduction in retrieval accuracy. Accuracy is measured in terms of mAP (mean average precision).This paper presents different techniques for the efficient searching of similar images from large database and handling of the mentioned issues. The main purpose of this paper is to review the related work on the various feature detection, extraction, transformation techniques, indexing and similarity measurement techniques for efficient image search.

II. RELATED WORK

Search results of the image search mainly depends on the techniques used in the searching process. The following section is surveyed with different techniques used for feature detection, extraction, transformation, indexing and similarity evaluation of image search. Table 1 shows various techniques used for the efficient image search.

Shiliang Zhang et.al [1] has proposed the usage of descriptive visual words(DVWs) and also the descriptive visual phrases(DVPs).From the classic visual words, the visual words that describes effectively certain objects are taken as DVW. The co-occuring visual word pairs are taken as DVP. VisualWordRank is used for the selection of DVW. After DVW and DVP are selected, they are combined to form a set. The set is first used in large-scale near duplicated image retrieval and inverted file structure is used for indexing and TF-IDF weighting is used for similarity measurement. The image ID and frequency for each of DVW and DVP are stored in index and hence it makes the index size more compact. The performance increases when the number of DVPs increases. The combination of 13057 DVWs and 9×10^6 DVPs perform better than the classic visual words by 19.5% in terms of mean average precision. The second application of DVW and DVP set is image search re-ranking. DWPRank is used for the re-ranking purpose. Using this method, the collected database images are sorted again based on the relevancy of the query image. By using DWPRank, DVW and DVP can be generated efficiently in less time. The third application of the set is recognition of object. The integration of DVW and DVP provides the best performance in recognition of objects because of their good discriminative power.

Romain Tavenard et.al [2] proposed K-means clustering approach to produce clusters for fast image search. But K-means approach results in producing imbalanced clusters with different cardinalities. This results in the variation of the response times. Partial balancing is an iterative method for the balancing process applied on the output generated by K-means approach. This method increases the distance between the centroids and the other points within the cluster and penalties are given to the distance based on the cluster population. Balancing process results in recall of about 0.33 and selectivity of about 0.0030.Balancing process results in producing clusters of equal size and reduces variation of response times.

Herve Jegou et.al [3] evaluated and compared various ways of converting the local descriptors of an image into vector representation such as BOW,VLAD representation and the Fisher Vector representation. Power normalization is used to normalize the FV and it increases the performance of FV. Principal Component Analysis(PCA) is a tool which is used for dimensionality reduction.PCA is applied to the SIFT descriptor to reduce its dimension from 128 to 64.By using PCA,FV outperforms BOW and VLAD.FV provides mean average precision of about 56.5%.Then the vector representation is converted into compact binary codes. The dimension of FV is reduced by using PCA. Encoding of vectors are performed by using Asymmetric Distance Computation (ADC) and indexing is performed by using IVFADC.ADC is combined with inverted file structure to produce IVFADC method. Memory size of 20 bytes is used to index an image. This method provides higher accuracy and results in mean average precision of about 6.6%.

Liang Zheng et.al [4] have implemented L_p-norm IDF technique to optimize the weight of the visual word. It is implemented by using L_p-norm pooling method with the traditional TF-IDF(Term Frequency-Inverse Document Frequency)method.TF-IDF is a weighting scheme that is used to measure the weight of the visual word. It has two disadvantages. It only makes rough estimation of frequency and contain the burstiness problem. Burstiness describe that repeated visual words are present and it affects the system performance. L_p-norm IDF reduces the weights of the visual word in bursts and solves the burstiness problem. Improvement in the performance is about 27.1% over the traditional approach. Since it is performed offline additional memory cost is not required.

Liang Zheng et.al [5] introduces a concept called the Visual Phraselet for image search. Visual phrases are grouped together to form visual phraselet in which it clearly specifies the spatial relationship of visual phrases. Visual phraselet eliminates the false matches of the visual phrases. Soft quantization approach is used to boost the identification of visual phraselets. It increases the discriminative power of visual word. The improvement in mean average precision is about 54.6% in case of one million dataset.

N.V.Murali Krishna Raja et.al [6] used K-means clustering approach to cluster the images based on color and to create index. Decorrelation stretching is used to separate color of an image. Euclidean distance metric is used to find the difference of two colors. Index is created only with the cluster value and hence it results in fast image search. Time complexity, space complexity, precision complexity and recall complexity can be greatly reduced.

Zhen Liu et.al [7] have implemented a Flexible SIFT Binarization (FSB) algorithm. Using this algorithm magnitude of each SIFT descriptor is converted into binary code. When image patches carry out some transformations, it results in the generation of noise. This noise affect the magnitude of SIFT descriptor. To avoid this problem, SIFT descriptor is first transformed by Principal Component Analysis before binarization process. Dimension reduction is also carried out by using PCA. Soft binarization method is used to reduce the binarization error. Cross indexing approach is used for indexing purpose in which code word expansion process and visual word expansion process are performed iteratively. Hamming distance is used for the similarity of binary SIFT codes. Memory size required is about 20 bytes. Mean Average Precision value is of 0.60.

Zhen Liu et.al [8] described contextual hashing method to perform binarization process of spatial context. Indexing with the information of spatial context leads to memory constraint and high retrieval time. To overcome this issue, the information of spatial context and multimode information are binarized and indexed. Multimode property describe that same point of an image contain different features. Multimode property is evaluated to improve the search accuracy. Geometric verification is performed by calculating the hamming distance between the binary codes and this method eliminates the false matches of image feature. This method provides mAP value of about 0.645 for holidays dataset and 0.443 for paris dataset. Each feature requires 12 bytes of memory.

Zhen Liu et.al [9] proposed K-means clustering approach to form feature group by grouping the local features of an image. It reduces time complexity in searching images. BVLAD (Balanced VLAD) representation is used to generate descriptor for the feature group. Search accuracy can be increased by using BVLAD. Product quantization approach is used to encode the group descriptor into binary representation of only 16 bytes. Hence memory size can be reduced. Dichotomizing search approach is used for similarity calculation between the query image and each of the database images. Search efficiency is improved by using this approach.

Wengang Zhou et.al [10] have implemented binary SIFT(BSIFT) concept. Scalar quantization approach is used to quantize a SIFT [16] descriptor into a bit vector, which is referred as BSIFT. Inverted file structure is used to create index and only the top 32 bits of BSIFT is used for indexing. For the reduction of quantization error, three approaches are used. First approach is to remove the unreliable SIFT [16] features which reduces the search accuracy by using their median values. This lessen the memory cost and improve precision. Second approach is using soft decision strategy in order to reduce quantization loss in generating code word since it eliminates some of the candidate features having some flipping bits. This approach is used to improve recall. Third approach is the generation of mask vector. It is used to protect some of the binary bits whose values can be changed during the quantization process. It is used to improve precision. BSIFT method results in higher search accuracy of about 0.59.Memory space required for each feature in indexing is about 32 bytes.

Table 1 Techniques used in reviewed papers

S.No	Author	Year	Proposed Method	Parameter used for Evaluation	Results
1.	**Shiliang Zhang et.al [1]**	**2011**	**Descriptive visual words and Descriptive visual phrases + VisualWordRank + Inverted file structure + TF-IDF weighting, DWPRank**	**Mean average precision**	**mAP is about 19.5%**
2.	**Romain Tavenard et.al[2]**	**2011**	**K-means clustering approach + Partial balancing**	**Recall and selectivity**	**Recall of about 0.33 and selectivity is 0.0030**
3.	**Herve Jegou et.al [3]**	**2012**	**Fisher vector representation + Power normalization + Principal Component Analysis, Asymmetric Distance Computatio-n(ADC) + IVFADC**	**Mean average precision, Memory size**	**mAP is 6.6% and memory size is about 20 bytes**
4.	**Liang Zheng et.al [4]**	**2013**	**L_p-norm IDF technique**	**Performance**	**Improvement in performance is about 27.1%**
5.	**Liang Zheng et.al [5]**	**2013**	**Visual Phraselet + Soft quantization approach**	**Mean average precision**	**Improvement in mAP is about 54.6%**
6.	**N.V.Murali Krishna Raja et.al [6]**	**2013**	**K-means clustering approach + Decorrelation stretching + Euclidean distance metric**	**Complexity**	**Reduced time, space, recall and precision complexity**
7.	**Zhen Liu et.al [7]**	**2014**	**Flexible Scale Invariant Feature Transform Binarization algorithm + Principal Component Analysis+ Soft binarization method + Cross indexing approach + Hamming distance**	**Mean average precision, Memory size**	**mAP is 0.60 and memory size is about 20 bytes**

8.	**Zhen Liu et.al [8]**	**2014**	**Contextual hashing method + Multimode property + Hamming distance**	**Mean average precision, Memory size**	**mAP value of about 0.645 for holidays dataset and 0.443 for paris dataset, memory size is 12 bytes**
9.	**Zhen Liu et.al [9]**	**2015**	**K-means clustering approach + BVLAD(Balanced VLAD) representation + Product quantization approach + Dichotomizing search approach**	**Memory size**	**Memory size is of 16 bytes**
10.	**Wengang Zhou et.al [10]**	**2015**	**Binary SIFT(BSIFT) + Scalar quantization approach + Inverted file structure**	**Accuracy and Memory size**	**Accuracy is about 0.59 and memory size is 32 bytes.**

III. CONCLUSION

In this paper, various techniques are reviewed for the efficient searching of the similar images from the large database. Various methods are analyzed for the binarization process and to reduce dimension. Clustering method is also reviewed which is used to improve speed of image retrieval. Different methods are evaluated to reduce memory required for indexing and to improve the accuracy in retrieving images.

REFERENCES

[1] S.Zhang, Q.Tian, G.Hua, Q.Huang and W.Gao, “Generating descriptive visual words and visual phrases for large-scale image applications,” IEEE Trans. Image Process., vol. 20,no. 9,pp. 2664-2677,Sep. 2011.
[2] R.Tavenard, H.Jegou and L.Amsaleg, “Balancing clusters to reduce response time variability in large scale image search,” in IEEE,CBMI’2011,pp. 19-24.
[3] H.Jegou, F.Perronnin, M.Douze and C.Schmid et al., “Aggregating local image descriptors into compact codes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34,no. 9,pp. 1704-1716,Sep. 2012.
[4] L.Zheng, S.Wang, Z.Liu and Q.Tian, “L_p-norm IDF for large scale image search,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1626-1633,Jun. 2013.
[5] L.Zheng and S.Wang, “Visual phraselet:Refining spatial constraints for large scale image search,” IEEE Signal Process. Lett., vol. 20,no. 4,pp. 391-394,Apr. 2013.
[6] N.V.Murali Krishna Raja and K.Shirin Bhanu, “Content bases image search and retrieval using indexing by Kmeans clustering technique,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 2,Issue 5,pp. 2181-2189,May 2013.
[7] Z.Liu, H.Li, L.Zhang, W.Zhou and Q.Tian, “Cross-Indexing of binary SIFT codes for large-scale image search,” IEEE Trans. Image Process., vol. 23,no. 5,pp. 2047-2057,May 2014.
[8] Z.Liu, H.Li, W.Zhou, R.Zhao and Q.Tian, “Contextual hashing for large-scale image search,” IEEE Trans. Image Process., vol. 23,no. 4,pp. 1606-1614,Apr. 2014.
[9] Z.Liu, H.Li, W.Zhou, R.Hong and Q.Tian, “Uniting Keypoints: Local visual information fusion for large-scale image search,” IEEE Trans. Multimedia, vol. 17,no. 4,pp. 538-548,April 2015.
[10] W.Zhou, H.Li, R.Hong, Y.Lu and Q.Tian, “BSIFT: Toward data-independent codebook for large scale image search,” IEEE Trans. Image Process., vol. 24,no. 3,pp. 967-979,Mar. 2015.
[11] Y.Zhang, Z.Jia and T.Chen, “Image retrieval with geometry-preserving visual phrases,” in Proc: IEEE Conf. Comput. Vis. Pattern Recog., pp. 809-816,Jun. 2011.
[12] L.Zhang and Y.Rui, “Image search from thousands to billions in 20 years,” ACM Trans. Multimedia Comput.,Commun.,Appl., vol. 9,no. 1,pp. 36:1-36:20,2013.
[13] R.Arandjelovic and A.Zisserman, “All about VLAD,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1578-1585,Jun. 2013.
[14] K.Mikolajczyk and C.Schmid, “An affine invariant interest point detector,” in Proc. Eur. Conf. Comput. Vis., pp. 128-142,2002.
[15] H.Bay, T.Tuytelaars and L.Van Gool, “Surf:Speeded up robust features,” in Proc. Eur. Conf. Comput. Vis., pp. 404-417,2006.
[16] W.Zhou, H.Li, Y.Lu and Q.Tian, “SIFT match verification by geometric coding for large-scale partial-duplicate web image search,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 9,no. 1,p. 4,2013.