# A Brief Review On Image Retrieval Techniques and its Scope

Shreya Bansod
PG Student, Dept of CSE,
DBACER,
Nagpur,India
*Shreyabansod05@gmail.com*

Asst. Prof. Dinesh.S. Gawande
Lecturer CSE Dept.
DBACER
Nagpur,India
*Gawande.dinesh@gmail.com*

Asst. Prof. Roshan Thakur
Lecturer CSE, Dept,
DBACER
Nagpur, India
*Roshanthakur11@gmail.com*

***Abstract***— This paper presents the novel approach for image retrieval. Image retrieval is an important problem in many applications, such as copyright infringement detection, tag annotation, commercial retrieval, and landmark identification. Image retrieval definition is given and the concept and significance of image retrieval is also provided. Various image retrieval techniques based on content based, sketch based, also based on image annotation is explained here. The last section includes the approach for retrieval is given as a problem formulation.

***Keywords***- *Image retrieval, problem formulation, content based.*

*****

## I. INTRODUCTION *(HEADING 1)*

We are interested in the problem of retrieving similar images in a large corpus of images. Similar images are defined as images that have the same scenes or same objects viewed under different photometric and geometric transformations [1], [2]. Image retrieval is an important problem in many applications, such as copyright infringement detection [2], tag annotation [3], commercial retrieval [4], and landmark identification [5].

**Figure 1: Sample Image retrieval results. Image on extreme left shows the input image and other six images are the image extraction results**

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotated words. Manual image annotation is time consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools. Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images similar to the query. The similarity used for search criteria could be meta tags, color distribution in images, region per shape attributes, etc.

Content based image retrieval (CBIR) is a technique used for retrieving similar images from an image database. The most challenging aspect of CBIR is to bridge the gap between the low-level feature layout and high-level semantic concepts. CBIR systems have used local colour and texture features [6–8] for increasing the retrieval accuracy, but nevertheless they have reached to the desirable goal. In the region-based image retrieval (RBIR) systems [9,10], the image is segmented into regions based on the colour and texture features. The regions are closer to the human perception and are used as the basic building blocks for feature computation and similarity measurement. RBIR systems have been proven to be more efficient than the CBIR systems in terms of retrieval performance. In [9], images are compared based on the individual region to region similarity. Precise image segmentation has still been an open area of research. For example, the integrated region matching (IRM) algorithm introduced in [9], proposes an image-to-image similarity based on combining all corresponding regions of two images. In [9], each region is assigned significance worth based on its size in the image. In [11], fuzzy features are used to capture the shape information. Shape signatures are computed from the blurred images and then the global invariant moments are used as the shape features. The retrieval performance of Hirremath and Pujari [11] has been shown to be better than several systems such as [9,12]. The studies mentioned in the above clearly indicate that, in CBIR, local features along with the shape information of objects play a significant role in determining the similarity of images. Precise segmentation is not only difficult to achieve but also is so critical in object shape determination.

Image retrieval must deal with the difference between the user's desire and the query example. This difference may be even more severe in sketch-formed queries, because of the ambiguousness in the query sketch caused by a lack of semantic information such as texture attributes [1] and luminance. A simple and similar image is needed for image-based retrieval. But for SBIR, results may vary dramatically if the user's drawing skills are not sophisticated, or if the target cannot be simply depicted using only lines. For example, if a user is looking for pictures of a pyramid but they can only draw a triangle, sketch-based retrieval becomes very challenging [2]. To address this problem, researchers proposed incorporating sketches and text descriptions to disambiguate the input. Lin et al. proposed a method that does not use lines to form the query sketch [4]. The sketch is a drawing that uses different words to represent diverse objects. Their locations and sizes are represented by the words. ith the help of these words, the approach first finds some corresponding exemplars, which is then used to search for objects in images. In this sense, it is like a concept-based image retrieval system instead of a sketch-based method.

## II. EXISTING APPROACHES

The research directions on this topic can be broadly categorized into four classes: a) those aiming to improve the visual vocabulary; b) those performing a query expansion; c) those optimizing efficiency and memory resources related to image representation; and d) those improving the matching process by taking into account geometric considerations. Regarding the first direction, several approaches in the literature have proposed the use of very large vocabularies (upto 16.7 M words). Using such large vocabularies improves the discrimination capabilities of the model by giving more importance to image details. The most common approach to build vocabularies is the well-known k-means clustering algorithm [2]; however, its results usually scale poorly with the size of the vocabulary. Consequently, more recent works have moved towards either hierarchical approaches such as trees), or approximate nearest neighbor techniques [3]. Furthermore, other authors have proposed a soft quantization approach for the representation of the local descriptors. In this manner, the actual distances between each descriptor and the closest words in the vocabulary are also taken into account. In [4], a soft quantization process was proposed that provided a notable increase in the system performance. It is also worth mentioning the approach suggested in[5], where a kernel-based density estimation was used to jointly address the quantization and the matching processes. With respect to the second direction, query expansion techniques, [6] and [4] used top-ranked images as new queries in order to perform several iterations of the matching process. A similar idea is explored in [7], where the authors use a visual query expansion method to enhance the ranking results of an initial text-based search. These methods achieved notable improvements in retrieval performance at the expense of an important increase in the computational time. The third direction involves obtaining compact image representations, either by reducing the number of detected features per image, or by using compact image representations, such as hashes [9], compressed representations of Fisher vectors [10], or binary vectors computed using a Hamming embedding of GIST descriptors [11]. Furthermore, we can also consider in this direction those approaches that use latent topic models to obtain compact higher-level representations, such as that presented in [12]. Finally, with respect to the use of geometric considerations, the most prevalent approach consists of a geometric-based post-processing step [3], [4]. However, other proposals have also been successful in taking geometric constraints into account. In [13], [14], the authors proposed a combined use of Hamming embedding and weak geometric consistency to improve the retrieval process. In [15], bundling features were proposed so that the large regions detected by the MSER detector [16] contained several local patches detected by the SIFT detector [17]. In doing so, these features combine the higher discrimination properties of larger regions with the repeatability and robustness to occlusions of local patches. In [18], geometry-pre-serving visual phrases were proposed that capture short- and long-range spatial relations between visual words.

Moreover, the inclusion of geometric considerations in the matching process generates so me spatial information that can help to detect the Region of Interest (ROI). In [3], for example, only those matches obeying a specific transformation were considered as true matches. This true/false match classification provided a segmentation mask that allowed for identifying the ROI of the query image. The model in [19] efficiently estimated an affine transformation between every two images by discretizing the transformation space, decomposing it into rotation, scaling and translation. This model was then utilized to generate spatial voting maps in the query, allowing for bounding-box based localization of the object of interest. Furthermore, in [20] a latent topic model named Geometric Latent Dirichlet Allocation was introduced that takes into consideration some geometric constraints. In particular, the topic model was used to unsupervisely model images as mixtures of topics hopefully associated with objects in the scene. Then, they evaluated the performance of this approach for image retrieval applications and, finally, demonstrated that is possible to automatically discover and segment regions-of-interest.

Most existing methods mainly use global features or divide images into blocks to represent the image [12]–[15]. These

methods do not work well because of the ambiguousness of sketches and shapes. Additionally, the incompleteness of a

user's drawing may also affect the results. Consequently, researchers proposed exploring the local saliency in SBIR. Chen et al. used a freehand sketch and some text labels to search for Internet images [16]. Although this method was very accurate, it was very computationally expensive. Thus, a SBIR with index structures is more appropriate for a large-scale image set, and achieves the best balance between the retrieval performance, and time and storage costs. The edged index approach is a shape-based indexing
method [2]. It solves the shape-to-image matching problem using pixel level matching. Oriented chamfer matching [17] is used to compute the distance between contours.

Object-based image retrieval (OBIR) is a typical kind of content-based image retrieval (CBIR), through which the user intends to find object images. According to the form of queries, it can be divided into inter-active OBIR and automatic OBIR. For interactive OBIR, the user is supposed to offer the object region manually, usually in the form of bounding-box. For example, image representations in [14] are based on image segmentations and the objects are modeled based on the regions using Latent Semantic Analysis (LSA). In [17], the authors represent the images as a bag of visual words, and employ language modeling to derive the ranking function. The authors also argue that the object context is as important as the object itself, thus visual words locating outside the object region should also be taken into consideration. For automatic OBIR, the query object region is supposed to be detected automatically by the search engine. Recently, a number of works utilize multi-instance learning (MIL) to model the retrieval process. For instance, in [19] and [17], the users are required to provide multiple query images at a time and model the queries as a multi-instance classifier. To achieve this, the authors of [19] propose a new kind of multi-instance Support Vector Machine (SVM) with a convex training method, while in [17] the authors bring in a novel feature representation scheme
using identified evidence ROI.

## III. PROBLEM FORMULATION

Given a query image 'I' and a set of reference images 'R', the objective of an image retrieval system is to compute a similarity measure between 'I' and each one of the reference images in order to generate a similarity ranking. The computation of this similarity measure involves several steps that are briefly reviewed next. Salient points (keypoints) are detected for every image. Subsequently, a descriptor is obtained for each keypoint. Each descriptor depicts the appearance of a local region around the corresponding keypoint. Then, a keypoint-level matching process is performed for each pair of images (the query and each one of the reference images). As a result, a set of N potential matches are generated between the query image and each reference image. This step usually relies on several thresholds on the (visual) distance between descriptors, so that non-likely matches are filtered out. Finally, these low-level matching results serve as the basis for computing the similarity measure. Usually this matching process is prone to false negative and false positives. In this context, we propose to use a generative model of the query image that allows us to incorporate some assumptions that make the matching model more robust.

The complete approach is divided into following steps.

a. **Image Enhancement :** In order to remove noise, each image is passed through a low-pass Gaussian filter. This stage is commonly used in most retrieval systems as the pre-processing step.

b. **Image Regioning (Segmentaion):** Each image is partitioned into non-overlapping tiles. In other words image is segmented based on region of interest.

c. **Feature set generation:** The feature set consists of texture and shape descriptors. Sometimes colour features can also be extracted.

d. **Image Matching:** Here a tile from the query image is allowed to be matched to any tile in the target image, but any tile of the query image will participate in the matching process only once. Although in the conventional methods, each tile of the query image is compared with all tiles of the target image.

**e. Image Retrieval**

The proposed model takes into account several properties of the matching process between two objects in different images, namely: objects undergoing a geometric transformation, typical spatial location of the region of interest, and visual similarity. The model assumes that has been generated as the mixture of components. The first part of the model defines the "apriori" probability of each component or, in other words, the mixture weights. The second part describes the location of each keypoint of the query image by means of an Affine transforma-tion that aims to capture the geometric transformation that each object undergoes to fit the same object in the query image (transformation-based location). The third part provides additional insight into the object location, but now according to the expected location of the object in the image (Spatial consistency-based location). Finally, the forth part considers the visual similarity itself by taking into account how likely the computed distance is, given each one of the potential objects (visual similarity).

f. **Image Enhancement :** In order to remove noise, each image is passed through a low-pass Gaussian filter. This stage is commonly used in most retrieval systems as the pre-processing step.

g. **Image Regioning (Segmentaion):** Each image is partitioned into non-overlapping tiles. In other words image is segmented based on region of interest.

ISSN: 2321-8169
455 – 459

h. **Feature set generation:** The feature set consists of texture and shape descriptors. Sometimes colour features can also be extracted.

i. **Image Matching:** Here a tile from the query image is allowed to be matched to any tile in the target image, but any tile of the query image will participate in the matching process only once. Although in the conventional methods, each tile of the query image is compared with all tiles of the target image.

**j. Image Retrieval**

The proposed model takes into account several properties of the matching process between two objects in different images, namely: objects undergoing a geometric transformation, typical spatial location of the region of interest, and visual similarity. The model assumes that has been generated as the mixture of components. The first part of the model defines the "apriori" probability of each component or, in other words, the mixture weights. The second part describes the location of each keypoint of the query image by means of an Affine transforma-tion that aims to capture the geometric transformation that each object undergoes to fit the same object in the query image (transformation-based location). The third part provides additional insight into the object location, but now according to the expected location of the object in the image (Spatial consistency-based location). Finally, the forth part considers the visual similarity itself by taking into account how likely the computed distance is, given each one of the potential objects (visual similarity).

## IV. CONCLUSION

This paper presents a novel approach for image retrieval system. Image retrieval is very important task in various applications. This paper describes the various image retrieval task, based on different techniques. Its starts its discussion with meaning of image retrieval definition followed by various approaches for image retrieval. In third section the problem formulation is given in detail for the system.

## REFERENCES

[1] H. Jegou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search," Int. J. Comput. Vis., vol. 87, no. 3, pp. 316–336, 2010.

[2] D.-Q.ZhangandS.-F.Chang, "Detecting image near-duplicate by stochastic attributed relational graph matching with learning," in Proc.ACM Conf. Multimedia.

[3] W.-L. Zhao, X. Wu, and C.-W. Ngo, "On the annotation of web videos by efficient near-duplicate search," IEEE. Trans. Multimedia , vol. 12, no. 5, pp. 448–461, Aug. 2010.

[4] W. Tang, R. Cai, Z. Li, and L. Zhan g, "Contextual synonym dictionary for visual object retrieval," in Proc. ACM Conf. Multimedia ,2011.

[5] D.Chen,G.Baatz,andK.Koeser et al. , "City-scale landmark identification on mobile devices," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2011.

[6] K. Barnard and D. Forsyth, "Learning thesemantics of words and pictures," in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 2. IEEE,2001, pp. 408-415.

[7] D. Joshi, J. Z.Wang, and J. Li, "The story picturing engine|a system for automatic text illustration," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), vol. 2, no. 1, pp. 68{89, 2006.

[8] J. Li and J. Z. Wang, \Real-time computerized annotation of pictures," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 6, pp. 985- 1002, 2008.

[9] A. Ghoshal, P. Ircing, and S. Khudanpur, "Hidden markov models for automatic annotation and content-based retrieval of images and video," in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005, pp. 544-551.

[10] F. Yu and H. H.-S. Ip, "Automatic semantic annotation of images using spatial hidden markov model," in Multimedia and Expo, 2006 IEEE International Conference on. IEEE, 2006, pp. 305-308.

[11] S. Wan and R. A. Angryk, "Measuring semantic similarity using wordnet-based context vectors," in Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. IEEE, 2007, pp. 908-913.

[12] M. W. Berry, S. T. Dumais, and G. W. O'Brien, \Using linear algebra for intelligent information retrieval," SIAM review, vol. 37, no. 4, pp. 573-595, 1995.

[13] H. Jégou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in Proc. Eur. Conf. Computer Vision , 2008.

[14] H. Jegou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search," Int. J. Comput. Vision vol. 87, pp. 316–336, 2010. [Online]. Available: http://dx.doi.org/10.1007/s11263-009- 0285-2.

[15] Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling features for large scale partial-duplicate web image search," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009 (CVPR 2009) , Jun. 2009, pp. 25–32. [Online]. Available: http://dx.doi.org/10.1109/CVPRW.2009. 5206566.

[16] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. British Machine Vision Conf. , London, U.K., 2002, vol. 1, pp. 384–393. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/sum-mary?doi=10.1.1.7.2484.

[17] D. G. Lowe, "Distinctive image features from scale-invariant key-points," Int. J. Comput. Vision , vol. 60, pp. 91–110, 2004.

[18] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry-pre-serving visual phrases," in Proc. IEEE Int.

Conf. Computer Vision and Pattern Recognition (CVPR) , 2011, pp. 809–816.

[19] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu, "Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking," inProc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) , 2012, pp. 3013–3020.

[20] J. Philbin, J. Sivic, and A. Zisserman, "Geometric latent dirichlet allocation on a matching graph for large-scale image datasets," Int. J. Comput. Vision vol. 95, no. 2, pp. 138–153, Nov. 2011. [Online]. Avail-able: http://dx.doi.org/10.1007/s11263-010-0363-5.

[21] Shiv Ram Dubey, Student Member, IEEE, Satish Kumar Singh, Senior Member, IEEE,andRajat Kumar Singh, Member, IEEE, "Local Diagonal Extrema Pattern: A New and Efficient Feature Descriptor for CT Image Retrieval", 2015.