International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 2 | Issue 2 | ISSN : 2456-3307

An Efficient Image Retrieval System Using Surf Feature Extraction and Visual Word Grouping Technique

S. Bhuvana, F. Ragini Felicia Suruti, R. Shariene Fathima, P. Vincy Roshalin, M. Radhey
Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu , India

ABSTRACT

Content Based Image Retrieval is widely used to find the location of images for many application scenarios. It is used to tag the images using large geo-tagged image set. In recent years, images are tagged based on their locations and geo-tagged images consumes more memory space. Therefore, the performance of the image location estimation can be improved by using visual word groups. The mean shift clustering algorithm and a position descriptor have been used to generate visual word groups. A fast indexing structure is build using document builder interface. Thus the drawbacks in the existing system have been over-come in the proposed system. The proposed system involves Speeded-Up Robust Transform (SURF) which is a feature detection and descriptor method which is used for object recognition. The modules include Feature extraction which is used to identify the interest points within the image, Indexing which builds an inverted file structure to reduce the image size, Image retrieval which compares the input image with the query image to give the resultant output and Re-Ranking which categorizes the top ranked images.

Keywords : Image Retrieval, Bag-Of-Visual Words, Spatial Constraint, Salient Area Detection, And Mean-Shift Clustering.

I. INTRODUCTION

1.1 Background

Image processing is the technique to analysis and manipulate digitized images in order to boost image quality. It is a methodology to convert a picture into digital type and perform some operations thereon, so as to urge increased image quality or to extract some helpful data from it. It's a kind of signal dispensation within which input is image, like video form or picture and output is also image or characteristics related to that image. Image processing mainly involves digital image processing, optical and analogue image processing.

1.2 Content Based Image Retrieval:

Content Based Image Retrieval (CBIR) is the method of retrieving images from a large dataset. The content may refer colors, shapes, textures that can be derived from the image itself. It is used to find a particular image from a collection of images which is shared by many professional groups. With massive image databases turning into a reality in both scientific and medical domains and in the advertising/marketing domain ways for organizing information of pictures for effective image retrieval have become important. CBIR plays a vital role for this purpose. CBIR mainly determines the locations of an input image using clustering techniques. It improves the image location estimation process.

1.3 SURF Feature Extraction:

Speeded up robust features (SURF) is a local feature detector and descriptor method. It can be used for tasks such as object survey or beholding, categorization or 3D reconstruction. It is followed by the scale invariant feature transform (SIFT) descriptor. The performance and speed measures of SURF is several times quicker than that of SIFT. It is also more robust than SIFT. SURF descriptors are widely intended to locate and identify objects, human faces and to trail objects. Feature detection is the procedure to inspect a picture to extricate features that are similar to the objects which are compared in the image, in such a manner that we are able to find an object, based on its characteristics in various images. This detection method should preferably be applicable when the image shows the object with various transformations, mainly scale and rotation, or when certain parts of the object are occluded. The main steps involved in SURF feature extraction involves:

Detection: Automatically identify interesting features such as colour, texture, interest points this must be done robustly. The same feature should always be detected regardless of viewpoint.

Description: Every interest point must have a specific description that does not rely on the features such as scale and rotation.

Matching: Given an input image, determine which objects it contains and matches, and possibly a transformation of the object, based on predetermined interest points. This paper involves various sections: Section I described the introduction of the paper, section II depicts the literature review, section III gives a brief detail about the proposed system, section IV explains the experimental results, section V gives the conclusion and section VI Describes the future work.

II. METHODS AND MATERIAL

Related works:

In this section, the existing solutions for the image retrieval technique and several state of image retrieval methods are reviewed. The related work includes several methods which are briefly explained below:

2.1 Bag of visual words method: In a recent paper, S. Zhang, Q.Tian ,G. Hua, Q. Huang, and S. Li [1] proposed a descriptive visual words and visual phrases for various image applications. The Bag-of-visual Words (BoW) image representation has been implemented for various scenarios and problems in the fields of multimedia and computer vision. The basic idea is to represent images as visual documents composed of repeatable and distinguishing visual elements, which are comparable to the words in texts. In this paper, Descriptive Visual Words (DVWs) and Descriptive Visual Phrases (DVPs) are suggested as the visual correspondences to text words and phrases, where visual phrases allude to the frequently co-occurring visual word pairs. Based on this idea, a general framework is proposed for creating DVWs and DVPs from classic visual words for several applications.

2.2 Location estimation method: J.Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei [4] proposed a GPS estimation method for places of interest to determine the location of images. In recent years people tend to share their photos, taken while they are traveling, to the social media sharing websites, such as Facebook. With the help of these images taken in places of interest in conjunction with the broad multimedia information realm, the task of automatic image GPS location estimation became possible. Typically, images taken from similar locations share some unique features, even when the images are taken from several viewpoints. These salient features play an important role in the image location estimation problem. Experiments on different datasets provide the effectiveness of the proposed method. Saliency of each BoW in determining the image will be measured. In this paper a method has been proposed to avoid duplication of images. Spatial constraint method is both efficient and effective. It also discovers false matches of certain features among images and improves image retrieval performance.

2.3 Object detection method: In this paper G. Cheng et al.,[3] proposed an effective representation and classification of scenes using very high resolution (VHR) for remote sensing images which cover a wide range of applications. Although low-level image features have been proven to be efficient for scene classification, they are not semantically useful and thus have difficulty to deal with challenging visual recognition tasks. In this study, the authors proposed a new and effective auto-encoder-based method to make use of a shared mid-level visual dictionary. This dictionary serves as a shared and universal basis to find mid-level visual elements. On the one hand, the mid-level visual dictionary learnt using machine learning technique is more discriminative and contains rich semantic information, compared with the traditional low-level visual words.

2.4 Image search using spatial coding method: In this paper W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian[8] have proposed the spatial constraints method which is used to find the images based on its latitudinal and longitudinal information. The geometric constraints among visual words in a picture is usually ignored or exploited for full geometric verification, which is computationally expensive. Estimation of the geometric constraints is a complicated task.

2.5 Landmark Retrieval Method: In this paper, E. Gavves, C. G. M. Snoek, and A. W. M. Smeulders[7] have addressed the incoherence problem of the visual words in bag-of-words vocabularies. Different from existing work, which assigns words based on closeness in descriptor space, focus is made to find pairs of independent, distant visual words – the visual synonyms – that are likely to host image patches of same visual reality. Image geometry is used to find those image features that lie in the nearly distinct physical location, yet are assigned to various other words of the visual vocabulary. Defined in this way, evaluation is done based on the validity of visual synonyms. Examining the closeness of synonyms in the L2 - normalized feature space is also done. Visual synonyms is widely used for vocabulary reduction.

2.6 Motivation: From the literature review it was observed that most of the image retrieval technique utilizes SIFT method where the computational cost is extremely high for model parameters training. Global Positioning System (GPS) estimation approach is also very computationally intensive when the scale of dataset is very large. To overcome these issues, the proposed system utilizes SURF method. The computational cost of SURF extraction is very low when compared to other extraction methods.

III. RESULTS AND DISCUSSION

Proposed Scheme:

The following block diagram depicts the overall image retrieval system. The various modules of the image retrieval system is mentioned in the below diagram.

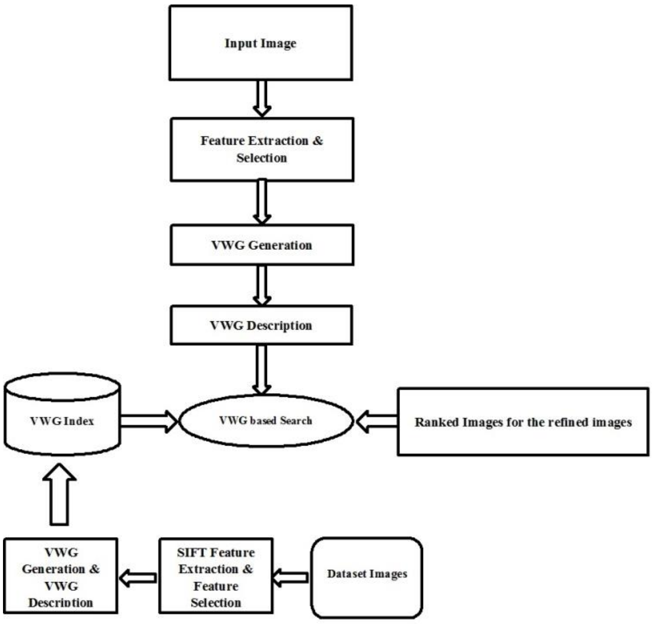

Fig 1. Block diagram of overall retrieval system

3.1 Feature extraction and selection:

Feature extraction involves a set of data within which the interest points are selected from the image for object comparison. Latent semantic analysis is the main process done in feature extraction and selection. Latent semantic analysis is the process which groups words that are similar in meaning various regions in each image is matched according to the similarity.

3.2 VWG description and generation:

Visual word group description and generation involves several steps:

Set up image category sets: Organize and separate the images into training and test subsets. Use the image data store function to store the images to use for training an image classifier. Grouping images into categories makes handling large sets of images much easier. Split Each Label function is used to split the images into training and test data.

Figure 2. VWG Generator

Create bag of features: Develop a visual vocabulary, by extricating feature descriptors from a given set of images of each category. The bag of Features object represents the features, by using the k-means clustering algorithm on the feature descriptors acquired from training sets. The algorithm iteratively groups the descriptors into k mutually exclusive clusters. Extricate features based on a feature detector, or define a grid to extract feature descriptors. Therefore, use the grid for images that do not have similar features, such as an image containing scenery. Using speeded up robust features (or SURF) detector higher scale invariance can be obtained.

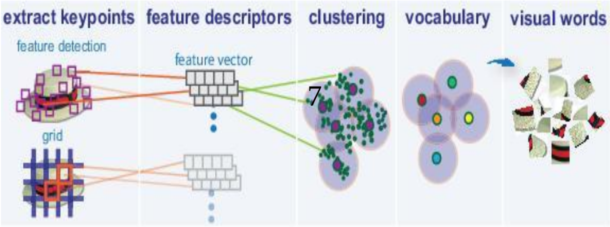

Figure 3. Train an Image Classifier with BoVW

The trainImageCategoryClassifier function gives an image classifier. The function trains a multiclass classifier by using the error correcting output codes (ECOC) framework. Along with binary support vector machine (SVM) classifiers. The trainImageCategoryClassifier function utilizes the bag of visual words returned by the bagofFeatures object to convert images in the image set into the histogram of visual words. The histogram of visual words are used as the positive and negative samples which is used to train the classifier. Use the bagofFeatures conversion method to convert each image from the training set. This function identifies and extracts features from the image and then makes use of the approximate nearest neighbor algorithm to build a feature histogram for each image. The function then increases histogram bins based on the proximity of the descriptor.

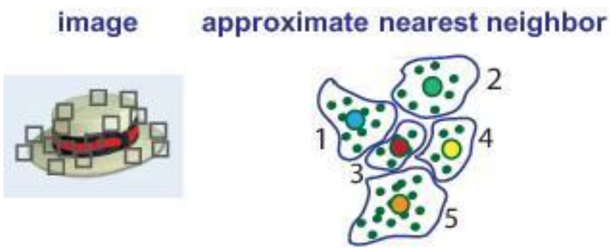

Figure 4. Encoding method

Repeat this for each image in the training set to develop the training data.

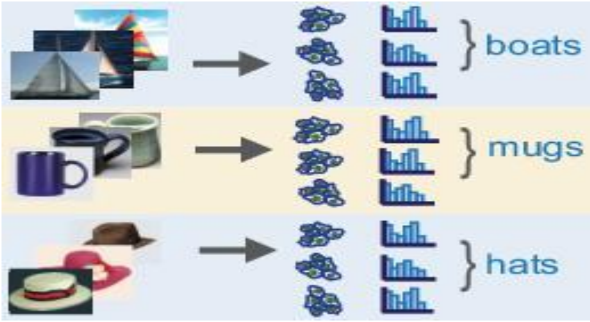

Figure 5. Classification of training data

Assess the quality of the classifier. The output confusion matrix depicts the analysis of the prediction. A perfect classification results in a normalized matrix which contains 1s on the diagonal. An incorrect classification will result fractional values.

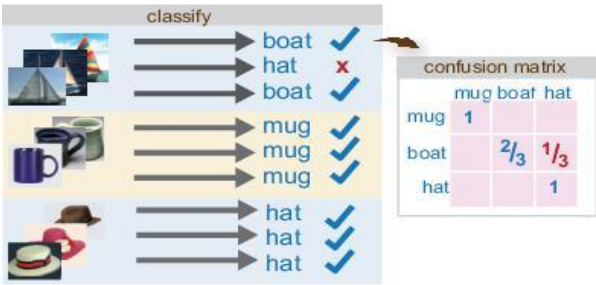

Figure 6. Construction of confusion matrix

3.3 Grouping images into clusters:

Images are clustered based on certain features. Through this feature clustering the whole dataset can be divided into several groups. K-means clustering is utilized to divide the dataset into M small clusters, denoted as Cn (n = 1, . . . , M).

Mean Shift Clustering:

1. Mean shift is a technique for detecting the maxima of a density function given discrete data sampled from that given function. It is useful for finding the modes of this density.
2. Consider a given set of points in a two-dimensional space.
3. Consider a circular window centered at C, having radius r as the kernel.
4. Mean shift is a hill climbing algorithm which mainly involves shifting this kernel iteratively to a higher density region till it achieves the convergence.
5. Every shift is given by a mean shift vector. The mean shift vector always points toward the direction of the maximum density.

6. At every iteration the kernel is shifted to the centroid. The method of calculating this mean mainly depends on the choice of the kernel.
7. At the point of convergence, no more direction will be found at which a shift can occupy more points inside the kernel.

3.4 Image Indexing

Image indexing builds inverted file structure for all the images to reduce the image size. The relative position descriptor of visual word is recorded. Each image contains different visual word groups. Sift points and visual word groups are detected and calculated. The position descriptor vectors are generated. Finally, the similarity between the input image and query image has been calculated.

3.5 Re-ranking of images

The images which are compared with the query image is then grouped based on its similarity. The images which are more similar to the query image is presented at the top while the other images are ranked according to the similarity. Thus the image retrieval becomes easier with the help of re-ranking technique.

4. Experimental results:

4.1 Image dataset: Image dataset involves several images contained in a data matrix. Variety of data sets are available for retrieval of images such as CALTECH dataset, CORAL dataset etc. A collection of 100 images of several category from WANG dataset has been made.

4.2 Performance analysis:

Images from the WANG dataset have been used. This dataset comprises of a large number of images ranging from animals to outdoor sports to natural images. Some researchers are of the opinion that the Corel database meets all the requirements to evaluate an image retrieval system, due to its large size. For our experiment, a collection 100 of several categories of images from WANG database has been made. These images are collected from several domains such as buses, dinosaurs, elephants, flowers etc. Each category has images with a resolution of either 256 384 or 384 256. For each query, the system collects database images with the shortest image matching distance computed for easy retrieval. If the retrieved image belongs to same category as that of the query image, then we can say that the system has appropriately identified the expected image, or else, the system has failed to find the expected image after comparison. The test image is matched with matched database to identify high frequency regions and matching. Precision is the fraction of retrieved documents that are relevant to the search. Precision takes all retrieved documents, but it can also be evaluated at a cut-off rank, considering only the top-ranked results returned by the system.

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{\{retrieved\ documents\}}$$

Eq.1

Recall in information retrieval is the fraction of the documents that are similar to the query image which are successfully retrieved.

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{\{relevant\ documents\}}$$

Eq. 2

The performance measure of the algorithms proposed is done with the calculation of the precision-recall.

4.3 Comparison Graph:

The below image represents the query image comparison between different image sets after calculating Image Retrieval Precision (IRP) and re-call category.

Table : Different type Query Image Category comparison

Image category	Scalable (HRPP)		Edge Histogram(HRPP)		SIFT	
Parameter	IRP	Recall	IRP	Recall	IRP	Recall
Butterfly	33	30	45	40	56	50
Sunrise	22	20	67	60	56	50
Rose	67	60	45	40	45	40
Car	45	40	67	60	33	30
Building	78	70	67	60	56	50
Flag	11	10	67	60	78	70
Tree	56	50	67	60	56	50
Average IRP	46	40	61	54	54	49

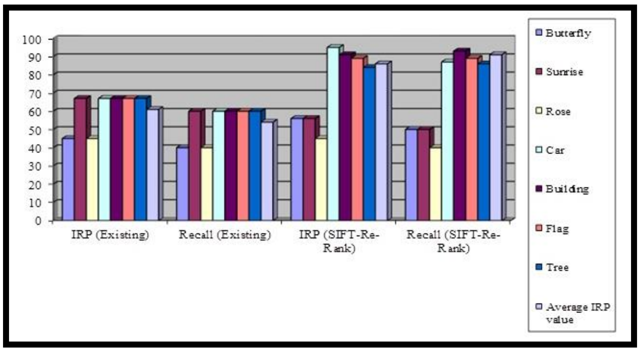

Figure 7. Compare different image with IRP and Recall

IV. CONCLUSION

The process of dealing with the manipulations of any image database has become extremely tedious due to its intense growth. Thus our approach dealt in this paper involves an efficient CBIR technique that focuses on low-level image features such as the color, texture and shape. The proposed system aims on an automatic image search and retrieval. This system incorporates a new indexing technique along with few other feature extraction techniques that extracts various multi-resolution SIFT features with re-ranking from the images in the database. The system also contains a very simple but very effective retrieval method and all these put together have reduced the number iterations that were there in the existing systems. Experiments were stimulated using WANG database and were proved that this system has minimized the effects of user's inaccurate relevance feedback and appropriate images were retrieved very efficiently with less iteration.

V. FUTURE WORK

The future works appear from the limitations and the difficulties when we develop our system. The following developments can be made in the future: It hope to build more generalized CBIR system which increase the system searching ability and provide more accurate results. To improve the retrieval results, the system has to take the feedback from the user. The user checks the results and comments on them by some way. Then, the system recalculates the results with the advantage of the feedback. To further improve the performance of the retrieval system, we hope to optimize some settings of the system architecture and techniques that used in this research. There exist some details setting that can be discussed and optimized with the images retrieval issues. This is like, the number of clusters k as an inputs for the K-means algorithm and input parameters of the genetic algorithm etc.

VI. REFERENCES

[1] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, “Descriptive visual words and visual phrases for image applications,” in Proc. 17th ACM MM, 2009, pp. 75–84.

[2] Y. Li, D. J. Crandall, and D. P. Huttenlocher, “Landmark classification in large-scale image collections,” in Proc. IEEE 12th ICCV, Sep. /Oct. 2009, pp. 1957–1964.

[3] G. Cheng et al., “Object detection in remote sensing imagery using a discriminatively trained mixture model,” ISPRS J. Photogram. Remote Sens., vol. 85, pp. 32–43, Nov. 2013.

[4] J. Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei, “GPS estimation for places of interest from social users’ uploaded photos,” IEEE Trans. Multimedia, vol. 15, no. 8, pp. 2058– 2071, Dec. 2013.

[5] C. Hauff and G.-J. Houben, “Placing images on the world map: A microblog-based enrichment approach,” in Proc. 35th ACM SIGIR, 2012, pp. 691–700.

[6] H. Liu, T. Mei, J. Luo, H. Li, and S. Li, “Finding perfect rendezvous on the go Accurate mobile visual localization and its applications to routing,” in Proc. 20th ACM Int. Conf. Multimedia, 2012, pp. 9–18.

[7] E. Gavves, C. G. M. Snoek, and A. W. M. Smeulders, “Visual synonyms for landmark image retrieval,” Comput. Vis. Image Understand., vol. 116, no. 2, pp. 238– 249, 2012.

[8] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, “Spatial coding for large scale partial-duplicate Web image search,” in Proc. ACM Int. Conf. Multimedia, 2010, pp. 511–520.

VII. Author Profile

Dr. S. Bhuvana Shanmugam is currently working as Assistant Professor in the department of Computer Science and Engineering, Sri Krishna College of Technology, Her research work in Image Retrieval. Her research interests includes Image Processing and Machine Learning.

Ms. F. Ragini Felicia Suruti is currently working as Assistant Professor in the department of Computer Science and Engineering, Sri Krishna College of Technology. Her area of interests includes Image Processing and Database management system.

Ms. P. Vincy Roshalin is currently pursuing Bachelor's degree in Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. Her area of interest includes Database Management System and Java programming.

Ms. M. Radhey is currently pursuing Bachelor's degree in Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. Her area of interest includes Database Management System and Java programming.

Ms. R. Shariene Fathima is currently pursuing Bachelor's degree in Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. Her area of interest includes Database Management System and Java programming.