Design and Implementation of an Efficient Image Retrieval System using Spatial Feature Reduction

Imran Naykude
Electronics and telecommunication
Shree L.R.Tiwari College of Engineering,Mira Road,Mumbai
[1] *immi.naykude@gmail.com*

Prof. V.B. Gaikwad
Computer Science
Terena College Of Engineering
vb_2k@rediffmail.com

Dr Ravish R. Singh
Electronics and telecommunication
Shree L.R.Tiwari College of Engineering,Mira Road,Mumbai

***Abstract*—Image segmentation is a very emerging and important area due to a large number of real life applications. Bag-of-features (BoFs) model is the one of the most successful algorithm used for the image classification. BoF methods are based on order-less collections of quantized local image descriptors; they discard spatial information and are therefore conceptually and computationally simpler than many alternative methods. Bag-of-features algorithm is used for the feature extraction from the image. BoFs model has several advantages like, scalability, simplicity, and generality. But BoFs have some disadvantages to and some of them are Time and Accuracy of the image classification process. In this paper Bag-of-features model is extend by using spatial pooling to improve the time and accuracy of the image classification model. In proposed method first the system is train by creating the database of the image feature for the evaluation process and then the evolution of the features is done for the input image by using clustering algorithm. In the proposed method KKN algorithm is used to find out the similar features from the database. Spatial pooling is then used to improve the performance of the existing system.**

Keywords— Image Segmentation, Image Classification, spatial pooling.

I. INTRODUCTION

The problem of image classification has drawn considerable attention in the Computer Vision community. The concentrated effort of the research community in the last few years resulted in many novel approaches for image classification, which progressed the field quickly in a few years. In general the image classification can be done by two ways. One is the Visual interpretation and other one is computer automatic classification. The computer automatic classification uses the pattern recognized technology and artificial intelligence technology for the classification of image. According to the land target's characteristics in the remote sensing image and the goal land targets interpretation experience and image formation rules in the library based on the computer system. Minimum distance from means, maximum-likelihood, cluster analysis and Bayesian classification are the simple and conventional classification method [1], [2] based on statical principals. Some of the other methods are also developed for image sensing classification, including machine learning, support vector machine, neural network, fuzzy set and genetic algorithm [3].

Having a good classification process opens the possibility of filtering out images from irrelevant classes and therefore will enhance the performance of image retrieval. The machine learning framework can be very useful in this situation. It allows for learning a semantic category associated with images based on low-level visual features of the images. Recent machine learning techniques have demonstrated their capability of identifying image categories from image features, like [4], [5], to mention a few.

The past decade has seen the rise of the Bag of Features approach in computer vision. Bag of Features (BoF) methods have been applied to image classification, object detection, image retrieval, and even visual localization for robots. BoF approaches are characterized by the use of an order less collection of image features. Lacking any structure or spatial information, it is perhaps surprising that this choice of image representation would be powerful enough to match or exceed state-of-the-art performance in many of the applications to which it has been applied. Due to its simplicity and performance, the Bag of Features approach has become well-established in the field. An early version was proposed in [1], and many improved versions are introduced since then [2] [3] [6].

In this paper the Bag-of-features model is extend by using spatial pooling to improve the time and accuracy of the image classification model. Proposed method is divided into two parts database creation and evaluation. KNN algorithm is used to evaluate the features from the database and the input image.

II. BAG OF FEATURES

Image classification and image search are topics of high interest because of the rapid growth of digital image and video collections. Classification and search algorithms heavily depend on image information extraction (which should focus on valuable information) and image representation (which should lead to efficient algorithms).

A Bag of Features method is one that represents images as order-less collections of local features. The name comes from the Bag of Words representation used in textual information retrieval. There are two common perspectives for explaining the BoF image representation. The first is by analogy to the Bag of Words representation. With Bag of Words, one represents a document as a normalized histogram of word counts. Commonly, one counts all the words from a dictionary that appear in the document. This dictionary may exclude certain non-informative words such as articles (like "the"), and it may have a single term to represent a set of synonyms. The term vector that represents the document is a sparse vector where each element is a term in the dictionary and the value of that element is the number of times the term appears in the document divided by the total number of dictionary words in the document (and thus, it is also a normalized histogram over the terms). The term vector is the Bag of Words document representation – called a "bag" because all ordering of the words in the document have been lost.

The Bag of Features image representation is analogous. A visual vocabulary is constructed to represent the dictionary by clustering features extracted from a set of training images. The image features represent local areas of the image, just as words are local features of a document. Clustering is required so that a discrete vocabulary can be generated from millions (or billions) of local features sampled from the training data. Each feature cluster is a visual word. Given a novel image, features are detected and assigned to their nearest matching terms (cluster centers) from the visual vocabulary. The term vector is then simply the normalized histogram of the quantized features detected in the image. The second way to explain the BoF image representation is from a codebook perspective. Features are extracted from training images and vector quantized to develop a visual codebook. A novel image's features are assigned the nearest code in the codebook. The image is reduced to the set of codes it contains, represented as a histogram. The normalized histogram of codes is exactly the same as the normalized histogram of visual words, yet is motivated from a different point of view.

There is a number of design choices involved at each step in the BoF representation. One key decision involves the choice of feature detection. Many use an interest point operator, such as the Harris-Affine detector [7] or the Maximally Stable External Regions (MSER) detector [8]. At every interest point, often a few thousand per image, a high-dimensional feature vector is used to describe the local image patch. Lowe's 128-dimension SIFT descriptor is a popular choice [9].

III. RELETED WORK

Image classification and object recognition are well studied areas with approaches ranging from simple patch based voting to the alignment of detailed geometric models. Here, in keeping with our approach to recognition, we provide only a representative random sample of recent work on local feature based methods. We classify these into two groups, depending on whether or not they use geometric object models.

The geometric approaches represent objects as sets of parts whose positions are constrained by the model. Inter-part relationships can be modeled pair wise [10], in terms of flexible constellations or hierarchies [11] [12], by co-occurrence [13] or as rigid geometric models [14] [15]. Such global models are potentially very powerful but they tend to be computationally complex and sensitive to missed part detections. Recently, "geometry free" bag-of-features models based purely on characterizing the statistics of local patch appearances have received a lot of attention owing to their simplicity, robustness, and good practical performance. They evolved when text on based texture analysis models began to be applied to object recognition. The name is by analogy with the bag-of-words representations used in document analysis (e.g. [16]): image patches are the visual equivalents of individual "words" and the image is treated as an unstructured set ("bag") of these.

Leung at al. [17] sample the image densely, on each patch evaluating a bank of Gabor-like filters and coding the output using a vector quantization codebook. Local histograms of such 'text on' codes are used to recognize textures. Text-on are also used in content based image retrieval, e.g. [18]. Lazebnik et al**.** [19] take a sparser bag-of-features approach, using SIFT descriptors over Harris-affine key points [20] and avoiding global quantization by comparing histograms using Earth Movers Distance [21]. Csurka et al [22] approach object classification using k-means-quantized SIFT descriptors over Harris-affine keypoints [20]. Winn et al. [23] optimize k-means codebooks by choosing bins that can be merged. Fergus *et al.* [24] show that geometry-free bag-of-features approaches still allow objects to be localized in images.

IV. KNN CLASSIFIER

The k Nearest Neighbour (KNN) is one of the most commonly used methods for pattern recognition [25], and has been applied in a variety of cases [26] [27] [28]. Its simplicity and relatively high convergence speed make it a popular choice.

KNN Classifier works as follows. First for each one of the training set elements a classification of it is performed based on various neighbourhoods. The *k* value that maximizes the DC of each classification is found. Therefore, for each training set there corresponds a particular k value which is considered the best available. Afterwards, for each unknown element, the nearest neighbour is found and its *k* value is assumed (based on the "optimum" *k* array). Then, the KNN classifier is applied on that test element, using that k value. As a concept, this is something similar to one of the ideas presented in [27].

K nearest neighbour algorithm is also called as lazy learning algorithm. This is so because it defers the decision to generalize till a new query is encountered. Whenever we have a new point to classify, we find its K nearest neighbours from the training data [28].

K nearest neighbours search algorithm maintains a priority queue. The entries of the queue are Minimum Bounding Rectangles (MBRs) and objects which will be examined by the algorithm and are sorted according to their distance from the query point. An object will be examined when it reaches the top of the queue. The algorithm begins by inserting the root elements of the R-tree in the priority queue. Then, it selects the first entry and inserts its children. This procedure is repeated until the first data object reaches the top of the queue. This object is the first nearest neighbour. The KNN Classifier algorithm is shown below. Therefore, each object of the test set is a query point and each object of the training set, contains an additional attribute which indicates the class where the object belongs to. The R-tree is built using the objects of the training set.

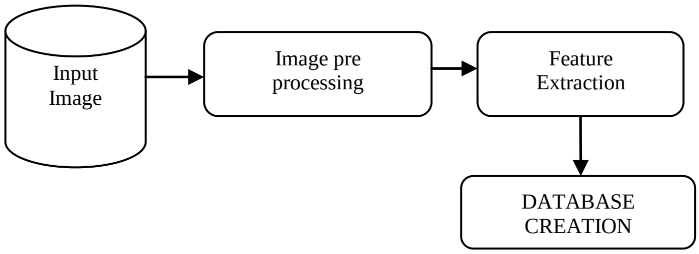
```
PriorityQueue.enqueue(roots children)
NNCounter = 0
while PriorityQueue is not empty and NNCounter · k do
    element = PriorityQueue.dequeue()
    if element is an object or its MBR then
        if element is the MBR of Object and PriorityQueue is not
        empty and objectDist(q, Object) > PriorityQueue.top then
          PriorityQueue.enqueue(Object, ObjectDist(q, Object))
        else
          Report element as the next nearest object (save the
          class of the object)
      NNCounter++
        if early-break conditions are satisfied then
          Classify the new object q in the class where the most
          nearest neighbors belong to and break the while loop. Q
          is classified using NNCounter nearest neighbors
        endif
    endif
  else if element is a leaf node then
      for each entry (Object, MBR) in element do
        PriorityQueue.enqueue (Object, dist(q, Object))
      endfor
      else /*non-leaf node*/
          for each entry e in element do
              PriorityQueue.enqueue(e, dist(q, e))
          endfor
      endif
end while
if no early-break has been performed then // use k nearest 0
    neighbors
Find the major class (class where the most nearest neighbors
   belong to)
Classify the new object q to the major class
endif
```

V. *PROPOSED WORK*

This paper proposed the method to improve the existing bag of features (BoFs) technique used for the image classification. Proposed Methodology is divided into two parts. They are as follows:

I. Training/Database Creation
II. Type Evaluation

The overall diagrammatic representation of the proposed work is as shown below:-



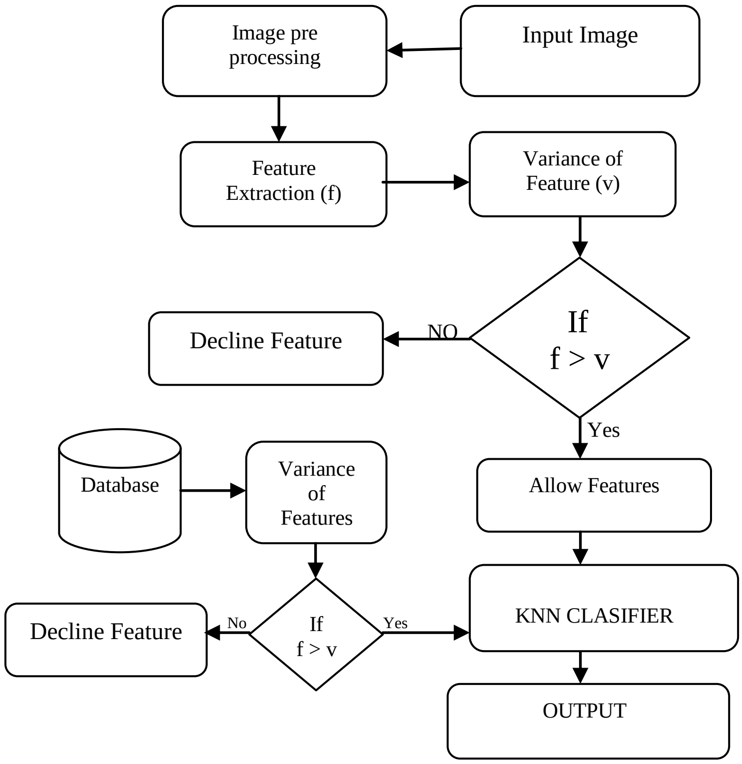


Figure 1 - Diagrammatic representation of work flow

The proposed methodology consists of two phases:-

1) Training Phase and
2) Evaluation Phase

A. Training Phase

In training phase firstly the complete system is trained. Input images are used to create the database, image pre processing. Different features of the image are calculated from the pre processing of the image. Features calculated from the images are Energy level, Edge Map, Extended Histogram and Heat map features. For calculating Heat-map features Silencing mapping technique is used. This features are then stored in the database for the training of the system. For the better result n numbers of image features are stored in the database.

B. Evaluation Phase

In the evaluation module input image is given to the system. The pre processing is done on the image to remove the noise from the image, and then features are calculated from the image using pre processing. Spatial pooling is applied on the input image feature to improve the classification results. Variance of features is calculated and if the value of feature is greater than value of variance then only feature is used for the classification otherwise feature is decline. Feature whose value is greater than variance pass to the KNN classifier. Database which is created in the training phase is also passing to KNN classifier for the evolution process. From the database of the features only those features are allow for classification whose values are more than variance and other features are decline. KNN algorithms take the input from both database and system and find the similar features from the database and generate output. Now the active features are given as input to the system as shown below:-

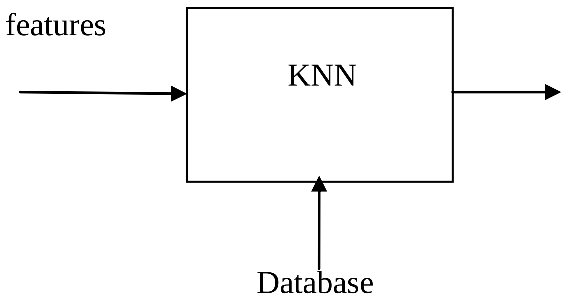


Figure 2 - Diagrammatic Representation of Evaluation Phase

VI. *EXPERIMENTAL RESULT*

Table 1- Evaluation Result of proposed and BoF method

No of Images	No. of Feature	Reduced Features	P_{BoF}	R_{BoF}	P_{OT}	R_{OT}	T_{BoF} (ms)	T_{OT} (ms)
10	600	455	0.81	0.8	0.9	0.92	30	10
20	600	350	0.82	0.8	0.91	0.93	35	15
30	600	320	0.84	0.83	0.94	0.95	40	24
40	600	310	0.87	0.85	0.96	0.98	50	30
50	600	305	0.88	0.89	0.98	0.98	70	39
60	600	290	0.9	0.88	1	1	100	48

Table 1 shows the complete evaluation results for the both bag of features classification method and proposed improve bag of features classification method with spatial pooling technique. From the results it is clear that proposed method improve the accuracy and time of the image classification process.

VII. *CONCLUSION*

In this paper the Bag-of-features model is extend by using spatial pooling to improve the time and accuracy of the image classification model. Main aim of the paper is to improve the accuracy and time consumption of the image segmentation and classification process. Experimental result showed in the table 1 proves that the proposed method is able to improve the accuracy and the time consumption of the image classification process. Time required by the Bag-of-features model is reducing up to 50 % in the proposed method.

REFERENCES

[1] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual Categorization with Bags of Keypoints,” in Proc. Workshop Statist. Learn. Comput. Vis., Eur. Con. Comput. Vis., 2004, pp. 1-22.
[2] M. Marszalek and C. Schmid, “Spatial weighting for bag-of-features,” in Proc. Comput. Vis. Pattern Recognit., 2006, pp. 2118-2125.
[3] E. Nowak, F. Jurie, and B. Triggs, “Sampling strategies for bag-of features image classification,” in Proc. Comput. Vis. Pattern Recognit.,2006, pp. 490–503.

[4] G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," Proc. ECCV Workshop Statistical Learning in Computer Vision, pp. 59-74, 2004
[5] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning Object Categories from Google's Image Search," Proc. 10th IEEE Int'l Conf. Computer Vision, vol. 2, pp. 1816-1823, 2005.
[6] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-offeatures," in Proc. Comput. Vis. Pattern Recognit., 2010, pp. 3352–3359.
[7] Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27(10):1615–1630.
[8] Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10):761–767.
[9] Lowe DG (2004) Distinctive image features from Scale-Invariant keypoints. International Journal of Computer Vision 60(2):91–110.
[10] Agarwal, S., Awan, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. PAMI 26 (2004) 1475–1490
[11] Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: CVPR03. (2003) II: 264–271.
[12] Bouchard, G., Triggs, B.: Hierarchical part-based visual object categorization. In: CVPR. Volume 1. (2005) 710–715.
[13] Agarwal, A., Triggs, B.: Hyperfeatures – multilevel local coding for visual recognition. In: ECCV. (2006)
[14] Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60 (2004) 91–110
[15] Leibe, B., Schiele, B.: Interleaved object categorization and segmentation. In: BMVC. (2003)
[16] Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: ECML-98, 10th European Conference on Machine Learning, Springer Verlag (1998) 137–142
[17] Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials
using three-dimensional textons. IJCV 43 (2001) 29–44
[18] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, D., Petkovic, D., Yanker, P.: The qbic project: Querying image by content using color, texture, and shape. SPIE 1908 (1993) 173–187
[19] Lazebnik, S., Schmid, C., Ponce, J.: Affine-invariant local descriptors and neighborhood
statistics for texture recognition. In: ICCV. (2003) 649–655
[20] Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: ECCV. (2002) I: 128.
[21] Rubner, Y., Tomasi, C., Guibas, L.: The earth mover's distance as a metric for image retrieval. IJCV 40 (2000) 99–121'
[22] Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: ECCV'04 workshop on Statistical Learning in Computer Vision. (2004) 59–74
[23] Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: ICCV. (2005)
[24] Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from google's image search. In: ICCV. (2005) II: 1816–1823.
[25] C. Hsieh, E. Lai, and Y. Wang, "Robust speaker identification system based on wavelet transform and Gaussian mixture model", Journal of Information Science and Engineering, 19, pp. 267-282, 2003.
[26] Moreno-Seco, F., Mico, L. and Oncina, J. A Modification of the LAESA Algorithm for Approximated k-NN Classification. Pattern Recognition Letters 24 (2003), pp. 47–53
[27] Abidin, T. and Perrizo, W. SMART-TV: A Fast and Scalable Nearest Neighbor Based Classifier for Data Mining. Proceedings of ACM SAC-06, Dijon, France, April 23-27, 2006. ACM Press, New York, NY, pp. 536-540
[28] Khan, M., Ding, Q. and Perrizo, W. K-Nearest Neighbors Classification of Spatial Data Streams using P-trees. Proceedings of the PAKDD, 2002, pp. 517-528