International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Spatial pooling of heterogeneous features for image classification using GPP

Charulata Rathi[1], S.V.Jain[2]

Student, Computer Science and Eng., ShriRamdeobaba College of Engineering and Management, Nagpur, India [1]
Assistant Professor, Computer Science and Eng., ShriRamdeobaba College of Eng. and Management, Nagpur, India [2]

Abstract—**Geometric phrase pooling (GPP) refers to extraction of spatial features of images in both x and y coordinate. It holds an important stake in the process of image classification. Application of GPP for image extraction is done using BOF model.The Bag-of-Features (BoF) is a widespread model that targets to epitomize images as a loose collection of features devoid of the use of some spatial statistics. As a result of its straightforwardness and decent enactment its attractiveness have significantly increased for image classification. BoF has progressed from method named as texton used in texture analysis. For execution of the BoF model four selections are necessary to be done, these selections are method to be used for sampling patches, procedures used to define them, characterization of the obtained spatial dissemination and finally classification of images built on the outcome. With the use of various methods such as Region of Interest (ROI) and extraction of multiple descriptors, disadvantages of BoF model can be neutralised. In spite of using all the above mentioned techniques rational assimilation of all the modules is not capable of resolving the disadvantages. And to overcome the above mentioned issue, this paper suggest an excellent framework with spatial assembling of heterogeneous features .BOF model is appliedby means ofsucceedingprocedures , speeded up robust feature (SURF) descriptor and canny edge map detector are used for extraction of spatial heterogeneous features, then feature pooling is used for construction of codebook, Support vector machine (SVM)is used for image classification and lastly color, texture and shape based techniques are used for image retrieval.**

Keywords— **Image classification, BoF Model, K-medoid clustering, Image matching, SURF.**

I. INTRODUCTION

Image classification has been one of the most prominent procedure in computer vision. It comprises of image processing and image analysis ,this technique primarily aims to convert input 2D image to another by performing from pixel wise operation such as enhancement in the contrast level , local or global descriptors extraction, noise removal or edge extraction. Some of the main applications of image classifications are scene matching, remote sensing, object detection, medical applications, Google goggle and many more.One of the prominent algorithm applied for image classification is Bag-Of-Feature, which recommends enhanced illustration of images by statistics-based model [1,2,3].To achieve this speeded up robust feature(SURF)is used. using SURF, descriptors from the input image are extracted by selecting some prominent features based on pixels depicting Swift changes in intensity values in both the planes. In SURF a histogram is constructed along the local neighbourhood of each key point by construction of a descriptor vector. [4,5]This paper proposes a technique for image retrieval using SURF, prominent features extraction is done. Robustness and reduction in run time are some of the primary advantages that SURF depicts over SIFT. For edge map extraction canny edge map detector is used. one of the many advantages of canny is its preciseness. Procedure involve in combining of SURF descriptor and canny edge map is called as feature pooling, which is used for codebook construction. SVM classifier is trained using training dataset, so as to label the images into restricted categories.[6]feature based extraction technique such as texture of the object, color of the image and shape of the object are used for retrieval of images.

II. RELATED WORK

In the year 2014, LingxiXie, Qi Tian, Senior Member, IEEE, Meng Wang, and Bo Zhang proposed Spatial Pooling of Heterogeneous Features for Image Classification[7]. In this paper texture and edge based local features of input image are extracted using SIFT (Scale Invariant feature Transform) descriptor, then midlevel image representation upon complementary features is build using geometric visual phrase and finally spatial weighting of the image is calculated using the smoothed edge map to capture the image saliency. Construction of codebook for database vocabulary is done using K-means algorithm. Although BoF model is successful using SIFT, some of the disadvantages of SIFT are it suffers from Synonymy and polysemy[7,8,9], time complexity is more as compared to SURF, also process complexity is

increased for calculating local and global features as two different algorithms are used. K-means algorithm also have some drawbacks such as it is not invariant to non-linear transformation which means with different representation of same data we get different results, unable to handle noisy data and outliners.

In the year 2011, Y. Zhang, Z. Jia, and T. Chen, suggested "Image retrieval with geometry-preserving visual phrases", Bag-of-visual-word (BoV) is one of the most popular methods of large scale image retrieval [10]. RANSAC is the spatial verification used to provide the ranking for the results acquired by adding spatial information as the post processing step. Due to its enormous cost computational techniques r spatial verification cannot be applied to whole hierarchy of images. Approach in this paper uses Geometry-preserving visual phrases (GVP) to encode more spatial information and hence reducing the cost spatial verification. Memory usage or computational time is increased by using GVP. Only disadvantage of the GVP method is that it is translation invariant.

In the year 2010, J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, proposed "Locality constrained linear coding for image classification," in Proc. Comput. Vis. Pattern Recognit. Better results than traditional BoF model can be obtained by using nonlinear classifiers which is the base concept of proposed paper[11]. Very effective yet simple coding scheme, Locality-constrained Linear Coding (LLC) is used replacing the traditional Vector Quotient (VQ). Final representation is generated by integrated max pooling of the projected coordinates, which are generated using locality constraints to project all the local descriptors into their local coordinate system. Constrained least square fitting problems are solved using K-nearest-neighbour search which process a fast approximated LLC. Bearing the advantage that several frames per second can be processed into the system even if the size of the code book is very high. Despite of the above mentioned advantage, disadvantage of the mentioned K-nearest-neighbour search are value of the parameter K prerequisites to be define first, training of which features to be used for different types of distance that means Distance based learning is not clear and working out cost is quite high because we requisite to calculate distance of each request case to all working out models.

III. PROPOSED WORK

In the proposed architecture improved BOF model will be used for image classification. Local descriptors such as SURF and global descriptors are extracted from input image, information extracted from various descriptors is used to build the codebook and noise depressing techniques are used resulting into formation of the dataset with compacted pictorial vocabulary. At last, codebook is raised by quantization of descriptor, and image representation is done with help of a statistical histogram derived from visual words. Image retrieval and image classification are some of the major application of BOF model. Following are the steps involve in the proposed work:

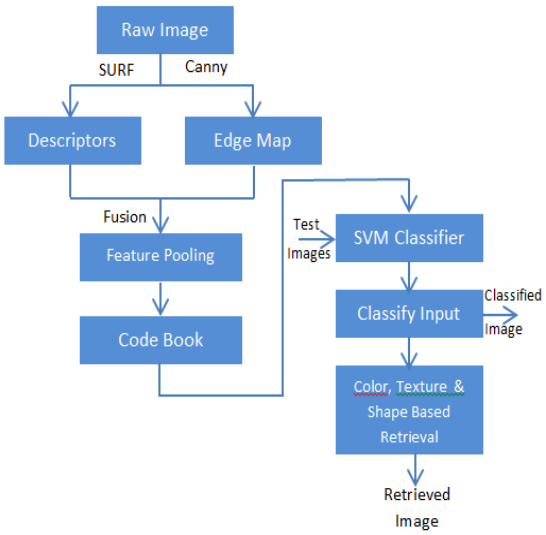

Fig.1. Flowchart of the proposed work

IV. IMPLEMENTATION

A. *Dataset*

Dataset used is Caltech 101 which contains 9144 images belonging to 102 classes which also include background category.Substantial distortion is found between dissimilar objects belonging to the same category.We have taken some categories from this dataset and divide images of all categories into training and testing dataset.

Fig.2. Example of classes of Caltech dataset

B. *Feature extraction*

Applications or algorithms that yield explanation of visual structures of any image or video are called as descriptors. Fundamental characteristics such as texture, shape, color or motion are captured using this descriptor. They are the primary step to find out the connection between pixels contained in a digital image.There are two types of descriptors, local and global.Purpose of the local detector is to match the features in order to offer a representation that allows matching the local structures between images. To fulfil this aim the local descriptor extractors much achieve two important norms:

- Repeatability and preciseness should be present in the procedure of feature extraction, as a result the same and correct features are extracted on the images displaying the identical object
- Simultaneously, distinctiveness should be present in the extracted features, as a result dissimilar image structures can be distinguished from each other.

The principal footstep of the local feature extraction is to discover a group of unique keypoints that can be dependably confined under fluctuating imaging situations, viewpointdeviations, and in the manifestation of noise. For local feature extraction SURF is used. The regular version of SURF is quite a few times faster than SIFT and more tough against altered image transformations than SIFT.[4]Primarily, the SURF key points are mined in both the base image and the real-time image respectively, and then, different methods are used to understand the equivalent of key points.

Scale space theory is adopted for the detection of the SURF key point which results in induction of exceptional scale invariant property in the extracted key points. In the proposed paper we are using SURF algorithm for local feature extraction.[4,5]SURF descriptor extracts strongest 64 features from the images. So the feature vector dimension for one particular image has become 64*64. This accounts to 4096 SURF features for every single image. For achieving the mentioned feature extraction blob detector created on Hessian Matrix is used, For an image I at any point P(x,y), the Hessian matrix H(p, σ) at point p and scale σ, is distinct as mentioned below:

$$H(p,\sigma) = \begin{pmatrix} L_{xx}\ (p,\sigma) L_{xy}(p,\sigma) \\ L_{xy}\ (p,\sigma) L_{yy}(p,\sigma) \end{pmatrix} \quad (1)$$

Where $L_{xy}\ (p,\sigma)$ etc. can be defined as grayscale image second order derivatives.

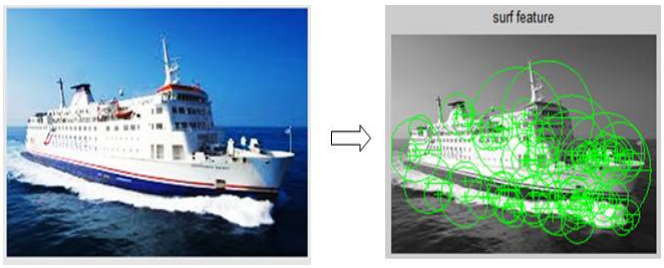

Fig.3. Example of local feature extraction using SURF

Global features are the edge map of the image. For extraction of the edge map canny edge detector is utilized. The principalobjective of edge detection in general is reduction in the amount of data obtain from the image, without disturbing spatial properties which are to be used for further procedure of image processing. Detection, localization and number of responses are the primary criteria for canny algorithm. In the propose paper canny algorithm is used for edge map extraction .canny detector extracts many features out of which prominent 64 features are taken into consideration for particular image.

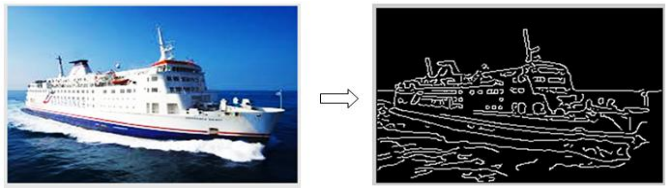

Fig.4. Example of global feature extraction using Canny

C. *Code book construction*

Quantization of the extracted descriptor is done to make them compact. To achieve this target codebook is constructed using the descriptors extracted from the whole dataset. Combination of both the descriptors extracted by using SURF and canny edge detector is used for the construction of codebook.

D. *SVM classifier*

One of the new controlled classification procedures to the image mapping community is Support Vector Machines (SVMs) [6] Accuracy, Robustness and effectiveness are the prominent features of SVM due to which it has gained reputation. Fundamentally SVM acts as the binary classifiers, though implementation can be done to switch the multiple classification tasks if required.SVM classifier is trained using the codebook .Input image from testing data set is provided to SVM classifier and SVM classifier classify the image by providing correct label to it.. One-Against-One (1A1) and One-Against-All (1AA) are the two commonly followed procedures in the multi-class SVM classifiers
1-against-all approach works on the rule set which distinguishes a single class from left over classes, decision functions "k", are calculated for each class. In 1-against-1 approach constructs k(k − 1)/2 classifiers where each one is trained on data from two classes. Proposed paper also uses Multi-class SVM.

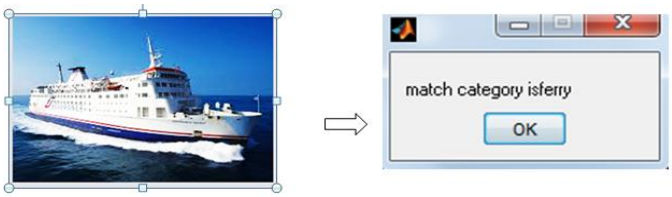

Fig.5. Example of labelled input image

E. *Color, Texture and Shape based extraction*

One of the prominent feature of Information source in an image document. In any communicating image classification and retrieval system, firstly color histogram feature and thenGray-Level Co-Occurrence Matrix (GLCM)are used to direct image contents,then clustering of images is done to classify them intonumerous classes on the basis of their visual features. Finally on the basis of the feature vectors keptin the database after matching alike images are retrieved. Calculation of Hue, Saturation and Value (HSV) color histogram and derivation of joint histogram is done on the basis of the hue and saturation histogram considered together. Joint histogram is used for the extraction of color. Invariance of the color with respect to the rotation of image, translation of image or scaling of image makes color as an important feature for image extraction. In the proposed approach color momentum for feature extraction is used for extraction of color. Once the color moment of the image is calculated, likeness score is computed to retrieve similr images from database using following function,

$$d_{mom}(H,I) = \sum_{i=1}^{r} w_{i1}|E_i^1 - E_i^2| + w_{i2}|\sigma_i^1 - \sigma_i^2| + w_{i3}|S_i^1 - S_i^2| \quad (2)$$

Where:

H and I are the color distributions of the under consideration
i is the channel index and r is the total number of channels
E_i^1 And E_i^2 are the first order moments computed for the image distributions.
σ_i^1And σ_i^2 are the second order moments computed for the image distributions.
S_i^1And$S_i^2$are the third order moments computed for the image distributions.
w_{i1}, w_{i2}And w_{i3} are weights, specified by the user, for each of the three color moments used.
Homogenous property that is not result of any single color creates visual patterns which are known as Texture.Density, Uniformity and contrast are some of the features considered under texture.Statistical model based and transform based are the two texture descriptors.Some of the important properties to be possessed by shape features are affine invariance, noise resistance, reliable, statistically independent, transitional, and rotational and scale invariance.Description of any shape feature is given by set of numbers called as shape descriptor. Main purpose of the Shape descriptor is to produce the output that would quantify the shape, so that quantified shape agrees with task specific requirement.Shape based image retrieval consist of measurement of the likeliness between two shapes based on their features. Above mentioned all three procedures are used in the proposed paper for completion of image retrieval module.

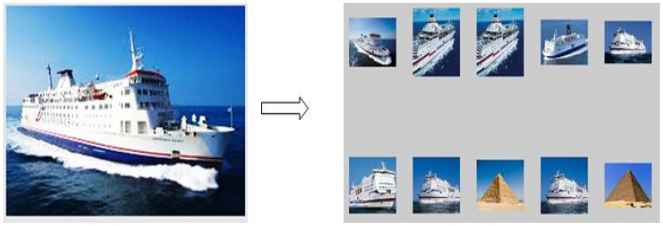

Fig.6. Example of retrieved image from database usingColor, texture and Shape based Extraction.

V. RESULT

Using the methodology proposed in this paper, 75 % is the image classification and retrieval accuracy rate. Accuracy rate has been increased sustainably. Time complexity involved in the extraction of descriptor have been reduced because of the use of SURF descriptor. Advantage SURF bears over SIFT is its robustness and scale invariant feature.

VI. CONCLUSION

In the area of image classification and retrieval there is still a huge scope of improvement. The result achieved using the proposed method can still be improved and sophisticated using more robust SURF and BOF model. Reduction in the time complexity for extraction of descriptors is due to the use of SURF algorithm .proper classification of the input

images is achieved using multiclass SVM. Using feature extraction techniques based on color, texture and shape,image with maximum alike features from the database is retrieved. More accuracy and stability in the result achieved using feature extraction techniques can be brought using more scale invariant techniques.

ACKNOWLEDGMENT

We sincerely thank the authors mentioned in the references for the valuable information and statistics obtained through their findings.

REFERENCES

[1] M. Marszalek and C. Schmid, "Spatial weighting for bag-of-features," in Proc. Comput. Vis. Pattern Recognit., 2006, pp. 2118–2125.
[2] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-offeatures image classification," in Proc. Comput. Vis. Pattern Recognit.,
2006, pp. 490–503.
[3] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-offeatures," in Proc. Comput. Vis. Pattern Recognit., 2010, pp. 3352–3359
[4]Su Juan1, XuQingsong, Zhu JinghuaXi'"A Scene Matching Algorithm Based on SURFFeature",in
an Research Institute of High Technology.

[5] zhihengzhout,xiaowenout,jing xu2 "SURF feature detection method used in object tracking",in
College of Electronics & Information Engineering,South China University of Technology.

[6] *Gidudu Anthony, * Hulley Greg and *MarwalaTshilidzi , "Classification of Images Using Support Vector Machines,"inDepartment of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, Private Bag X3, Wits, 2050, South Africa

[7] Spatial Pooling of Heterogeneous Features for Image Classification LingxiXie, Qi Tian, Senior Member, IEEE, Meng Wang, and Bo Zhang.
[8] D. G. Lowe, "Distinctive image features from scale-invariant key points," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.
[9] L. Xie, Q. Tian, and B. Zhang, "Spatial pooling of heterogeneous features for image applications," in Proc. 20th ACM Multimedia, 2012,
pp. 539–548.
[10] Image Retrieval with Geometry-Preserving Visual Phrases Yimeng Zhang ZhaoyinJiaTsuhan Chen School of Electrical and Computer Engineering, Cornell University.

[11] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality constrained linear coding for image classification," in Proc. Comput.Vis. Pattern Recognit., 2010, pp. 3360–3367.
[12]Sridhar, Gowri "Color and Texture Based Image Retrieval" Department of Information Science and Technology College of Engineering, Guindy Anna University Chennai .

[13] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349–1380, 2000
[14] CharulataRathi, S.V.Jain2, " A study of Image classification using BoF model",ijarcce publication volume 5,issue 2.

BIOGRAPHIES

CharulataRathi completed her Bachelor of Engineering in Computer science and Engineering in 2014.She is pursuing her Masters in Technology in Computer Science and Engineering from ShriRamdeobaba College of Engineeringand Management, Nagpur-440013. Herareas of interest include Image Classification, Information Retrieval.

Professor Shweta Jain received the Masters in Technology in Computer Science and Engineering from Nagpur University in 2009 as a first merit holder. She is currently Assistant professor in computer science and engineering department at ShriRamdeobaba College of Engineering and Management Nagpur. She has a total teaching experience of around 13 Years. Her research interests include Pattern Recognition, Digital Image Processing and Machine Learning.