FUSION OF DEMANDS IN REVIEW OF BAG-OF-VISUAL WORDS

Silkesha Thigale[1], A.B Bagwan[2]

[1]*Department of Computer Engineering, JSPM's RSCOE, University of Pune*
[2]*Department of Computer Engineering, JSPM's RSCOE, University of Pune*

Abstract
This paper presents the review related to the contributions of the work in Bag-of-Visual words applications. State-of-the-art has adopted the BoW model for retrieving the images on large scale, as it proves to be reliable. Visual words are the end results of the quantization. BoW model is used in Content Based Image Retrieval (CBIR). It system operates as per the user's demand for the relevant image retrieval. BoW is primarily used for the local descriptors.

Keywords— *Code word, retrieval of information, classification*

--***--

1. INTRODUCTION

Users in many different professional fields are lucky enough to access the web based images remotely. The term 'Bag-of-Visual Words' (BoVW) is very popular in Content based Image Retrieval (CBIR). CBIR have always being the active areas for all types of researchers. A CBIR's mainstream is from the field of image processing. Image processing covers the sub areas like image enhancement, compression, interpretation, and transmitting. BoVW is used as the representation method for image classification. It had prominently used in the fields of image processing, computer vision, image categorization, object recognition and annotation. This helps to solve the problems related to vision in the images. BoW gives excellent performance in the task of image retrieval. As per the query formats, the retrieval is categorized into text and image contents. Retrieving via text is by using keywords. Due to the simplicity of BoW model it is widely been used.

Term Frequency –Inverse Document Frequency (TF-IDF) is a weighing scheme, which has been used in for weighing the visual words either low/ high. There are problems associated with searching of an image as per its semantic content like performance, resolution, illumination. BoW treats image features as words. Bag-of-Visual Words are nothing but the sparse vector which actually counts the occurrence The design of BoW or bag-of-features model is essentially for the purpose of local descriptors in the images which are used to detect the key points in the images. Scale Invariant Feature Transform (SIFT) is one of the local descriptor. Then the quantization of local descriptors into "visual words" is done. This method also named as building of visual vocabulary. Thereafter, the representation of the each image is done by the vector of words as shown in fig 1. Above. It is possible to get the size of visual vocabulary by using clustering methods.

2. FACE DETECTION

Face plays a crucial role in the identification of any human. To extract any information from the image, users are required to understand the content of the image. Face detection technique is the way of detecting the human faces with the combination of use of computer technology. The orientation of this face detection is from the branch of Computer vision and Human Computer Interaction (HCI). Today's world deals the digital images for which number of devices are used like digital cameras, surveillance system etc.
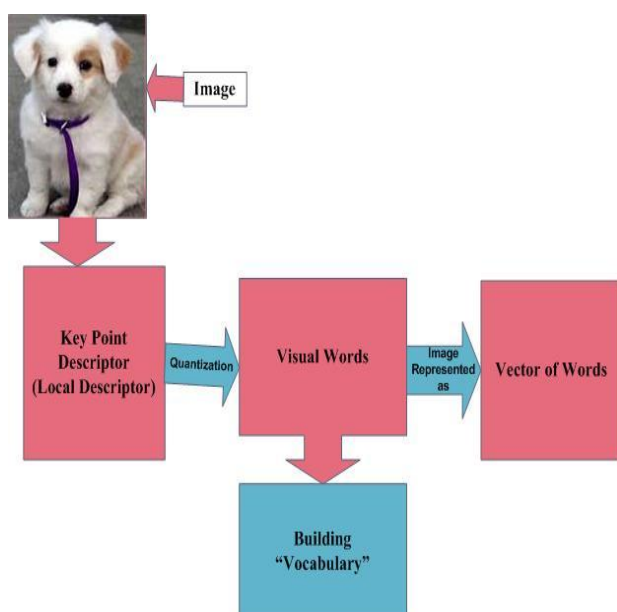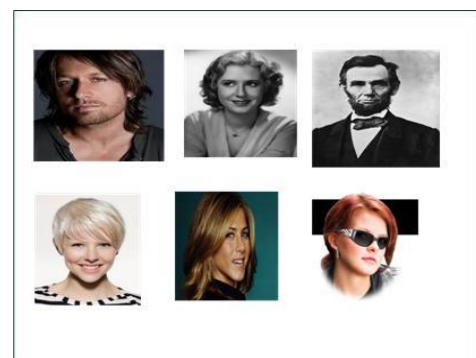

Fig 1 'Steps in BoW Model'

Fig 2 'Example of images of faces'

Face is having local features as hair, eyes, nose, lips, mouth etc. All the images related to the face belong to same category which is 'face'. The generation "codebook" is when the numerical vector are converted to "code words" i.e clusters. Mapping of image face features with codeword is through the process of clustering which then are ultimate represented in the form of histogram. The descriptors used here are Histograms of Oriented Gradients (HOG). [1]

3. SCENE CLASSIFICATION

In recent decades, there were handmade paintings, now-a-days there is a drastic growth in digital images in all different fields like advertising, designing, education etc. [9] The technology, like in the form of 'photography' played a crucial role in capturing the images. The scenes images are categorize as indoor /outdoor. [2] Its a challenging job to classify a scene into one of the categories. For example: classifying a photograph which may contain the outdoor scene like (beach, forest, and high way etc). Bag-of-feature methods are more suitable for the whole-image categorization as it represent an order- less collection of local features .The 'regions' here are nothing but meaningful division of images for which 2 approaches are used. Features are collected like colour then codebook is generated by region clustered. Fig 2 shows below different categorizations of scenes in columns. A "bag-of-individual regions", are associated with region labels. The regions which meet the previous criteria are referred as "bag of region pairs".
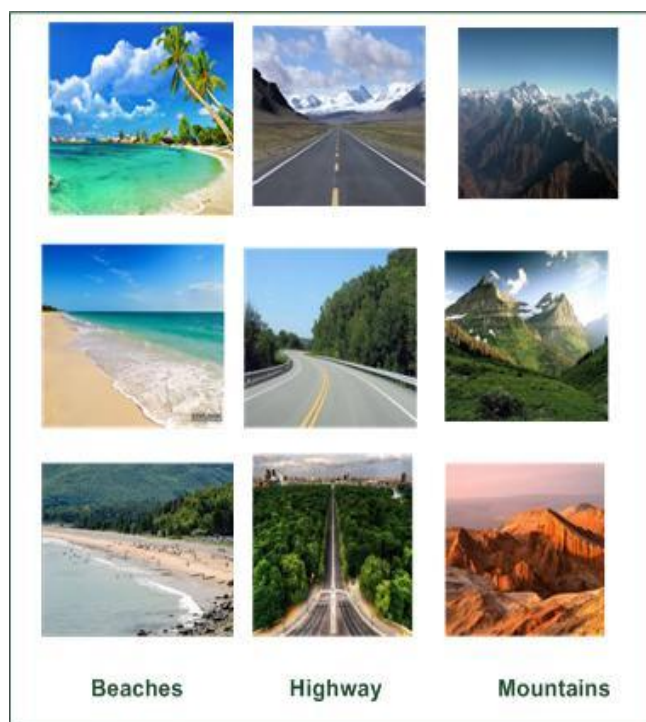

Fig 3 "Example of Categorization of Outdoor Scenes"

Pattern recognition works on the principal of selection of feature. The images often viewed in a specific class but they may not be or say occasionally exists in other class. The technique can also be used in remote sensing.

4. COLOUR IMPACT

The usage of colour improves performance for detection. For the purpose of improving colour information in BoW model there are some approaches here. These approaches are used in the field of object as well as scene recognition. In "Feature detection" improvement can be done with the help of colour based regions which reveals more information for guiding. For this initially it is required to detect the region or say selection. Representation of Local descriptors is through the selected regions. Thereafter quantization is done into codebook form of visual words. The images are then represented in the form of histogram. The proportion of pixels in a given image is well identified by use of colours. In "Feature Description" the centre of attention is the 'shape' [3]

Table 1: 'BoW Applications'

Sr. No	Techniques	Dataset used
1	Face Detection	i)AT&T ii) FEI
2	Scene Classification	MIT LabelMe
3	Colour Impact	i) Flower ii) Pascal VOC 2009
4	Language for BoVW	i) Caltech-101
5	Forests and Ferns	i) Caltech-101 ii) Caltech-256

In concern with shapes, to match a particular 3-D objects are a tough task. Converting gray to colour is somewhat difficult. The above table shows the different techniques which make use of BoW model. The dataset used in these techniques are actually used in the experiments performed by the contributors to the papers respectively as per the references given in this paper. Even though there is a utilization of 'colours' in numerous techniques they still exists depends on the variations of basic ideas.

5. LANGUAGE FOR BoVW

BoW model was first launch for video retrieval. [8] It has been proved very efficient and effective. Difficulties in the representation of BoW are due to loss of spatial information of visual words and noisy visual words. Language modelling tool is used here which have gained its popularity from speech recognition and text analysis. [4] The approaches in relation with the image classification and object recognition rely on BoW model and visual words as these techniques proved to gives the assuring results.
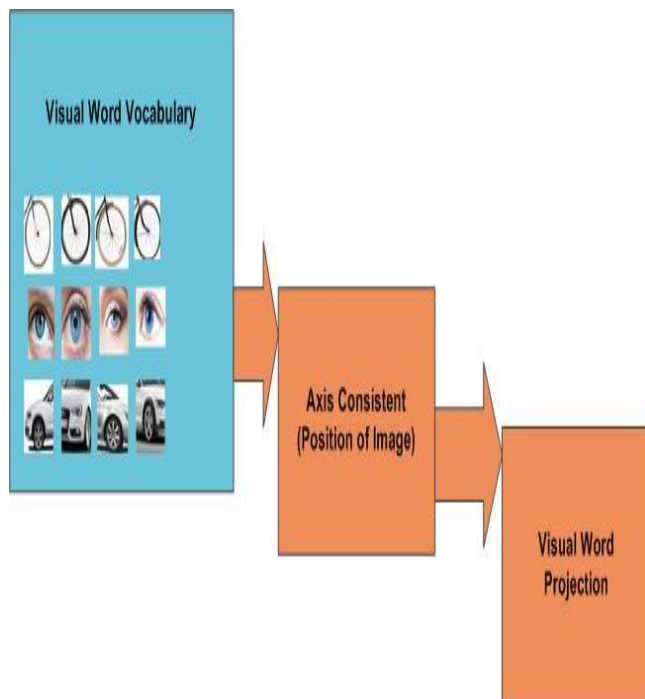
Fig 4 'Steps in Visual Sentences: Language'

Due to some limitations in BoW, as in vocabulary building process there is existence of some noisy words due to coarseness. BoW use clustering algorithms like k-means in which the words having same meaning creates ambiguity. New approach commence here is visual sentences for representation of image.

Key point detector commercially popular are for instance: Harris Laplace, Hessian Affine etc. Visualization is the activity of creative process. Earlier the images were in the form of pictures, 2-D sketches, graphics, and photographs. Photographs are used in different areas like**:** advertising, publishing newspaper etc.

6. CLASSIFYING IMAGE: FORESTS AND FERNS

In object recognition rather than using the complete image, new method introduces here is of automatic selection of Region of interest (ROI). [5]

6.1 Matching of Images

The representation of the image is by making use of spatial pyramid which is related to appearance and shape. 'Matching' means verifying the similarity between the images I and J. Kernel function is used to compute
PHOW/PHOG (Pyramid Histogram of visual words: for descriptor appearance / pyramid (HOG) descriptor for shape)

6.2 Auto ROI Selection

The fig 5. Given below shows the technique for automatic selection of Region of interest (ROI). The selected area in the figure is shown by black frame on the image.
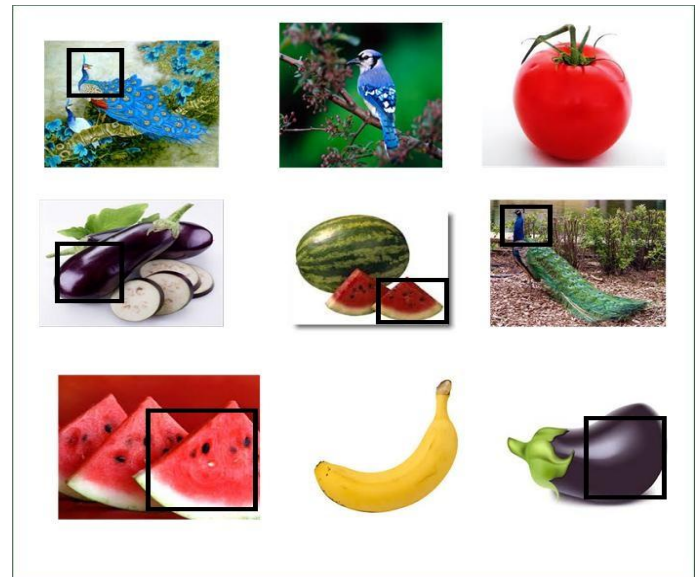

Fig 5 'Example**:** Automatic ROI selection'

7. CONCLUSIONS

Motivation to this work is due to interest to work in Content Based Image Retrieval (CBIR). There is excellent performance shown by BoW in Content Based Image Retrieval (CBIR). Query Expansion can also be used in BoW model. The techniques described above are advantageous to the images, as it keeps on varying as per the computations that takes place.

ACKNOWLEDGMENTS

REFERENCES

[1]   L.R. Cerna, G. Camara- Chavez, D.Menotti, "Face Detection: Histogram of Oriented Gradients and Bag of Feature Method",

[2]   D. Gokalp, S. Aksoy, "Scene Classification Using Bag-of-Regions Representations", Proceedings of CVPR, 2008

[3]   David A. Rojas Vigo, Fahad Shahbaz Khan, Joost van de Weijer and Theo Gevers "The Impact of Color on Bag-of-Words based Object Recognition,"

[4]   Pierre Tirilly, Vincent Claveau, Patrick Gros, "Language Modeling for Bag-of-Visual Words Image Categorization",

[5]   A.Bosch, A.Zisserman and X. Munoz, "Image Classification Using Random Forests and Ferns", Proceedings of ICCV, 2007

[6]   J.Sivic, B.C Russell, A.A Efros, A.Zisserman, and W.T Freeman, "Discovery Object Categories in Image collections", Proceeding of ICCV, 2005

[7]   S.Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", In CPVR, vol 2. 2006

[8]   J. Sivic and A. Zisserman "Video Google: A text retrieval approach to object matching in videos", proceeding ICCV, vol. 2, 2003.

[9]   A. Bosch, A. Zisserman, and X. Munoz, "Scene classification via pLSA", proceedings ECCV, 2006

[10]   G. Salton, M. McGill. Introduction to Modern Information Retrieval, McGraw-Hill, 1983

BIOGRAPHIES

Silkesha S.Thigale[1] I am currently pursuing my Master's (M.E). in Computer Engineering from J.S.P.M's RSCOE, Pune from 2012 till now. I have received my B.E. (IT) from PVG's COET, Pune in 2007 and Diploma in Information Technology from SSPP, MIT, Pune in 2004. My interest lies in Image Processing, Data Mining, and Computer Vision.

Dr.A.B Bagwan[2] He is currently working as Professor and Head at RSCOE Tathawade, Pune. He completed his B.Tech Electrical Engineering from IIT Kanpur in 1973 and M.Tech in Computer Science from IIT Madras is 1977. He finished his PhD in Computer science and Engineering from Bundalchand University, Jhansi in 2012.