Effect of Salient Features in Object Recognition

Kashif Ahmad
University of Engineering & Technology, Peshawar, Pakistan

Nasir Ahmad
University of Engineering & Technology, Peshawar, Pakistan

Kamal Haider
Ghandhara institute of Science & Technology Peshawar, Pakistan

Muhammad Jawad Ikram
University of Bradford, UK.

ABSTRACT

With the SIFT and SURF based recognition, the paper presents the impact of salient features in object recognition. We use the two well-known image descriptors in the bag of words framework on five online available standard datasets. Experiments show that by introducing saliency in the bag of words model, state-of-the-art performance can still be retained while reducing considerable amount of data processing and thus achieving faster execution times.

General Terms

Object Classification, saliency, image description.

Keywords

Salient features, Saliency, Object recognition, SIFT, SURF, Interest point detectors, Feature points.

1. INTRODUCTION

Object recognition is one of the most active areas in computer vision and image processing due to its wide range of applications. For this purpose a number of techniques have been proposed over the years. The basic goal of all the proposed techniques is to reduce the processing time with significant amount of accuracy. The use of interest point detectors is a part of such effort as discussed in [1]. These interest point detectors are based on repeatability, however their repeatability do not provide any information that the feature is salient i.e. the probability of the features to be correctly matched. Here the goal is to select the salient features to reduce the processing time and retain the sate-of-art performance. We introduce saliency in the bag of words model with two well-known image descriptors, SIFT [2] and SURF [3], and test the performance on five standard datasets.

2. RELATED WORK

Object recognition and salient features detection has been an active area of research for many years. In this section we will cover object recognition, salient features detection and the connection between them. We will discuss all the notable work done in this area.

Object recognition: object recognition or object categorization is the process of identifying an instance of an object category in an image. Different techniques of object recognition are available. Most of the proposed methods of object recognition use training data to obtain the visual dictionary. One such kind of notable work presented in [4] by Csurka. Key points are extracted from the training data, and an image descriptor known as SIFT by Lowe [2] is used for representing local information in neighbor hood of these key points. Clustering of K-means with descriptors is used to produce a visual dictionary. [5] and [6] represents the similar approach of visual dictionary. All the above mentioned papers relay on the histogram of visual words for the classification of images, and this technique is known as bag of features model.

There are some other approaches presented in [7, 8-10], that uses statistical models for generalization and extraction of semantic context for object recognition. Wolf and Bileschi in [11] used the training data for obtaining the semantic layers, from which they obtained the semantic context. Each semantic layer represents an object category in an image. Each pixel is assigned a label v. v =1 represents that the pixel belongs to the object in layer and v = 0 represents that it does not belong to object. So the corresponding values of pixels are used to determine the occurrence of pixel in an object. Object recognition model presented in [12] is based on annotating image regions with words. They used a number of features for the categorization of segmented regions into region types. EM based method is used for learning generalizing the mapping between region types and key words.

Hall et al in [13] discussed the importance of salient features in object recognition and saliency under scale changes. They compared and evaluated the performance of three well-known interest point detectors including Harris corner and edge detector, Iindeberg point detector and Harris-Laplacian interest point detector, in terms of selecting salient features in an image. The authors in [14] have provided a detailed study of detecting salient object in an image. For describing the salient object they proposed a number of features such as color and spatial distribution, multi-scale contrast and centre-surround histogram. They extend their approach of detecting salient object to sequential images too. [15] Presents a multi-scale algorithm that relates saliency, scale selection and content description for the selection of salient regions in an image.

3. SALIENT FEATURES

Salient features are those features that discriminate a feature from other features. Different researchers have defined the term “salient features” in different ways as presented in [16, 17, 18, and 19]. A feature which distinguishes an object from other objects in an image is said to have maximum saliency, so in other words we can say that it is the measure of discrimination power of an image feature.

4. OVERVIEW OF DESCRIPTORS

4.1 SIFT

An image detector and descriptor by Lowe [2] enjoying a wide range of applications. It is implemented in four stages. It uses DOG [20] for the collection of the feature vector and Best-bin-first [21] for the matching and indexing. Cluster identification is performed with Hough transformation, and uses Linear Least Square solution for relating the model with image. SIFT has been proved to be the best image descriptor among the all in terms of scale invariance in [22]. However it is slower then its competitor SURF.

4.2 SURF

Introduced by Herbert Bay in [3], SURF is a robust image detector and descriptor which uses integral images [23]. It is based on Hessian matrix and Haar wavelet transform. SURF is available in 64 and 128 dimensions. However Lau Juan proved in [24] that increasing SURF dimension does not help in improving the quality by a sufficient amount.

5. DATASETS

We used five standard datasets including Caltech, UIUC, TUDarmstadt VOC2005-1 and VOC2005-2, for the experimentation. From Caltech and UIUC datasets we used 200 images for training and 50 images for test, while from TUDarmstadt, VOC2005-1 and VOC2005-2 80,160 and 60 images for training and 20,40 and 15 images for test respectively. Figure.1 to Figure.5 shows the sample images from these datasets. In Caltech dataset we have four categories of objects Airplanes, Cars, Faces and Motorbikes, and in UIUC we have just one category of cars. TUDarmstadt has three categories including cars, cows and bikes. VOC2005-1 and VOC2005-2 both has four categories including cars, bikes, persons and cycles.

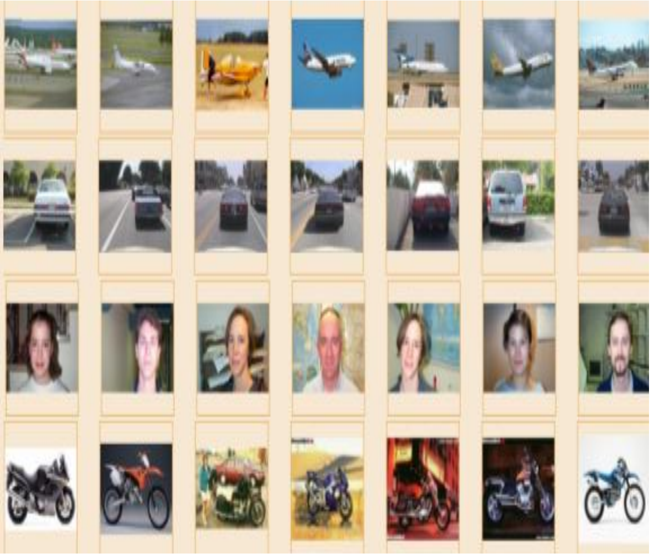

Fig 1: Sample images from Caltech Dataset

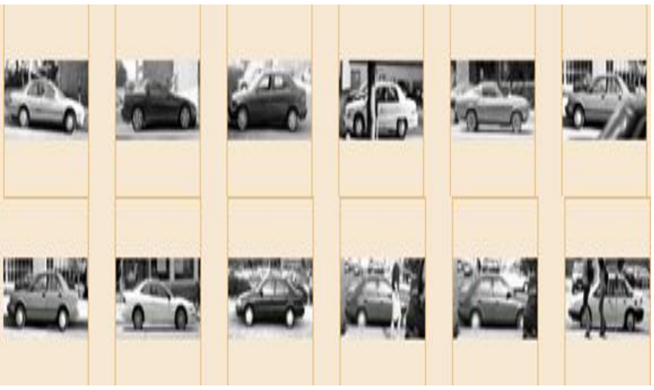

Fig 2: Sample images from UIUC Dataset

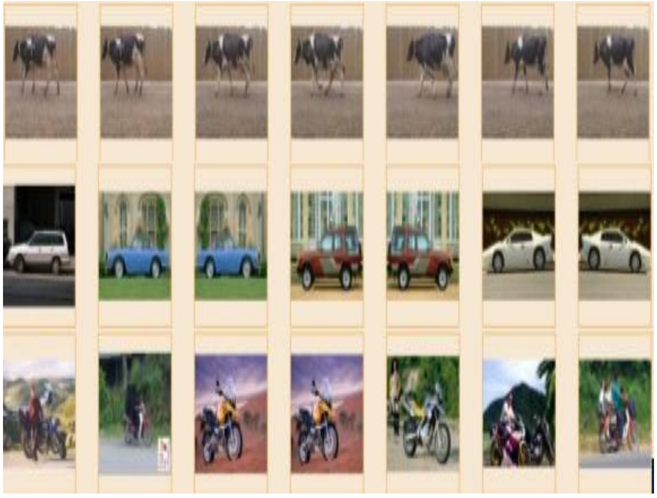

Fig 3: Sample images from TUDarmstadt Dataset

Fig 4: Sample images from VOC2005-1 Dataset

Fig 5: Sample images from VOC2005-2 Dataset

6. FLOW CHART

The figure.6 shows the flow chart of our work. In 1st step images are taken from the database, saliency is calculated in 2nd step, followed by the image descriptors (both SIFT and SURF). After extracting these feature points, K-means clustering is used to cluster the same feature pints and histogram representation of visual words is used which is known as bag of features. In the next step the algorithm uses a number of images for training. Testing is based on KNN search algorithm. At the end evaluation is carried out to get desired results.

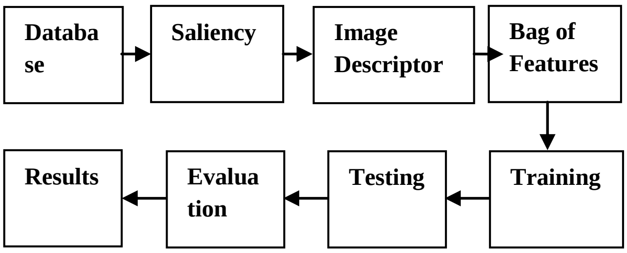

Fig 6: Flow Chart of Proposed Work

7. EXPERIMENTS AND RESULTS

In the experimentation process, first we measure the performance of both image descriptors i.e. SIFT and SURF on all five datasets in bag of feature framework. We are interested in two things, the number of objects correctly recognized and the number of feature points per image. In second phase, we introduce saliency in the bag of feature framework. Images are passed through saliency before applying the image descriptors and measuring the number of correctly recognized objects and the number of feature points per image. Figure.7 shows three types of images, the original image, Ltti Koch [25] saliency map and the image resulted after applying the saliency algorithm on input image.

It can be clearly seen in figure.6 that the background in saliency applied images is removed and only the object of interest or in other words only the salient features are selected. The results of the experimentation process are given in Table.1 to Table.10, which shows the number of objects correctly recognized for every category in all datasets. Table.11 and Table.12 shows the average number of feature points per image, detected by SURF and SIFT for all datasets with and without saliency respectively.

We used 200 images for training and 50 images for test from Caltech and UIUC datasets. TUDarmstadt has 80 images for training and 20 for testing purposes, while from VOC2005-1 and VOC2005-2 160 and 60 for training and 40 and 15 images for test.

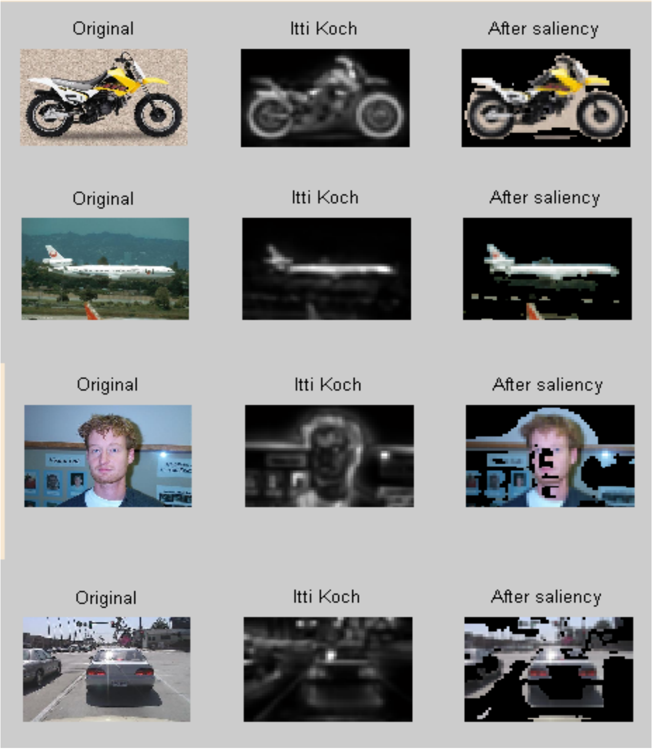

Fig 7: Images during different phases of experiment

Table 1: SURF Results with and without saliency on Caltech dataset. Number of images used for training=200, and number of images used for test=50

Category	SURF with out Saliency	SURF with Saliency
Bikes	48	44
Faces	41	37
Cars	39	35
Airplane	49	47

Table 2: SIFT Results with and without saliency on Caltech dataset. Number of images used for training=200, and number of images used for test=50

Category	SIFT without Saliency	SIFT with Saliency
Bikes	47	41
Faces	46	45
Cars	42	38
Airplanes	47	45

Table 3: SURF Results with and without saliency on UIUC dataset.Number of images used for training=200, and number of images used for test=50

Category	SURF without Saliency	SURF with Saliency
Cars	15	16

Table 4: SIFT Results with and without saliency on UIUC dataset. Number of images used for training=200, and number of images used for test=50

Category	SIFT without Saliency	SIFT with Saliency
Cars	13	20

Table.5: SURF Results with and without Saliency on TUDarmstadt Dataset.Number of images used for training=80, and number of images used for test=20

Category	SURF without Saliency	SURF with Saliency
Bikes	18	17
Cars	18	16
Cows	15	13

Table 6: SIFT Results with and without saliency on TUDarmstadt Dataset.Number of images used for training=80, and number of images used for test=20

Category	SIFT without Saliency	SIFT with Saliency
Bikes	**14**	**11**
Cars	**15**	**11**
Cows	**19**	**19**

Table 7: SURF Result with and without saliency on VOC2005-1 Dataset.Number of images used for training=60, and number of test images=15

Category	SURF without Saliency	SURF with Saliency
Bikes	27	24
Persons	15	12
Cars	7	8
Bi-Cycles	15	18

Table 8: SIFT Results with and without saliency on VOC2005-1 Dataset. Number of images used for training = 160, and number of images used for test=40

Category	SIFT without Saliency	SIFT with Saliency
Bikes	26	12
Persons	11	22
Cars	5	8
Bi-cycle	9	11

Table 9: SURF Results with and without saliency on VOC2005-2 Dataset. Number of images used for training=60, and number of images used for test=15

Category	SURF without Saliency	SURF with Saliency
Bikes	6	6
Pedestrians	7	8
Cars	8	6
Bi-Cycles	8	5

Table 10: SIFT Results with and without saliency on VOC2005-2 Dataset. Number of images used for training=60, and number of images used for test=15

Category	Sift without saliency	Sift with saliency
Bikes	8	5
Pedestrians	5	3
Cars	6	7
Bi-cycles	3	8

Table 11: Feature Points detected by SURF with and without saliency for all datasets

Dataset	Total Images	Avg:Feature points without saliency	Avg:Feature points with saliency
Caltech	1200	272.35	238.75
UIUC	250	8.55	6.49
TUDarmstadt	320	198.67	78.63
VOC2005-1	300	751.44	239.60
VOC2005-2	800	601.65	270.07

Table 12: Feature Points detected by SIFT with and without saliency for all datasets

Dataset	Total Images	Avg:Feature points without saliency	Avg:Feature points with saliency
Caltech	1200	69.56	34.07
UIUC	250	10.33	9.23
TUDarmstadt	320	68.01	31.06
VOC2005-1	300	187.82	40.22
VOC2005-2	800	145.45	27.64

8. CONCLUSION

In experimentation process we focused on two things, Number of the objects correctly recognized and the average number of feature points or key points detected per image. From the results shown above it can be concluded that by introducing saliency in bag of feature framework, the state-of-art performance can be retained with a tremendous improvement in execution time by reducing data processing. Both SIFT and SURF produced better performance with saliency in terms of execution time and accuracy.

9. ACKNOWLEDGMENTS

We are very thankful to Dr. Rehan Ullah khan (from SUIT), for his help throughout this research work. We are also thanking to the authorities of Department of computer systems Engineering, University of Engineering and Technology Peshawar, Pakistan, for their support during this research work.

10. REFERENCES

[1] C. Schmid and R. Mohr. “Local greyvalue invariants for image retrieval.” TPAMI, 1997.

[2] D. Lowe.,“ Object recognition from local scale-invariant features”,ICCV1999.

[3] Herbert Bay , Tinne Tuytelaars, Luc Van Gool, “SURF: Speeded Up Robuts Features”, ECCV.2006 .

[4] Csurka, G., Dance, C., Fan, L., Williamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: ECCV’04 workshop on Statistical Learning in Computer Vision. (2004) 59–74

[5] Lowe, D. “Distinctive image features from scale-invariant keypoints.” IJCV 60 (2004)

[6] Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from google’s image search. In: ICCV. (2005) II: 1816–1823

[7] Leung, T., Malik, J. “Representing and recognizing the visual appearance of materials using three-dimensional textons.” IJCV 43 (2001) 29–44

[8] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, S. Belongie, "Objects in context", ICCV, 2007.

[9] A. Torralba, "Contextual priming for object detection", International Journal of Computer Vision (IJCV) 53 (2) (2003) 153–167

[10] J. Verbeek, B. Triggs, Scene segmentation with CRFs learned from partially labeled images, in: NIPS, vol. 11, 2008.

[11] L.Wolf, S. Bileschi, "A critical view of context", International Journal of Computer Vision (2006).

[12] P. Duygulu, K. Barnard, J. F. G. de Freitas and D. A. Forsyth, "Object recognition as Machine translation: Learning a Lexicon for a fixed image vocabulary" Computer Vision ECCV2002.

[13] Hall, D., Leibe, B., Schiele, B.: "Saliency of interest points under scale changes". In: BMVC. (2002)

[14] Tie Liu, "Learning to detect a salient object" Pattern Analysis and Machine Intelligence, IEEE Transactions on Feb.2011

[15] Timor Kadir and Michael Brady "Saliency, scale and image description" Jurnal of computer vision (2001)

[16] P.J. Flynn. "Saliencies and symmetries: Toward 3d object recognition from large model databases". In CVPR'92, pages 322–327, 1992

[17] B. Schiele and J. L. Crowley. "Probabilistic object recognition using multidimensional receptive field histograms". In ICPR96, Vienna, Austria, 1996.

[18] N. Sebe and M.S. Lew. "Salient points for content-based retrieval." In BMVC'01, pages 401–410, 2001.

[19] K. N. Walker, T.F. Cootes, and Chris Taylor. "Locating salient object features". In BMVC'98,pages 557–566, 1998.

[20] Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., Poggio, T., "A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex", MIT-CSAIL-TR-2005-082.

[21] Beis, J., and Lowe, D.G "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces", Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 1000–1006.

[22] K. Mikolajczyk and C. Schmid. "A Performance Evaluation of Local Descriptors". In Interna-tional Conference on Computer Vision and Pattern Recognition, volume 2, pages 257–263 jun 2003

[23] P.A. Viola and M.J. Jones. "Rapid object detection using a boosted cascade of simple features". In CVPR (1), pages 511 –518, 2001

[24] Luo Juan, Oubong Gwun, "A Comparison of SIFT, PCA-SIFT and SURF"

[25] Itti, L., Koch, C., & Niebur, E. "A model of saliency-based visual attention for rapid scene analysis". IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 1254–1259.1998.