Sindh Univ. Res. Jour. (Sci. Ser.) Vol.45 (3) 492-495 (2013)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Evaluation of SIFT and SURF Using Bag of Words Model on a Large Dataset

K. AHMAD[++], N. AHMAD, R. KHAN, J. KHAN, A.U.REHMAN, S. R. HASSNAIN

Department of Electronics Engineering, University of Engineering and Technology, Peshawar, Pakistan

Received 1st January 2013 and Revised 25th August 2013

Abstract: In this paper, the objective is image classification analysis based on the well known image descriptors, the Scale Invariant Feature Transform (SIFT) and the Speeded up Robust Features (SURF) on five online available standard datasets. For the classification framework, we adopted the visual words approach. For SIFT, we use the Lowe's implementation and for Speeded up Robust Features (SURF), the Herbert Bay's implementation is used. Extensive experimentation using five datasets shows that SURF is a better choice compared to Scale Invariant Feature Transform (SIFT).

Keywords: SIFT, SURF, Image Classification, Supervised and Un-supervised classification, Image Detector, Image Descriptor

1. INTRODUCTION

Image classification has a wide range of applications in image processing and computer vision. Two commonly used categories of image classification are supervised classification and unsupervised classification. Supervised classification uses the training data, which is not employed in the other category of image classification. In this paper we are using a supervised classification framework, called bag of words model. Bag of words is based on the order less collection of image features and does not consider spatial information which leads to simplicity both in computation and concepts. Bag of words model has shown tremendous improvements in the field of object recognition and image retrieval and has shown better performance over other competitors. Due to its simplicity and better performance over the other techniques, it has been an active area of research. Bag of feature uses SIFT presented in (Lowe, 1999) as image descriptor, in this paper we evaluate the performance of SIFT with another well know descriptor presented in (Bay *et al*., 2006) called SURF. We also try out the results by varying the values of radius around the corners for calculating SIFT vector, and radius of Harris edge detector (Harris and Stephens, 1998).

2. RELATED WORK

(Bauer *et al*., 20010) compare and evaluate different implementations of SIFT and SURF. He evaluated the performance of three implementations of SIFT including the original one by David Lowe, and two different parameter settings of SURF. He tested and evaluated their invariance on a dataset of natural outdoor scenes against

- Scale changes
- Rotation
- Image Noise
- Change in lighting conditions
- Change of view point

The algorithms were applied both on the initial image and the secondary image i.e. altered according the above given tests. The matching policy was to find the descriptor from the original image that had the smallest Euclidean distance to the secondary image's descriptor. His evaluation criterion was to calculate the total number of matches as well as the ratio of the correct and incorrect matches. Total number of images corresponds to the key points detected where as the ratio of correct and incorrect matches are concerned with the accuracy of the algorithm. The runtime efficiency was compared by measuring the processing time taken by all the algorithms. (Juan and Gwun, 2009) described three robust feature detection methods including:

- Principal Component Analysis (PCA)-SIFT
- Scale Invariant Feature Transform (SIFT)
- Speeded Up Robust Features(SURF)

They used KNN (K-Nearest Neighbor) along with Random Sample Consensus (RANSAC) to analyze the applications of these methods in recognition .KNN is used for finding the matches while RANSAC is used for rejecting those matches that are inconsistent. They compared these algorithms for illumination changes, affine transformation, scale changes, rotation and blurring. Their evaluation criterion is based on the number of correct matches. They concluded that the SIFT is stable but much slower compared to the other two algorithms. PCA-SIFT is more promising against rotation and changes of illumination source.

[++]Corresponding Author Email: kashif05_uet@yahoo.com Phone No. 091-9216590

(Lankinen *et al*., 2012) Presented another paper in which the authors compared different kind of feature detectors and descriptors, they also compared the performance of SURF and SIFT being among the best descriptors. They used two implementations of SIFT, one the original implementation ((Lowe *et al*., 1999), with another (Zhao *et al*., 2008). They proved both the algorithms to be best and reliable in terms of repeatability rate. In (Lei *et al*., 2010) authors compared three image descriptors, including Speeded up Robust Features SUFR, SIFT and Daisy descriptor. They concluded that SURF is fastest among them, SIFT shows better performance in terms of invariance, and daisy is suitable in describing non-extreme features. The authors in (Barazzett *et al*., 2010) have provided an overview of feature detection, and both descriptors SURF and SIFT and also have a comparison of these two image descriptors. They discussed two strategies of comparison of image descriptors. One is called quadratic matching and other is kd-tree procedure. (Ballesta and Gil, 2010) Represented the comparison of some well-known local image descriptors, including SIFT, SURF, Gray level patch and Zernike moments in terms of SLAM. In this paper the authors studied the performance of these image descriptors on several images in 2D and 3D. They also provided an overview of these algorithms. They used three implementations of SURF, the original one by Bay, and e-SURF and u-SURF.

3. MATERIALS AND METHODS

It is an image detecting algorithm, and is widely used in computer vision applications such as image detection, object recognition and 3D modeling. SIFT is composed of four stages: scale invariant feature detection, feature matching and indexing, cluster identification by Hough transforms voting and model verification. In first stage the image is transformed to a collection of feature vectors with the help of DoG (Serr *et al*., 2005) function. These feature vectors are invariant to scaling, distortion, rotation and illumination changes. Best-bin-first search (Beis and Lowe, 1997) is used for feature matching and indexing. Hough transform is used in the 3rd stage for cluster identification. Linear Least Square Solution is performed on the identified clusters for relating the model to image. SIFT has outstanding performance compared to other descriptors (Mikolajczyk *et al*., 2005). Its ability of mixing the local information with gradient related features makes it a better choice. Different implementation of SIFT are available. The original implementation of SIFT by Lowe is unable to handle high resolution images which was later on handled by Novozin implementation. PCA based SIFT (Sukhtankar *et al*., 2004), which uses Principal component analysis for the dimensionality reduction of high resolution images.

SURF

Surf is a robust image detector and descriptor (Bay *et al*., 2006). It is much faster than SIFT and make use of integral images (Viola and Jones, 2005). SURF detector is based on Hessian matrix measures and uses 2D Haar wavelet transform for descriptor employing only 64 dimensions which leads to fast feature computation and matching. Its dimensions can be increased to 128 however later it has been proved (Juan *et al*., 2009) that it does not add much to the speed. SURF sometimes provides with more than 10% improvements compared to other descriptors (Bay *et al*., 2006).

4. DATASET

We used around 2570 images from five online standard datasets for the evaluation of both algorithms. Mostly in literature Caltech dataset is used for the evaluation of bag of features model, however this paper presents the evaluation results of both algorithms on four other datasets along with Caltech, providing a complete set of images to evaluate an algorithm's performance. These datasets include: UIUC, TUDarmstadt, VOC2005-1 and VOC2005-2. **(Fig.1 to Fig.5)** shows sample images from the Caltech, UIUC, TUDartmstdt, VOC2005-1 and VOC2005-2 datasets respectively.

Fig. 1. Sample Image from Caltech Dataset

Fig.2. Sample Image 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

5. EXPERIMENTS AND RESULTS

The evaluation criterion used for the evaluation of both algorithms is based on the total number of correctly recognized objects. During the experimentation process, the algorithm is trained and tested on a large number of images in two phases.

First experimental phase is about analyzing the performance of bag of features Model at different values of SIFT's parameters i.e. radius around the corners for calculating the SIFT feature vector and the Harris radius. SIFT results are measured at ten different values of radius around the corners and Harris radius on Caltech dataset to find the best possible combination of its parameter's values. **(Table.1,2)** shows the SIFT performance at ten different values of radius around the corners and the Harris's radius respectively. The original value of the SIFT radius is 6 and Harris radius is 3, which are highlighted in the table.1 and table.2.

The results shows that although the variation in parameter values effect the performance of SIFT but there is no such a combination that boosts the SIFT's performance.

In the second phase, the performance of SIFT and SURF is measured on all five datasets. **(Table.3 to Table.7)** shows SURF and SIFT performance on Caltech, UIUC, Darmstadt, VOC2005-1 and VOC2005-2 datasets respectively. In Caltech and UIUC datasets we used 200 images for training and 50 images for testing form each category. From TUDarmastadt we have 80 images for training and 20 for testing.VOC2005-1 and VOC2005-2 have 160 and 60 images from training respectively. While the number of testing images for VOC2005-1 is 40 and for other one we used 15 images for testing purposes.

Table.1: SIFT Results at different values of the radius around the corners.(Caltech Dataset) number of images used for test=200, number of images used for test=50.

Value of radius	2	3	4	5	6	7	8	9	10
Bikes	43	46	48	47	49	49	46	49	50
Faces	42	48	37	36	47	39	43	46	45
Cars	41	42	43	44	45	47	47	48	46
Air planes	38	43	47	46	49	47	47	46	47

Table.2: SIFT Results at different values of Harris radius.(Caltech Dataset) Number of images used for training=200, number of images used for test=50.

Value of R	2	3	4	5	6	7	8	9	10
Bikes	47	49	50	49	46	47	48	46	48
Faces	43	47	43	45	39	44	42	40	43
Cars	45	45	46	48	44	43	46	47	40
Airplanes	46	49	49	45	45	48	45	48	47

Table.3: Evaluation of SURF and SIFT on Caltech Dataset. Number of images used for training=200, Number of images used for test=50.

Category	SURF	SIFT
Bikes	50	49
Faces	49	47
Cars	46	45
Airplane	50	49

Table.4: Evaluation Results of SURF and SIFT on UIUC Dataset. Number of images used for training=200, Number of images used for test=50.

Category	SURF	SIFT
Cars	15	13

Table.5: Evaluation of SURF and SIFT on TUDarmstadt Dataset. Number of images used for training=80, Number of images used for test=20.

Category	SURF	SIFT
Bikes	18	14
Cars	18	15
Cows	15	19

Table.6: Evaluation of SURF and SIFT on VOC2005-1 Dataset.Number of images used for training=160, number of images used for test=40.

Category	SURF	SIFT
Bikes	27	26
Persons	15	11
Cars	7	5
Bi-Cycles	15	9

Table.7: Evaluation Results of SURF and SIFT on VOC2005-2 Dataset.Number of images used for training=60, number of test images=15.

Category	SURF	SIFT
Bikes	6	8
Pedestrians	7	5
Cars	8	6
Bi-Cycles	8	3

6. CONCLUSIONS

Experimental results demonstrate that the SURF has better performance than SIFT in terms of number of correctly recognized objects. Being a fast and accurate feature point descriptor, SURF further enhances the object recognition ability of bag-of-features framework. As far as the first experiment is concerned, it can be concluded that although the performance of the bag-of-features framework vary with variation of its parameters values, however there is no such unique combination of its parameters that give better results.

.

REFERENCES:

Ballesta M., and A. Gil (2010) Oscar Martınez Mozos, Oscar Reinoso, "Local Descriptors for SLAM", Machine Vision and Application, (21), **(6):** 905-920.

Barazzetti L., M. Scaioni F. Remondino (2010) "Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation", Article first published in Photogrammetric Record (online), (25), **(132):** 356-381.

Bauer J., N. Sunderhauf P. Protzel (2010) "Comparing Several Implementations of two Recently Published Feature Detectors", In Proc. of the International Conference on Intelligent and Autonomous Systems, IAV, Toulouse, France, 143-148.

Bay H., T. Tuytelaars and L. Gool (2006) "SURF: Speeded Up Robuts Features", Computer Vision ECCV, 3951, 404-417.

Beis J. and D.Lowe (1997) "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces", IEEE Computer Socity Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1000–1006.

Harris C. and M. Stephens (1998) "A combined corner and edge detector", In: Proceedings of the Alvey Vision Conference, 147–151. Plessey Company pic.

Juan L and L.Gwun (2009) "A Comparison of SIFT, PCA-SIFT and SURF", International Journal Image Processing, (3), **(4):** 143-157.

Lankinen J., V. Kangas and J. Kamarainen (2012) "A comparision of local feature detector and descriptors for visual object categorization by intra-class repeatability and matching." International conference on Pattren Recognition (ICPR), 780-783. 11-15. Nov. 2012

Lei, Lan-Yi-Fei, and L. Hai-Tao (2010) "Comparison of local image features ", Jisuanji Yingyong / Journal of Computer Applications, (30), no. SUPPL.2, 50-53.

Mikolajczyk K., and C. Schmid (2005) "A performance evaluation of local descriptors", PAMI, **(27):** 1615–1630.

Lowe D. (1999) "Object recognition from local scale-invariant features", proceedings of 7th International Conference on Computer Vision, (2), 1150-1157. 20-27 Sep. 1999.

Serre, K. T., M. Cadieu, C. Knoblich, D. Kreiman, and T. Poggio (2005) "A Theory of Object Recognition: Computations and Circuits in the Feed forward Path of the Ventral Stream in Primate Visual Cortex", MIT-CSAIL-TR-082.

Viola P. A. and M. J. Jones (2005) "Rapid object detection using a boosted cascade of simple features", In conference of Computer Vision and Pattren Recognition, (1), 511–518.

Zhao, L. (2008) "Local Interest point extraction toolkit", retrieved on 20th january 2013 from: http://vireo.cs.cityu.edu.hk.