AN EVALUATION OF LOCAL IMAGE FEATURES FOR OBJECT CLASS RECOGNITION

Saiful Islam and Andrzej Sluzek
Centre for Computational Intelligence, School of Computer Engineering
Nanyang Technological University, Singapore 639798, Singapore

Keywords: Local Image Feature, Object Class Recognition, Nearest Neighbour Classification.

Abstract: The use of local image features (LIF) for object class recognition is becoming increasingly popular. To better understand the suitability and power of existing LIFs for object class recognition, a simple but useful method is proposed in evaluation of such features. We have compared the performance of eight frequently used LIFs by the proposed method on two popular databases. We have used *F-measure* criterion for this evaluation. It is found that the individual performance of SURF and SIFT features are better than that of the global features on ETH-80[*] database with considerably lower number of training objects. However, it may not be good enough for more challenging object class recognition problem (e.g. Caltech-101[+]). The evaluation of LIFs suggests the requirement for further investigation of more complementary LIFs.

1 INTRODUCTION

Object class recognition is one of the central issues in many practical applications. For the recognition of a partially occluded object in cluttered environment, locally defined low level image feature known as local image feature (LIF) is preferable (Lowe, 2004; Mikolajczyk and Schmid, 2001). A large number (order of hundred) of LIFs could be extracted from an image and each of them are represented as a vector by computing different kinds of descriptors of such an image patch. An up-to-date review of different methods of local feature detection and description can be found in (Li and Allinson, 2008).

Apparently a LIF captures description about the appearance of an image patch. Most of these features are primarily intended to be used for wide baseline matching. The suitability of appearance based existing LIFs for object class recognition of previously unseen object is rather vague. Despite this limitation some desperate affords of object class recognition are made using LIF (Boiman, Shechtman and Irani, 2008; Mikolajczyk, Leibe and Schiele, 2006; Stark and Schiele, 2007). However, it is generally agreed (Boiman et al, 2008; Zhang, Berg, Maire and Malik, 2006) that LIF-based nearest neighbor classification method can overcome the intra-class variation of objects in a class to some extent.

The main purpose of this work is to investigate the suitability of local image features currently available in the literature for object class recognition. In this paper, we have proposed a method in the evaluation of existing LIFs. We evaluated performance of some of the popular and frequently used LIFs. Some previous works of evaluation of LIF can be found in (Asbach, Hosten and Unger, 2008; Stark and Schiele, 2007; Zhang and Marszalek, 2006). Asbach et. al. (2008) evaluated a few LIFs for face detection. Another method of evaluation for geometric object class recognition was reported by Stark and Schiele (2007). Zhang and Marszalek, (2006) evaluated the performance of kernel based method on three LIFs. However, in this paper we have proposed a more generic method and systematically evaluated some of the existing LIFs used for different object recognition problems.

We have nominated the simple nearest neighbor based method. Eight existing LIFs are primarily evaluated on two landmark databases: ETH-80 and Caltech-101. We have used robust *F-measure* criterion for this evaluation. It is found that SURF and SIFT are the two best LIFs for object class

[*] http://tahiti.mis.informatik.tu-darmstadt.de/oldmis/Research/Projects/categorization/eth80-db.html

[+] http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Islam S. and Sluzek A. (2010).
AN EVALUATION OF LOCAL IMAGE FEATURES FOR OBJECT CLASS RECOGNITION.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 519-523
DOI: 10.5220/0002844805190523

recognition with *F-measure* of 0.89 and 0.84 respectively for ETH-80 database with nine labelled objects. The individual performance of SURF and SIFT are better than that of the global features on ETH-80 database (Leibe and Schiele, 2003) with considerably lower number of labelled (training) objects.

The outline of the paper is as follows. In Section 2 we have given a review of existing LIFs. We have listed the LIFs considered for evaluation in this paper. In section 3, we have described the proposed method for evaluation. In Section 4 we have presented experimental results of evaluation of the LIFs.

2 REVIEW OF LIFS

The first step of LIF extraction is the detection of interest regions. Many scale and affine invariant region detectors have been recently proposed. Mikolajczyk, Tuytelaars, Schmid, Zisserman, Matas, Schaffalitzky, Kadir and Gool, (2005) has systematically evaluated six detectors. These detectors are generally evaluated by repeatability and overlapping accuracy for different kinds of transformations such as viewpoint change, rotation, scale change, illumination change, and image blur etc. (Bay, Ess, Tuytelaars and Gool, 2008; Mikolajczyk et al., 2005). It was found that the Hessian-Laplace detector has higher localization and scale selection accuracy. In this work, we have used this region detector to produce all the experimental results.

There are a good number of possible descriptors which are based on different image properties like intensities, color, texture, edges etc. Many different descriptors of local image patch have been proposed in recent years. Mikolajczyk and Schmid (2005) reviewed different distribution-based (e.g. SIFT, GLOH, PCA-SIFT), differential-based (e.g. local jet, steerable filter, complex filter), and moment-based (e.g. gradient moment) descriptors and evaluated their performance. The evaluation is based on ground truth, i.e. matching of LIFs from a same scene under different viewing condition (such as viewpoint change, rotation, scale change, illumination change, image blur etc.) It was found that GLOH (Mikolajczyk and Schmid, 2005) and SIFT (Lowe, 2004) are two best performing high dimensional LIFs whereas gradient moments GM (Mikolajczyk and Schmid, 2005) and steerable filters SF (Freeman and Adelson, 1991) are two low dimensional LIFs. In the proposed evaluation work, we have considered these four LIFs together with another medium dimensional descriptor named as SURF (Bay et al., 2008) based on Haar wavelet transformation. Local shape context (SC) feature as in (Mikolajczyk and Schmid, 2005) was used by quantizing the edge points in log-polar coordinate system centered at the center of interest region. We have also implemented two invariant moment based descriptors such as CMI (color moment invariant based descriptor of dimension 18) (Mindru et al., 2004) and RMI (revised moment invariant based descriptor of dimension 15) (Reiss, 1993). In total we considered to evaluate eight LIFs. In order to extract LIFs, we used the executables available online for six features and implemented the CMI and RMI based feature extraction methods

3 EVALUATION METHOD

The evaluation of different kinds of LIFs for object class recognition is based on a simple nearest neighbor (NN) classification method. Our objective is to observer the inter-class discriminative and intra-class generalization power of LIFs but not the efficiency of the method. As suggested in (Boiman et al., 2008), we have used the non-parametric nearest neighbor classification method. In this method we do not need any explicit training phase.

Suppose, we have total c classes and for each class a set of labelled (training) images are given. Now, the problem of object class recognition is to assign a previously unseen object to a particular class. There are two phases of a supervised object classification method: training and classification. In the training phase the classifier is trained with features extracted from the labelled images. In the classification phase the extracted features of a query image are used by the classifier to decide the appropriate class.

Section 3.1 discusses the NN-based minimum risk Bayes classification method. Section 3.2 gives a brief introduction of two databases. The evaluation criterion is explained in section 3.3. Finally, in Section 3.4 the results of empirically verification of the robustness of the method for wide baseline matching is presented.

3.1 NN Classification

Suppose D_i is the set of LIFs extracted from all the labelled images of class ω_i; $i = 1\ldots c$. The basic idea of the classification method is as follows. Suppose $p(\omega_i|Q)$ is the probability that a query image Q belongs to class ω_i. With the assumption that the prior $p(\omega_i)$ is uniform, Bayes maximum a-posteriori classifier becomes:

$$\omega = \arg\max_i p(\omega_i \mid Q) = \arg\max_i p(Q \mid \omega_i) \qquad (1)$$

Suppose, we extract n LIFs $d_1, \ldots, d_n$ from the query image. With Naïve-Bayes assumption (LIFs $d_1, \ldots, d_n$ are independent given its class ω_i) it can be show that

$$p(Q \mid \omega_i) = \arg\max_i p(d_1,\ldots,d_n \mid \omega_i) = \arg\max_i \prod_{j=1}^{n} p(d_j \mid \omega_i) \qquad (2)$$

Taking the log probability the decision rule becomes:

$$\omega = \arg\max_i p(\omega_i \mid Q) = \arg\max_i \frac{1}{n}\sum_{j=1}^{n} p(d_j \mid \omega_i) \qquad (3)$$

We need to estimate the probability density $p(d_j|\omega_i)$. As the number of LIFs is large the estimation can be approximated by nonparametric nearest neighbor method (Boiman et al., 2008). In the following we redefine the NN method in terms of minimum risk classification.

For a LIF d we can find the nearest neighbor NN(d, D_i) in class ω_i. In fact the distance ||d - $NN(d, D_i)$|| between d and $NN(d, D_i)$ defines the 'feature-to-class' distance. Thus, the 'image-to-class' (δ_i) distance can be defined by:

$$\delta(Q,\omega_i) = \delta_i = \sum_{j=1}^{n} \| d_j - NN(d_j, D_i) \| \qquad (4)$$

Obviously we would prefer to minimize the 'image-to-class' distance. So taking the 'image-to-class' distance as the risk we can obtain the following minimum risk classification.

When we assign a query to a particular class there is a risk $R(\omega_i|Q)$. Bayes minimum risk classification tries to minimize the risk.

$$\omega = \arg\min_i R(\omega_i \mid Q) = \arg\min_i \delta_i; \quad i = 1,\ldots,c \qquad (5)$$

We expect that an ideal LIF minimizes image-to-class distance for the intended class and maximizes it for other classes.

3.2 Databases

There are a quiet good number of databases publicly available for object class recognition. We have selected two of them for the evaluation. The first one is ETH-80 and the other is Caltech-101. The main motivation behind the selection of these two databases is that it would be possible to carry out object class recognition experiments with different difficulty levels.

In ETH-80 database, there are eight classes of objects. In each of the classes there are ten objects. Each object is represented by 41 views spaced evenly over the upper viewing hemisphere. Figure 1 shows the classes of this database. On the other hand, there are 101 classes in Caltech-101 database with large intra-class appearance and shape variability. In our experiments we have used 85 classes having at least 40 images.

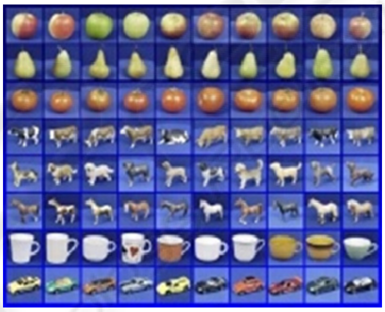

Figure 1: Classes of ETH-80 databases.

3.3 Evaluation Criterion

A robust measure is used as the criterion for evaluation of the performance of different LIFs for object class recognition. *F-measure* is based on the number of true positive (*TP*), false positive (*FP*) recognitions with respect to number of positive (*P*) examples. In a classification process we present some positive (*P*) and some negative examples of each class to the classifier and count the number of true positive (*TP*) and false positive (*FP*) classifications. The *F-measure* can be computed form TP and FP as follows:

$$F - measure = \frac{2 \times TP}{P + TP + FP} \qquad (6)$$

3.4 Empirical Verification

As the LIFs are originally designed for matching of wide baseline images, we wish to empirically verify the performance of the proposed method for recognition of object using test images which are essentially the images of all the same object with a different viewpoints. For this experiment we use the ETH-80 database. Here we set aside four images from each of 10 objects (in total 40 images) for each class as a classification set. Rest of the images is used as the labelled images. Table 1 gives the average *F-measure* for each of the class. It can be seen that on the average SURF has the highest score (0.984) and SIFT has the second highest score (0.981) of *F-measure*. For other LIFs the performances are inferior. However, the result indicates that proposed evaluation framework perform well with wide base-line classification set given an ideal LIF.

Table 1: *F-measure* for wide baseline matching for all classes of ETH-80 database.

Object	SIFT	SURF	GLOH	GM	CMI
Apple	0.97	0.99	0.93	0.91	0.79
Car	0.99	0.96	0.97	0.96	0.71
Cow	0.98	0.98	0.95	0.82	0.53
Cup	0.99	1.00	0.97	0.97	0.81
Dog	0.97	0.99	0.92	0.87	0.53
Horse	0.98	0.97	0.94	0.87	0.62
Pear	0.98	0.99	0.95	0.95	0.55
Tomato	0.98	1.00	0.96	0.88	0.93
Average	**0.981**	**0.984**	**0.95**	**0.90**	**0.68**

4 RESULTS OF EVALUATION

We have implemented the evaluation method using c++. All the LIFs extracted from the labelled images are indexed in k-d trees. For each class we use a separate k-d tree. Now, for each image of a classification set we extract LIFs. For each of the LIFs the k-d tree is used to search the nearest neighbor using Euclidian distance measure. One of the important objectives of this evaluation is to examine the effect of number of labelled (training) objects. For that we carried out experiments with different number of labelled object for each class. We have observed the performance of all the LIFs with such experiments. In the following two subsections we have discussed the experimental results for two databases.

4.1 ETH-80 Database

For ETH-80 database, we carried out experiments with different size of labelled image set. We leave one object, five objects and nine objects out for each class for three different sets of experiments. The remaining objects in each class are used as classification set for the respective experiment. For each class we get some positive examples and some negative examples. For instance, with the nine objects as labelled images, we presented 41 images as positive examples and 41×7 images as negative examples for each class. We counted the true positives and false positives for each class. We carried out multiple such tests for a particular size of labelled images and computed the average of true positives and false positives for each class. Finally we computed *F-measure* from the averages using Equations (6).

We repeated the experiments for all eight LIFs. Figure 2 shows the *F-measures* for different number of objects as labelled images. From the figure, it can be seen that generally the performance improve with increasing number of labelled objects. Individually SURF is best feature followed by SIFT and GLOH. Using SURF we get 0.90 of *F-measure*, which is better than that of any individual global features as reported in (Leibe and Schiele, 2003). Here we use considerably smaller number of labelled objects. However, it can be observed that RMI just performs randomly.

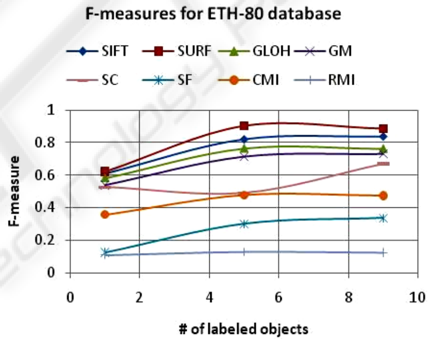

Figure 2: *F-measures* for ETH-80 database.

4.2 Caltech-101 Database

For Caltech-101 database we also carried three sets of experiments with different size (such as 10, 20 and 30) of labelled image set. We considered all 85 classes having at least 40 images each. The images for a labelled set are randomly selected form the first 40 images of the class and the remaining of the 40 images are used as classification set. As in ETH-80, we extract all LIFs of labelled image set for each class and use a separate k-d tree. In an experiment we presented some positive examples and some negative examples. We counted the number of true positives and false positives for each class. We carried out several such experiments for a particular size of labelled images and computed the average of true positives and false positives for each class. Finally we computed the *F-measure* from the averages using Equations (6).

Here we considered only best four LIFs obtained from experiments on ETH-80. Figure 3 shows the *F-measures* for all the classes. As before, generally the performance is improved with increased number of labelled objects. Individually the performance for SURF is the best followed by SIFT, GLOH, and GM as well. However, the *F-measure* drops sharply with respect to ETH-80 dataset (e.g. to 0.38 from 0.89 for experiments with 30 labelled images).

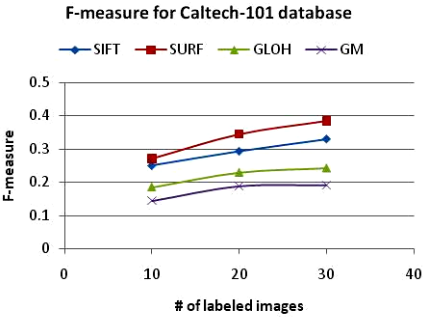

Figure 3: *F-measures* for Caltech-101 database.

5 CONCLUSIONS

In this paper, we have proposed a useful method in evaluation of existing local image features for object class recognition. The proposed method is based on a simple nearest neighbor method. In this work, eight prominent and frequently used local features are evaluated using two popular databases. We have used *F-measure* criterion to analyze the performance of the LIFs

It is found that average individual performance for SURF and SIFT are quite satisfactory (with *F-measure* of 0.89 and 0.84 respectively) on ETH-80 database. They outperform the individual performers of different global features as considered in (Leibe and Schiele, 2003). Here we used considerably lower number of labelled images. GLOH and GM features are the next best features for object class recognition. However, on Caltech-101 database this performance drops sharply (e.g. to 0.29 from 0.84 for SIFT). This may caused by different reasons. Most obvious among them is that without any quantization the feature space gets more crowded with the increase of object class and thereby the chance of misclassification increases. However, it the evident that we need to extract more complementary image features or alternatively to combine several features for better performance of object class recognition.

ACKNOWLEDGEMENTS

The research presented in this paper is a part of A*STAR Science \& Engineering Research Council grant 072 134 0052. The financial support of SERC is gratefully acknowledged.

REFERENCES

Asbach, M., Hosten, P. and Unger, M. (2008). An Evaluation of Local Features for Face Detection and Localization. *Ninth Int. Workshop on Image Analysis for Multimedia Interactive Services*.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V. (2008). Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding* 110, 346-359

Boiman, O., Shechtman, E. and Irani, M. (2008). In Defense of Nearest-Neighbor Based Image Classification. *IEEE Conf. on CVPR*, 1 - 8.

Freeman, W. and Adelson, E. (1991). The Design and Use of Steerable Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*: 13, 891-906.

Leibe, B. and Schiele, B. (2003). Analyzing Appearance and Contour Based Methods for Object Categorization. *IEEE Conf. on CVPR,* Wisconsin.

Li, J. and Allinson, N. M. (2008). A Comprehensive Review of Current Local Features for Computer Vision. Neurocomputing, 71, 1771-1787.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*: 60, 91-110.

Mikolajczyk, K., Leibe, B. and Schiele, B. (2006). Multiple Object Class Detection with a Generative Model. *IEEE Conf. on CVPR*, 26 - 36

Mikolajczyk, K. and Schmid, C. (2005). A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*: 27, 1615-1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L. V. (2005). A Comparison of Affine Region Detectors *IJCV*: 65, 43-72

Mindru, F., Tuytelaars, T., Gool, L. V. and Moons, T. (2004). Moment Invariants for Recognition under Changing Viewpoint and Illumination. *Computer Vision and Image Understanding*: 94, 3-27.

Reiss, T. H. (1993). *Recognizing Planar Objects Using Invariant Image Features*, Springer-Verlag.

Stark, M. and Schiele, B. (2007). How Good Are Local Features for Classes of Geometric Objects. *IEEE 11th International Conference on Computer Vision*, 1 - 8.

Zhang, H., Berg, A. C., Maire, M. and Malik, J. (2006). Svm-Knn: Discriminative Nearest Neighbor Classification for Visual Category Recognition. *IEEE Conf. on CVPR*, 2126- 2136.

Zhang, J. and Marszalek, M. (2006). Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. *IJCV*: 73, 213 - 238.