This article can be cited as B. Gajjar, H. Mewada and A. Patani, Parameterizing SIFT and Sparse Dictionary for SVM Based Multi-class Object Classification, International Journal of Artificial Intelligence, vol. 19, no. 1, pp. 95-108, 2021.

Parameterizing SIFT and Sparse Dictionary for SVM Based Multi-class Object Classification

Bhavinkumar Gajjar[1], **Hiren Mewada**[2] **Ashwin Patani**[3]

[1]Department of Electronics and Communication Engineering
Indus University
Rancharda, Ahmedabad-382115, India
bhavingajjar.rs@indusuni.ac.in

[2]Department of Electrical Engineering
Prince Mohammad Bin Fahd University
P.O. Box 1664, Al Khobar 31952, Saudi Arabia
hmewada@pmu.edu.sa

[3]Department of Electronics and Communication Engineering
Indus University
Rancharda, Ahmedabad-382115, India
ashwinpatani.ec@indusuni.ac.in

ABSTRACT

To build a general-purpose object recognizer, capable to recognize many different classes of objects is the most challenging task. Spatial pyramid matching (SPM), an extension and revision of bag-of-feature (BoF) computes histograms of native features at various levels of resolution. The support vector machine (SVM) using SPM gained large popularity in object classification. However, its uses are trivial due to the large computational cost and limited classification accuracy for very large classes of objects. This paper presents an extension of SPM based on tuneable SIFT sparse code using multi-class SVM to enhance the performance by encoding the SIFT features into sparse code and tuning its parameters. The various experimental studies are presented in the paper to investigate the effect of SIFT parameters to achieve better sparsity in the dictionary. During the number of experimentation, parameter tuning in SIFT feature extraction with linear-kernel SVM has improved the recognition accuracy for the datasets having a large number of classes and a huge number of images for each class. The comparison with state-of-art concludes that optimum tuning of the parameters of SIFT can minimize the feature vector size reducing the computational cost and achieving higher classification accuracy.

Keywords: sparse, classification, SVM, SIFT, SPM.

2000 Mathematics Subject Classification: 68T10, 68T45, 68Q25.

1 Introduction

Object classification is an example of the extraction of useful information from image data. Machine learning plays an eminent role in the classification of objects from real-life images. The powerful learning capability of supervised machine learning algorithms like neural network (NN) and support vector machine (SVM) has effectively improved the object classification accuracy. Recent trends in object classification using neural networks proved benchmark results over the traditional algorithms. SVM works in two stages: feature extraction and classification whereas, NN integrates these two stages into a whole. But NN is a complete black box and its computational power, amount of data, selection of proper network and feature interpretability create difficulties in faster development of the solution to any problem. Moreover, the approximation of the complex function in NN is impossible (Liu and An, 2020). Whereas traditional algorithms like decision tree, Naive Bayes, SVM, etc in object classification have an easy understanding of features to control overall performance. Another alternate to this machine learning classifiers is fuzzy classifiers resulting small model size (Precup and Tomescu, 2015). However, building optimum fuzzy rules need automated learning approach i.e. Juang and Chen used SVM to design these fuzzy rules in the classifier (Juang and Chen, 2018).
Researchers have implemented many models on the bases of the famous classical BoF model for object classification. Though it performs a remarkably over previous methods (Hofmann, 2001) (Fergus, Perona and Zisserman, 2005) for image classification but it fails to discriminate key features for image representation. To address this issue extension BoF was proposed named Spatial Pyramid Matching (SPM) (Lazebnik, Schmid and Ponce, 2006). SPM claimed many utmost challenging object classifications. Over the years SPM reformed to multiple extensions using sparse coding (Yang, Yu, Gong, Huang et al., 2009) (Oliveira, Nascimento, Vieira and Campos, 2012) (Balasubramanian, Yu and Lebanon, 2016) and pooling strategy. Scale-invariant feature transform (SIFT) produces high-dimensional features from image patches in the spatial domain. It has been proved that the SIFT descriptor with SPM outperforms any traditional methods in image classification (Yang et al., 2009). Unfortunately, the procedure involved in codebook formation and feature quantization in the SIFT descriptor has a high computational cost in the spatial domain. Yang et al (Yang et al., 2009) presented a sparse based SPM (ScSPM) to reduce the complexity in feature quantization and overlooked the codebook formulation. In this paper, we revisited the work of sparse coded SPM (ScSPM) to investigate the effect of codebook formation and SIFT key feature's orientation. The parameters involved in SIFT can be altered to check its behavior. Corresponding choice and amendment of carefully chosen values can be used to improve the performance of the algorithm. We change the SIFT codebook formulation and tuning of feature orientation in the algorithm presented by Yang et al (Yang et al., 2009) and performance for object classification is examined. Discriminative power in the large database increases with the complexity of the SIFT descriptor. We propose tuning of SIFT parameters to improve the performance of the ScSPM method. The comparison of schematic presentation of the proposed algorithm with ScSPM (Yang et al., 2009) and standard SPM approach is presented in Fig-1.
In section-2, we reported related work on sparse coding (SC) and revised the ScSPM(Yang et al., 2009). Section-3 explains the implementation of the proposed algorithm with a discus-

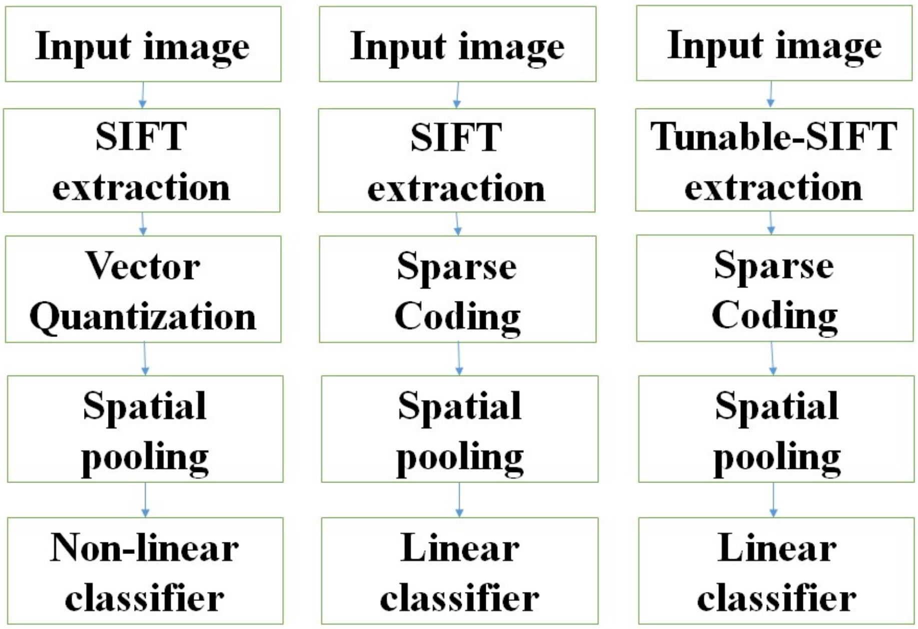

Figure 1: (a) Non-linear SPM (b) ScSPM (c) Proposed tunable SIFT ScSPM

sion on SIFT parameters. Section-4 provides experiment results and compares the obtained results with SPM approaches i.e. Kernel SPM (KSPM), Linear SPM (LSPM), and Sparse Coded SPM (SPM). At last, the conclusion of our work with some future scope for further improvement is presented in section-5.

2 Related Work

Sparse code expresses the vectors using a linear combination of a small number of weighted basis functions. Sparse coding learns compact and concise representation using a few numbers of coefficients to represent vectors. And dictionary formation using a linear combination of sparse atoms is a suitable method for unknown model (Tang, Panahi, Krim and Dai, 2019). The sparse coding and dictionary-based image classification approach have attracted many researchers. Many supervised (Aharon, Elad and Bruckstein, 2006) (Boureau, Bach, LeCun and Ponce, 2010) (Jiang, Lin and Davis, 2011) (Zhang, Berg, Maire and Malik, 2006) (Zhang and Li, 2010) and unsupervised (ShenghuaGao and Liang-TienChia, 2010) (Lazebnik et al., 2006) (Wang, Yang, Yu, Lv, Huang and Gong, 2010) (Yang et al., 2009) techniques are proposed based on sparse dictionary to solve multiclass reorganization problem. Based on this state-of-art are presented results on standard benchmark datasets i.e. Caltech-101 (L. Fei-Fei and Perona, 2004), Caltech-256 (Griffin, Holub and Perona, 2007) and Scene-15 (Ali and Zafar, 2018). Where spatial pyramid integrates local and global representation and hence it is very efficient in scene recognition with large datasets.
The vector quantization to sparse code with max-pooling was presented in (Yang et al., 2009). This approach exceptionally reduces the computation complexity of SVM to O(n) from O(n^2). In this technique, a linear SPM kernel based on SIFT sparse codes were used. To speed up

the process and to remove the constraint of dictionary learning Hu et al (Hu and Guo, 2012) proposed a new descriptor in SPM for object classification. This descriptor combines local binary patterns (LBP) and three-patch local binary pattern (TBLBP) in SPM. Oliveira et al (Oliveira et al., 2012) proposed a new off-line method entitled orthogonal class learning (OCL) derived on support vector decomposition (SVD) for high dimensional feature space. Based on spatial constraint on the coding stage this method named spatial sparse coding (SSC) and it obtained better accuracy in comparison with accuracy stated in (Yang et al., 2009). In (Seidenari, Serra, Bagdanov and Del Bimbo, 2013) multi-resolution pyramids in SIFT (P-SIFT) feature space were introduced for matching multilevel detail locally during learning and recognition stages. This P-SIFT showed promising results in streamline work.
In (Balasubramanian et al., 2016), authors proposed two techniques called kernel smoothing and marginal regression for SC where non-parametric kernel smoothing presents feature similarity or temporal information in dataset more flexible. Feature space partitioning for analogous category images into clusters of visual prototypes was proposed by Alameen et al(Najjar, Ogawa and Haseyama, 2015). In which Bregman co-clustering applied offline on training data. That method also achieved good results on standard datasets but the computational cost was higher because of the high dimension of feature, dictionary, and pre-processing data to generate clusters. In all these studies, authors have not reported the effect of SIFT parameters in their algorithms. Table 1 list the parameters controlling the SIFT features. Major experiments in the literature use default values without tuning them according to the task.

Table 1: Parameters involved in SIFT features extraction process

Parameter	**Description**
Octaves	No of Gaussian function producing set of scale-space image
Sigma	Amount of blur in an image
Scale	Down-sample factor of an image
Orientation of histogram bins	Number of bins in histogram to assign orientation
Orientation radius	Radius of the region in orientation assignment
Feature vector	Dimension of the feature vector

In the section, we presented the empirical study on SIFT parameters on sparse based dictionary approach for image classification as suggested by Yang et al (Yang et al., 2009). Effect of other parameters like the size of a dictionary, pooling method, and Kernels affecting the accuracy of classification algorithms was discussed in (Yang et al., 2009).

3 SIFT features analysis and Proposed Implementation

For the given dictionary V best coefficient U for signal X can be found using sparse coding. This paper adopts this encoding of SIFT features into sparse code and investigates the performance by tuning the parameters. In the proposed algorithm, the conversion of the SIFT feature's vector quantization to the sparse code is solved using the following equation:

$$\underset{U,V}{min} \sum_{m=1}^{M} \left\| x_m - u_m V^2 \right\|^2 + \lambda \left| u_m \right| \tag{3.1}$$

subject to $\|v_k\| \leq 1, \forall k = 1, 2, ...K$. Where V is NxK size over-complete (K>N) dictionary and u_m is sparse coefficient matrix for signal x_m.

In this unit L2-norm on V and L1-norm on u_m is typically applied with regularization parameter λ. The problem in equation-(3.2) is convex in V and U simultaneously. This can be solved for the fixed number of iteration to achieve optimization over V or U while fixing any other. By fixing codebook V equation-(3.3) can be solved as a linear regression problem with L1-norm regularization on sparse coefficients.

$$\underset{u_m}{min} \left\| x_m - u_m V \right\|_2^2 + \lambda \left| u_m \right| \tag{3.2}$$

And fixing U, the same problem will be transformed to least square with quadratic constraints:

$$\underset{V}{min} \left\| X - UV \right\|_F^2 \tag{3.3}$$

$s.\ t\ \|v_k\| \leq 1, \forall k = 1, 2, ...K$. Lagrange dual (Lee, Battle, Raina and Ng, 2007) can cleanly deal with it. The success of the object classification relies on the highly descriptive capacity of the dictionary or sparse representation. Therefore, the first phase of the experiment involves a study on dictionary formation. The second phase involves a study on the effect of SIFT parameters on image classification accuracy. These two phases are explained as follows:

Phase 1: Effective sparsity generation in the dictionary is an important aspect of the image classification algorithm. The ratio of the number of zero elements to the number of non-zero elements referred to as sparsity of the dictionary and proportion of the number of non-zero elements to total elements referred as average coefficients generated by the sparse dictionary. The lower the average coefficients provide better sparsity (Cai, Wang and Zhang, 2014) and consecutively low arithmetic operation and less memory requirement. The studies on the selection of the dictionary formation methods were presented in (Patel and Mewada, 2018). The comparative analysis of the sparse dictionary formation using wavelet transform, discrete cosine transform, and Kronecker delta was presented. Wavelet features found better in comparison with other features. However, to achieve more sparsity, better localization is required and therefore SIFT becomes one of the suitable choices amongst all. However, a large number of SIFT features are the problem for multi-class scene classification in terms of scalability and speed. In (Liu, Yu, Chen, Li and Fan, 2019), the removal of redundant features using the SIFT features' selection algorithm was presented. In the proposed approach, a correct tuning of the parameters is presented providing a balance between the speed and accuracy. Designing a comprehensive model for inferring a discriminative visual dictionary is the key to an efficient image retrieval framework(Arun and Govindan, 2015). In the proposed experiment, 200000 patches are used randomly to train the k-SVD and to create a sparse dictionary. The experiments are performed on the standard datasets containing a large number of classes i.e. Caltech-256, Caltech-101, and Scene-15. Experiments are repeated with 30 iterations to find the appropriate selection of the number of iteration and size of the dictionary achieving maximum sparsity. In this experiment we fixed Dictionary dimension (NxK) to some standard dimensions 256, 512, 1024 and 2048 which are used in other literature(Yang et al., 2009) (Zhang, Wang, Shi, Gong, Xia and Zhanga, 2018) (Huang, Mu and Zeng, 2016). Figure-2, Figure-3, and Figure-4 shows the average coefficient value achieved in each iteration for three datasets. We can analyze that for the first ten to twelve iterations it converges fast then there is a marginal decrement in the average number of coefficients. This suggests 30 iterations are sufficient to achieve the maximum sparsity in the dictionary.

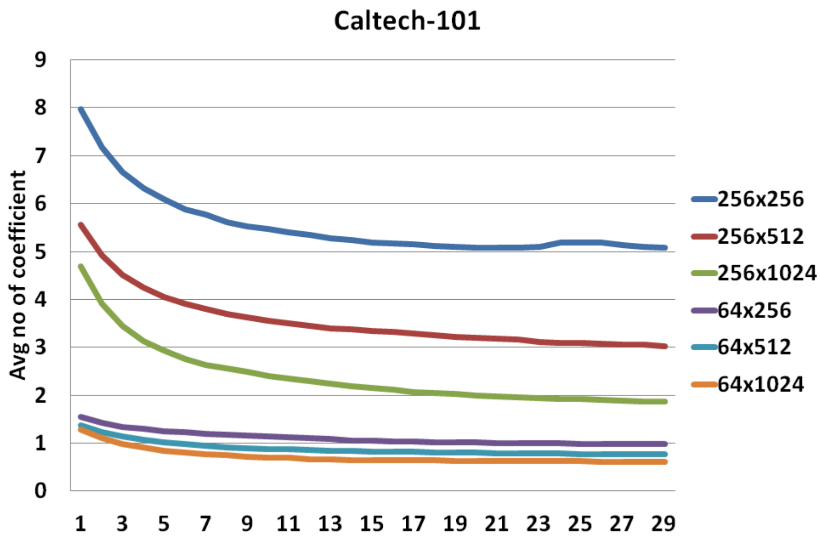

Figure 2: Average coefficient value for each KSVD iteration on Caltech-101 dataset

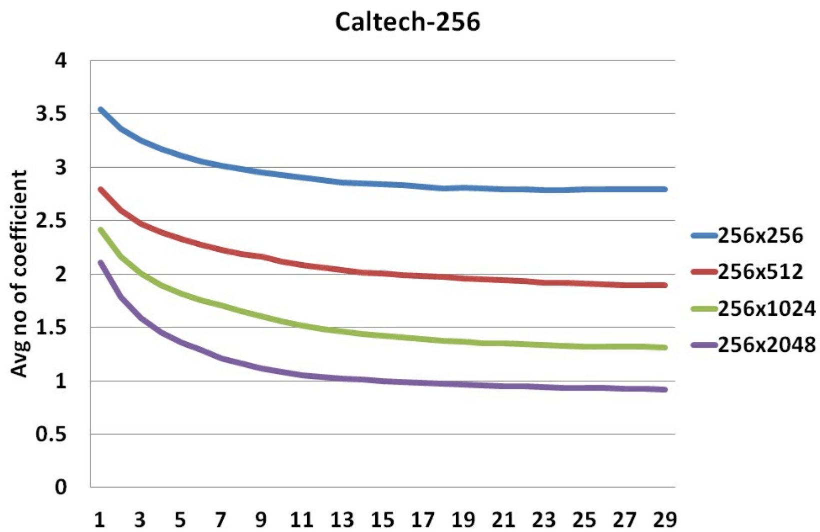

Figure 3: Average coefficient value for each KSVD iteration on Caltech-256 dataset

Phase 2: The average number of coefficients has been achieved in phase 1 after 30 iterations during the generation of a sparse dictionary. This phase investigates the effect of these dictionary sizes on overall classification accuracy for the given dataset. Multiclass classification usually disintegrates into a group of a binary 1,-1 problem that can easily accommodate

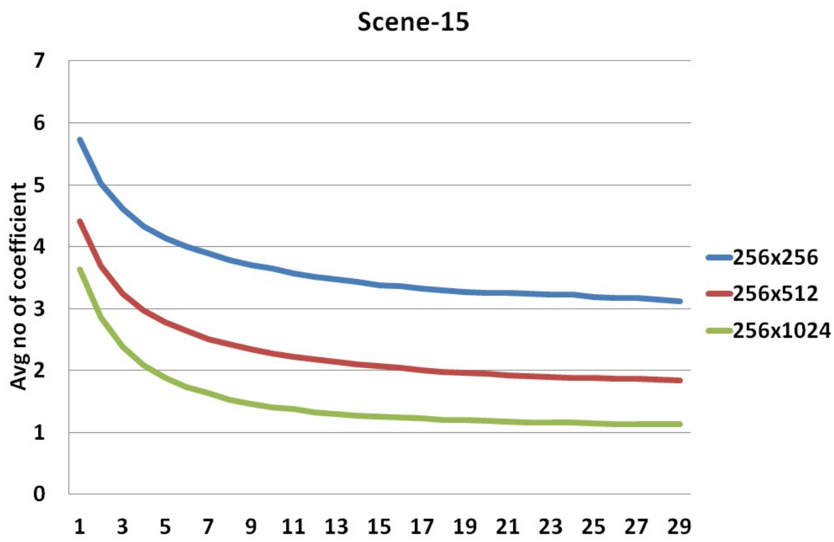

Figure 4: Average coefficient value for each KSVD iteration on Scene-15 dataset

the functionality of standard SVM. There are two famous approaches one-versus-rest and one-versus-one. In this experiment we have implemented the SVM classifier as proposed in (Aharon et al., 2006) to solve the following unconstraint convex optimization problem.

$$\underset{w_c}{min}\left\{J(w_c) = \|w_c\|^2 + C\sum_{i=1}^{n} l(w_c; y_i^c, z_i))\right\} \tag{3.4}$$

Where y_i^c = 1 if y_i = c else y_i^c = -1 and hinge loss $l(w_c; y_i^c, z_i)$ utilize differentiable quadratic form as shown below equation (3.5)

$$HingLoss = \left[max(0, w_c^T z \cdot y_i^c - 1)\right]^2 \tag{3.5}$$

According to the equation-(3.3), Dimension of feature vector U and Dictionary V must match for matrix multiplication. SIFT feature dimension is proportional to its number of bins and angles as shown in Fig-5 and the discrimination between the images depends on this feature vector (Lowe, 2004). The dictionary formation with variation in size using the different combinations

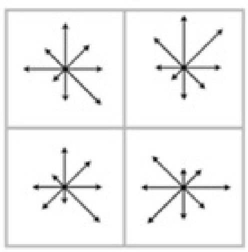

Figure 5: 2x2 SIFT descriptor with 8 orientations

of orientation and number of bins in each orientation is presented in Table 2.
Using these different sizes of a sparse dictionary, the investigation of the classification accuracy for each database is carried out in the experiment. The sample results for CalTech-101 dataset

Table 2: SIFT descriptor and Dictionary size

Angles(a)	**Bins**(b)	**SIFT descriptor size**(ab^2)	**Dictionary size**
1	8	64	64x256,
4	4	64	64x512,
16	2	64	64x1024
1	16	256	
4	8	256	256x1024
16	4	256	
64	2	256	

is presented in Fig-6 and Fig-7.

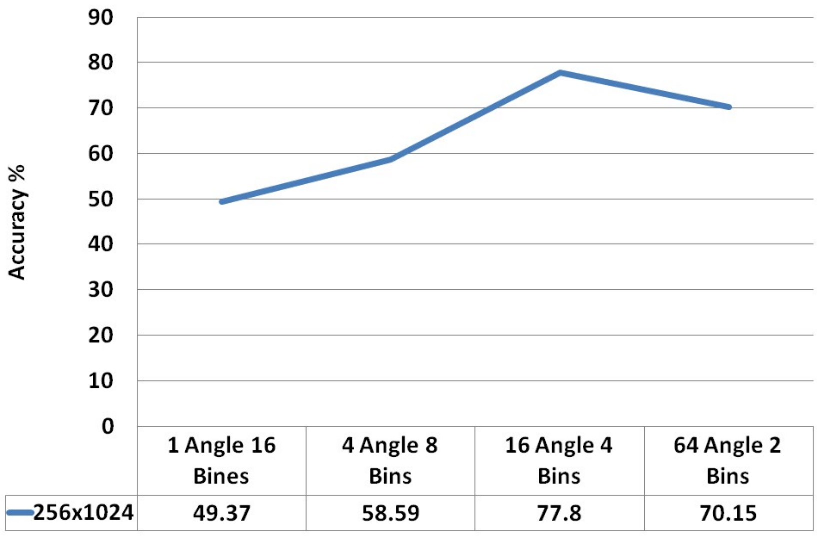

Figure 6: Results for 256 SIFT descriptor size on Caltech-101

It has been analyzed that lower angle and higher bins or vice versa fail to bind quality local details and hence accuracy for classification is lower. Hence, the classification accuracy is less for such a combination. Based on this we fixed our descriptor size 256 with 16 orientations and the width is 4 throughout the remaining experiments.

4 Experiments and Results

This experiment was conducted to find the finely tuned parameters of ScSPM. In this experiment, we used three benchmark datasets named Caltech-101, Caltech-256, and Scene-15. This experiment was focused on ScSPM (Yang et al., 2009) work and exclude the comparisons of other types of SPM(KSPM and LSPM) discussion in detail by referencing their results only for comparison. ScSPM uses linear kernel on spatial-pyramid pooling for sparsified SIFT features. By detailing the work for ScSPM(Yang et al., 2009) we selected three possible tuning

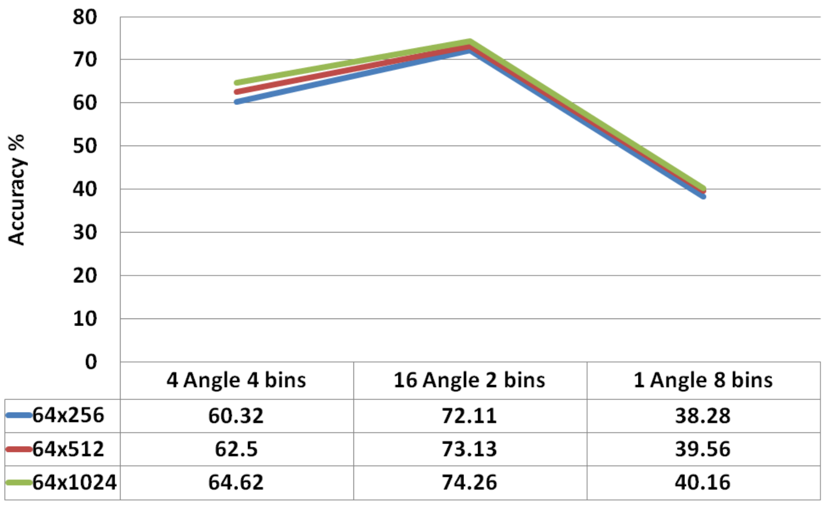

Figure 7: Results for 64 SIFT descriptor size on Caltech-101

Table 3: Results on Caltech-101 dataset

Algorithm	**Dictionary Size**	**15 training**	**100 training**
Zhang et al(Zhang et al., 2006)		59.10±0.60	66.20±0.50
NBNN(Boiman, Shechtman and Irani, 2008)		65.00±1.14	70.40
ML+CORR (Jain, Kulis and Grauman, 2008)		61.00	69.60
KC (Van Gemert, Geusebroek, Veenman and Smeulders, 2008)		-	64.14±1.18
KSPM(Lazebnik et al., 2006)		56.40	64.40±0.80
KSPM(Yang et al., 2009)		56.44±0.78	63.99±0.88
SIFT-WCS-LTP (Huang et al., 2016)	1024	65.59±0.005	73.05±0.010
PD-KSVD (Wang, Tu and Chiang, 2019)	6120		67.02
DASDLp (Yang, Chang and Luo, 2017)	1024		75.54
N^3SC (Zhang et al., 2018)	2048	67.45±0.97	73.87±1.06
Zhang et al(Zhang and Lu, 2018)	2048+1024		73.05±0.010
LSPM(Yang et al., 2009)	512	53.23±0.65	58.81±1.51
ScSPM(Yang et al., 2009)	1024	67.00±0.45	73.20±0.54
Proposed method	256	66.94±0.54	74.12±0.41
	512	68.32	76.12±0.57
	1024	70.55±0.40	**77.08±0.31**

areas that need to be revisited for accuracy enhancement were patch size reference to dictionary size, the number of training and testing samples and SVM itself after literature survey.

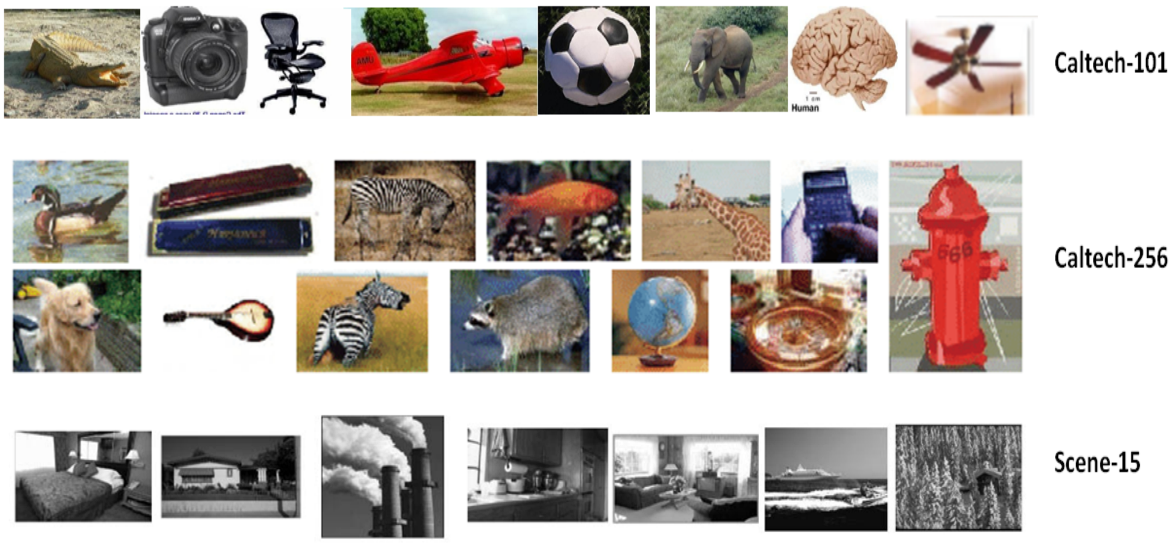

Figure 8: Example Images of datasets used in this experiment.

As discussed in section-II we were interested in parameter tuning in this experimental so we fixed the multiclass-SVM algorithm. Here dictionary size referenced to patch size was very important to match dimensions when sparse coding is used[equation (3.2)]. In this experiment, we used the same size of patches used in reference work 16x16 but dictionary sizes were 256x256, 256x512, and 256x1024 used for experimenting their effects on overall accuracy change. Each dictionary was trained for 30 iterations for KSVD.

Each sparsified SIFT features sets for Caltech-101 and Caltech-256 were divided for 30 and 15 training samples and Scene-15 for 100 and 150 training samples. Training was done with multiclass SVM for the linear kernel. We tested the performance for 5 independent runs and noted average accuracy achieved after 5 runs.

Caltech-101: This dataset contains 9145 images for 101 different classes in a variety of object categories like animals, instruments, vehicles, flowers, plants, etc. As shown in Table 3 ScSPM has accuracy higher than its baseline work SPM. This experiment shows that by selecting a proper number of bins and dictionary size accuracy can also be increased by $\sim$5 percent. We noted that even with the small size of dictionary 256 atoms classification accuracy with 15 training samples is comparable to reference work of ScSPM with 1024 atom and 30 training samples as shown in Table 3.

Caltech-256: This dataset contains 30000 images for a wide range of object categories in 256 classes. Each class has a higher intra-class covariance than Caltech-101. As discussed for Caltech-101 results were quite better than KSPM, LSPM, and ScSPM for the small size of the dictionary and less number of training samples of Caltech-256 dataset shown in Table 4.

Scene-15: This dataset contains main categories of indoor and outdoor scenes like the kitchen, living room, offices, etc. Though the number of classes is less bus it has low interclass covariance which makes it difficult to achieve high accuracy. A total of 4000 images are available in 15 classes ranging from 200 to 400. Again the same improvement was achieved with a small size of the dictionary and 100 training samples from each class shown in Table 5.

Table 4: Results on Caltech-256 dataset

Algorithm	Dictionary Size	15 training	100 training
KSPM(Lazebnik et al., 2006)		-	34.10
KC(Van Gemert et al., 2008)		-	27.17±0.46
KSPM(Yang et al., 2009)		23.34±0.42	29.51±0.52
SIFT-WCS-LTP (Huang et al., 2016)	256	26.75±0.003	
PD-KSVD (Wang et al., 2019)	15420		31.03
N^3SC (Zhang et al., 2018)	2048		34.09±0.75
LSPM(Yang et al., 2009)	512	13.20±0.62	15.45±0.37
ScSPM(Yang et al., 2009)	1024	27.73±0.51	34.02±0.35
	256	25.91±0.42	-
Proposed	512	-	27.38±0.24
method	1024	29.05±0.52	**34.23±0.31**
	2048	28.02	31±0.29

Table 5: Results on Scene-15 dataset

Algorithm	Dictionary Size	100 training
KC(Van Gemert et al., 2008)		76.67±0.39
KSPM(Yang et al., 2009)		76.73±0.65
LSPM(Yang et al., 2009)	512	65.32±1.02
ScSPM(Yang et al., 2009)	1024	80.28±0.93
Proposed	256	76.64
method	512	78.13
	1024	**81.13±0.53**

5 Conclusion and Future Work

This paper presents the investigation of SIFT features and dictionary size for object/scene classification accuracy. The large sparsity in the formed dictionary requires less memory to store the coefficients and fewer arithmetic operations. This leads to a reduction in computational cost. The size of the dictionary and its sparsity is depended on the selection of the SIFT parameters. Therefore, the effect of orientation and orientation bins describing feature vector size and the sparsity of the dictionary is presented in the paper. After getting maximum sparsity of the dictionary, the effect of dictionary sizes on overall classification accuracy is presented. The study concludes that 30 iterations are sufficient to achieve the maximum sparsity in the dictionary by reducing the average number of coefficients. Further investigation concludes that the classification accuracy will be less for the low value of either orientation or orientation bins in histogram formation. Therefore, the appropriate selection of these two parameters leads to boost the performance as presented in section 4. This empirical study was carried out using linear kernel SVM. In the future, the fusion of features (Wang et al., 2019), evolutionary optimization of regularization parameter λ in equation-3.1 using heuristic algorithms (Téllez, Jimeno, Salazar and Nino-Ruiz, 2018) (Nino-Ruiz, Ardila and Capacho, 2018) (Precup, David, Petriu, Szedlak-Stinean and Bojan-Dragos, 2016) or integration of multiple kernels in SVM (Tang, Chen, Zhao, Huang and Luo, 2019) using concluding remarks of the paper can be used to boost the accuracy further.

References

Aharon, M., Elad, M. and Bruckstein, A. 2006. ”k-svd: An algorithm for designing overcomplete dictionaries for sparse representation”, *IEEE Transactions on signal processing* **54**: 4311–4322.

Ali, N. and Zafar, B. 2018. 15-scene image dataset.

Arun, K. and Govindan, V. 2015. Optimizing visual dictionaries for effective image retrieval, *International Journal of Multimedia Information Retrieval* **4**(3): 165–185.

Balasubramanian, K., Yu, K. and Lebanon, G. 2016. Smooth sparse coding via marginal regression for learning sparse representations, *Artificial Intelligence* **238**: 83–95.

Boiman, O., Shechtman, E. and Irani, M. 2008. In defense of nearest-neighbor based image classification, *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1–8.

Boureau, Y.-L., Bach, F., LeCun, Y. and Ponce, J. 2010. Learning mid-level features for recognition, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Citeseer, pp. 2559–2566.

Cai, P., Wang, G. and Zhang, H. 2014. K-svd with reference: an initialization method for dictionary learning, *Neural computing and applications* **25**: 1263–1274.

Fergus, R., Perona, P. and Zisserman, A. 2005. A sparse object category model for efficient learning and exhaustive recognition, *CVPR (1)*, pp. 380–387.

Griffin, G., Holub, A. and Perona, P. 2007. Caltech-256 object category dataset.

Hofmann, T. 2001. Unsupervised learning by probabilistic latent semantic analysis, *Machine learning* **42**(1-2): 177–196.

Hu, J. and Guo, P. 2012. Combined descriptors in spatial pyramid domain for image classification, *arXiv preprint arXiv:1210.0386* .

Huang, M., Mu, Z. and Zeng, H. 2016. Efficient image classification via sparse coding spatial pyramid matching representation of sift-wcs-ltp feature, *IET image processing* **10**: 61–67.

Jain, P., Kulis, B. and Grauman, K. 2008. Fast image search for learned metrics, *2008 IEEE Conference on computer vision and pattern recognition*, IEEE, pp. 1–8.

Jiang, Z., Lin, Z. and Davis, L. S. 2011. Learning a discriminative dictionary for sparse coding via label consistent k-svd, *CVPR 2011*, IEEE, pp. 1697–1704.

Juang, C.-F. and Chen, G.-C. 2018. Fuzzy classifiers learned through svms with application to specific object detection and shape extraction using an rgb-d camera, *Computational Intelligence for Pattern Recognition*, Springer, pp. 253–274.

L. Fei-Fei, R. F. and Perona, P. 2004. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories.

Lazebnik, S., Schmid, C. and Ponce, J. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2, IEEE, pp. 2169–2178.

Lee, H., Battle, A., Raina, R. and Ng, A. Y. 2007. Efficient sparse coding algorithms, *Advances in neural information processing systems*, pp. 801–808.

Liu, J.-e. and An, F.-P. 2020. Image classification algorithm based on deep learning-kernel function, *Scientific Programming* **2020**.

Liu, Y., Yu, D., Chen, X., Li, Z. and Fan, J. 2019. Top-sift: the selected sift descriptor based on dictionary learning, *The Visual Computer* **35**(5): 667–677.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints, *International journal of computer vision* **60**(2): 91–110.

Najjar, A., Ogawa, T. and Haseyama, M. 2015. Bregman pooling: feature-space local pooling for image classification, *International Journal of Multimedia Information Retrieval* **4**(4): 247–259.

Nino-Ruiz, E. D., Ardila, C. and Capacho, R. 2018. Local search methods for the solution of implicit inverse problems, *Soft Computing* **22**(14): 4819–4832.

Oliveira, G. L., Nascimento, E. R., Vieira, A. W. and Campos, M. F. 2012. Sparse spatial coding: A novel approach for efficient and accurate object recognition, *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 2592–2598.

Patel, H. and Mewada, H. 2018. Dictionary properties for sparse representation: Implementation and analysis, *Journal of Artificial Intelligence* **11**: 1–8.

Precup, R.-E., David, R.-C., Petriu, E. M., Szedlak-Stinean, A.-I. and Bojan-Dragos, C.-A. 2016. Grey wolf optimizer-based approach to the tuning of pi-fuzzy controllers with a reduced process parametric sensitivity, *IFAC-PapersOnLine* **49**(5): 55–60.

Precup, R.-E. and Tomescu, M. L. 2015. Stable fuzzy logic control of a general class of chaotic systems, *Neural Computing and Applications* **26**(3): 541–550.

Seidenari, L., Serra, G., Bagdanov, A. D. and Del Bimbo, A. 2013. Local pyramidal descriptors for image recognition, *IEEE transactions on pattern analysis and machine intelligence* **36**(5): 1033–1040.

ShenghuaGao, I.-H. and Liang-TienChia, P. 2010. Local features are not lonely–laplacian sparse coding for image classification.

Tang, T., Chen, S., Zhao, M., Huang, W. and Luo, J. 2019. Very large-scale data classification based on k-means clustering and multi-kernel svm, *Soft Computing* **23**(11): 3793–3801.

Tang, W., Panahi, A., Krim, H. and Dai, L. 2019. Analysis dictionary learning based classification: Structure for robustness, *IEEE Transactions on Image Processing* **28**: 6035–6046.

Téllez, N., Jimeno, M., Salazar, A. and Nino-Ruiz, E. 2018. A tabu search method for load balancing in fog computing, *Int. J. Artif. Intell* **16**(2).

Van Gemert, J. C., Geusebroek, J.-M., Veenman, C. J. and Smeulders, A. W. 2008. Kernel codebooks for scene categorization, *European conference on computer vision*, Springer, pp. 696–709.

Wang, H.-H., Tu, C.-W. and Chiang, C.-K. 2019. Sparse representation for image classification via paired dictionary learning, *Multimedia Tools and Applications* **78**: 16945–16963.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y. 2010. Locality-constrained linear coding for image classification, *2010 IEEE computer society conference on computer vision and pattern recognition*, Citeseer, pp. 3360–3367.

Yang, J., Yu, K., Gong, Y., Huang, T. S. et al. 2009. Linear spatial pyramid matching using sparse coding for image classification., *CVPR*, Vol. 1, p. 6.

Yang, M., Chang, H. and Luo, W. 2017. Discriminative analysis-synthesis dictionary learning for image classification, *Neurocomputing* **219**: 404–411.

Zhang, H., Berg, A. C., Maire, M. and Malik, J. 2006. Svm-knn: Discriminative nearest neighbor classification for visual category recognition, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2, IEEE, pp. 2126–2136.

Zhang, H. and Lu, Y. 2018. Image classification by combining local and mid-level features, *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence*, ACM, pp. 101–105.

Zhang, Q. and Li, B. 2010. Discriminative k-svd for dictionary learning in face recognition, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 2691–2698.

Zhang, S., Wang, J., Shi, W., Gong, Y., Xia, Y. and Zhanga, Y. 2018. Normalized non-negative sparse encoder for fast image representation, *IEEE Transactions on Circuits and Systems for Video Technology* .