Journal of ELECTRICAL ENGINEERING, VOL 72(2021), NO6, 374–380

sciendo

PAPERS

Sparse coded spatial pyramid matching and multi-kernel integrated SVM for non-linear scene classification

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

Support vector machine (SVM) techniques and deep learning have been prevalent in object classification for many years. However, deep learning is computation-intensive and can require a long training time. SVM is significantly faster than Convolution Neural Network (CNN). However, the SVM has limited its applications in the mid-size dataset as it requires proper tuning. Recently the parameterization of multiple kernels has shown greater flexibility in the characterization of the dataset. Therefore, this paper proposes a sparse coded multi-scale approach to reduce training complexity and tuning of SVM using a non-linear fusion of kernels for large class natural scene classification. The optimum features are obtained by parameterizing the dictionary, Scale Invariant Feature Transform (SIFT) parameters, and fusion of multiple kernels. Experiments were conducted on a large dataset to examine the multi-kernel space capability to find distinct features for better classification. The proposed approach founds to be promising than the linear multi-kernel SVM approaches achieving 91.12 % maximum accuracy.

K e y w o r d s: multiple kernel learning, support vector machine, classification, SIFT, spatial pyramid matching (SPM)

1 Introduction

The use of the kernel in the machine learning approaches is playing a significant role in various computer vision applications. The kernel is playing a significant role in various classification algorithms including SVM, regression algorithms, kernel-based principle component analysis, convolution neural network, *etc.* The CNNs are helpful for high-dimensional data and are increasingly used in image classification and pattern recognition applications. Deep CNNs require a lot of processing power and take a long time to train, making it challenging to extensively examine, repeat, and refine their accuracy. Even though CNN has gained attraction in the classification problems, it is more susceptible to trap in local minima and over-fitting is the major concerned in contrast to support vector machine. In addition, the support of wide varieties of the kernel in SVM can assist the classification of the data more accurately for the dataset where the class labels and their features are related nonlinearly. Therefore, this paper proposes the use of SVM in multiclass scene classification applications.

For multiclass classification problems, it is difficult to claim accuracy using a single kernel because there is large interclass relation may present, and its limited feature map of two similar type categories makes it separation difficult. To address this issue multi-kernel approach is necessary for better bifurcation of features derived from objects. Proper kernel formation is complex for the multiclass problem. The goal of the prediction task is to identify a subset of relevant features optimally, and it is a base for multiple kernel learning (MKL) [1]. However, MKL is computationally very expensive so it limits its use in the huge size of the feature set. In our study, we focus on sparse coded spatial pyramid matching (ScSPM) features to be used for multi-kernel SVM. Types of MKL and its performance are not a scope of this study. We focus on the accuracy enhancement of multiclass SVM using the simple MKL approach.

This paper studies and compares overall performance with baseline [26] work and a few famous SVM-based MKL methods. As there is no universal solution that states that a particular multi-kernel works well with all multiclass scene data [2], we proposed our kernels based on simple MKL to combine standard SVM's kernels Gaussian and polynomial. Differentiation between classes can be more accurate based on finding the best combinations of kernel functions using multiple kernel learning (MKL).

We report on related work on classification and multiple kernel learning, and explain the implementation of the proposed algorithm with a discussion on the ScSPM feature. After providing experimental results and comparing the obtained results with sparse coded SPM and other MKL approaches, we conclude with future scope for further improvement.

[1] Department of Electronics and Communication Engineering, Indus University, Rancharda, Ahmedabad, 382115, India, [2] Department of Electrical Engineering, Prince Mohammad Bin Fahd University, PO Box 1664, Al Khobar 31952, Saudi Arabia, * corresponding author bhavingajjar.rs@indusuni.ac.in

https://doi.org/10.2478/jee-2021-0053, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X

2 Related work

Vision classification and recognition have gained more importance in the past few years. It consists of three components: point of interest detection, description of the region of interest, and classification. Description of the region includes the extraction of various features discriminating various scenes like local binary pattern, histogram of gradient visual words, *etc.* These robust and powerful feature sets are used in the categorization of the scene. Further improvement in the classification can be achieved using either multiple feature sets or multiple kernels in SVM. Scene classification plays an essential role in urban planning, land management, environment monitoring, and exploration, and object classification. This section presents the study of single and multi-kernel SVM for multiclass classification applications.

The combination of multiple descriptors using multiple kernel SVM was proposed in [3] and showed remarkable improvement in varied scene classification. The multi-label least-squares SVM method was proposed in [4]. They used multi-kernel RBF-based SVM for multi-label scene classification problem. The algorithm was validated on a mixture of four datasets achieving maximum accuracy of 85 %. Effect of kernel in SVM is verified by Kancherla *et al* [5]. They used different feature set with various linear kernel SVM and simulated the algorithm with a 3 to 4 class dataset. They found that the RBF kernel provides a better classification rate of 82.06 % on the MIT dataset than other kernels.

A comparative analysis of SVM with the decision tree and k-nearest neighbors (KNN) algorithm for satellite scene classification is presented [6]. Speeded up Robust features (SURF) and bag of visual words (BOVW) models were used as image futures in SVM in classification. SVM-based scene classification model for robotic application was presented in [7]. The robotic application requires fast execution. Therefore, heuristic metric-based key points were identified from the captured scene and used in the SVM model. They conclude that the integration of local binary pattern and SURF features with SVM received better accuracy in comparison with VGG based neural network model. A combined model of SVM and DNN is presented for acoustic scene classification [8]. A discordancy using scene utterance was created for the SVM model. Dimensionality is reduced using support vector decomposition. Their hybrid DNN model achieved a 66.1 % classification rate. Nazir *et al* [9] extracted features using the ResNet network, and these features were used to classify the action from the videos. They used the RBF-based multi-kernel SVM with L2 regularization function achieving 70 % classification accuracy. A hybrid approach of spatial, spectral, and semantic features was proposed in [10] to classify the hyperspectral images. Gabor-based structural features are integrated with morphological-based spatial features and K-means and entropy-based semantic features. Later a composite kernel is created in corresponds to these three features in SVM-based classification achieving an average accuracy of 98 %.

A multi-label scene classification using multi-instance learning was proposed in [11], where images are categorized into several graphs using muti-instance learning. Then RBF kernel with variation in its parameters was associated with each graph, and a composite kernel was created for SVM-based classification. They obtained a 61 % F1 score on the scene dataset. Further comparison between the SVM and CNN network was established by Hasan *et al* [12]. In the SVM approach, they used a combination of linear and RBF kernels to classify the features from the hyperspectral images. The features were optimized using principal component analysis. They presented that SVM using PCA optimization outperform the CNN network with 98.84 % accuracy, and CNN obtained 94 % accuracy. A multi-kernel SVM based on a fuzzy algorithm was proposed [13]. The histogram of oriented gradients (HOG) features are extracted from the hand digit images and a multi-kernel SVM was proposed using the fuzzy triangle membership function. Patel and Mewada [14] analyze various large multi-class scene classification algorithms and concluded that the demerit of over-fitting and poor convergence problem of CNN algorithm causes limitation in scene classification. In contrast, the geometrical interpretation of the features in SVM outperforms the NN for a dataset with large classes. The fusion of features like dense SIFT, color SIFT, and structure similarity was used along with localized multi-kernel SVM for real-world scene classification [15]. They achieved 81.92% and 42.12% accuracy for CalTech101 and Caltech256 dataset respectively. A sparse dictionary approach was presented in [14] to reduce the features and use robust features in SVM. Initially, a rigorous study on various dictionaries, *ie* DCT, DWT, K-SVD, was presented, and the effect of patches in the features was analyzed. Later a reduced feature-based multi-kernel SVM was proposed to classify scenes.

In summary, the multi-kernel SVM has played an essential role in many recognition and classification applications. The study proposed that even though multi-kernel outperformed the latest CNN approaches to classify scenes amongst a large number of categories, a further improvement is required to reduce the miss-classification rate for the databases containing large number classes. Moreover, this can be achieved if robust features were used by reducing redundancy and designing an SVM kernel with optimum parameters that correspond to these feature sets.

3 ScSPM features and MKL implementation

A complete workflow of our algorithm shown in Fig. 1. We used pre-trained KSVD dictionary for sparsifying features. 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

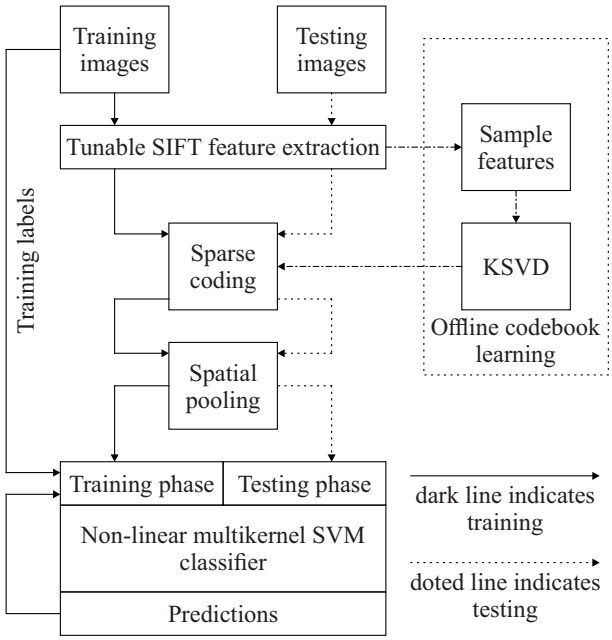

Fig. 1. Work flow of proposed algorithm

and investigates the performance by tuning the parameters. In the proposed algorithm, the conversion of SIFT features vector quantization to the sparse code is

$$\min_{U,V} \sum_{m=1}^{M} \|x_m - u_m V^2\|^2 + \lambda |u_m|, \tag{1}$$

subject to $\|v_k\| \leq 1,\ k = 1, 2, \ldots, K$, where V is $N \times K$ 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 (2) is convex in V and U simultaneously. This can be solved to fix the iteration number to achieve optimization over V or U while fixing any other. Fixing codebook V , (3) can be solved as a linear regression problem with L1-norm regularization on sparse coefficients

$$\min_{u_m} \|x_m - u_m V\|_2^2 + \lambda |u_m| \,. \tag{2}$$

Fixing U, same problem will be transformed to least square with quadratic constraints

$$\min_{V} \|X - UV\|_F^2, \tag{3}$$

subject to $\|v_k\| \leq 1,\ \forall k = 1, 2, \ldots, K$. Lagrange dual [16] can cleanly deal with it.

ScSPM feature is computed by the histogram pooling method

$$z = \frac{1}{M} \sum_{m=1}^{M} u_m \,. \tag{4}$$

For the pre-chosen pooling function F sparse matrix U will result ScSPM feature X

$$Z = F(U), \tag{5}$$

where the max pooling function F is applied on each column of absolute sparse code U as

$$z_j = \max\{|u_{1j}|, |u_{2\ j}|, \ldots, |u_{Mj}|\}, \tag{6}$$

where z_j is the j-th element of z, u_{ij} is the matrix element at i-th row and j-th column of U, and M is the number of local descriptors in the region.

Multiclass classification usually disintegrates into groups of binary $1, -1$ problem that can easily accommodate the functionality of standard SVM famous approaches one-versus-rest and one-versus-one. In this experiment, we have implemented SVM as proposed in [17] to solve the following convex optimization problem using kernel weights d_m.

$$J(d) = \sum_{p \in P} J_p(d), \tag{7}$$

where P is the set of all pairs to be considered, and $J_p(d)$ is the binary SVM objective value shown in (8) for the classification problem pertaining to pair p.

$$J(d) = \begin{cases} \max_\alpha -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \sum_m d_m K_m(z_i, z_j) \\ \text{with } 0 \leq \alpha_i \leq \frac{1}{vl}\ \forall i \\ \sum_i \alpha_i = 1 \end{cases}, \tag{8}$$

where α_i is Lagrange multiplier. The gradient of the given function in (8) can be found as

$$\frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{p \in P} \sum_{i,j} \alpha_{i,p}^* \alpha_{j,p}^* y_i y_j K_m(z_i, z_j) \quad \forall m, \tag{9}$$

where $\alpha_{i,p}$ is the Lagrange multiplier of the i-th example involved in the p-th decision function. For every pair of example Lagrange multiplier is obtained independently.

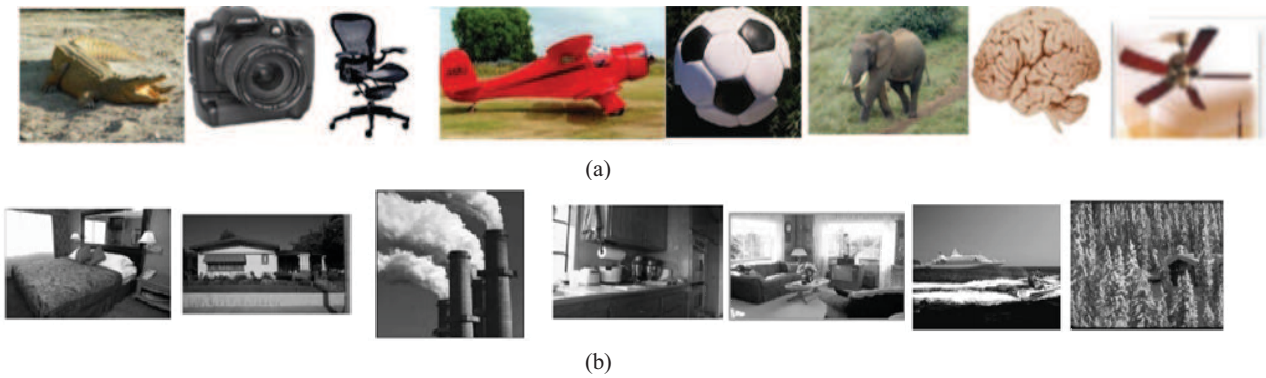

Fig. 2. Sample images of datasets used in this experiment: (a) – Caltech-101, (b) – Scene-15

Table 1. Kernels and their parameter's value

MKL	Base kernel	Coefficient value	Number of kernels
K1	Polynomial	(1,2,3)	3
K2	Gaussian	(0.5,1,2,5,7,10,12, 15,17,20)	10
K3	Gaussian and polynomial	(0.5,1,2,5,7,10,12, 15,17,20) (1,2,3)	13

4 Experiments and results

The improvement in the classification accuracy in SVM requires fine-tuning of the kernel within SVM. Therefore, a different fusion of the kernel is used, and classification accuracy is analysed rigorously. A benchmark dataset named Caltech-101 [27] and Scene-15 [28–30] is used in the experiment. The Caltech-101 dataset contains 9145 images for 101 different classes with various object categories like animals, instruments, vehicles, flowers, plants, *etc.* The dataset contains 40 to 800 images per class. The scene-15 dataset contains main indoor and outdoor scenes like the kitchen, living room, offices, *etc.* Though the number of classes is less, it has low inter-class covariance, making it difficult to achieve high accuracy. A total of 4000 images are available in 15 classes ranging from 200 to 400. Sparse coding of features and multi-resolution approach using spatial pyramid matching robust to local spatial translation [18] are used. That reduces the training complexity from $O(n^3)$ to $O(n)$ and keeps the testing complexity constant.

Each sparsified SIFT features sets for Caltech-101 was divided into 30 and 15 training samples and the remaining for testing. For this experiment, Scene-15 dataset was divided into 50 and 100 training images per class, and the remaining were left for testing. In this experiment, we extracted SIFT feature as per our previous work on parametrizing SIFT and sparse dictionaries [26]. Total six parameters are involved in the extraction of SIFT features, including the number of Gaussian functions, its variance, amount of image scaling, histogram bins' orientation, and its radius and features vector size. The empirical study presented in [26] suggests that the size of SIFT feature depends on the number of bins and angles. And analysis propagates that large bins with less angle failed to extract local features and hence classification accuracy. Therefore, the proposed method uses 16 orientations and four bins for SIFT features. We excluded the detailed discussion of other types of SPM (KSPM and LSPM) and MKL and their results are used in the comparison. ScSPM [18] used a linear kernel on spatial-pyramid pooling for sparsified SIFT features whereas, in the proposed experiment, the linear kernel is replaced by MKL as suggested in [17].

By the rigorous study of literature and detailing the work for ScSPM, it has been observed that the patch size reference to dictionary size, number of training and testing samples, and SVM contributes magnificently in the improvement of classification rate.

The patch size reference to dictionary size contributes to the sparsity of the features as expressed in (3). In the proposed experiment, 256×1024 and 16×16 sizes are used for dictionary and patch respectively. Dictionary is trained for 30 iterations for KSVD. The average coefficients value for the learned KSVD dictionary over 30 iterations is shown in Fig. 3. The One-Versus-Rest SVM approach is used in training. The fusions of the kernels with the values of their coefficients are listed in Tab. 1. We tested the performance for 5 independent runs and noted average accuracy achieved after five runs. This experiment was conducted with Intel Core i3 of 2.50 GHz, 8 GB RAM, and Windows-10 of 64 bit machine configurations.

Table 2 presents a comparison of the obtained results with other state-of-art methods.

The proposed method uses the same size as the dictionaries used in [26] and [18]. Wang *et al* [31] presented an SVM-based scene classification model where images are characterized using SIFT features obtained from the

Table 2. Comparison with other SPM or multi-kernel approaches

Dataset: Caltech–101			
Algorithm	Average accuracy (%)	Training images	Method name
ScSPM [18]	67.0 ± 0.45 73.02 ± 0.54	30	SPM sparse coding
BOW(400), [19]	72.02	30	Bag of words
BOW(1000), [19]	70.11		
BOW(4000), [19]	71.24		
NBNN, [20]	70.4	15	Naive-Bayes nearest-neighbor
LVFC-HSF, [21]	70.7 78.7		Local visual feature coding based on heterogeneous structure fusion
CLGC(RGB-RGB) (22)	72.6	30	Concatenation of local and global color
CSAE, [23]	64.0 71.4	15	Convolutional sparse auto-encoder
LMMK, [24]	62.3	30	Large margin multiple kernel
Parameterizing ScSPM (26)	77.08 ± 0.31		Parameterizing SPM sparce coding
Proposed method	79.29 ± 0.43 85.06 ± 0.31	15 30	kernel K1
	79.87 ± 0.36 85.72 ± 0.47	15 30	kernel K2
	78.96 ± 0.24 84.97 ± 0.21	15 30	kernel K3
Dataset: Scene –15			
Algorithm	Average accuracy (%)	Training images	Method name
ScSPM, [18]	87.28 ± 0.93	100	SPM sparse coding
LVFC-HSF, [21]	87.23		Local visual feature coding based on heterogeneous structure fusion
OVH, [25]	87.07		Orthogonal vector histogram
Parameterizing ScSPM, [26]	81.13 ± 0.53		Parameterizing SPM sparce coding
Proposed method	81.94 ± 0.54 89.12 ± 0.41	50 100	kernel K1
	83.32 91.12 ± 0.57	50 100	kernel K2
	85.55 ± 0.40 90.52 ± 0.21	50 100	kernel K3

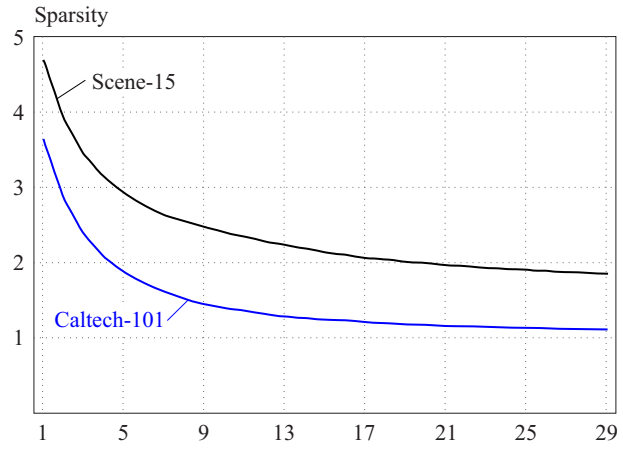

Fig. 3. Average coefficient value for each KSVD iteration on Caltech-101 and Scene-15 dataset

spatial pyramid. For the Caltech-256 dataset, their accuracy is limited to 31 % only due to a single kernel SVM in classification. In the proposed algorithm, selective sparsified features with Gaussian kernels combination in SVM can transform data in more separable dimension space. So performance is better than SPM based SIFT approach [18] where only local features of different scales are concatenated. Similarly, our previous work [26] shows the SIFT feature tuning can outperform over ScSPM. In this work, we used the tuned SIFT feature with MKL so better grouping between homogenous features can achieve. Our method shows ~ 8 % higher accuracy than [26]. LVFC-HSF offers greater optimization between local and global features but it cannot achieve more linearity in higher dimension space than our algorithm. In the LMKL method [24] higher weight is assigned to feature which make more discrimination in classes. They calculated base kernels by ten different image descriptors given in [24] using Gaussian function, increasing the com-

putation complexity for higher dimensional feature vector. Overall accuracy they reported $\sim 3\,\%$ higher than our but in their experiment test images from each class are fixed to 15. For the scene-15 dataset, OVH [25] calculates a global rotation invariant geometric visual word to relate with BoVW as special information but cannot take advantage of distinct local information.

The proposed approach increases the accuracy to $85.72\,\%$ for a large multiclass dataset of CalTech-101 and $91.12\,\%$ for the scene-15 dataset. The features classification using kernels K1, K2, and K3 provides a better classification rate than large margin multiple kernel (LMMK). The sparse-based learning model's improvement depends on the configuration of the parameters in the sparse dictionary and the extraction of robust feature sets. Also, the optimum selection of dictionary size, patch size, and integration of the kernel plays a vital role in classifying a large confusing multi-label dataset. The non-linear nature of polynomial and Gaussian kernel helped distinguish these features in SVM and hence proposed model achieved a better classification rate.

5 Conclusions

CNN has obtained large popularity in the classification models at the cost of large training time and increased computation cost. In comparison with CNN, SVM is found to have greater flexibility in characterization if an appropriate kernel is used for challenging datasets. The single kernel limits its application for datasets having linear classification. Therefore, a multi-kernel SVM has experimented again with the aim of optimization in the selection of the kernels and study of various parameters affecting the kernel performance in classification. The investigation of simple MKL over ScSPM features for classification accuracy is presented initially and the role of various parameters has been explored to minimize the redundant features. Then a sparse-dictionary is created for minimizing the features size.

After getting the maximum sparsity of the dictionary, the effect of MKL on overall classification accuracy is presented. We noted that even with a minimal combination of a single type kernel like Polynomial as shown in Tab. 2 accuracy will be higher than the single kernel SVM algorithm. Multiple combinations of Gaussian kernels lead to an increase in the classification accuracy to $85.72\,\%$ for 101 class datasets. We observe that training time and storage requirement also increases with a higher number of the Gaussian kernel which makes difficult to work on large dataset like Caltech-256 using minimum hardware requirements. Hence we conclude that with good features and Multi kernels, object recognition is still an open area to work. Coral reef classification using image augmentation in [32] gives promising results in the limited dataset. In that work, they used RGB and gray colours as a feature for predicting most corals that have the most similarity. In this work, there are many classes in a dataset that have this kind of similarity *ie* sunflowers, water lily in Caltech-101 and bedroom, living room in Scene-15 dataset, *etc.* In the future scope, we will examine the effect of this feature on alike classes.

References

[1] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering", *IEEE Transactions on knowledge and data engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[2] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition" A review", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1354–1369, 2013.

[3] M. Varma and D. Ray, "Learning the discriminative power invariance trade-off", in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[4] S. Xu and X. An, "Ml2s-svm: multi-label least-squares support vector machine classifiers", *The Electronic Library*, 2019.

[5] D. Kancherla, J. D. Bodapati, and N. Veeranjaneyulu, "Effect of different kernels on the performance of an svm based classification", *Int. J. Recent Technol. Eng*, no. 5, pp. 1–6, 2019.

[6] S. Bouteldja and A. Kourgli, "A comparative analysis of svm, k-nn, and decision trees for high resolution satellite image scene classification", in *Twelfth International Conference on Machine Vision (ICMV 2019)*, vol. 11433, p. 114331I, International Society for Optics and Photonics, 2020.

[7] D. Santos, E. Lopez-Lopez, X. M. Pardo, R. Iglesias, S. Barro, and X. R. Fdez-Vidal, "Robust and fast scene recognition in robotics through the automatic identification of meaningful images", Sensors, vol. 19, no. 18, p. 4024, 2019.

[8] X. Bai, J. Du, Z.-R. Wang, and C.-H. Lee, "A hybrid approach to acoustic scene classification based on universal acoustic models", in *Interspeech*, pp. 3619–3623, 2019.

[9] S. Nazir, Y. Qian, M. Yousaf, S. A. V. Carroza, E. Izquierdo, and E. Vazquez, "Human action recognition using multi-kernel learning for temporal residual network", 2019.

[10] Y. Wang, W. Yu, and Z. Fang, "Multiple kernel based svm classification of hyperspectral images by combining spectral, spatial, and semantic information", *Remote Sensing*, vol. 12, no. 1, p. 120, 2020.

[11] C. Tong-Tong, L. Chan-Juan, Z. Hai-Lin, Z. Shu-Sen, L. Ying, and D. Xin-Miao, "A multi-instance multi-label scene classification method based on multi-kernel fusion", in *2015 SAI Intelligent Systems Conference* (IntelliSys), pp. 782–787, IEEE, 2015.

[12] H. Hasan, H. Z. Shafri, and M. Habshi, "A comparison between support vector machine (svm) and convolutional neural network (cnn) models for hyperspectral image classification", in *IOP Conference Series: Earth and Environmental Science*, vol. 357, p. 012035, IOP Publishing, 2019.

[13] A. Sampath and N. Gomathi, "Fuzzy-based multi-kernel spherical support vector machine for effective handwritten character recognition", *Sādhanā*, vol. 42, no. 9, pp. 1513–1525, 2017.

[14] H. Patel and H. Mewada, "Analysis of machine learning based scene classification algorithms and quantitative evaluation", *International Journal of Applied Engineering Research*, vol. 13, no. 10, pp. 7811–7819, 2018.

[15] F. Zamani and M. Jamzad, "A feature fusion based localized multiple kernel learning system for real world image classification", *EURASIP Journal on image and Video processing*, vol. 2017, no. 1, pp. 1–11, 2017.

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

[17] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl", *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[18] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification", in *2009 IEEE Conference on computer vision and pattern recognition*, pp. 1794–1801, IEEE, 2009.

[19] H. Liao, J. Xiang, W. Sun, and S. Yu, "Adaptive aggregating multi-resolution feature coding for image classification", *Mathematical Problems in Engineering*, vol. 2014, 2014.

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

[21] G. Lin, C. Fan, H. Zhu, Y. Miu, and X. Kang, "Visual feature coding based on heterogeneous structure fusion for image classification", Information Fusion, vol. 36, pp. 275–283, 2017..

[22] L. Kabbai, M. Abdellaoui, and A. Douik, "Image classification by combining local and global features", *The Visual Computer*, vol. 35, no. 5, pp. 679–693, 2019.

[23] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, "Convolutional sparse autoencoders for image classification", *IEEE transactions on neural networks and learning systems*, vol. 29, no. 7, pp. 3289–3294, 2017.

[24] B. Hosseini and B. Hammer, "Large-margin multiple kernel learning for discriminative features selection and representation learning", in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2019.

[25] B. Zafar, R. Ashraf, N. Ali, M. Ahmed, S. Jabbar, and S. A. Chatzichristofis, "Image classification by addition of spatial information based on histograms of orthogonal vectors", *PloS one*, vol. 13, no. 6, p. e0198175, 2018.

[26] B. Gajjar and H. M. A. Patani, "Parameterizing sift and sparse dictionary for svm based multi-class object classification", *International Journal of Artificial Intelligence*, vol. 19, pp. 95–108, 2021.

[27] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories", *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[28] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope", *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[29] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories", in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05), vol. 2, pp. 524–531, IEEE, 2005.

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

[31] H.-H. Wang, C.-W. Tu, and C.-K. Chiang, "Sparse representation for image classification via paired dictionary learning", *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16945–16963, 2019.

[32] S. Sharan, S. Kininmonth, U. V. Mehta, *et al*, "Automated cnn based coral reef classification using image augmentation and deep learning", *International Journal of Engineering Intelligent Systems*, vol. 29, no. 4, pp. 253–261, 2021.

Received 3 May 2021

Bhavinkumar Gajjar is a holder of MTech degree in Communication System Engineering from Gujarat Technological University, India. His field of interest is Image processing, Computer Vision and Optimization Algorithms. Currently he is working on accuracy enhancement for multiclass classifications techniques as a research scholar in Indus University. He is professional software developer and working in Arohi Operations Pvt Ltd. He has 7 years of academic and 3.5 years of industrial experience. He has six international and two national publications in reputed journals/conferences.

Hiren Mewada has obtained his MTech and PhD degree from Sardar Vallbhbhai National Institute of Technology-Surat, Gujarat, India. Presently he is Assistant Research Professor at Prince Mohammad Bin Fahd University, Kingdom of Saudi Arabia. Previously he was associate professor at Charotar University of Science and Technology, Gujarat, India. He has more than 17 years teaching experience. His current areas of interest are computer vision, signal processing, machine learning and Embedded System design. He has published more than 60 research papers and completed several funded research projects. He is coauthor of one book and published five book chapters. He is member of IETE and ISTE.

Ashwin Patani has obtained his MTech from Gujarat university, Gujarat and PhD degree from meghalaya university Meghalaya, India. Presently he is senior Assistant Professor at Indus University, Ahmedabad Gujarat. He has more than 15 years teaching experience. His current areas of interest are sensors & networks, machine learning and Embedded System design. He has published more than 20 research papers. He is author of one book. He is member of IETE and ISTE.