Aircraft Type Recognition in Remote Sensing Images using Mean Interval Kernel

Jaya Sharma[1], Rajeshreddy Datla[1,2], Yenduri Sravani[1], Vishnu Chalavadi[1] and Krishna Mohan C.[1]
[1]*Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, Hyderabad, India*
[2]*Advanced Data Processing Research Institute (ADRIN), Department of Space, Secunderabad, India*

Keywords: Remote Sensing Images, Aircraft Type Recognition, Structural Information Model, Scale-invariant Feature Transform (SIFT), Dynamic Kernels.

Abstract: Structural characteristics representation and their fine variations are crucial for the recognition of different types of aircrafts in remote sensing images. Aircraft type classification across different sensor remote sensing images by spectral and spatial resolutions of objects in an image involves variable length spatial pattern identification. In our proposed approach, we explore dynamic kernels to deal with variable length spatial patterns of aircrafts in remote sensing images. A Gaussian mixture model (GMM), namely, structure model (SM) is trained over aircraft scenes to implicitly learn the local structures using the spatial scale-invariant feature transform (SIFT) features. The statistics of SM are used to design dynamic kernel, namely, mean interval kernel (MIK) to deal with the spatial changes globally in the identical scene and preserve the similarities in local spatial structures. The efficacy of the proposed method is demonstrated on the multi-type aircraft remote sensing images (MTARSI) benchmark dataset (20 distinct kinds of aircraft) using MIK. Also, we compare the performance of the proposed approach with other dynamic kernels, such as supervector kernel (SVK) and intermediate matching kernel (IMK).

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

With the current earth observation abilities at pixel level, the fine details of the ground are illustrated and the spatial content is gathered from high resolution remote sensing images. Along with their sub-parts, these details facilitate the computer vision community in exploring even the small man made objects. The inherent fine variations within their sub-parts reveal unique characteristics, which are helpful for object recognition tasks. In remote sensing images, aircraft type recognition is one such task that distinct characteristics in deciding an aircraft type. Aircraft type recognition includes various applications such as status monitoring (Zhong et al., 2018) , airport surveillance analysis (Chen et al., 2014), and aircraft identification (He et al., 2018), (Chen et al., 2014). In particular, an aircraft type and its dynamics remarkably help in examining the battlefields thereby to formulate rapid strategic military decisions (Liu et al., 2012), (Zhong et al., 2018).

Typically, the background pixel occupancy of an aircraft is high in a remote sensing aircraft scene when compared to the existence of an aircraft in the image. The fine variations among different types of aircrafts such as nose, empennage, fuselage, wings, engines, etc., have challenges in perceptibility due to their fewer pixel occupancy in the images resulting in low classification performance. So, the different aircraft types with identical backgrounds as shown in Fig. 1 cause low inter-class variation. Also, the various types of aircrafts such as B-52 & C-135, KC-10 & Boeing, and A-26 & P63, etc., are similar to each other visually. Along with distinct background of an aircraft, acquisition image factors such as spatial resolution, manifest shadows caused by illumination condition, variations in scale, view-angle, occlusion of sub-parts, also cause high intra-class variations (the aircraft of type Boeing, P-63, C-5, C-135, etc. from Fig. 1).

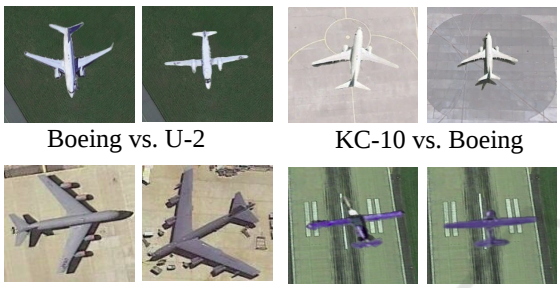
In addition to the remote sensing image characteristics, any equipment for ground handling, e.g., refuelers, aircraft service stairs, dollies, aero bridges, and ground power units that are near the aircraft disrupt an aircraft structure. Moreover, such equipment presents an additional or a change in the existing structure of an aircraft. These changes in the structure further pose the challenges in demonstrating the fine varia-

Sharma, J., Datla, R., Sravani, Y., Chalavadi, V. and C., K.
Aircraft Type Recognition in Remote Sensing Images using Mean Interval Kernel.
DOI: 10.5220/0011062600003209
In *Proceedings of the 2nd International Conference on Image Processing and Vision Engineering (IMPROVE 2022)*, pages 166-173
ISBN: 978-989-758-563-0; ISSN: 2795-4943
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

tions among the different aircraft types.

C-5 Boeing C-135 P-63

(a) Intra-class diversity



Boeing vs. U-2 KC-10 vs. Boeing

C-135 vs. B-52 P-63 vs. A-26

(b) Inter-class variation

Figure 1: Typical characteristics of multi-type aircraft remote sensing images (MTARSI) benchmark dataset (Wu et al., 2020). (a) Intra-class diversity: First, second, third, and fourth column represent aircraft types, such as C-5, Boeing, C-135, and P-63, respectively. (b) Inter-class similarity: First row shows the similarity of Boeing vs. U-2 and KC-10 vs. Boeing; Second row: C-135 vs. B-52; and P-63 vs. A-26 (best viewed in color).

In literature, for recognition of an aircraft type in the remote sensing images, the methods based on hand engineered features (Cheng et al., 2017) are not vigorous as their features are designed specific to the attributes of an image. The template matching-based methods (Wu et al., 2014) (Xu and Duan, 2010) (Liu et al., 2012) have some predefined templates which are applied in the recognition process. These methods have certain limitations on generalizing the process of recognition for the similar type of aircrafts but with different sizes, if they are not incorporated in the templates. Though the high level features from the data are automatically learned by CNN (Zhao et al., 2017), they do not focus on the finer variations in different aircrafts. Hence, the performance of CNN-based approaches are not satisfactory in recognising various types of aircrafts in remote sensing images.

In this work, we propose an approach to obtain an efficient representation for the classification of different types of aircrafts in remote sensing images. We employ scale-invariant feature transform (SIFT) (Ha and Moon, 2011) to capture the local structures such as nose, empennage, fuselage, wings, engines, etc., of aircrafts in remote sensing images. To encode these variable length local SIFT features, a single Gaussian mixture model (GMM) also known as structure model (SM) is trained. The statistics of SM are employed to estimate the likeness between any two images by computing the distance between them. A better separability is achieved by kernel methods for various image classes by mapping distances to different space (Smola and Schölkopf, 2004). Although, most of the approaches are suited to deal with specified length patterns, these are limited when compared to variable number of local features between two images. So, to handle the patterns of variable length we use dynamic kernels to select the best local features or projecting into fixed-length patterns.

In dynamic kernels, the use of base kernels is helpful in measuring the likeness among two images by computing the distance between their local features. In kernels which are based on probability, the base kernel is computed using the posterior probability of GMM. Whereas, in matching based kernels, the local features that are alike to GMM means are included in the kernel computation. This helps in retaining key local structures including the spatial patterns during the computation of base kernel. Some of the scenes in the images, such as rounded road connectedness in collateral images, row houses in dense residential, are considered as the significant spatial patterns. Hence, dynamic kernels become suitable option in order to represent the likeness among the images.

The contributions done by this paper can be summarized as:

- The similarities in the spatial patterns of local structures of the aircraft are retained by training structure model (SM).
- To deal with variable length spatial features of images, dynamic kernels are explored to retain the local structures and capture the global variations in an image
- The efficacy of our approach is demonstrated on a large aircraft recognition dataset: multi-type aircraft remote sensing images (MTARSI), a benchmark dataset consisting of 20 different types of aircrafts.

2 RELATED WORK

The existing approaches related to the image classification in remote sensing images are discussed in this section. Also, we summarize the techniques that han-

dle the variable length patterns based on the dynamic kernels.

The low-level, mid-level, and high-level features are explored in the existing methods for classification of different objects in remote sensing images (Cheng et al., 2017) (Azam et al., 2021) (Lin et al., 2018). The representations encoded on the low-level features relying heavily on the hand-crafted features primarily focuses on the precise attributes of the images. In their design, the most prevalent spatial features used are: shape, color, structural details, spatial and texture, etc. However, due to the remote sensing image characteristics, these spatial feature combinations are usually difficult to achieve. Some global attributes like texture descriptors and color histograms are used in object classification tasks (Cheng et al., 2017)– (Li et al., 2018). But, for encoding the local properties, we need an extra mechanism to describe an entire image. Hence, the mid-level features are used by transforming the local features into the global features to describe an image completely. In locality constrained linear coding (LLC) methods, the combination of BOVW and spatial pyramid matching (SPM) (Cheng et al., 2017), (Yang and Newsam, 2008), and bag-of visual-words (BOVW) with scale-invariant feature transform (SIFT) features (Lowe, 2004), of remote sensing images are used in image classification. To obtain effective sparselets (Cheng et al., 2015a)– (Cheng et al., 2015b), part detectors are explored by employing feature descriptors of histogram of oriented gradients (HOG) for image classification. A consolidated framework for joint super resolution and aircraft recognition (Joint-SRARNet) is proposed by Tang, Wei, et al.(Tang et al., 2020), that tries to enhance the recognition performance by generating discriminative, high-resolution aircraft from low-resolution remote sensing images. Technically, this network integrates super resolution and recognition tasks into the generative adversarial network (GAN) framework through a joint loss function.

The methods employing convolutional features have demonstrated better classification performance when compared to mid-level or hand-crafted feature based methods. This is because the convolutional features are able to provide good discrimination along with better generalization. Also, the effectiveness of fine-tuned and pre-trained versions of GoogLeNet, VGGNet16, and AlexNet, (Cheng et al., 2018), (Wu et al., 2020), (He et al., 2018) in the classification of objects in remote sensing images were demonstrated. The ensemble of CNNs is utilized to improve the classification performance over pre-trained CNN models (Zhao et al., 2017), (Chang and Lin, 2011). The object classification performance is further improved by accumulating, integrating, or fusing numerous attributes of CNN (Zhong et al., 2018)–(Chaib et al., 2017). In (Sitaula et al., 2020), the combination of object-based and scene-based attributes from both region level as well as scene level are used for image depiction. To address the problem of inter and intra class dissimilarities in object classification, an objective function is augmented along with the features of CNN (Cheng et al., 2018).

Later, a key filter bank based CNN (KFBNet) (Li et al., 2020) is used from the key locations of each image by assimilating the class-specific features in order to preserve global information for image classification. Another approach (Zhao et al., 2017) is explored to model the unrevealed ontological formation by using the multi-granularity canonical appearance pooling from the remote sensing images. A siamese network is explored to obtain CNN attributes and to determine the structure at its granule level. By calculating the second order statistics from the obtained CNN attributes, Gaussian covariance matrices are derived. The better classification performance is obtained by suitable normalization of the covariance matrices during training.

The combination of hidden markov model (HMM) and Gaussian mixture model (GMM) for the variable length pattern representation is explored in various application domains such as, image, speech, video, and music analysis. Dynamic kernels (Dileep and Sekhar, 2013) (Perveen et al., 2020) (Boumeddane et al., 2019) are important for obtaining a fixed length feature vector from patterns of variable length. To construct a probabilistic sequence kernel (PSK), instead of generative features which produce discriminative features, Lee et al. (Lee et al., 2007) estimated the Gaussian densities. Bhattacharyya distance-based calculation is employed between the GMM mixtures to incorporate both the first and the second-order GMM statistics (You et al., 2009) in order to boost the computational performance of PSK. A single universal background model (UBM) is constructed to model the features from multiple speakers to be modeled. From the covariances and mean of UBM obtained in mean interval supervectors, the covariances and means are adjusted for each speaker. The kernel resulting from a supervector is specified as the Gaussian mean interval kernel for categorization using support vector machine (SVM). Rather than mean or covariance transformation, intermediate matching kernels (IMK) (Boughorbel et al., 2005) uses virtual feature vector sets based on GMM mixtures for the nearest local feature vectors to be selected from each image. As the selected features acquired from a clip are lesser than the local features (Dileep and Sekhar,

2013) from the probabilistic sequence kernel (PSK) and the Gaussian mean interval kernel, IMK is computationally more efficient. Also, by the virtual features optimal selection, it was demonstrated that further depletion in computation time is possible.

3 PROPOSED METHOD

A dynamic kernel, namely, mean interval kernel (MIK) for aircraft type recognition in remote sensing images is discussed in this section. Figure 2 gives the proposed approach which consists of three modules, namely, feature extraction, structure model (SM), and construction of dynamic kernels for classification in better kernel space.

3.1 Feature Extraction

The scale-invariant feature transform (SIFT) (Lowe, 2004) features are extracted to represent the local structural information such as engines, wings, nose, etc. For this, the key points are sampled in linear scale space to recognise the locations that are invariant to occlusion, view-point, and scale. A 16×16 region is extracted around these key locations to compute the magnitude and angle of the gradient. This region is again divided into 4×4 sub-regions resulting in computing 8 orientation histograms of 128 feature dimensions. The features $N \times 128$ around these key points N describe the information of local structures for discriminating different aircrafts. These SIFT features from all the remote sensing aircraft scenes are used to construct the structure model, which is explained in the next sub-section.

3.2 Structure Model (SM)

The structure model (SM) is a Gaussian mixture model (GMM) that encodes the SIFT features in order to capture the local structures responsible for differentiating the types of aircrafts. These local structures, also known as attributes like nose, engine, fuselage, etc., collectively form an aircraft. The SM is trained on SIFT features of all aircraft remote sensing images to model these local structures of aircrafts. The SM with weights w_q, means μ_q, and covariances σ_q is given by

$$p(\mathbf{x}_k|(w_q,\mu_q,\Sigma_q)) = \sum_{q=1}^{Q} w_q \mathcal{N}(\mathbf{x}_k|\mu_q,\Sigma_q), \quad (1)$$

The SM is trained on SIFT feature descriptor, $\mathbf{x}_k$ using expectation maximization (EM) algorithm. We assume that each GMM component captures an attribute of the aircraft and the variance of each mixture determines the variations of spatial patterns in different aircrafts. The parameters of SM are adapted using maximum a posterior (MAP) adaptation to enhance the contribution of each component. These parameters of SM are computed by

$$n_q(\mathbf{x}) = \sum_{k=1}^{K} p(q|\mathbf{x}_k), \quad (2)$$

$$\mathbf{F}_q(\mathbf{x}) = \frac{1}{n_q(\mathbf{x})} \sum_{k=1}^{K} p(q|\mathbf{x}_k)\mathbf{x}_k, \quad (3)$$

and

$$\mathbf{S}_q(\mathbf{x}) = \frac{1}{n_q(\mathbf{x})} \sum_{k=1}^{K} p(q|\mathbf{x}_k)\mathbf{x}_k^2. \quad (4)$$

where $p(q|\mathbf{x}_k) = \frac{w_q p(\mathbf{x}_k|q)}{\sum_{q=1}^{Q} w_q p(\mathbf{x}_k|q)}$ is the posterior probability of SM component q and $p(\mathbf{x}_k|q)$ is the likelihood of feature vector $\mathbf{x}_k$. These SM parameters are used to obtain a compact representation for handling variable-length patterns using various dynamic kernels.

3.3 SM-Mean Interval Kernel (SM-MIK)

Dynamic kernels are kernel functions that map the variable length spatial features to either constant length feature vectors (kernels based on probability) or choosing the optimal features (matching based kernels). SM-MIK is a probaility based kernel that incorporates the additional information captured by second-order statistics, along with means of SM. The adapted means and covariances is given by

$$\hat{w}_q = \alpha n_q(\mathbf{x})/K + (1-\alpha)w_q, \quad (5)$$

$$\hat{\mu}_q(\mathbf{x}) = \alpha \mathbf{F}_q(\mathbf{x}) + (1-\alpha)\mu_q, \quad (6)$$

$$\hat{\Sigma}_q(\mathbf{x}) = \alpha \mathbf{S}_q(\mathbf{x}) + (1-\alpha)(\Sigma_q + \mu_q^2) - \hat{\mu_q^2}. \quad (7)$$

The mean supervector using the adapted means and covariances is given by

$$\varphi_q(\mathbf{x}) = \left(\frac{\hat{\Sigma}_q(\mathbf{x}) + \Sigma_q}{2}\right)^{-\frac{1}{2}} (\hat{\mu}_q(\mathbf{x}) - \mu_q). \quad (8)$$

Later, the SM-MIK is calculated to measure the similarity between two remote sensing image $\mathbf{x}_m$ & $\mathbf{x}_n$ by

$$K_{mv}(\mathbf{x}_m,\mathbf{x}_n) = \Phi_{mv}(\mathbf{x}_m)^T \Phi_{mv}(\mathbf{x}_n). \quad (9)$$

Computational time of SM-MIK is $O(QL + Q(K_l^2 + K_l) + K_s^2)$. The time complexity of SM-MIK is high due to calculation of first & second order statistics of SM.

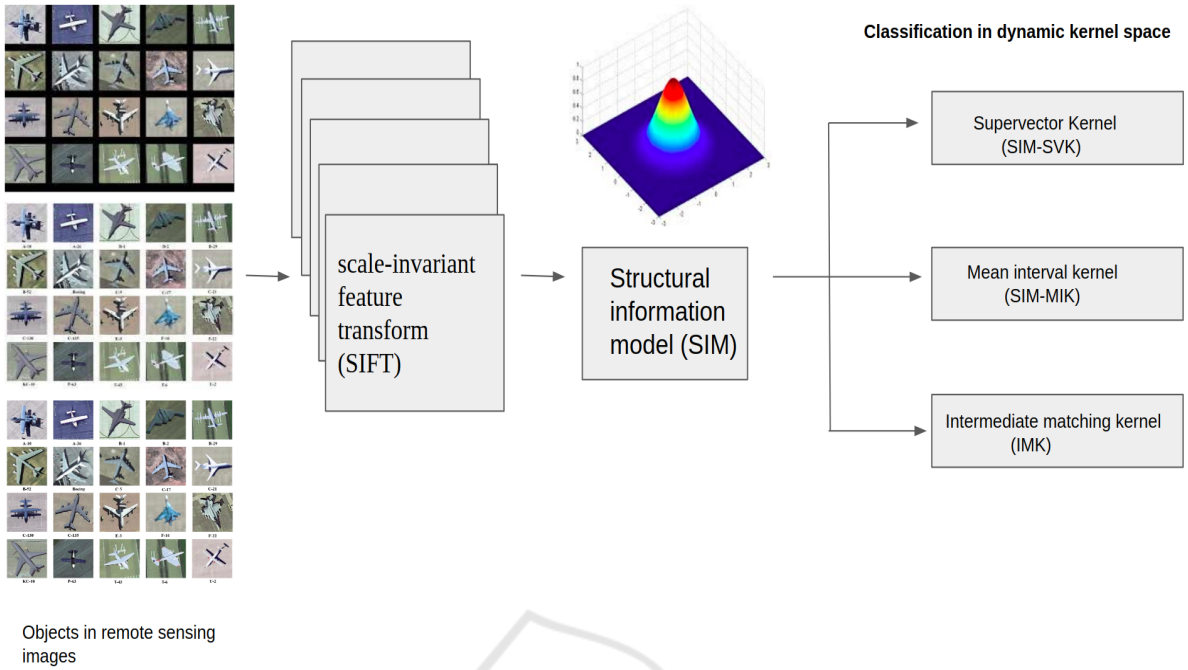


Figure 2: Block diagram of the proposed method for aircraft type recognition.

3.4 SM-Supervector Kernel (SM-SVK)

The SM-SVK kernels computes the similarity between two aircraft scenes by measuring the adapted means of each image w.r.t the means of SM. The adapted means is given by

$$\hat{w}_q = \alpha n_q(\mathbf{x})/K + (1-\alpha)w_q, \tag{10}$$

$$\hat{\mu}_q(\mathbf{x}) = \alpha \mathbf{F}_q(\mathbf{x}) + (1-\alpha)\mu_q, \tag{11}$$

The SM supervector $\varphi_q(\mathbf{x}) = [\sqrt{w_q}\Sigma_q^{-\frac{1}{2}}\hat{\mu}_q(\mathbf{x})]^T$ is obtained by concatenating the adapted means of all mixtures of SM. The SM-SVK constructed using the $\varphi_q(\mathbf{x})$ is given by

$$K_{sv}(\mathbf{x}_m, \mathbf{x}_n) = \Phi_{sv}(\mathbf{x}_m)^T \Phi_{sv}(\mathbf{x}_n). \tag{12}$$

The computation time of SM-SVK is $O(QL + QK_l^2 + K_s^2)$. Where, Q is number of mixtures in SM, L represents the number of local feature. K_l gives the dimension of SIFT feature vector and K_s denotes supervector dimension.

3.5 Intermediate Matching Kernel (IMK)

The IMK is a matching based kernel that calculates the similarity by choosing the nearest local features. The IMK matches the local features from $\mathbf{x}_m$ and $\mathbf{x}_n$ with a virtual features set $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_Q\}$ by calculating

$$\mathbf{x}^*_{mq} = argmin_{x \in \mathbf{x}_m} D(\mathbf{x}, \mathbf{v}_q), \tag{13}$$

and

$$\mathbf{x}^*_{nq} = argmin_{x \in \mathbf{x}_n} D(\mathbf{x}, \mathbf{v}_q), \tag{14}$$

Here, the $D(.)$ is the similarity function that finds the spatial patterns in local features similar to patterns learnt from the particular SM mixture. The virtual features represented with the posterior probability can effectively incorporate information about component coefficients, means, and covariances. It is given by

$$\mathbf{x}^*_{mq} = argmax_{x \in \mathbf{x}_m} p(q|\mathbf{x}_k), \tag{15}$$

and

$$\mathbf{x}^*_{nq} = argmax_{x \in \mathbf{x}_n} p(q|\mathbf{x}_k). \tag{16}$$

The computational time of IMK is $O(QL)$. The complexity is lower than other dynamic kernels due to the selection of best feature vectors.

4 EXPERIMENTAL RESULTS

The effectiveness of the proposed SM-MIK based method is evaluated on multi-type aircraft remote sensing images (MTARSI) dataset. We also compare SM-MIK with other dynamic kernels, namely, SM-SVK and IMK in this section.

4.0.1 Dataset

We consider a challenging dataset, multi-type aircraft remote sensing images (MTARSI) (Wu et al., 2020) in recognizing the different aircraft types to demonstrate the effectiveness of the proposed method. From Google Earth satellite imagery constituting 20 different aircraft types, this dataset consists of 9,385 aircraft scenes. Each of the aircraft scenes contains exactly one aircraft and the number of aircraft scenes vary with different types of aircraft 230 to 846. From different airports around the world with spatial resolution between 0.3 m and 1.0 m, the scenes of a specific aircraft type are obtained. In addition to high within class variation and inter-class similarity, this dataset illustrates rich image variations. For each class in the experimental settings, this dataset is randomly splitted into 80%-20% training-testing ratio.

For training of GMM, the feature vectors of SIFT, which describes both global and local semantics are extracted from each image. A single GMM is trained for 5 different mixtures on MTARSI dataset.

4.0.2 Evaluation of Dynamic Kernels

In Table 1, the performance of various kernels are presented on MTARSI dataset, by formulating kernel based SVM classifiers using LibSVM (Chang and Lin, 2011). The dynamic kernel performance is improved with the SIFT features and it is examined that beyond 128 the SM components do not participate in the advancement of classification achievement. Also, it is noticed that SM-SVK and SM-MIK provide improved classification performance than IMK. This is due to the incorporation of SM statistics (1^{st} & 2^{nd}-order) in SM-MIK and SM-SVK which can efficiently model the crucial information across the spatial patterns of variable length entities. Though the MIK is not efficient computationally than IMK, we can select the suitable kernel based on the use-case.

Table 1: Comparison of classification performance for various dynamic kernels with structural information model mixtures of $\{2^{\ell}\}_{\ell=5}^{9}$ on MTARSI benchmark dataset.

SM mixtures	SM-SVK	SM-MIK	IMK
32	75.32	88.64	73.59
64	83.54	89.79	82.13
128	89.52	**90.87**	89.79
256	86.89	90.54	85.76
512	86.21	89.32	84.43

4.0.3 Comparison with State-of-the Art Methods

The Table 2 gives the comparison of the existing approaches with the proposed method on MTARSI dataset. The proposed SM-MIK based approach outperforms the current state-of-the-art methods on MTARSI data. Also, the performance of SM-MIK on MTARSI dataset is 90.87%. This is because SM-MIK can capture global dissimilarities efficiently by modelling the variable length spatial patterns of aircrafts in the images and conserving local formations. This shows that SM-MIK is able to apprehend global dissimilarities successfully. Thus, using the mean and covariances of SM, we are able to capture the global spatial features for the aircraft type recognition more efficiently than the local spatial features modelled by SIFT.

Table 2: Comparison of the proposed method with existing approaches on MTARSI benchmark dataset.

Method	Accuracy(%)
SIFT (Ha and Moon, 2011)+BOVW	59.02
HOG (Dalal and Triggs, 2005)+SVM	61.34
ScSPM (Yang et al., 2009)	60.61
LLC (Yu et al., 2009)	64.93
AlexNet (Krizhevsky et al., 2012)	85.61
GoogleNet (Szegedy et al., 2015)	85.61
VGG (Simonyan and Zisserman, 2014)	87.56
ResNet (He et al., 2016)	89.61
DenseNet (Huang et al., 2017)	89.15
EfficientNet (Tan and Le, 2019)	89.79
SM-SVK	89.52
IMK	89.79
SM-MIK	**90.87**

5 CONCLUSION

In this paper, a mean interval kernel (SM-MIK) based-method is presented to obtain an effective representation for recognizing different aircraft types from remote sensing images. To model varying length spatial features of the aircrafts while capturing global variations, we construct SM-MIK over the trained Gaussian mixture model. To deal with the varying length spatial features of objects in scenes of remote sensing images, the SM-MIK has demonstrated to be better than the other kernels. In the calculation of SM-MIK, the utilization of first-order and second-order statistics of the Gaussian mixture model contributes useful information for the aircraft type classification task. Though IMK are not much discriminative in comparison to SM-MIK, the IMK have better computational time complexity. The effectiveness of the proposed approach is demonstrated on the challenging large-scale MTARSI benchmark dataset for aircraft type recognition.

REFERENCES

Azam, F., Rizvi, A., Khan, W. Z., Aalsalem, M. Y., Yu, H., and Zikria, Y. B. (2021). Aircraft classification based on pca and feature fusion techniques in convolutional neural network. *IEEE Access*, 9:161683–161694.

Boughorbel, S., Tarel, J. P., and Boujemaa, N. (2005). The intermediate matching kernel for image local features. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 889–894. IEEE.

Boumeddane, S., Hamdad, L., Dabo-Niang, S., and Haddadou, H. (2019). Spatial kernel discriminant analysis: Applied for hyperspectral image classification. In *ICAART (2)*, pages 184–191.

Chaib, S., Liu, H., Gu, Y., and Yao, H. (2017). Deep feature fusion for vhr remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4775–4784.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27.

Chen, J., Zhang, B., and Wang, C. (2014). Backscattering feature analysis and recognition of civilian aircraft in terrasar-x images. *IEEE Geoscience and Remote Sensing Letters*, 12(4):796–800.

Cheng, G., Han, J., Guo, L., and Liu, T. (2015a). Learning coarse-to-fine sparselets for efficient object detection and scene classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1173–1181.

Cheng, G., Han, J., and Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883.

Cheng, G., Yang, C., Yao, X., Guo, L., and Han, J. (2018). When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. *IEEE transactions on geoscience and remote sensing*, 56(5):2811–2821.

Cheng, G., Zhou, P., Han, J., Guo, L., and Han, J. (2015b). Auto-encoder-based shared mid-level visual dictionary learning for scene classification using very high resolution remote sensing images. *IET Computer Vision*, 9(5):639–647.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee.

Dileep, A. D. and Sekhar, C. C. (2013). Gmm-based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines. *IEEE Transactions on Neural Networks and Learning Systems*, 25(8):1421–1432.

Ha, S.-W. and Moon, Y.-H. (2011). Multiple object tracking using sift features and location matching. *International Journal of Smart Home*, 5(4):17–26.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer.

He, N., Fang, L., Li, S., Plaza, A., and Plaza, J. (2018). Remote sensing scene classification using multilayer stacked covariance pooling. *IEEE Transactions on Geoscience and Remote Sensing*, 56(12):6899–6910.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.

Lee, K.-A., You, C., Li, H., and Kinnunen, T. (2007). A gmm-based probabilistic sequence kernel for speaker verification. In *Eighth Annual Conference of the International Speech Communication Association*. Citeseer.

Li, F., Feng, R., Han, W., and Wang, L. (2020). High-resolution remote sensing image scene classification via key filter bank based on convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):8077–8092.

Li, P., Ren, P., Zhang, X., Wang, Q., Zhu, X., and Wang, L. (2018). Region-wise deep feature representation for remote sensing images. *Remote Sensing*, 10(6):871.

Lin, J., Li, X., and Pan, H. (2018). Aircraft recognition in remote sensing images based on deep learning. In *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 895–899. IEEE.

Liu, G., Sun, X., Fu, K., and Wang, H. (2012). Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. *IEEE Geoscience and Remote Sensing Letters*, 10(3):573–577.

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

Perveen, N., Roy, D., and Chalavadi, K. M. (2020). Facial expression recognition in videos using dynamic kernels. *IEEE Transactions on Image Processing*, 29:8316–8325.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sitaula, C., Xiang, Y., Basnet, A., Aryal, S., and Lu, X. (2020). Hdf: hybrid deep features for scene image representation. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR.

Tang, W., Deng, C., Han, Y., Huang, Y., and Zhao, B. (2020). Srarnet: A unified framework for joint superresolution and aircraft recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:327–336.

Wu, Q., Sun, H., Sun, X., Zhang, D., Fu, K., and Wang, H. (2014). Aircraft recognition in high-resolution optical satellite remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 12(1):112–116.

Wu, Z.-Z., Wan, S.-H., Wang, X.-F., Tan, M., Zou, L., Li, X.-L., and Chen, Y. (2020). A benchmark data set for aircraft type recognition from remote sensing images. *Applied Soft Computing*, 89:106132.

Xu, C. and Duan, H. (2010). Artificial bee colony (abc) optimized edge potential function (epf) approach to target recognition for low-altitude aircraft. *Pattern Recognition Letters*, 31(13):1759–1772.

Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. In *2009 IEEE Conference on computer vision and pattern recognition*, pages 1794–1801. IEEE.

Yang, Y. and Newsam, S. (2008). Comparing sift descriptors and gabor texture features for classification of remote sensed imagery. In *2008 15th IEEE international conference on image processing*, pages 1852–1855. IEEE.

You, C. H., Lee, K. A., and Li, H. (2009). Gmm-svm kernel with a bhattacharyya-based distance for speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1300–1312.

Yu, K., Zhang, T., Gong, Y., et al. (2009). Nonlinear learning using local coordinate coding. In *NIPS*, volume 22, pages 2223–2231. Citeseer.

Zhao, A., Fu, K., Wang, S., Zuo, J., Zhang, Y., Hu, Y., and Wang, H. (2017). Aircraft recognition based on landmark detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 14(8):1413–1417.

Zhong, Y., Ma, A., soon Ong, Y., Zhu, Z., and Zhang, L. (2018). Computational intelligence in optical remote sensing image processing. *Applied Soft Computing*, 64:75–93.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS