

Selecting principal attributes in multimodal remote sensing for sea ice characterization

Eduard Khachatrian, Saloua Chlaily, Torbjørn Eltoft, and Andrea Marinoni
Department of Physics and Technology, UiT the Arctic University of Norway, NO-9037 Tromsø, Norway.
E-mails: {eduard.khachatrian, saloua.chlaily, torbjorn.eltoft, andrea.marinoni}@uit.no

Abstract

Automatic ice charting can not be achieved using only SAR modalities. It is fundamental to combine information from other remote sensors with different characteristics for more reliable sea ice characterization. In this paper, we employ principal feature analysis (PFA) to select significant information from multimodal remote sensing data. PFA is a simple yet very effective approach that can be applied to several types of data without loss of physical interpretability. Considering that different homogeneous regions require different types of information, we perform the selection patch-wise. Accordingly, by exploiting the spatial information, we increase the robustness and accuracy of PFA.

1 Introduction

Recently, sea ice research has become a hot topic in Earth observation disciplines, as it plays a vital role in the polar ecosystem, and it is one of the main indicators of global climate change. Indeed, it affects several anthropogenic activities in the Arctic region, such as oil and gas industry, fisheries, shipping, and tourism. It also affects the lifestyle and welfare of indigenous population. All these factors make sea ice monitoring of primary interest to protect the Arctic and to ensure safe and effective commercial activities and polar navigation [1].

Due to the remoteness and extreme weather conditions, remote sensors (especially synthetic aperture radar (SAR)) are the primary source of information about the Arctic region. At present, the automatic interpretation of remote sensing data is challenging and relies on the expert's analysis. However, automatic sea ice analysis is required to perform robust near real-time investigation on a global scale [2].

Various remote sensing systems (*modalities*) grasp different aspects of sea ice by using different physical principles. Thus, integrating relevant information from multiple modalities enables better characterization of sea ice [3]. Nevertheless, although different modalities provide complementary information, they can also be redundant, corrupted, or irrelevant for a particular task. Hence, combining all available information may deteriorate the analysis. Moreover, it increases the algorithm's complexity. Therefore, the selection of relevant information is an essential step of multimodal data fusion for reliable and efficient performance [4].

Relevant information retrieval can be achieved by dimensionality reduction methods, such as feature selection and

feature extraction. Feature extraction methods transform the original set of data into a lower dimensional space. On the other hand, feature selection methods identify the most relevant elements of the dataset according to a given criterion, such as maximum variance, minimum dependency, or maximum correlation. As opposed to feature extraction methods, feature selection approaches preserve the physical interpretability of the results. The aforesaid methods can be classified as supervised, if they require training data, or unsupervised. The supervised methods, can be accurate if the training data are rich and reliable. However, when considering sea ice characterization, scarce training sets (either in terms of quality or size) are typically available. Therefore, supervised methods can be hardly employed to obtain accurate and reliable results.

Principle component analysis (PCA) is one of the most used information extraction techniques being efficient and easy to implement. PCA is unsupervised and applies an orthogonal transformation to convert a set of observations of potentially correlated variables into a smaller set of linearly uncorrelated variables, called principal components [5]. In [6], Lu et al. proposed a variation of PCA for feature selection. Principal feature analysis (PFA) exploits the same tools as PCA to generate a new representation of the data. This new representation, as opposed to PCA, can be mapped back to the original domain and hence preserves the physical interpretation of the dataset (essential for sea ice analysis).

In this paper, we use PFA for multimodal information selection for remote sensing of sea-ice. In contrast to the classic PFA, we perform the selection in a patch-wise manner. In this fashion, we fully exploit the potential of each modality, since only the deficient or irrelevant parts of an image are discarded. Moreover, we take into account the particularity of each object in the observed scene,

since our approach only chooses the relevant attributes to its characterization. Experimental tests carried out on several instances of multivariate remote sensing datasets, acquired over the Arctic region, show that the proposed approach increases the efficiency and accuracy of sea ice multimodal analysis.

The remainder of this paper is organized as follows: Section 2 recall the principles of PFA approach. Section 3 introduces the superpixel PFA. Section 4 delivers the experimental results of our method on different datasets. Finally, the conclusions are presented in Section 5.

2 Principal feature analysis

Let M the number of all attributes retrieved from remotely sensed images, e.g., polarization intensities, spectral channels, and textural features, etc. The images acquired by different sensors may have different physical units, different resolutions, and different coordinate systems. The first step of our analysis consist of making the data comparable, by means of normalization, subsampling, and alignment on the same coordinate system. We denote by $\mathbf{x}_i \in \mathbb{R}^M$ the set of M attributes associated to the i -th pixel, and by $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$ the matrix of observations, where N is the number of pixels.

In order to increase the separability of attributes and unveil their hidden structure, PFA generates a new representation using the eigenvectors of their covariance matrix. The sample covariance matrix of the attributes can be written as follows:

$$\Sigma \approx \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T, \quad (1)$$

where $\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$, and \cdot^T denote the transpose operator. Using singular value decomposition [7], we obtain:

$$\Sigma = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T, \quad (2)$$

where $\mathbf{V} \in \mathbb{R}^{M \times M}$ is a unitary matrix, whose columns are the eigenvectors of Σ . $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_M\}$ is a diagonal matrix, whose elements are the eigenvalues of Σ , such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$.

Σ depicts the second-order statistical relationships between the attributes. Moreover, the eigenvectors of Σ can be geometrically interpreted as the axes that best fit the data [8]. Their corresponding eigenvalues reflect the variability of the attributes along the axes. Accordingly, eigenvectors with large eigenvalues reveal most of the information about data and are hence more illustrative of its variance [8].

PCA uses the K first principal eigenvectors, corresponding to the largest eigenvalues, as the basis of a lower-dimensional space onto which the dataset is projected. As opposed to PCA, PFA uses the rows of the principal eigenvectors as the new representations of the attributes. This new representation accounts only for the largest eigenvalues and exploits the linear dependency of attributes.

Hence, it presents a strong discrimination power compared to the original set of attributes.

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1K} & \dots & v_{1M} \\ v_{21} & v_{22} & \dots & v_{2K} & \dots & v_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ v_{M1} & v_{M2} & \dots & v_{MK} & \dots & v_{MM} \end{pmatrix} \quad (3)$$

Using this new representation, and by means of clustering algorithms such as k -means, the attributes are arranged in different groups according to their similarity. Accordingly, constituents of one cluster depict similar information, while members of different clusters represent different information. Therefore, the attributes corresponding to the centroids of the clusters are chosen to preserve the information content of the original set.

In fact, PFA is a graph clustering approach using eigenvectors of the similarity matrix, where the attributes constitute the vertices of the graph, and the similarity matrix is defined using covariance.

3 Superpixel PFA

The homogeneous parts of a region of interest have different properties. Accordingly, they require distinct types of attributes to be characterised. In order to reflect the particularity of each homogeneous part, we perform the selection at a superpixel level. Specifically, we partition the images into L homogeneous patches, i.e., superpixels, using Watershed [9] segmentation, however other methods such as Simple linear Iterative Clustering (SLIC) [10] can be also applied. To each superpixel, we apply PFA to select the relevant attributes to its characterization. The different steps of superpixel PFA are shown in **Algorithm 1**.

Algorithm 1 Attributes selection at the l -th superpixel

Input: Attributes of l -th patch - \mathbf{X}_l , Number of selected features - $K < M$.

Output: Subset of K Attributes

1. Compute the covariance matrix of \mathbf{X}_l , Σ_l ,
 2. Compute the first K largest eigenvectors of Σ_l , $\mathbf{v}_{l1}, \dots, \mathbf{v}_{lK}$.
 3. Form $\mathbf{U}_l = [\mathbf{v}_{l1}, \dots, \mathbf{v}_{lK}] \in \mathbb{R}^{M \times K}$.
 4. Cluster the rows of \mathbf{U}_l using k -means.
 5. Assign the i -th attribute to the same cluster as the i -th row of \mathbf{U}_l
 6. Return the centroids of the clusters.
-

4 Experiment

The following section reports the performance analysis of superpixel PFA, as well as comparison results with other existing dimensionality reduction methods using several multimodal sea ice datasets.

Dimensionality reduction methods can be used as a preprocessing step of various remote sensing applications. In this work, we use it to improve sea ice classification accuracy. To validate the performance of superpixel PFA method, we use one of the widely applied supervised classifiers in remote sensing, support vector machine (SVM).

SVM is a classification method that determines a set of hyperplanes that separate the dataset into different classes [11]. To perform a non-linear classification, we choose as a kernel the radial basis function (RBF).

In all the experiments, we randomly choose 20% of the samples from each label as the training set. The remaining 80% of samples are used as a test set. To quantitatively estimate the classification result, we use the overall accuracy (OA) index and Cohen's kappa statistic (Kappa).

4.1 Datasets

To evaluate the performance of superpixel PFA for multimodal remote sensing, we used two multisensor and multiband datasets obtained from various satellite platforms.

Both datasets were acquired in April 2018 on the North-East from Svalbard. They consist of SAR and optical data obtained from Sentinel-1 and Sentinel-2 for the first dataset and Radarsat 2 and Landsat 8 for the second. Datasets were downsampled to the same resolution by the nearest neighbor resampling method and projected onto the same WGS 84 / Arctic Polar Stereographic - EPSG:3995 coordinate system. Both datasets were labeled by sea ice experts and include several sea ice types along with open water. **Figure 1** shows the region of interest for two datasets. In the remaining of this section, we refer to the datasets as S1/S2 and R2/L8.

Along with the optical bands and SAR polarizations, we expand the dataset by extracting Gray-Level Co-Occurrence Matrix (GLCM) textural features [12, 13]. Textural features were extracted for each original attribute, i.e., optical band and SAR polarization. **Table 1** illustrates the extracted features as well as their mathematical definitions. Specifically, $g_{i,j}$ denotes the element of the GLCM matrix \mathbf{G} . Q is the number of gray levels used, and $\mu = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} i g_{i,j}$ and $\sigma^2 = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} (i-\mu)^2 g_{i,j}$ are, respectively, the GLCM mean and variance. ASM refers to angular second momentum. Finally, S1/S2 dataset includes 84 attributes (14 SAR and 70 optical) and R2/L8 consist of 91 attributes (14 SAR and 77 optical).

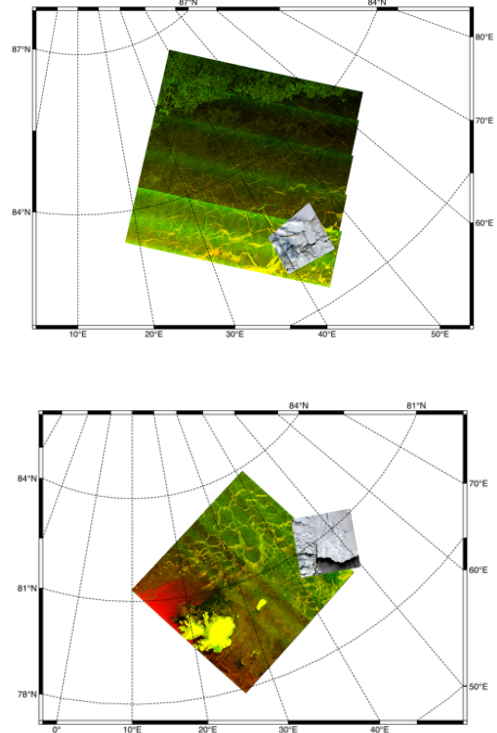


Figure 1 Region of interest maps for both multimodal datasets. **Top:** S1/S2. **Bottom:** R2/L8.

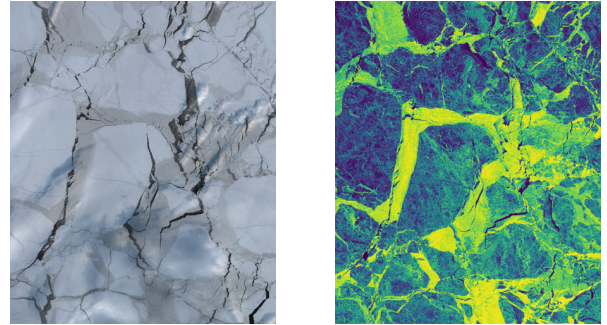


Figure 2 Overlapping area of S1/S2. **Left:** Natural color composite from Sentinel-2. **Right:** HH polarisation from Sentinel-1.

4.2 Parameter Sensitivity Analysis

The number of selected attributes and the size of superpixels are two parameters that may affect the performance of the superpixel PFA method.

Figure 4 shows the overall accuracy as a function of the number of selected attributes for both datasets. It is possible to appreciate that the use of all available attributes does not surely lead to the best classification accuracy since OA becomes stable or even slightly decreasing after a particular point. Additionally, for R2/L8 dataset, maximum classification accuracy was achieved with less than the third of the original feature set. This shows the ability of superpixel PFA for relevant information selection. For

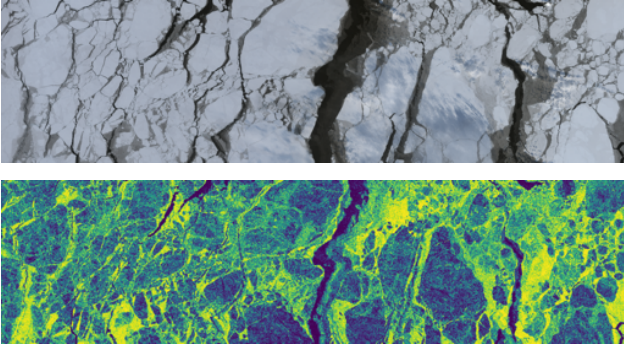


Figure 3 Overlapping area of R2/L8. **Top:** Natural color composite from Landsat 8. **Bottom:** HH polarisation from Radarsat 2.

Table 1 GLCM Features.

Features	Definition
Contrast	$\sum_{i,j=0}^{Q-1} g_{i,j} (i-j)$
Dissimilarity	$\sum_{i,j=0}^{Q-1} g_{i,j} i-j $
Homogeneity	$\sum_{i,j=0}^{Q-1} \frac{g_{i,j}}{1+(i-j)^2}$
ASM	$\sum_{i,j=0}^{Q-1} g_{i,j}^2$
Energy	\sqrt{ASM}
Correlation	$\sum_{i,j=0}^{Q-1} g_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right]$

both datasets the optimal number of attributes is equal to 30.

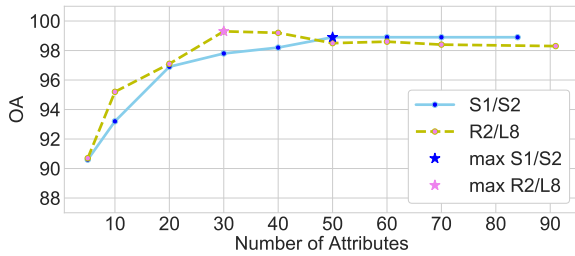


Figure 4 Overall accuracies of superpixel PFA over a different number of selected attributes for two datasets.

A superpixel based approach is affected by the size of the homogeneous areas. Small superpixels hold consistent information, while large sizes, including more data, are more accurate. **Figure 5** illustrates overall accuracy with respect to the number of superpixels for both datasets. A large number of superpixels implies superpixels of small size. It is evident from the curves that the number of superpixels does not affect the classification accuracy significantly. Although accuracy does not vary much for both datasets, the slightly more accurate result was achieved using 1000 superpixels, thus we use this value for subsequent analysis.

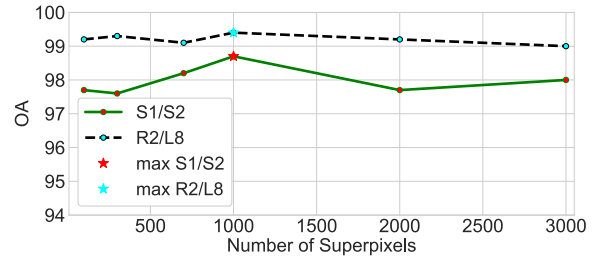


Figure 5 Overall accuracies of superpixel PFA over a different number of superpixels for both multimodal datasets using SVM classifier. Note that the vertical scale is Small.

4.3 Analysis of Selected Attributes

Figure 6 shows the number of attribute occurrences for S1/S2 dataset, i.e., the number of times an attribute was selected in all superpixels. SAR and optical attributes are shown in red and green color, respectively. The histogram shows a clear predominance of optical attributes, which is due to their large number compared to SAR (14 SAR attributes and 70 optical). It is evident from the histogram that the proposed method mainly selects data attributes (polarization intensities when considering SAR and reflectances when considering optical). The histogram shows the relevance of multimodal data since both SAR/optical attributes are selected by the superpixel PFA method without clear priority. It means that both datasets contain valuable, unique, and complementary information that can improve further applications.

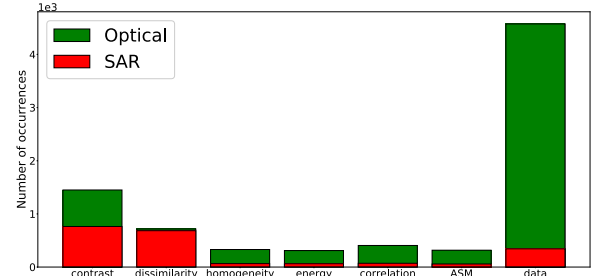


Figure 6 Number of occurrences for SAR and optical attributes of S1/S2 multimodal dataset.

4.4 PFA versus superpixel PFA

On **Figure 7** we show the overall accuracy with respect to the number of selected attributes for PFA and superpixel PFA. We remark that superpixel PFA outperforms the classic PFA in accuracy. Moreover, while superpixel PFA shows a stable behaviour, PFA is extremely affected by the number of selected attributes. That is because the superpixel analysis improve the separability of data which makes it less affected by the high variance induced by increasing the number of attributes.

Figure 8 and **Figure 9** shows the classified maps for both datasets using optimal number of attributes selected by proposed method. Sea ice labels used in this work differ from WMO Sea Ice Nomenclature [14], since we use multisens-

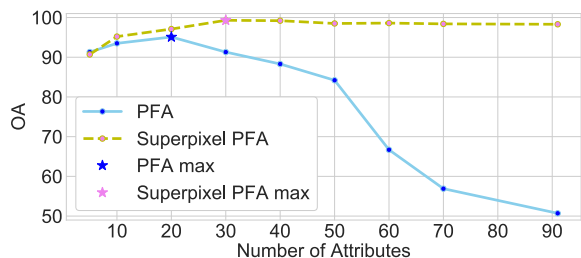


Figure 7 Overall accuracies of PFA and superpixel PFA over a different number of attributes for R2/L8 dataset.

tor data (SAR and optical) that can provide different information about the same region. Therefore it is complicated to determine exactly the same labels that will correspond both to WMO and radar classes simultaneously. Thus the sea ice types that are thicker than Nilas are labeled as 1-Thin and 2-Thick, which corresponds to 1-different young ice types and 2-various first-year ice types, respectively.

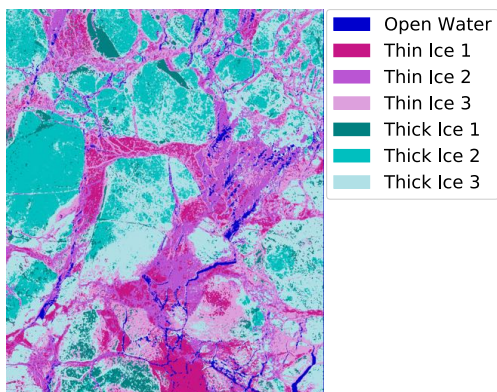


Figure 8 Classified map for S1/S2 dataset with optimal number of attributes selected by means of the proposed method.

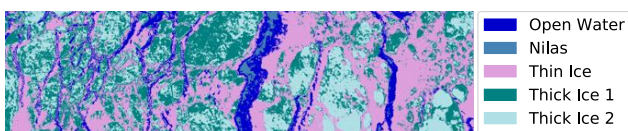


Figure 9 Classified map for R2/L8 dataset with optimal number of attributes selected by means of the proposed method.

4.5 Comparison with other Methods

Now, we compare the achieved results with other six dimensionality reduction algorithms, namely three feature extraction methods: principal component analysis (PCA), decision boundary feature extraction (DBFE), Fisher information feature extraction (FIS) and three feature selection methods: forward feature selection (FS), branch and bound (OBB) and genetic algorithm (GA). FIS uses Fisher information for data transformation [15]. DBFE is a supervised method that extracts information and exploits the geometrical properties of decision boundaries [16]. FS

starts with a minimum number of features and with each new step it adds one feature that improves classification the most in terms of accuracy [17]. OBB is a backtracking feature selection algorithm that is based on the assumption that the adopted criterion function fulfills the monotonicity condition at which a straightforward application of this property many feature subset evaluations may be omitted [18]. GA is an adaptive algorithm that finds the global optimum solution for an optimization problem, based on the mechanics of natural genetics and biological evolution [19].

Table 2 demonstrates the OA and Kappa of R2/L8 among different dimensionality reduction methods. It is evident that superpixel PFA outperforms the other methods in terms of classification accuracy. This is due to the inability of some approaches to process multimodal dataset, in addition to the effectiveness of a superpixel PFA to select the best descriptive features for each homogeneous patch.

Table 2 Performance comparison among different dimensionality reduction methods for R2/L8 dataset. The number of attributes are automatically chosen by each method. N refers to the number of selected attributes.

Method	N	SVM	
		OA	Kappa
PCA	16	86.2	86.1
DBFE	15	85.4	85.0
FIS	17	84.7	84.4
FS	16	83.4	82.9
OBB	15	90.3	89.5
GA	14	89.3	88.7
PFA	20	95.1	92.9
Superpixel PFA	10	95.2	93.9
Superpixel PFA	30	99.3	98.7

5 Conclusions

In this paper, we employed PFA, being flexible and efficient, for multimodal remote sensing information selection. PFA combines the accuracy of feature extraction and the interpretability of feature selection. We improved the robustness of PFA by proposing a superpixel based approach. Hence, selecting the best descriptive features for each superpixel.

The superpixel selection can be used not only to select the relevant information but also to understand the information pertinent to characterize different objects or regions of interest in the polar areas. This will improve accordingly several sea ice applications, such as sea ice types classification, sea ice deformation, sea ice drift, and iceberg detection, which in turn can be useful for ice charting and modeling services.

6 Acknowledgements

This work is funded by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Research Council of Norway (RCN Grant no. 237906).

7 Literature

- [1] S. Sandven, O. M. Johannessen, and K. Kloster, "Sea Ice Monitoring by Remote Sensing," *Encyclopedia of Analytical Chemistry*, vol. 1993, no. March 1993, pp. 1–43, 2006.
- [2] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Satellite sar data-based sea ice classification: An overview," *Geosciences*, vol. 9, no. 4, 2019.
- [3] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, pp. 1585–1601, Sep. 2015.
- [4] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, pp. 331–342, Feb 2012.
- [5] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*. Orlando, FL, USA: Academic Press, Inc., 4th ed., 2008.
- [6] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proceedings of the 15th ACM International Conference on Multimedia*, MM '07, (New York, NY, USA), pp. 301–304, ACM, 2007.
- [7] P. Olver and C. Shakiban, *Applied Linear Algebra*. Prentice Hall, 2006.
- [8] I. Jolliffe and Springer-Verlag, *Principal Component Analysis*. Springer Series in Statistics, Springer, 2002.
- [9] S. Beucher and C. D. M. Mathmatique, "The watershed transformation applied to image segmentation," in *Scanning Microscopy International*, pp. 299–314, 1991.
- [10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," 2010.
- [11] S. R. Gunn, "Support vector machines for classification and regression," 1998.
- [12] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.
- [13] U. Kandaswamy, D. A. Adjeroh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2075–2083, 2005.
- [14] A. Bushuyev, "WMO sea-ice nomenclature, terminology, codes and illustrated glossary," pp. WMO /OMM /BMO 259, TP 145, 1970.
- [15] Z. Pei, P. Zhang, S. Sun, and J. Gu, "Fisher Information Analysis for Matching Feature Extraction," *Proceedings - 2009 International Conference on Information Technology and Computer Science, ITCS 2009*, vol. 1, pp. 425–428, 2009.
- [16] S. B. Serpico, M. D'Inca, F. Melgani, and G. Moser, "Comparison of feature reduction techniques for classification of hyperspectral remote sensing data," *Image and Signal Processing for Remote Sensing VIII*, vol. 4885, no. March, pp. 347–358, 2003.
- [17] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty, "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2824–2831, 2015.
- [18] P. Somol, P. Pudil, F. J. Ferri, and J. Kittler, "Fast branch & bound algorithm in feature selection," *Proc. SCI 2000 Conf., Orlando, FL*, vol. IIV, pp. 646–651, 2000.
- [19] S. Sivakumar and C. Chandrasekar, "Feature Selection Using Genetic Algorithm with Mutual Information," vol. 5, no. 3, pp. 2871–2874, 2014.