Master thesis topic (Masterproefonderwerp)	Sparse subspace clustering for large scale hyperspectral data
Promoters (Promotoren)	prof. dr. ir. Aleksandra Pižurica – Department of Telecommunications and Information Processing
Mentors (Begeleiders)	dr. Shaoguang Huang – Department of Telecommunications and Information Processing Srđan Lazendić – Department of Electronics and Information Systems
Contact person (Contactpersoon)	Srđan Lazendić
Number of master theses (Aantal masterproeven)	1
Number of students (Aantal studenten)	1
Academic year (Academiejaar)	2021-2022
Keywords (Trefwoorden)	Machine learning, clustering, dictionary learning, unsupervised classification
Location – (Locatie)	Technicum, De Sterre, thuis

References (Referenties):

1. C. N. Matcha. *A 2021 guide to semantic segmentation*, 2021, https://nanonets.com/blog/semantic-image-segmentation-2020/
2. J. Hu, Z. Luo, X. Wang, S. Sun. *End-to-end multimodal image registration via reinforcement learning*, 2021.
3. Casanova, P. O. Pinheiro, N. Rostamzadeh, C. J. Pal. *Reinforced active learning for image segmentation*, 2020.
4. F. Ryosuke, N. Inoue, T. Yamasaki. *PixelRL: Fully convolutional network with reinforcement learning for image processing*, 2019.

|	
Description of the problem (Probleemstelling)	Supervised classification methods such as the classical support vector machine (SVM) and the modern convolutional neural network (CNN) require labeled training samples to train the classification model. In some applications, labeled data are rather scarce or not available, either because data labeling is labor intensive and time consuming or simply because not enough examples of a particular phenomenon of interest have been recorded yet. Clustering, as an unsupervised approach, partitions data points into different clusters (classes) without any labeled data. Thus, clustering approaches are especially interesting in cases where supervised classification is not applicable or not reliable enough due to the lack of sufficient annotated data. Such cases arise often in dynamic scenarios such as monitoring forest fires, disaster damages, land use / cover change detection and trajectory data mining. We focus on the subspace clustering approach that yields state-of-the-art clustering performance in computer vision, image processing, remote sensing and pattern recognition. The main idea is to model the input data by a union of subspaces and uncover the cluster structure in lower-dimensional subspaces, as shown in Fig. 1. Compared with the classical fuzzy c-means and k-means methods, subspace clustering approaches are able to unveil more precisely the data correlations, leading to superior clustering performance. We are particularly interested in the processing of hyperspectral images (HSIs) in remote sensing, where the goal in this proposal is to cluster pixels or HSI into different groups using their spectral signatures, as shown in Fig. 2. Despite the excellent clustering accuracy of subspace clustering techniques, their high computational complexity limits their applicability in real applications involving big data sets, especially in real-time processing tasks. It is therefore important to reduce the computational complexity of these clustering models and to develop scalable subspace clustering methods for large-scale data. Important aspects in addressing this problem are understanding representation learning (including dictionary learning and subspace representation) and efficient algorithm design. Research group GAIM has rich experience in this domain and will provide full support in programming, model construction, optimization algorithms and experiment validation based on the well-founded expertise.

|	
Goal of the thesis (Doelstelling)	The goal of this Master's thesis is to advance further the current subspace clustering methods in particular by reducing their computational complexity such that they can be applied to large-scale hyperspectral data. The concrete objectives are: • Studying recent representation learning approaches from the literature, focusing on the design of discriminative dictionaries from input data. • Implementing one of these approaches and integrating it into the framework of sparse subspace clustering. • Conducting experiments on real hyperspectral remote sensing images. The students will start from the current subspace clustering code of the best available GAIMs technique and will also have other useful GAIMs techniques and source codes for dictionary learning, sparse coding and optimization algorithms. Motivated students will be encouraged to participate with the developed techniques in data classification challenges of the IEEE Geoscience and Remote Sensing community.
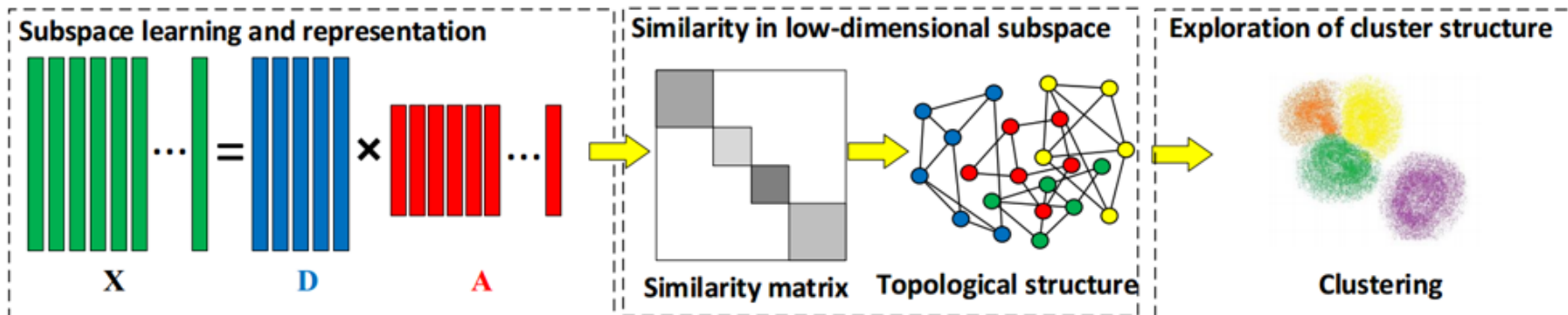
Figure 1: The framework of a typical subspace clustering method, which includes subspace learning and representation, graph construction and spectral clustering. X is an input matrix with each column representing a data point; D is a dictionary that models the underlying subspaces; A is the corresponding subspace representation matrix with respect to D.
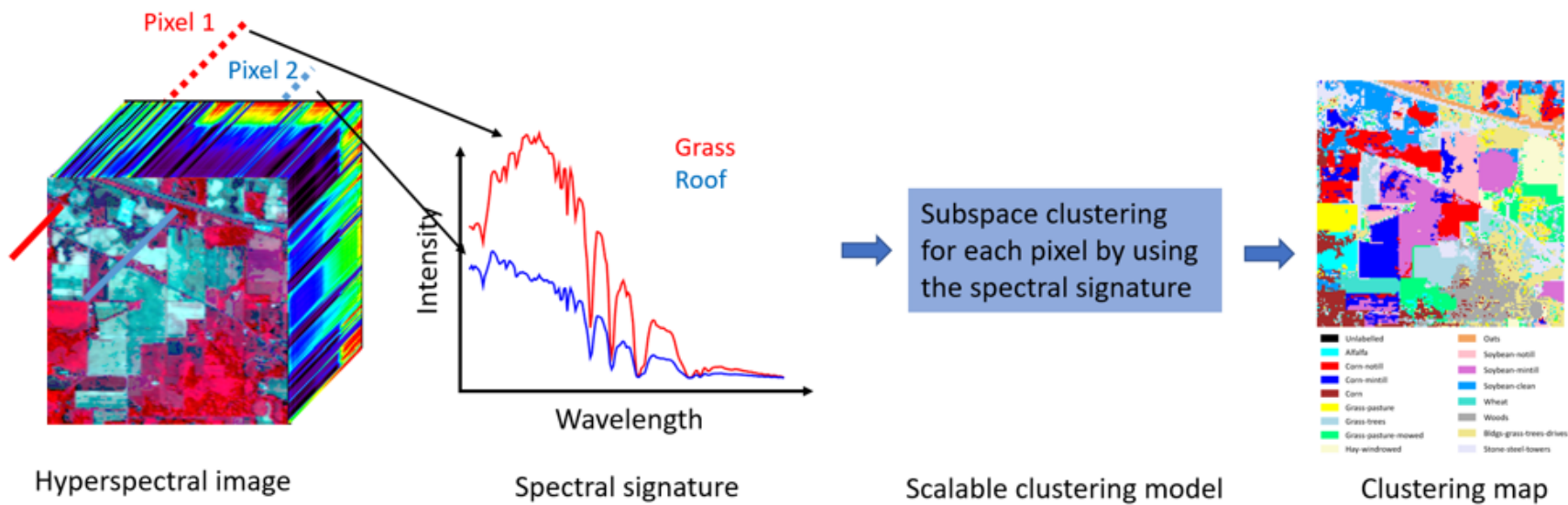
Figure 2: An illustration of subspace clustering in the application or hyperspectral remote sensing images.