Hierarchical Clustering Approach of Multisensor Data Fusion: Application of SAR and SPOT-7 Data on Korean Peninsula

Sang-Hoon Lee and Hyun-Gi Hong

Dept. of Industrial Engineering, Kyungwon University
San 65 Bokjeong-dong, Sujeong-gu, Seongnam-si, Kyunggi-do 461-701, Korea
Tel.: +82-31-750-5367 Email: shl@mail.kyungwon.ac.kr

In remote sensing, images are acquired over the same area by sensors of different spectral ranges (from the visible to the microwave) and/or with different number, position, and width of spectral bands. These images are generally partially redundant, as they represent the same scene, and partially complementary. For many applications of image classification, the information provided by a single sensor is often incomplete or imprecise resulting in misclassification. Fusion with redundant data can draw more consistent inferences for the interpretation of the scene, and can then improve classification accuracy.

The common approach to the classification of multisensor data as a data fusion scheme at pixel level is to concatenate the data into one vector as if they were measurements from a single sensor. The multiband data acquired by a single multispectral sensor or by two or more different sensors are not completely independent, and a certain degree of informative overlap may exist between the observation spaces of the different bands. This dependence may make the data less informative and should be properly modeled in the analysis so that its effect can be eliminated. For modeling and eliminating the effect of such dependence, this study employs a strategy using self and conditional information variation measures. The self information variation reflects the self certainty of the individual bands, while the conditional information variation reflects the degree of dependence of the different bands.

One data set might be very less reliable than others in the analysis and even exacerbate the classification results. The unreliable data set should be excluded in the analysis. To account for this, the self information variation is utilized to measure the degrees of reliability. The team of positively dependent bands can gather more information jointly than the team of independent ones. But, when bands are negatively dependent, the combined analysis of these bands may give worse information. Using the conditional information variation measure, the multiband data are split into two or more subsets according the dependence between the bands. Each subsets are classified separately, and a data fusion scheme at decision level is applied to integrate the individual classification results.

In this study, a two-level algorithm using hierarchical clustering procedure is used for unsupervised image classification. Hierarchical clustering algorithm is based on similarity measures between all pairs of candidates being considered for merging. In the first level, the image is partitioned as any number of regions which are sets of spatially contiguous pixels so that no union of adjacent regions is statistically uniform. The regions resulted from the low level are clustered into a parsimonious number of groups according to their statistical characteristics. The algorithm has been applied to satellite multispectral data and airbone SAR data.