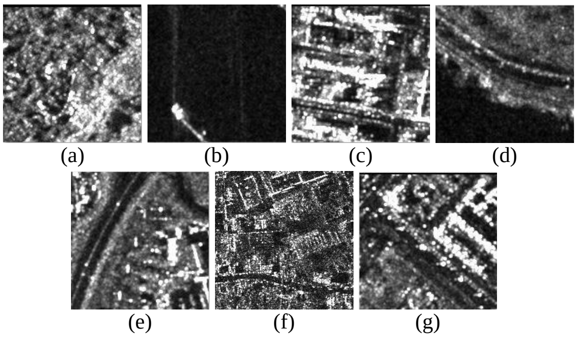


Figure 1: EO data set of 1230 SAR images grouped in seven non-equal size classes. (a) forest, (b) water, (c) medium density urban area, (d) forest + water, (e) roads, (f) high density urban area, and (g) urban area + roads).

aspects which let the other aspects to be still unknown. Therefore, studying the influences of different feature descriptors in understanding of the structure of data by computers is highly demanded.

In our paper, we explore the structure of the feature space, built by the feature descriptors extracted from EO images, by internal evaluation of clustering performed on the data. The unsupervised nature of clustering allows computers to find the structure of the feature space without being affected by any prior knowledge. In our experiments, we apply K-means, as the clustering technique, to a collection of SAR images for different number of clusters. Then we evaluate the clusters using an internal cluster evaluation method, namely S_Dbw (Halkidi and Vazirgiannis, 2001), to find the optimum clusters. The desired clusters are compact with a small density of feature points between clusters. Since there is no prior labeling considered in internal cluster evaluation, the clustering is not penalized for discovering new clusters or finding a different structure to the annotation labels (FÃrber et al., 2010).

Furthermore, we perform an external cluster evaluation, namely ARI (Hubert and Arabie, 1985), to compare the clusters to the prior annotation. This shows the difference between the understanding of the images by computers and the users' perceptions. Moreover, one can study from which aspects users annotating the given collection of images.

4 RESULTS AND DISCUSSION

4.1 EO image data set

In our experiments, we use a collection of EO data contains 1230 Synthetic Aperture Radar (SAR) images of size 160×160, Figure 1. The images are grouped in seven non-equal size classes (e.g., forest (198 images) , water (210 images), medium density urban area (204 images), forest + water (114 images), roads (67 images), high density urban area (279 images), and urban area + roads (158 images)). The images in the classes are rather homogeneous which allow to study the difference between the annotation and the resulting clusters.

4.2 Feature descriptors

In order to be processed by computers, the images are represented by a discrete set of extracted features. In this paper, we study the structure of feature space provided by three different feature descriptors, e.g., Gabor, WLD, and Random Features. In the feature space, each image is represented by a feature vector extracted from the entire image.

Gabor wavelet descriptors, proposed for texture analysis, are achieved by filtering a given image using a set of linear band-pass filters, the so called Gabor filters (Manjunath and Ma, 1996). These filters are generated by scaling and rotating a mother wavelet filter. The impulse response of this filter is a 2D modulated Gaussian function. The final Gabor feature vector is constructed by using the means (μ_{sr}) and the standard deviations (σ_{sr}) of the image filtered by S number of scales and R number of rotations, $F_{Gabor} = [\mu_{11}\ \sigma_{11}\ \mu_{12}\ \sigma_{12}\ ...\ \mu_{SR}\ \sigma_{SR}]$.
In our experiments, the Gabor features are constructed for 3 scales and 6 rotations which leads to a vector of 36 dimensions.

WLD is a feature descriptor developed based on Weber's law, a psychological law (Chen et al., 2010). According to this law, human notices the change in a stimuli as a valid signal if its ratio to the original intensity of the stimuli is above a certain constant value. WLD is constructed by a 2D histogram of: 1) Differential Excitation, the ratio between intensity difference between each pixel x and its neighbors; 2) Orientation, which is the gradient orientation of each pixel x. The final feature vector is constructed by building a 1D histogram of the computed 2D histogram, after quantizing to M number of excitations and T number of orientations.
In our experiments, we set M and T equal to 6 and 8, respectively, which results in a feature vector of 144 elements.

In order to use directly the intensity values, usually histogram of pixel values is constructed (e.g., color histogram in multimedia color images). Since the range of the intensity values of SAR data is rather wide which results in a very large vector, constructing the histogram of the pixel values is not trivial. Thus, in our experiments, we put all the pixel values of each given image $I_{m\times n}$ in a vector of size $d = m \times n$. Although in this way the size of the resulted vector is smaller, it is still too large to be used efficiently by the clustering techniques. Therefore, we used the idea of **Random Features (Rand_Feat)** (Liu and Fieguth, 2012) to decrease the dimensionality of the resulted feature vector d to a lower dimensional vector of size $\tilde{d}$. In this method, we compute the product of the high dimensional feature vector to a $d \times \tilde{d}$ random matrix.
In our experiments, we decrease the dimensionality of feature vectors to 32.
Figure 2 shows the 3D visualization of feature space built by the three feature descriptors. The prior annotation is also illustrated by different colors.

4.3 Internal evaluation of clusters

Internal evaluating of the clusters allows to investigate the structure of the given data, represented by feature descriptors, regardless of any prior knowledge.
In our experiments, S_Dbw is used to internally evaluate the clustering on three different feature descriptors (e.g., Gabor, WLD, and Rand_Feat) for different number of clusters (Figure 3c). Since S_Dbw is achieved by combining the average scattering and the average density between clusters, we show the two values as well in Figures 3a and 3b. As the results show, scattering monotonically decreases by increasing the number of clusters; however, after a certain number of clusters the change is not significant. Comparing the three feature descriptors, scattering decreases more for Rand_Feat than Gabor and WLD which means the structure of feature points is sparser in Rand_Feat.

As Figure 3b shows, average density between clusters does not change monotonically by increasing the number of clusters. Moreover, the general behaviors of the curves are rather different for different feature descriptors. It means that the average density

(a) WLD (b) Gabor (c) Random Feature

Figure 2: 3D visualization of feature space for three kinds of feature descriptors (e.g., WLD, Gabor, Rand_Feat) including the prior annotation of the data. Different colors represent different prior clusters (e.g.,forest, water, medium density urban area, forest + water, roads, high density urban area, and urban area + roads).

between clusters highly depends on the structure of the feature space. The optimum clustering provides clusters with minimum density of the points between clusters. For WLD we have the minimum density for two clusters which clearly indicates the feature space should be structured with two highly separated mass of points (Figure 4a). Furthermore, by increasing the number of clusters to three, the average density significantly increases due to splitting one of the separated masses into two clusters which come up with a large interfacing region (illustrated in Figure 4b); however, increasing the number of clusters to four, decreases the average density to some extent, as a newly added cluster introduces a small interface to the other clusters which leads to a decrease in the average density (illustrated in Figure 4c). For Gabor and Rand_Feat the behaviors of the curves provide general intuitions about the structure of corresponding feature spaces as well (some illustrations are shown in Figures 4d to 4i).

Combining the two criteria, average scattering and average density between clusters, results in S_Dbw measure. As it is illustrated in Figure 3c, for WLD and Gabor the optimality does not necessarily increases by increasing the number of the clusters; they have multiple local minimum which means there are more than one optimum clustering (e.g., for WLD we have optimum in 2, 4, 9, and 13 clusters). However, the behavior is rather different for Rand_Feat. The S_Dbw monotonically decreases which demonstrates that the feature space is not well structured leading to have the optimum number of clusters equal to the number of feature points.

4.4 External evaluation of clusters

As the results for internal clustering show, the optimum number of clusters is not necessarily equal to the number of the prior classes. Moreover, comparing the clusters, Figure 4, and the prior classes, Figure 2, illustrates that the points in one class does not group necessarily in one cluster. This demonstrates that the human semantic annotation does not necessarily correspond to what feature descriptors represent from the data. Therefore, we perform an external evaluation of clusters, which allows to compare the clusters to the prior annotations, using ARI. As Figure 5 illustrates, generally the structure represented by WLD and Gabor are closer to what users percept from the images in annotation time. In other words, users discriminate the images mostly based on the textures than the average intensity values. Moreover, comparing Figures 3c and 5, illustrates that the highest similarity of the clusters and the prior annotation does not necessarily occur at the optimum clusters.

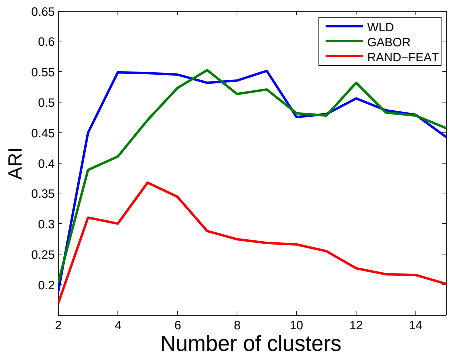
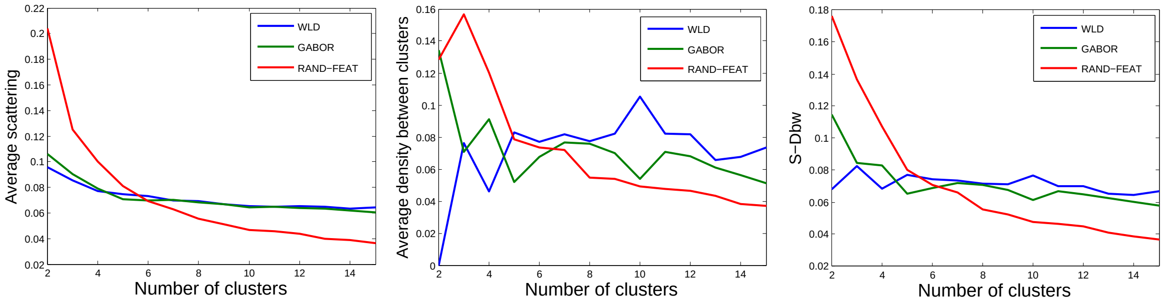


Figure 5: Adjusted Rand Index for three feature descriptors (e.g., WLD, Gabor, Rand_Feat).

5 CONCLUSIONS AND FUTURE WORK

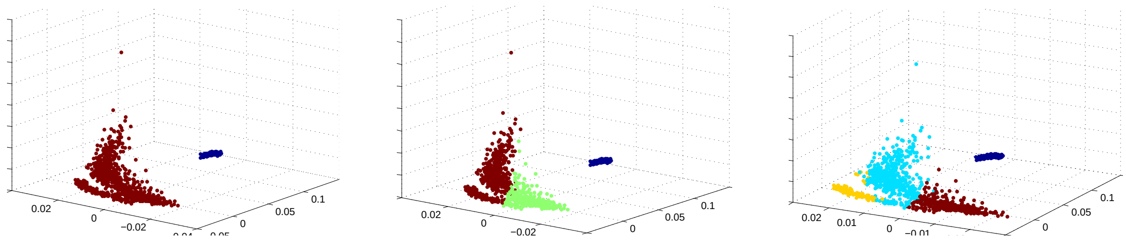
In this paper, we introduced a clustering-based approach to evaluate EO image indexing. In our proposed approach, the images are represented by feature descriptors. Then we perform clustering on the feature space, built by feature descriptors, for different number of clusters. The resulted clusterings are then evaluated both internally and externally (i.e., without and with using prior annotation, respectively). While internal evaluation demonstrate the structure of the feature space by finding the optimum clusters, external evaluation allows us to compare the image understanding by users and computers.
Experimental results show that the entire information provided by an image collection cannot be represented by only one kind of feature descriptors. Moreover, multiple optimum clusters can be found which corresponds to different levels of understanding of the images' contents (e.g., from low-level shape and texture to higher-level concepts).
External evaluation demonstrates that the structure discovered by computers (i.e, represented by feature descriptors) does not necessarily correspond to the prior annotation provided by users. Moreover, comparing the clusters for different feature descriptors shows that users consider some aspects of features in the images more than the others (e.g., consider textural features more than average intensity).

Evaluation of feature space provides users with better understanding of image collections. This allows the users to not only provide richer annotation, but also validate the annotation afterwards. Further, the study of feature space for different feature descriptors allows us to develop more sophisticated feature descriptors which not only group image collections to meaningful categories, but also provide relevant results to the users' queries.

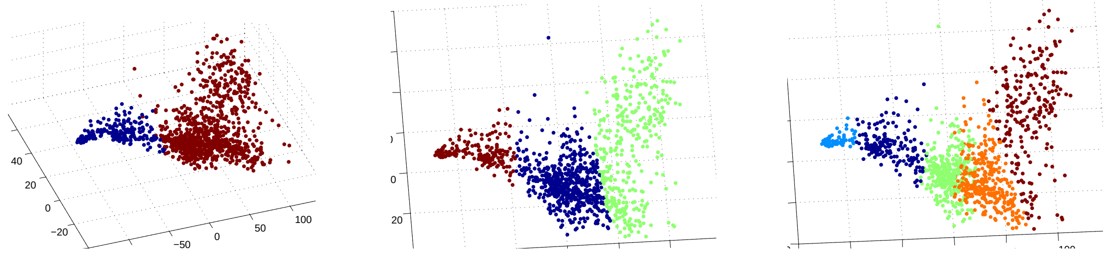


(a) Average scattering (b) Average density between clusters (c) S_Dbw measure

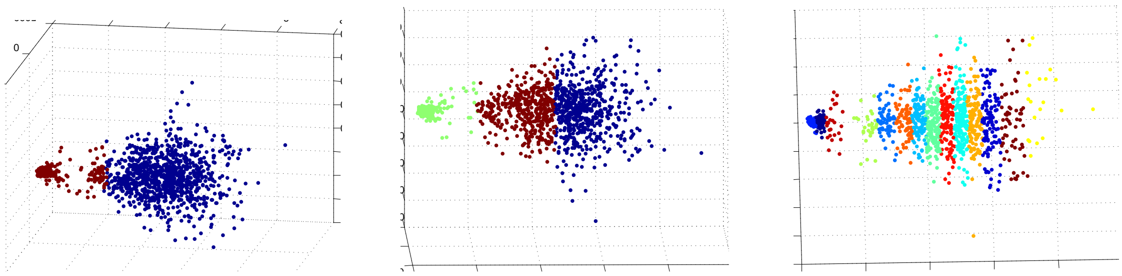
Figure 3: Internal evaluation of clusters for three kinds of feature descriptors (e.g., WLD, Gabor, Rand_Feat) and different number of clusters.



(a) WLD, # clusters = 2 (b) WLD, # clusters = 3 (c) WLD, # clusters = 4



(d) Gabor, # clusters = 2 (e) Gabor, # clusters = 3 (f) Gabor, # clusters = 5



(g) Rand_Feat, # clusters = 2 (h) Rand_Feat, # clusters = 3 (i) Rand_Feat, # clusters = 13

Figure 4: 3D visualization of feature space for three kinds of feature descriptors (e.g., WLD, Gabor, Rand_Feat) including colored labeling for different clusters.

REFERENCES

Bahmanyar, R. and Datcu, M., 2013. Measuring the semantic gap based on a communication channel model. In: IEEE International Conference on Image Processing, ICIP 2013.

Caliński, T. and Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods 3(1), pp. 1–27.

Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X. and Gao, W., 2010. Wld: A robust local image descriptor. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, pp. 1705 –1720.

Davies, D. L. and Bouldin, D. W., 1979. A cluster separation measure. Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-1(2), pp. 24–227.

FÃrber, I., GÃnnemann, S., Kriegel, H.-p., KrÃger, P., MÃller, E., Schubert, E., Seidl, T. and Zimek, A., 2010. On using class-labels in evaluation of clusterings.

Halkidi, M. and Vazirgiannis, M., 2001. Clustering validity assessment: finding the optimal partitioning of a data set. In: Data Mining, ICDM 2001, Proceedings IEEE International Conference on, pp. 187–194.

Hubert, L. and Arabie, P., 1985. Comparing partitions. Journal of Classification 2(1), pp. 193–218.

Larsen, B. and Aone, C., 1999. Fast and effective text mining using linear-time document clustering. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '99.

Liu, L. and Fieguth, P., 2012. Texture classification from random features. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(3), pp. 574–586.

Liu, Y., Li, Z., Xiong, H., Gao, X. and Wu, J., 2010. Understanding of internal clustering validation measures. In: Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10, IEEE Computer Society, Washington, DC, USA, pp. 911–916.

Manjunath, B. and Ma, W., 1996. Texture features for browsing and retrieval of image data. Pattern Analysis and Machine Intelligence, IEEE Transactions on 18(8), pp. 837–842.

Rendón, E., Abundez, I. M., Gutierrez, C., Zagal, S. D., Arizmendi, A., Quiroz, E. M. and Arzate, H. E., 2011. A comparison of internal and external cluster validation indexes. In: Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications, AMERICAN-MATH'11/CEA'11, World Scientific and Engineering Academy and Society (WSEAS), pp. 158–163.

Tan, P.-N., Steinbach, M. and Kumar, V., 2005. Introduction to Data Mining, (First Edition). Addison-Wesley Longman Publishing Co., Inc.

Vinh, N. X. and Epps, J., 2009. A novel approach for automatic number of clusters detection in microarray data based on consensus clustering. 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE) 0, pp. 84–91.

Vinh, N. X., Epps, J. and Bailey, J., 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. J. Mach. Learn. Res. 9999, pp. 2837–2854.