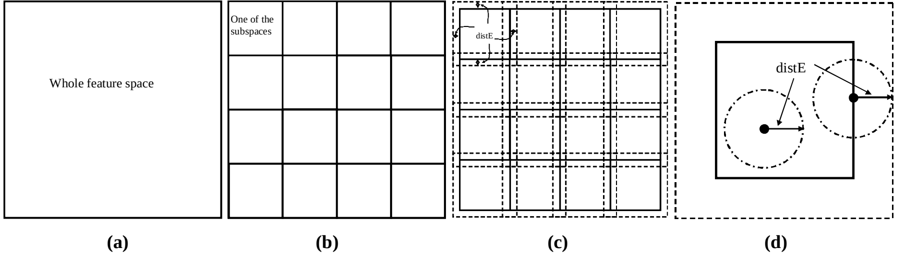
Higher is Better: High Dimensional Feature Space Analysis for Tissue Segmentation

P. A. Narayana[1], R. He[1]

[1]Radiology, University of Texas at Houston, Houston, Texas, United States

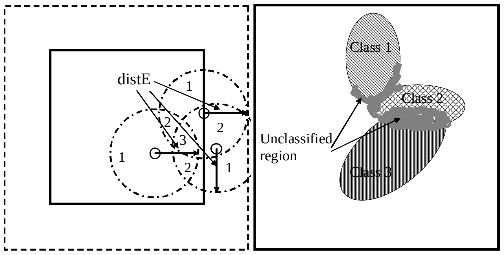
Background: Feature map-based segmentation techniques are attractive because of their speed and simplicity. However, feature space analysis is most commonly limited to two-dimensions and rarely to three dimensions. Full exploitation of MR images with multiple contrasts for accurate classification of heterogeneous pathology requires high dimensional ((≥4) feature space analysis. However, generation of high dimensional feature maps poses severe problems because of the increased memory requirements and the associated computational complexity. KNN (k nearest neighbor) scheme is commonly used for generation of feature maps. However, existing fast KNN methods can not be directly applied to higher dimensional feature space. In this work we present a method for generation and classification of high dimensional feature space that can be implemented on conventional PC's and applied it segment four echo based MR images of multiple sclerosis (MS) brains.

Methods: The feature space is divided into multiple hypercubes to reduce the memory requirements. This can be understood using a 2D case as shown in Fig. 1(a). The feature space is divided into subspaces (smaller squares as shown in Fig. 1(b). The classification of one subspace would be affected by the training seeds in and around that subspace. Applying the effective distance of the seed (distE) we form the overlap between subdivisions (Fig. 1 (c)). In Fig. 1(d), the feature vector (solid dot) inside the solid square can be classified by the training seeds contained within the circle of radius distE which is centered on the feature vector. In order to classify the feature vectors that are close to the boundary of the subspace, we need to examine the seed points outside the solid square (Fig. 1d), but confined within the overlapping region. This substantially reduces the data size that needs to be processed simultaneously. To simplify the computational complexity of KNN, three rules are applied. 1) Detect the number of seeds within the circle of radius distE and centered on a feature vector; discard if the number of seed points in this circle is smaller than K. 2) If the number of seeds within the circle is equal to K and if the number of points belonging to the same class is > K/2, then the point is automatically classified (winner takes all). 3) If the number of points, K', is greater than K but if there is one class that contains greater than K/2 points (winner), the class is automatically determined to be the same as the winner. Therefore, only those points in the feature space that have at least K seed prototypes need to be classified. Since the whole feature space is divided, we can concentrate only on one subspace at a time. Using circles of radius, distE, centered on each of the seed prototypes (small solid circles in Fig. 2(a)), we can determine the feature vectors that need to be classified by retaining those which are contained within the solid square and overlapping regions covered at least by K circles according to rule 1 above. This concept is described in Fig. 2 which shows three seed prototypes in three circles. In this figure "1" represents the region covered by only single circle, "2" denotes the overlapping region covered by two circles, and "3" denotes the region of intersection of all the three circles. If K equals to 1, then the region to be classified will be all those points with a value of at least 1. If K equals 2, then the region to be classified will be all those points with a value of at least 2. In this way, the number of feature vectors to be classified can be reduced considerably. Rules 2 and 3 are applied on those feature vectors that are retained after applying rule 1. If there are exactly K circles that cover this feature vector, then the feature vector can be automatically classified by counting the number in each class of those seeds and assigning it to the same class as the majority of the seed prototypes. If there are K'> K circles covering the feature vector to be classified, the feature vector is assigned to the same class as those seeds that are at least (K'-K/2) in number. Based on these criteria, the numbers of feature vectors that still need to be classified are restricted to those located at the interface (Fig 2(b)).



(a) (b) (c) (d)

Fig. 1 (a) Whole feature space inside solid line, (b) subdivision of feature space. (c) Seed prototypes inside the overlapped dash line regions will affect the boundaries of a subdivision of the feature space. (d) The region required for determining the class of feature vector inside one subspace. Black dots represent feature vectors.



(a) (b)

Fig. 2 (a) Procedures for reducing the number of feature vectors for KNN classification, seed prototypes are indicated using small circles, and (b) results after the reduction of feature vectors in whole subsapace.

Then we are left with the classification of vectors that are at the interface (unclassified region in Fig. 2(b)). Thus the only seed prototypes that need to be considered are those located in the region that is expanded by distE on the unclassified regions (see Fig. 3, regions confined in dash line). Therefore both the feature space that needs to be scanned for classification and the seed involved in classification is greatly reduced. Fast KNN is applied for all reset feature vectors using rest seeds for each subspace. The above algorithm was implemented on a standard Dell Dimension 8300 PC, equipped with a 1.8 GHz processor and 1Gb RAM. The total computational time for generating the four-dimensional feature map was about two hours.

Results: The above analysis was applied to the 4 echo MR images of MS brain acquired using the *affirmative* pulse sequence [2]. The segmented image along with the four affirmative images are shown in Fig. 4. In the segmented image, pink, gray, blue, and yellow represent WM., GM, CSF, and lesions, respectively. Using this method, we are currently working on the classification of MS lesions.

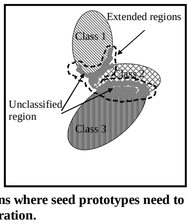
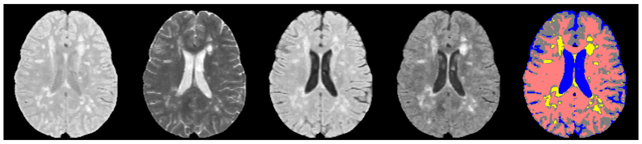


Fig. 3 Regions where seed prototypes need to be taken into consideration.



(a) (b) (c) (d) (e)

Fig. 4 AFFIRMATIVE images and segmentation, (a) echo1, (b) echo 2, (c) echo 3, (d) echo4, and (e) The segmentation using 4D feature map.

References: Cline HE et al, *JCAT*, 14(6):1037–1045, 1990.
Warfield S, *Pattern Recognition Letter*, 17:713-721, 1996
Bedell BJ and Narayana PA, *Magn. Reson. Med*. 39:961-969, 1998.
Clarke LP et al, *Magnetic Resonance Imaging*, 11:95–106, 1993.