Annales UMCS Informatica AI 6 (2007) 73-84

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI

http://www.annales.umcs.lublin.pl/

Hierarchical cluster analysis methods applied to image segmentation by watershed merging

Jakub Smołka*

Institute of Computer Science, Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract

A drawback of watershed transformation is over-segmentation. It consists in creating more classes than there are objects present in the image. Over-segmentation partially results from the fact that the transformation extracts almost all edges present in the image, even those which are very weak. To alleviate this problem images are preprocessed: blurring (or selectively blurring) filter is applied before the edge detection performed by a gradient filter. Additionally, the resulting image may be thresholded in order to eliminate small gradient values.

This paper presents an alternative solution to this problem. The solution uses the hierarchical cluster analysis methods for joining similar classes of the over-segmented image into a given number of clusters. First, it calculates attribute values for each class. Second optionally, the values are standardized. Third, cluster analysis is performed. The resulting similarity hierarchy allows for simple selection of the number of clusters in the final segmentation.

Several clustering methods, including the Complete Linkage and Ward's method along with many similarity/dissimilarity measures have been tested. The selected results are presented.

1. Introduction

Watershed transformation is a method that mimics pouring water onto a landscape created on a basis of a digital image. Unfortunately, it produces a region for each of the image's local minima so, usually, the number of these regions called watersheds or catchment basins are significantly larger than the number of objects depicted in the image – the image is over-segmented [1]. However, the advantage of transformation is a very good edge extraction. The hierarchical clustering methods allow for grouping objects using various criteria and kinds of attributes. This paper presents an attempt to solve the over-segmentation problem by using the cluster analysis methods.

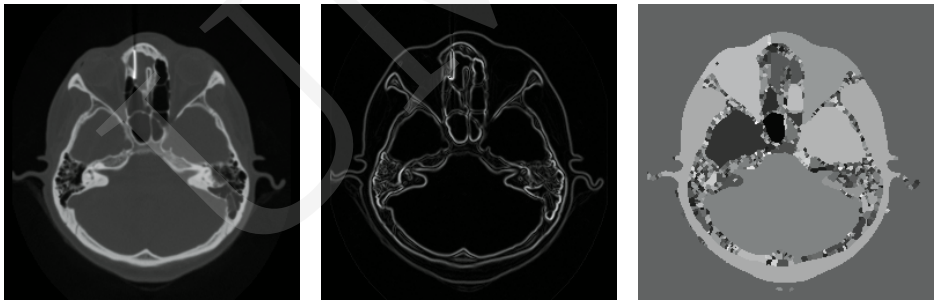
* *E-mail address*: Jakub.Smolka@pollub.pl

2. Watershed transformation

Watershed transformation (also called watershed segmentation) was introduced by Beucher and Lantuejoul in 1979 [2]. Its principle is very straightforward and can be easily explained by analogy to rain pouring onto a landscape. When rain pours onto a landscape, it flows with gravity to collect in low catchment basins. The size of those basins grows as the amount of precipitation they receive increases. At a certain point basins start spilling into one another, causing small basins to merge together into larger basins. To prevent them from merging, the transformation algorithm starts building dams between them.

As mentioned above the watershed transformation treats an image $I(x)$ as a height function that describes a landscape. It assumes that higher pixel values indicate the presence of boundaries in the original image $f(x)$.

The gradient operator satisfies this requirement and for this reason is often used for obtaining the height function $I(x) = |\nabla f(x)|$. An example is given in figures 1a and 1b.



(a) (b) (c)

Fig. 1. Obtaining the height function I and over-segmentation: (a) original image, (b) height function, (c) over-segmented image

The gradient operator is very sensitive to noise. Applying watershed transformation to the gradient image without prior preprocessing would cause significant over-segmentation. This is why the original image is usually smoothed before the use of the gradient filter [3-5]. An edge preserving smoothing filter [3-5] such as a gradient diffusion filter [6] is good for this purpose. Despite smoothing, the height function $I(x)$ usually still has many local minima (each local minimum corresponds to one catchment basin) and causes over-segmentation to occur [1,7]. Over-segmentation is illustrated in figure 1c. To alleviate this problem, one can establish a minimum watershed depth or threshold the gradient image $I(x)$ prior to segmentation, using a small threshold

value [1]. These solutions, however, are very simple and usually do not give satisfactory results.

3. Hierarchical cluster analysis in watershed merging

The hierarchical cluster analysis methods are very flexible and can be used to solve a great variety of problems [8,9], including those in computer graphics. Watersheds can be treated as objects with a set of attributes whose similarity can be measured with a number of coefficients. They, in turn, can be used to perform clustering.

Watershed merging by means of hierarchical clustering methods begins with specifying a set of attributes that will be the basis for grouping watersheds. There are a number of different attributes that can be used. In this preliminary study four different kinds have been used, as described below. Once the attributes to be used are known, they are calculated for each watershed. Due to the fact that the attributes usually represent different characteristics and as a consequence, their values may differ by orders of magnitude, they need to be standardized. This prevents one attribute from dominating others. Four different standardization equations were used for testing and are described below. Once all the data has uniform values, the clustering method can be executed. Such a method creates a similarity hierarchy which can be represented as a tree. Each node in the tree represents a merger of two clusters and has their value of similarity (or dissimilarity) measure associated with it. The clustering method stops executing when all watersheds are in one cluster. Of course, a single cluster is not what one usually wants. However, with the similarity tree available, obtaining the required number of classes comes down to picking the appropriate value of similarity (or dissimilarity) measure. This can be thought of as cutting the tree in two parts: top and bottom. The top part (containing final clustering steps) is left out and the bottom determines how the watersheds are merged. A piece of software that allows the user to interactively select the number of classes (after the execution of a clustering method) was created for testing this concept.

4. Attributes of watershed

In the initial research on the usability of hierarchical cluster analysis for watershed merging only four different attributes were used. Future work will include more kinds of attributes. There are plans to include more texture sensitive and some shape dependent characteristics.

The simplest of all the attributes is *watershed size*. It is expressed as the number of pixels the watershed consists of. It is included because, in some cases, not all classes in the image are over segmented. Figure 2b shows an example of

a CT scan where mostly bone is over segmented. Clustering based on watershed size should allow fragments of bone tissue to be grouped properly. The second kind of watershed attribute is *mean value*. An example is shown in figure 2c. The last two attributes are *variance*

$$\sigma_i^2 = \sum_{j=1}^{t} \frac{\left(X_{ij} - \bar{X}_i\right)}{\left(t-1\right)} \tag{1}$$

and *standard deviation* σ_i. Since they are similar they will be described together. These two attributes are, to some extent, sensitive to image texture. Even regions that are almost identical with respect to their mean value (figure 2c) may be quite different when their variance or standard deviation is taken into account (figures 2d and 2e).

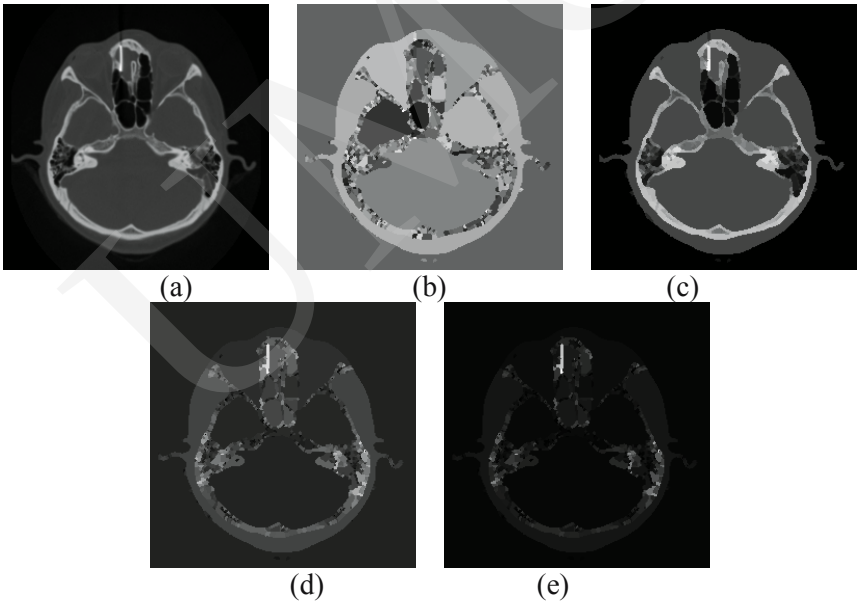


Fig. 2. Attributes: (a) original, (b) watersheds, (c) mean values, (d) standard deviation, (e) variance

5. Standardization methods

Attributes may take values from different ranges. They may differ even by several orders of magnitude. To prevent one attribute from dominating others standardization is needed. Four standardization methods were tested with clustering watersheds. In the equations below the following symbols are used: *j, k* – number of object, *t* – number of objects, X_{ij} – value of *i-th* attribute of *j-th* object, X_{ik} – value of *i-th* attribute of *k-th* object, $X_{min\ i}$, $X_{max\ i}$ – minimum and maximum values of *i-th* attribute.

Using the first equation:

$$Z_{ij} = \frac{\left(X_{ij} - \bar{X}_i\right)}{S_i}, \tag{2}$$

$$\bar{X}_i = \sum_{j=1}^{t} \frac{X_{ij}}{t}, \tag{3}$$

$$S_i = \sqrt{\sum_{j=1}^{t} \frac{\left(X_{ij} - \bar{X}_i\right)^2}{\left(t-1\right)}}, \tag{4}$$

causes the standardized attribute data to have a mean value of $\bar{Z}_i = 0$ and a standard deviation of $\bar{S}_i = 1$ [8]. The next two equations

$$Z_{ij} = \frac{X_{ij}}{X_{max\ i}}, \tag{5}$$

and

$$Z_{ij} = \frac{\left(X_{ij} - X_{min\ i}\right)}{\left(X_{max\ i} - X_{min\ i}\right)}, \tag{6}$$

scale the data [8] to the range of [0,1]. In the first case the maximal value in the original data corresponds to 1 after standardization. In the case of the second equation not only the largest value corresponds to 1 but also the lowest value corresponds to 0. The last equation

$$Z_{ij} = \frac{X_{ij}}{\sum_{j=1}^{t} X_{ij}}, \tag{7}$$

normalizes the data so that it sums up to 1 [8].

6. Similarity and dissimilarity measures

As was mentioned earlier, several similarity and dissimilarity measures have been used with the clustering methods for grouping watersheds. In the equations below the following symbols are used: *j, k* – numbers of object, *n* – number of attributes, X_{ij} – value of *i-th* attribute of *j-th* object, X_{ik} – value of *i-th* attribute of *k-th* object.

The *Euclidean distance coefficient* [8] given by

$$e_{jk} = \sqrt{\sum_{i=1}^{n} \left(X_{ij} - X_{ik}\right)^2} \tag{8}$$

is a dissimilarity measure. It represents the distance between two points in the n-dimensional space.

The *average Euclidean distance coefficient* [8]

$$d_{jk} = \sqrt{\sum_{i=1}^{n} \frac{\left(X_{ij} - X_{ik}\right)^2}{n}} \tag{9}$$

differs from e_{jk} in that it is able to compensate for missing values. If an object is missing an attribute value, then the attribute of the other object (even if present) is left out and n (the number of attributes) is decreased accordingly.

Another dissimilarity measure is the *coefficient of shape difference* [8]

$$z_{jk} = \sqrt{\frac{n}{(n-1)}\left(d_{jk}^2 - q_{jk}^2\right)}, \tag{10}$$

where

$$q_{jk}^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} X_{ij} - \sum_{i=1}^{n} X_{ik}\right)^2 \tag{11}$$

and d_{jk} is the average Euclidean distance. It compares profiles of two objects. However, it is insensitive to additive translation.

The *cosine coefficient* [8]

$$c_{jk} = \frac{\sum_{i=1}^{n} X_{ij} X_{ik}}{\sqrt{\sum_{i=1}^{n} X_{ij}^2 \sum_{i=1}^{n} X_{ik}^2}} \tag{12}$$

is a similarity that can be viewed as a cosine of an angle between two vectors in n-dimensional space and since the length of vectors is irrelevant this coefficient ignores proportional translations.

The *correlation coefficient* [8]

$$r_{jk} = \frac{\sum_{i=1}^{n} X_{ij} X_{ik} - \frac{1}{n}\sum_{i=1}^{n} X_{ij} \sum_{i=1}^{n} X_{ik}}{\sqrt{\left[\sum_{i=1}^{n} X_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_{ij}\right)^2\right]\left[\sum_{i=1}^{n} X_{ik}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_{ik}\right)^2\right]}} \tag{13}$$

is also known as the *Pearson product-moment correlation coefficient* and is the only one (among those mentioned in this paper) to ignore both additive and proportional translations.

The *Canberra metric coefficient* [8]

$$a_{jk} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|X_{ij} - X_{ik}\right|}{X_{ij} + X_{ik}} \tag{14}$$

equalizes the contribution of each attribute to overall similarity.

The last dissimilarity measure presented here is the *Bray-Curtis coefficient* [8]:

$$b_{jk} = \frac{\sum_{i=1}^{n}\left|X_{ij} - X_{ik}\right|}{\sum_{i=1}^{n}\left(X_{ij} + X_{ik}\right)} \tag{15}$$

Unlike the Canberra metric it allows one attribute to be dominant.

7. Hierarchical cluster analysis methods

Four clustering methods were tested: *single linkage* (SLINK), *complete linkage* (CLINK), *unweighted pair-group method using arithmetic averages* (UPGMA) and *Ward's minimum variance* method. The first three are very similar [8]. The following provides a generic description of these methods.

```
c:=getNumberOfWatersheds(W);
W:=standardizeData(c,W);
//dissimilarity matrix may be computed instead
S:=computeSimilarityMatrix(c,W);
while (c>1)
   //saves results to first, second and d
   findMostSimilarWatersheds(S,first,second,d);
   //updates the similarity hierarchy (combines two clusters)
   addToTree(first,second,d);
   //calculates similarity measures for the new cluster
   updateSimilarityMatrix(S,first,second);
   c:=c-1;
end while;
```

where: W – array containing attributes of watersheds, c – current number of clusters (initially number of watersheds), S – similarity (or dissimilarity) matrix, *first/second* – numbers of clusters to be combined, d – similarity (or dissimilarity) measure of clusters to be combined

First, the algorithm determines the number of watersheds to be grouped in clusters and, if needed, standardizes the data using the appropriate method. Second, it computes the similarity (or dissimilarity) matrix using one of the

coefficients described above. The coefficients can easily be exchanged. At the beginning, each cluster consists of one watershed. If there are more than one watershed, the algorithm begins clustering. It finds the two most similar clusters and the value of their similarity (or dissimilarity) measure. Third it merges those clusters by updating the tree which represents the clustering hierarchy. Fourth the algorithm updates the similarity (or dissimilarity) matrix. This process is continued until all watersheds are in the same cluster. The SLINK, CLINK and UPGMA methods differ only in the way distances between the newly created cluster and the remaining clusters are determined. With the SLINK method the distance between two clusters is the same as the distance between two of their most similar components (i.e. watersheds). The CLINK method is the opposite of SLINK because it assumes the distance between the two clusters to be the distance between the two most dissimilar components. UPGMA gives the intermediate distance values because it averages the similarity measures of all possible pairs of components (a pair must consist of components belonging to different clusters). The following table summarizes the differences between the three methods.

Table 1. Differences between clustering methods

	dissimilarity measure	similarity measure
SLINK	$d_{C_1C_2} = \min_{\substack{m\in C_1 \\ n\in C_2}} d_{nm}$	$s_{C_1C_2} = \max_{\substack{m\in C_1 \\ n\in C_2}} s_{nm}$
UPGMA	$d_{C_1C_2} = \frac{1}{n_{C_1} \cdot n_{C_2}} \sum_{\substack{m\in C_1 \\ n\in C_2}} d_{nm}$	
CLINK	$d_{C_1C_2} = \max_{\substack{m\in C_1 \\ n\in C_2}} d_{nm}$	$s_{C_1C_2} = \min_{\substack{m\in C_1 \\ n\in C_2}} s_{nm}$

where: C_1, C_2 – clusters, m – object belonging to cluster C_1, n – object belonging to cluster C_2, d_{nm} – dissimilarity measure of objects m and n, s_{nm} – dissimilarity measure of objects m and n, $d_{C_1C_2}$ – dissimilarity measure of clusters C_1 and C_2,– $s_{C_1C_2}$ similarity measure of clusters: C_1 and C_2, n_{C_1} – number of objects in cluster C_1, n_{C_2} – number of objects in cluster C_2.

Ward's minimum variance method differs more significantly from SLINK, CLINK and UPGMA. It is a method in which the fusion of two clusters is based on the size of an error sum of squares criterion [9]. Instead of searching for most similar clusters, the algorithm attempts to find a merger that will cause a minimal increase in the total within-cluster error sum of squares, E, given by

$$E = \sum_{m=1}^{c} E_m \,, \tag{16}$$

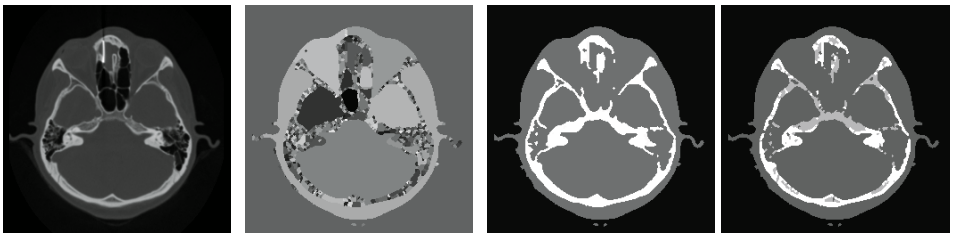
$$E_m = \sum_{l=1}^{t_m} \sum_{i=1}^{n} \left(X_{ilm} - \bar{X}_{im} \right)^2 \tag{17}$$

where: E – total error sum of squares, E_m – error sum of squares of *m-th* cluster, c – number of clusters, n – number of attributes, t_m – number of objects in *m-th* cluster, $\bar{X}_{im}$ – the average value of *i-th* attribute in *m-th* cluster, X_{ilm} – value of the *i-th* attribute of the *l-th* object in the *m-th* cluster.

As a result, in each step the algorithm has to check all possible mergers. In addition this method has a built-in dissimilarity measure and does not require any of the above-mentioned coefficients.

8. Results

The following presents selected preliminary results obtained with software created for the purpose of testing the concept described above. Four different types of images were used: CT, MRI T1, MRI T2 and MRI PD scans. All images were preprocessed with a gradient diffusion filter [6] in order to reduce noise without blurring the edges. The gradient filter was applied to extract edges. The resulting image was, in turn, thresholded with a threshold value equal to 5% (for CT) or 2% (for MRI) of maximal brightness in the gradient image. Despite these steps, the images were still over-segmented after the watershed transformation. This has confirmed the need to use cluster analysis.



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

Fig. 3. CT scan (the scan comes from The Visualization Toolkit's example data package) (a) original, (b) over-segmented image (885 watersheds), (c) CLINK method with Euclidean distance coefficient, standardization eq. 5, 4 classes, attributes: average, standard deviation, size, (d) Ward's method, standardization eq. 2, 4 classes, attributes: average, variance, size

During testing the complete linkage (CLINK) and Ward's methods usually gave better results than the single linkage (SLINK) and the unweighted pair-group method using arithmetic averages (UPGMA). SLINK failed in almost all cases whereas CLINK and Ward's methods were frequently successful. In most

successful segmentations, the CLINK method was used in combination with the Euclidean distance coefficient although the shape difference, Bray-Curtis and cosine coefficients produced their share of good results (sometimes with the UPGMA method). Other coefficients failed frequently.

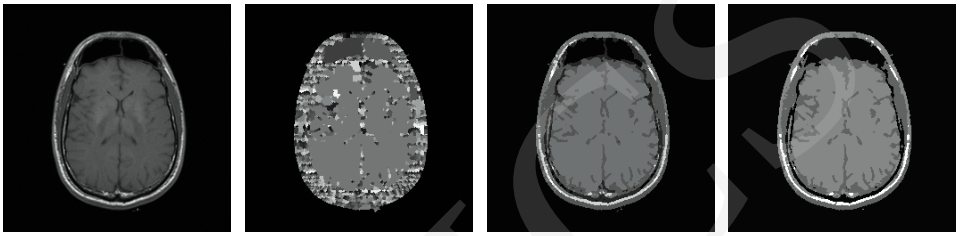


Fig. 4. MRI T1 scan (m_vm10xx data set from the Visible Human Project was used) (a) original, (b) over-segmented image (1466 watersheds), (c) CLINK method with the Euclidean distance coefficient, no standardization, 6 classes, attributes: average, (d) Ward's method, no standardization, 4 classes, attributes: average

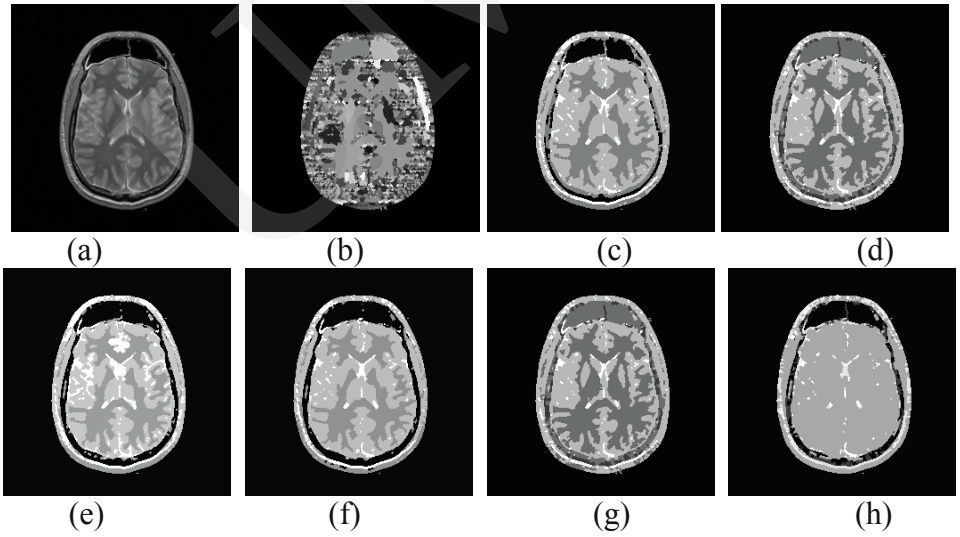


Fig. 5. MRI PD scan (m_vm10xx data set from the Visible Human Project was used) (a) original, (b) over-segmented image(2079 watersheds), (c) CLINK method with the Euclidean distance coefficient, standardization eq. 5, 8 classes, attributes: average, variance, size, (d) CLINK method with the average Euclidean distance coefficient, standardization eq. 5, 6 classes, attributes: average, standard deviation, size, (e) Ward's method, no standardization, 4 classes, attributes: average, (f) Ward's method, standardization eq. 5, 4 classes, attributes: average, variance, size (g) CLINK method with the Euclidean distance coefficient, standardization eq. 5, 6 classes, attributes: average, standard deviation, size, (h) CLINK method with the Euclidean distance coefficient, standardization eq. 5, 6 classes, attributes: average, standard deviation

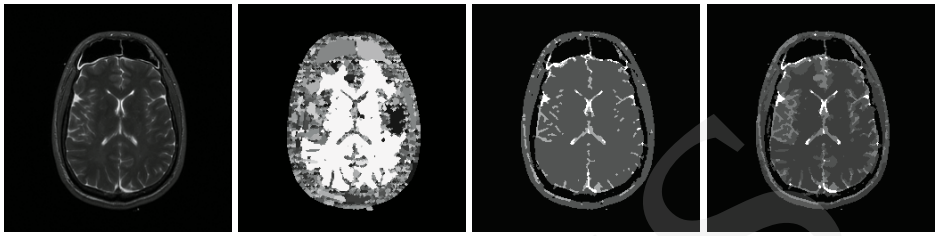


Fig. 6. MRI T2 scan (m_vm10xx data set from the Visible Human Project was used) (a) original, (b) over-segmented image (1773 watersheds), (c) Ward's method, standardization eq. 5, 4 classes, attributes: average, variance, (d) Ward's method, standardization eq. 6, 6 classes, attributes: average, variance

Four different attribute sets were used in testing: (1) watershed's average, (2) average, standard deviation, size, (3) average, variance, size, (4) average, variance. Adding certain attributes may or may not improve segmentation quality. A case where adding an attribute had a positive effect is illustrated in figures 5g and 5h. One of these segmentations included size, the other did not. The other parameters remained the same. When multiple attributes were used, the data was standardized. Equations 5 and 6 were used most frequently, but equation 2 was utilized for performing successful segmentation as well.

9. Summary

The results presented above show that the hierarchical cluster analysis methods can be used for watershed merging. The use of watersheds ensures that the boundaries of objects in the image overlap those of the classes in the final segmentation. The cluster analysis methods can take into account many different attributes. The attribute sets can easily be modified. Such modification may have significant influence on the segmentation quality (figures 5g and 5h). Among the clustering methods tested, CLINK and Ward's methods (combined with the Euclidean distance coefficient and standardization equation 5) give good results most frequently. However, the UPGMA method, the Bray-Curtis and the cosine coefficients are also useful. Future plans include: 1) introducing more kinds of watershed attributes, 2) statistical analysis of the usability of different clustering methods, 3) testing of different standardization/ similarity measure/attribute set combinations.

References

[1] Ibanez L., Schroeder W., Ng L., Cates J., *The ITK Software Guide*. Kitware Inc., (2003).
[2] Beucher S., Lantuejoul C., *Use of watersheds in contour detection*. International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, (1979).

[3] Gonzalez R. C., Woods R. E., *Digital Image Processing*. Addison-Wesley Publishing Company, (1993).
[4] Myler H. R., Weeks A. R., *The Pocket Handbook of Image Processing Algorithms in C*. Prentice Hall PTR.
[5] Seul M., O'Gorman L., Sammon M. J., *Practical Algorithms for Image Analysis: Description, Examples, and Code*. Cambridge University Press, (2000).
[6] Perona P., Malik J., *Scale-space and edge detection using anisotropic diffusion*. IEEE Transactions on Pattern Analysis Machine Intelligence, 12 (1990) 629.
[7] Beucher S., *The watershed transformation applied to image segmentation*. Scanning Microscopy International, 6 (1992) 299.
[8] Romesburg Ch., *Cluster Analysis for Researchers*. Lulu Press, (2004).
[9] Everitt B. S., Landau S., Leese M., *Cluster Analysis*. Arnold a member of the Hodder Headline Group, (2001).