ACCURACY ANALYSIS OF CLUSTERING ALGORITHMS FOR SEGMENTATION OF INDUSTRIAL CT IMAGES

Abstract: Image analysis today plays an important role in the field of industrial engineering for image processing. The use of different tools in order to extract important information from industrial CT images is an active field of research that attracts the attention of many researchers. In this paper two methods will be presented that are used for segmentation of CT images. Those methods are fuzzy C-means clustering and K-means clustering. Comparative analysis will be performed in the form of accuracy of segmented binary images. Presented analysis leads to the conclusion that fuzzy C-means clustering has greater accuracy than hard K-means clustering approach.

Key words: Segmentation, Industrial CT, fuzzy clustering, hard clustering

1. INTRODUCTION

The application of modern computer – aided technologies in different fields allows vast growth, large and increasing number of industrial applications of X-ray Computed Tomography (CT) in the manufacturing industry as well as in other industries such as aviation, railway, military industry and others [1,2].

CT technology is used to acquire two dimensional (2D) images of an object and then from these images a three dimensional (3D) model of the object’s external and internal structure is reconstructed and which can then be analyzed. Therefore, CT became very popular for its use in material analysis because of its non-destructive testing (NDT) ability and the detection of flaws such as voids and cracks, and particle analysis in materials. CT is the only technology able to measure as well the inner as the outer geometry of a component without need to cut it through and destroy it. As such, it is very useful for industrial quality control of workpieces having non-accessible internal features (e.g. components produced by additive manufacturing) or multi-material components (e.g. two-component injection molded plastic parts or plastic parts with metallic inserts) [3].

Many different methods for automatic extraction of important information from industrial CT images have been proposed. Feature extraction from CT images is mostly done through a preliminary segmentation step using methods such as region growing [4], fuzzy clustering [5] water shedding [6], morphological filter operations [7] and others. As a result, these methods provide a binary classification of individual volume voxels into segmented regions, meaning that a whole voxel either belongs to the region of a specific feature, or not.

2. METHODS

The segmentation process is widely spread and is applied in many applications such as algorithms for robot vision, registration of facilities, analysis of geographical areas and analysis of CT images [8,9]. In a more classical definition, segmentation is defined as a process of dividing image on non-overlapping, consistent homogenous regions, in accordance with certain characteristics such as intensity of grayscale or texture. As a result of the segmentation process a binary image is generated where the value 1 represents the region of interest, and 0 represents the background that has been excluded [10].

Cluster analysis is based on partitioning a collection of data points into a number of clusters, where the points inside a cluster show a certain degree of closeness or similarity. Clustering methods can be considered as either hard (crisp) or fuzzy, depending on whether a pattern belongs exclusively to a single cluster or to several clusters with different degrees of belongings. In hard clustering each point (voxel) of the dataset belongs to exactly one cluster, a membership value of zero or one is assigned to each pattern, whereas in fuzzy clustering, a value between zero and one is assigned to each point by a membership
K-means clustering method (also known as Lloyd's algorithm) is one of the most popular methods used for unsupervised classification of data. It is working on the principle of classification data in a predefined number of clusters (shown in figure 1) [11].

Fuzzy C-means clustering algorithm is widely used in the field of image processing. It is used for clustering, or grouping of data where each data belongs to a cluster with a certain degree of belonging, unlike hard clustering, where each data belongs to only one cluster [12]. In the clustering process of CT images the most common parameter that is taken into account is the pixels intensity displayed in grayscale. In this way the FCM objective function is minimized when high membership values are assigned to pixels whose intensities are close to the centroid of its particular class, and low membership values are assigned when the point is far from the centroid [13]. With fuzzy c-means, the centroid of a cluster is computed as being the mean of all points, weighted by their degree of belonging to the cluster. The degree of being in a certain cluster is related to the inverse of the distance to the cluster. By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the "right" location within a data set [12].

It is necessary to define input parameters such as the number of clusters $C$ and weighted exponent $m$ which defines the degree of fuzzification or affiliation of pixels to clusters. When the weighted exponent $m$ is equal to 1 it is called a crisp clustering, and when $m>1$ the degree of fuzziness increases among points in the decision space. Figure 2 shows the basic scheme of fuzzy clustering [12].

The main difference between fuzzy and hard clustering is that with fuzzy clustering any pixel in the image can belong to several clusters, but with a certain degree of belonging to a given cluster, which moves in the interval from 0 to 1. This is an important trait in image analysis in order to maximize sensitivity.

3. RESULTS

Experimental results were carried out on six CT images which were taken from a stack of CT images of a motor block [14] shown in figure 3 and figure 4. The two presented algorithm (named FCM and HARD) have been implemented in Matlab software, on a computer (Fujitsu CELSIUS M470-2) with Intel (R) Xeon (R) CPU E5645, 2.40GHz processor and 16 GB of RAM.

![Fig. 3. 3D model of a motor block](image)

![Fig. 4. CT images used for analysis](image)
No additional postprocessing was carried out in order to show the real results in the application of these algorithms for processing of industrial CT images.

Input parameters that were used in these two algorithms are a number of clusters $C$ which are defined by the user manually, termination criterion $\varepsilon$ (accuracy), the maximum number of iterations and weighting exponent $m$ (only FCM algorithm). In order to extract as much information from the images in both approaches, after a short experimental analysis, optimal segmentation parameters in each of these industrial CT images were considered.

Input parameters that were used in these two algorithms are a number of clusters $C$ which the user enters manually, termination criterion $\varepsilon$ (accuracy), the maximum number of iterations and weighted exponent $m$ (only with FCM algorithm). In order to extract as much information from the images in both algorithms, after a short experimental analysis, optimal parameters were considered in each of these three industrial CT images.

In order to be able to perform accuracy analysis of the segmented images, each of these CT images was manually segmented. These manually segmented images were then used as reference images in relation to which were both fuzzy clustering and hard clustering segmented images compared (figure 5).

Parameters that are commonly used for the statistical analysis of segmented binary images are Jaccard index [15] and Dice coefficient [16]. They are used as evaluators of the segmented CT images. In these evaluators extent overlapping ranges from 0 to 1, where 1 defines a complete overlap of the segmented image with the reference image.

These evaluators are defined according to the following formula:

Jaccard index

$$JI = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Dice coefficient

$$DK = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

where $A$ is the result of segmentation of the reference image, and $B$ is the segmentation result of FCM/HARD algorithm that was compared with $A$.

Figure 5 shows the results of segmentation results by FCM and HARD algorithm.

On figure 6 and figure 7 are shown evaluation results by Jaccard index and Dice coefficient for all six CT images, segmented by both FCM and HARD algorithm, which were compared to their segmented reference images.

4. DISCUSSION

Based on the presented results it can be seen that the segmentation results using the FCM algorithm show better results compared to the results obtained...
using HARD algorithm. The results of evaluation by Jaccard index for all six CT images segmented by FCM algorithm are 0.961, 0.9984, 0.9657, 0.9439, 0.9481 and 0.977, respectively. And the results of evaluation by Jaccard index for all six CT images segmented by HARD algorithm are 0.9576, 0.9515, 0.9536, 0.9434, 0.9407 and 0.9199, respectively. Results of the evaluation by Dice coefficient for all six CT images with FCM algorithm are 0.9801, 0.9992, 0.9825, 0.9711, 0.9734 and 0.9884, respectively. And the results of evaluation for all six CT images segmented by HARD algorithm are 0.9783, 0.9515, 0.9763, 0.9709, 0.9694 and 0.9583, respectively. However, in cases 2 and 6 it can be seen that the results by Jaccard are much better for FCM than HARD algorithm. The reason why this could be is that by applying fuzzy clustering on these images, due to the vague areas on CT image, the FCM algorithm was able to extract more information from the images unlike HARD algorithm, which essentially separates data by them either belonging to certain cluster or not. This can be seen on results from both Jaccard and Dice.

5. CONCLUSION

In this paper is presented a comparative analysis of two different clustering methods for segmentation of industrial CT images. Segmentation was performed on six industrial CT images, and those results were compared with the results of the manual segmentation. As segmentation evaluators were used Jaccard index and Dice coefficient. Results for all six CT images segmented by FCM algorithm are showing very good results, thus showing the advantage of fuzzy clustering over hard clustering for segmentation of industrial CT images.

6. REFERENCES


Authors: Assist. M.Sc. Mario Sokac, Assoc. prof. dr Igor Budak, Prof. dr. Ralević Nebojša, M.Sc. Željko Santosić, University of Novi Sad, Faculty of Technical Sciences, Institute for Production Engineering, Trg Dositeja Obradovića 6, Novi Sad, Serbia, Phone: +381 21 485 2332, Fax: +381 21 454-495.
E-mail: marios@uns.ac.rs
budakj@uns.ac.rs
ralevic@uns.ac.rs
zeljkos@uns.ac.rs

Prof. dr. Mirko Soković. University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva 6, 1000 Ljubljana, Slovenia, Phone: +386 1 4771214
E-mail: mirko.sokovic@fs.uni-lj.si.

Note: The part of results presented in this paper have been obtained in the framework of the project "Research and development of modeling methods and approaches in manufacturing of dental recoveries with the application of modern technologies and computer aided systems,” TR - 35020, funded by the Ministry of Education, Science and Technological Development of Republic of Serbia.