

# Comparative analysis of segmentation algorithms for the allocation of microcalcifications on mammograms

Yu A Podgornova<sup>1</sup>, S S Sadykov<sup>1</sup>

<sup>1</sup>Murom Institute (branch) Federal state budgetary Educational Institution of Higher Education" Vladimir State University named after Alexander Grigoryevich and Nickolay Grigoryevich Stoletovs", Orlovskaya street, 23, Murom, Russia, 602264

e-mail: yuliyabulanova@yandex.ru

**Abstract.** Breast cancer is the most common disease of the current century in the female population of the world. The main task of the research of most scientists is the detection of this pathology at an early stage (the tumor size is less than 7 mm) when a woman can still be helped. An indicator of this disease is the presence of small-point microcalcifications, located in groups within or in the immediate circle of the tumor. Microcalcification is a small-point character at cancer, reminding grains of sand of irregular shape which sizes are from 100 to 600 microns. The probability of breast cancer increases with the increase in the number of microcalcifications per unit area. So, the probability of cancer is 80% if more than 15 microcalcifications on 1 sq. cm. The microcalcifications are often the only sign of breast cancer, therefore, their detection even in the absence of a tumor node could be a harbinger to cancer. Image segmentation is one way to identify microcalcifications. The conducted research allowed us to choose the optimal segmentation algorithms of mammograms to highlight areas of microcalcifications for further analysis of their groups, sizes, and so on.

## 1. Introduction

The mammary gland is a complex, sensitive organ that requires constant monitoring due to the annual increase in the incidence of breast cancer and its "rejuvenation" [1]. Often a symptom of serious diseases is small-point calcification, called microcalcifications (deposits of calcium) [1]. Usually, isolated microcalcifications or clusters are small in size, so they do not self-identify. For detection of this pathology is required to carry out hardware diagnostics such as ultrasound and mammography.

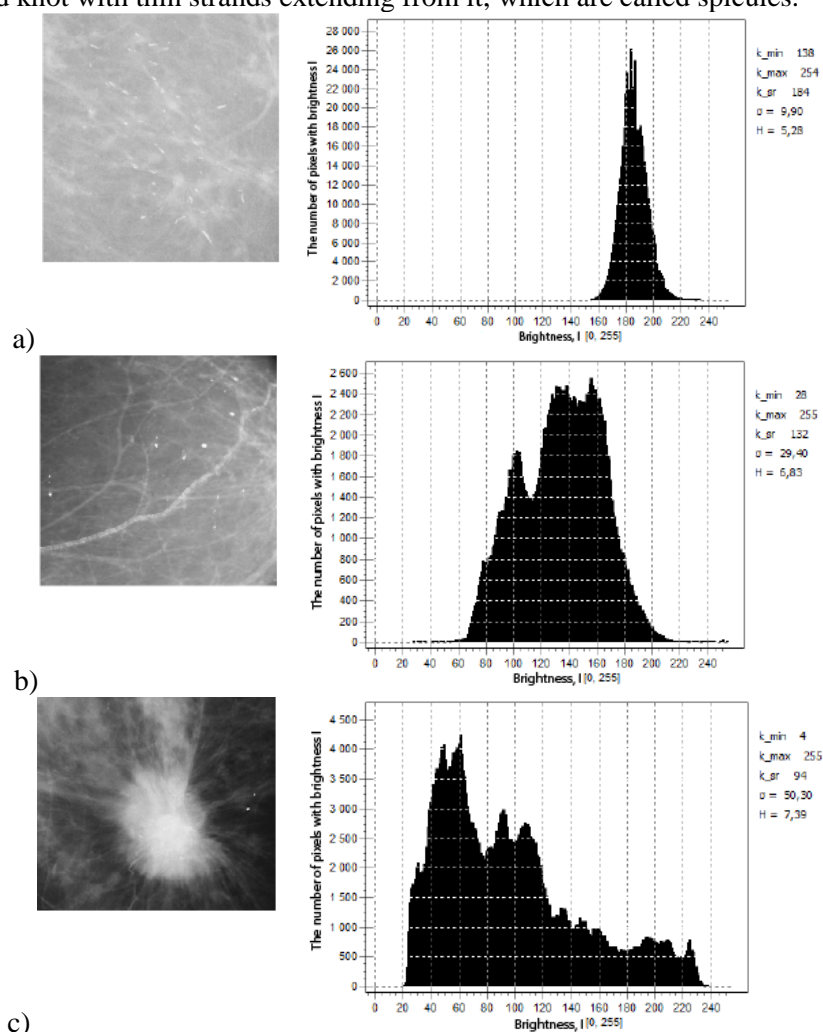
Mammography is a noninvasive method for the detection of pathologies of mammary glands [2].

Microcalcinates differ in localization, size, shape, concentration, quantity. Examples of microcalcifications are presented in figure 1. To assess all of these parameters will need to find and select the picture mammography. To evaluate all the specified parameters, it is necessary to find and highlight them in the mammography image. The process of finding homogeneous areas in an image is called segmentation. It is the first step in image analysis. Thus, segmentation [3] plays an important role in the processing of medical images. The main idea of the segmentation process is as follows: each pixel of the image can be associated with some visual properties, such as brightness, color, and texture. Within one object or one part of an object, these attributes change relatively little, whereas when crossing the border from one object to another, there is usually a significant change in the above attributes.

At the moment many image segmentation algorithms are developed [4], therefore the main task of this paper is the analysis of existing methods of segmentation and selection of an optimal algorithm for detection of microcalcifications in mammographic images.

There is the following classification of image segmentation algorithms [5]: threshold methods, region-based methods, edge detection methods, and clustering-based algorithms. Edge detection methods are not used in this work, as most are used to highlight the contours of the image. This article explores the following segmentation algorithms: algorithm FloodFill, the watersheds, MeanShift, and k-means.

Figure 1 shows examples of mammograms with different microcalcifications, in Figure 1(c), in addition to microcalcifications, there is also a malignant neoplasm that has fuzzy spiciform contours, i.e. a star-shaped knot with thin strands extending from it, which are called spicules.



**Figure 1.** Examples of forms, localization, and the number of microcalcifications on mammograms.

## 2. Overview of segmentation algorithms

### 2.1. Watershed method

The concept of watershed [7, 8, 9, 10] is based on the representation of the image as a three-dimensional surface defined by two spatial coordinates and the level of brightness as the height of the surface (relief). In this "topographical" interpretation, three types of points are considered: (a) points of the local minimum; (b) points located on the slope, i.e. from which water rolls down to the same local minimum; and (c) points located on the crest or peak, i.e. from which water is equally likely to roll down more than one such minimum. When applied to a specific local minimum, a set of points

that satisfy condition (b) is called a basin (or catchment area) of this minimum. The sets of points that satisfy condition (c) form the ridge lines on the surface of the relief and are called the watershed lines.

One of the most important applications of segmentation by watersheds is the selection of objects of uniform brightness in the background (in the form of spots). Areas characterized by small changes in brightness have small gradient values. Therefore, in practice, there is often a situation when the method of segmentation by watershed is applied not to the image itself, but to the gradient of this image. Under such conditions, the local minima of the basins agree well with the small gradient values, which usually corresponds to the objects of interest.

## 2.2. MeanShift segmentation

The main idea of this method [11, 12] is that the input image can be used to construct a nuclear estimate for the probability density of data distribution in the RGBXY feature space. Next, a natural assumption is made that the local maxima of the probability density corresponding to the cluster centers. From the necessary condition of a local extremum, an expression is determined for the shift vector  $m(p)$  of the feature space  $p \in \text{RGBXY}$ , applying which iteratively to the point  $p$  we get a sequence of points converging to the local maximum of the probability density estimate (i.e. to the center of the nearest cluster):

$$m(p) = \frac{\sum_{i=1}^n p_i \cdot g_i}{\sum_{i=1}^n g_i} - p,$$

where  $g_i = g\left(\left\|\frac{(p - p_i)}{h}\right\|^2\right)$ ,  $g(v) = -k'(v)$ ,  $h$  is smoothing parameter,  $K(v) = ck(v)$  is kernel estimates of a density.

## 2.3. FloodFill algorithm

Using the FloodFill method [11, 13, 14] it is possible to select areas of uniform color. To do this, select the starting pixel and set the interval for changing the color of neighboring pixels relative to the original. The interval can be asymmetric. The algorithm will combine pixels into one segment (filling them with one color) if they fall within the specified range. The output is a segment filled with a certain color and its area in pixels.

Such an algorithm can be useful for filling an area with weak color swings with a homogeneous background. One of the ways to use FloodFill is to detect damaged edges of the object. For example, if the algorithm fills neighboring regions by filling homogeneous areas with a certain color, then the integrity of the border between these areas is violated.

## 2.4. k-means algorithm

k-means segmentation [15, 16, 17, 18] is the most popular clustering method. The algorithm is aimed at minimizing the total quadratic deviation of cluster points from the centers of these clusters. Thus, this is an iterative algorithm that divides a given set of pixels into  $k$  clusters of points, which are as close as possible to their centers, and the clustering itself occurs due to the displacement of these same centers. It is necessary to take into account the fact that the k-means method is very sensitive to noise, which can significantly distort the results of clustering.

## 3 Experimental results

As criteria for evaluation of work of algorithms of segmentation, it is possible to use quality of background suppression and selection of objects in the form of connected areas.

Because microcalcifications are a complex object, it is impossible to demand accurate determination of the object consisting of several parts of different brightness as a single connected region.

For the analysis methods were taken real pictures of microcalcifications on mammograms. Images differ in the number and type of microcalcifications, brightness of the background and objects, and the presence of repetitive textures.

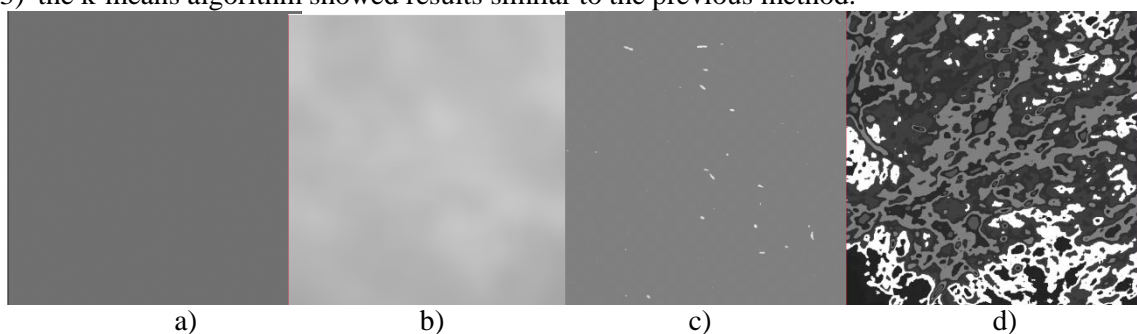
Figures 2-4 show the results of the algorithms on the original images. Only the FloodFill algorithm coped with the allocation of microcalcifications, all other algorithms have identified too many connected regions. Such experimental results suggest the need for using pre-processing methods before using segmentation.

Next to the images were applied contrasting methods, described in detail in [2].

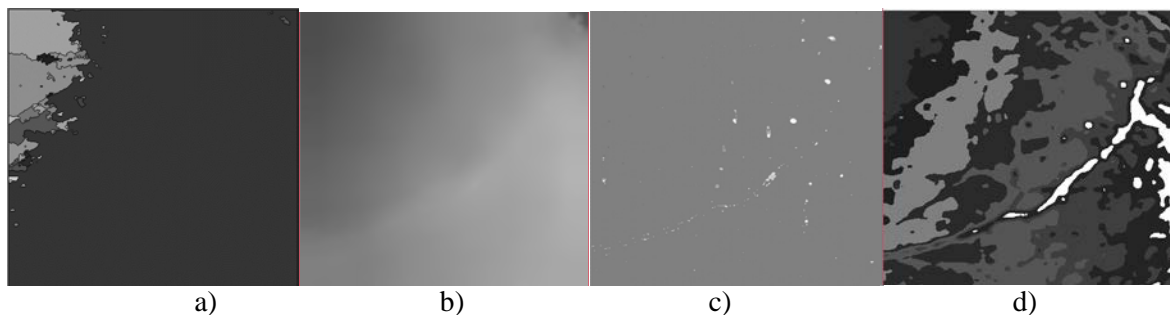
Figures 5-7 show examples of studies of segmentation algorithms on contrasted mammograms.

The results of the experiments are as follows:

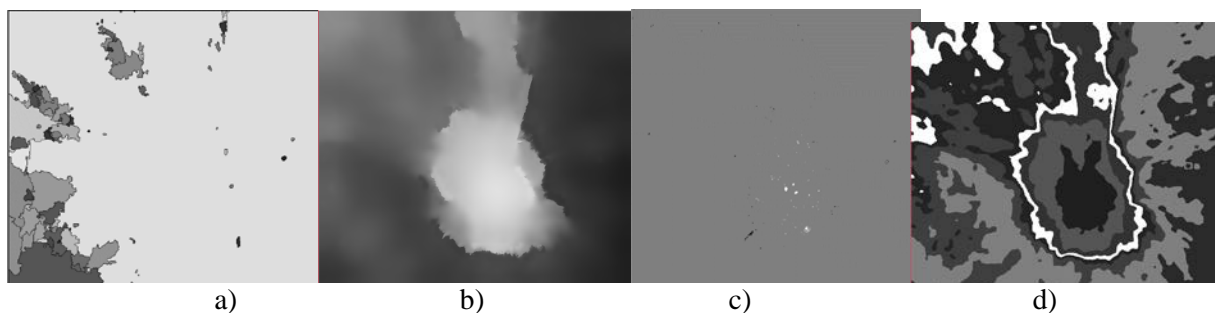
- 1) the watershed algorithm is not suitable for the solution of a task at all;
- 2) the MeanShift algorithm is able to allocate the required objects only in images without tumors on the background of fatty involution;
- 3) the k-means algorithm showed results similar to the previous method.



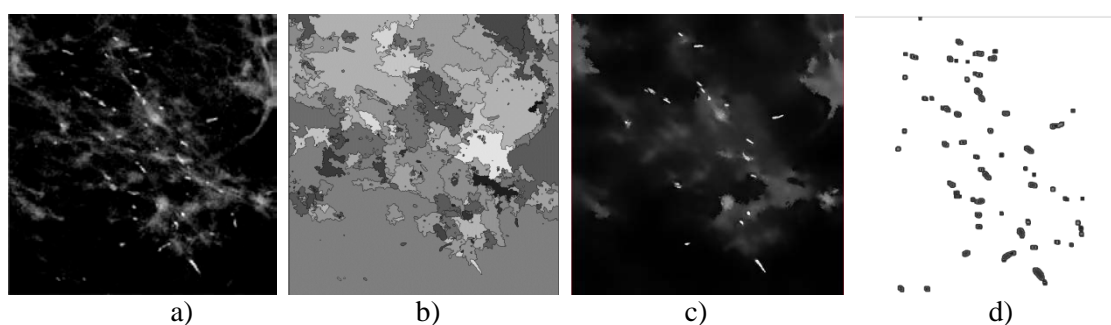
**Figure 2.** Examples of image segmentation shown in Figure 1(a): (a) watershed method, (b) MeanShift algorithm, (c) FloodFill algorithm, (d) k-means segmentation.



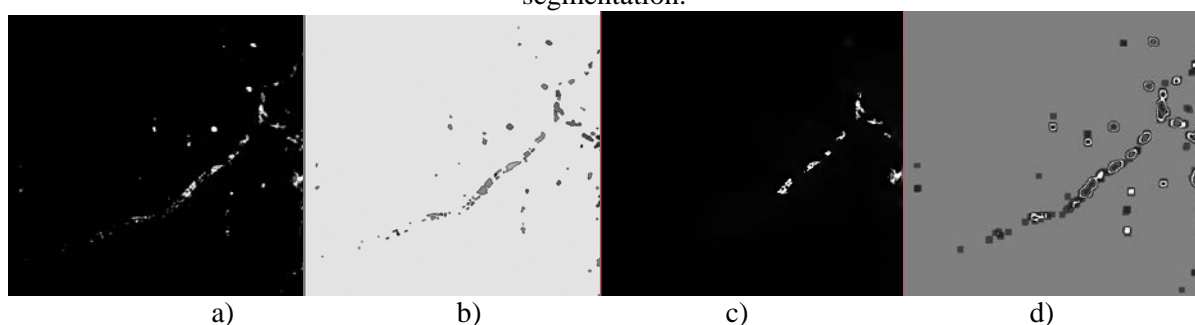
**Figure 3.** Examples of image segmentation shown in Figure 1(b): (a) watershed method, (b) MeanShift algorithm, (c) FloodFill algorithm, (d) k-means segmentation.



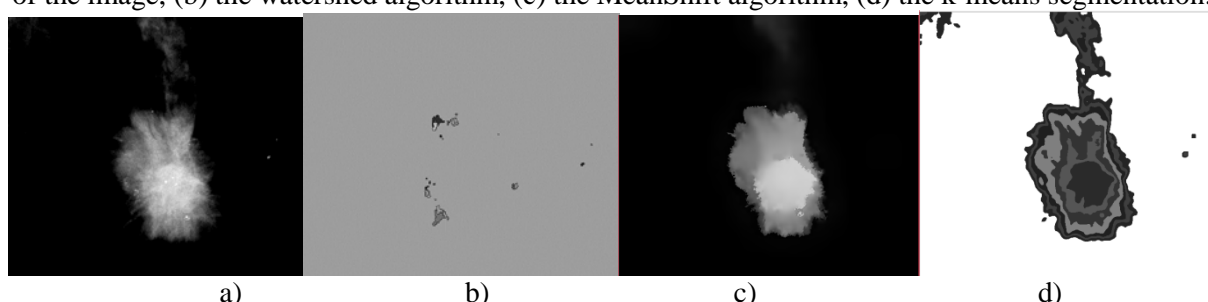
**Figure 4.** Examples of image segmentation shown in Figure 1(c): (a) watershed method, (b) MeanShift algorithm, (c) FloodFill algorithm, (d) k-means segmentation.



**Figure 5.** Examples of image segmentation shown in Figure 1(a): (a) changing the brightness/contrast of the image, (b) the watershed algorithm, (c) – the MeanShift algorithm, (d) the k-means segmentation.



**Figure 6.** Examples of image segmentation shown in Figure 1(b): (a) changing the brightness/contrast of the image, (b) the watershed algorithm, (c) the MeanShift algorithm, (d) the k-means segmentation.



**Figure 7.** Examples of image segmentation shown in Figure 1(c): (a) changing the brightness/contrast of the image, (b) the watershed algorithm, (c) the MeanShift algorithm, (d) the k-means segmentation.

The study of the work of the algorithms was carried out on 250 mammograms from the MIAS [2] database. Empirically managed to achieve the results presented in Table 1 and 2.

**Table 1.** The results of the segmentation algorithms on the original mammograms.

Algorithm	Number of correct detections (% of total images)	Number of false detections (% of total images)
Watershed algorithm	15 (6%)	235 (94%)
MeanShift algorithm	213 (85,2%)	37 (14,8%)
FloodFill algorithm	178 (71,2%)	72 (28,8%)
k-means segmentation	89 (35,6%)	161 (64,4%)

**Table 2.** The results of the segmentation algorithms on the processed mammograms.

Algorithm	Number of correct detections (% of total images)	Number of false detections (% of total images)
Watershed algorithm	45 (18%)	205 (82%)
MeanShift algorithm	98 (39,2%)	152 (60,8%)
k-means segmentation	107 (42,8%)	143 (57,2%)

#### 4. Conclusion

Comparative analysis of different methods of image segmentation applied to the problem of allocation of microcalcifications in mammographic images. To compare segmentation methods, criteria were used based on the expert's assessment (based on visual analysis) of the quality of background suppression and the selection of objects as connected areas. Through experimental studies, it was found that the method of watersheds incorrectly finds the boundaries of objects and is not acceptable in solving the problem.

The best segmentation result was obtained using the FloodFill algorithm, which consists of the selection of areas of uniform color. During the experiments, it was found that to improve the quality of mammogram segmentation, it is advisable to pre-process the images. It provides a reduction in the number of analyzed areas by combining segments and removing irrelevant fragments from the point of view of the problem. Using the same segmentation algorithms after processing the images showed that the MeanShift algorithms and k-means are able to highlight microcalcifications only on the images without tumors on the background of fatty involution.

It should be noted that further research is needed to improve the methods of thematic segmentation, taking into account the spatial properties of areas and providing the best compromise between insufficient and excessive segmentation.

The obtained results allow us to outline the prospects of using segmentation algorithms in the construction of automatic cancer detection systems on mammograms at an early stage.

#### 5. References

- [1] Korzhenkova G P 2004 *Comprehensive X-ray Sonographic Diagnosis of Breast Diseases* (Moscow: Firma STROM) p 128
- [2] Sadykov S S, Bulanova Yu A and Zaharova E A 2014 Computer diagnosis of tumors in mammograms *Computer Optics* **38(1)** 131-138
- [3] Eddaoudi F and Rezagui F 2011 Microcalcifications detection in mammographic images using texture coding *Applied Mathematical Sciences* **5** 381-393
- [4] Panchenko D S and Putyatin E P 1999 Comparative analysis of image segmentation methods *Radio electronics and informatics* **4** 109-114
- [5] Whitey D J and Koles Z J 2008 A review of Medical Image segmentation: Methods and available software *International Journal of Bioelectromagnetism* **10** 125-148
- [6] Doskolovich L L, Kharitonov S I, Petrova O I and Soifer V A 1998 A gradient method for design of multiorder varied-depth binary diffraction gratings – a comparison *Opt. And Lasers in Eng.* **29** 249-259
- [7] Sadykov S S, Bulanova Yu A, Zaharova E A and Yashkov V S 2013 Marker watershed study to isolate the breast cancer area *Algorithms, methods and data processing systems* **1** 56-64
- [8] Hagyard D and Razaz M 1996 Analysis of watershed algorithms for gray scale images *IEEE conf on image processing* **3** 41-44
- [9] Gauch J M 1999 Image segmentation and analysis via multiscale gradient watersheds *IEEE trans on image processing* **8** 69-79
- [10] Myasnikov E V 2017 Hyperspectral image segmentation using dimensionality reduction and classical segmentation approaches *Computer Optics* **41(4)** 564-572 DOI: 10.18287/2412-6179-2017-41-4-564-572
- [11] Comaniciu D, Ramesh V and Meer P 2000 Real-Time Tracking of Non-Rigid Objects Using Mean Shift *Conference on CVPR* **2** 1-8
- [12] Dingding Liu, Bilge Soran, Gregg Petrie, and Linda Shapiro 2019 *A Review of Computer Vision Segmentation Algorithms* (Washington: University of Washington)
- [13] Comaniciu D and Meer P 2002 Mean Shift: A Robust Approach toward Feature Space Analysis *IEEE Trans. Pattern Analysis and Machine Intelligence* **24** 603-619
- [14] Charles J J, Kuncheva L I, Wells B and Lim I S 2006 An Evaluation Measure of Image Segmentation Based on Object Centres *International Conference Image Analysis and Recognition ICIAR 2006: Image Analysis and Recognition* (Springer: Verlag Berlin Heidelberg)

- [15] Inaba M, Katoh N and Imai H 1994 Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering *Proceedings of 10th ACM Symposium on Computational Geometry* 332-339 DOI:10.1145/177424.178042
- [16] Steven J 2002 Acceleration of K-Means and Related Clustering Algorithms *Lecture Notes in Computer Science* (Springer Berlin Heidelberg) 166-177 DOI:10.1007/3-540-45643-0\_13
- [17] Nameirakpam D 2015 Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm *Procedia Computer Science* **54** 764-771
- [18] Tatarnikov V V, Pestunov I A and Berikov V B 2017 Centroid averaging algorithm for a clustering ensemble *Computer Optics* **41(5)** 712-718 DOI: 10.18287/2412-6179-2017-41-5-712-718

### **Acknowledgments**

Podgornova Yu A expresses her sincere gratitude to her academic advisor, Dr. of Tech. Sci., Professor Sadykov S S for his help at the research conduct, valuable recommendations in relation to their planning and article preparation as well as for his moral support.