steerable filters, mammograms,
moment and texture features,
feature extraction, classification

Ryszard S. CHORAŚ[1]

ANALYSIS OF SUSPICIOUS LESIONS IN DIGITAL MAMMOGRAMS

The system using steerable filters for analysis suspicious lesions in mammograms is proposed. This system is based on moments and texture features.

The set of well defined and classified suspicious lesions regions from mammograms database are used as a reference pattern. The similarity measure for reference pattern image and patient mammogram is found by computing the distance between their corresponding feature vectors. The Euclidean distance metric is used to finding the nearest class to patient feature vector what in result mark the automatically classify this mammograms.

1. INTRODUCTION

Mammography is currently a reliable method for early detecting of breast cancer. This technique is non invasive and make possible to detect lesions in the breast. Digital mammograms are among the most difficult medical images to be read due to their low contrast and differences in the types of tissues. The huge number of their requires CAD (computer-aided diagnosis) systems for detecting and analysis signs of breast cancer. These systems should be able to distinguish between the normal tissues and the different types of tumors and to differentiate between benign and malignant tumors. The critical points of such approach are:

- localization in mammograms the suspected area. Locating these areas is difficult as difference between image intensities of normal area ("cancer free") and abnormal suspicious area can be minimal,
- a feature extraction for locating suspicious,
- a feature analysis and the classification of regions as benign or malignant area (classifying tumors).

Basically three types of mammograms are distinguished (Fig.1) – normal, benign and malignant. Benign tumors are noninvasive, slowly growing and sparsely a threat to life. Malignant tumors grows rapidly, send branch into the normal tissue and require difficult and complicated surgery.

The methodology presented in this paper is based on following steps:

- Preprocessing and breast area ROI detection,
- Steerable filtering ROI,
- Feature extraction,
- Classification.

[1] University of Technology & Life Sciences, Department of Telecommunications & EE, ul. Kaliskiego 7, 85-796 Bydgoszcz, email: choras@utp.edu.pl.

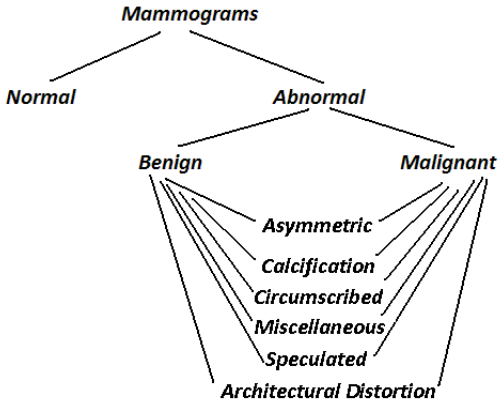

Fig. 1. Basically types of mammograms.

We used images from the MIAS digital mammography database which includes a description of the locations and types of the abnormalities.

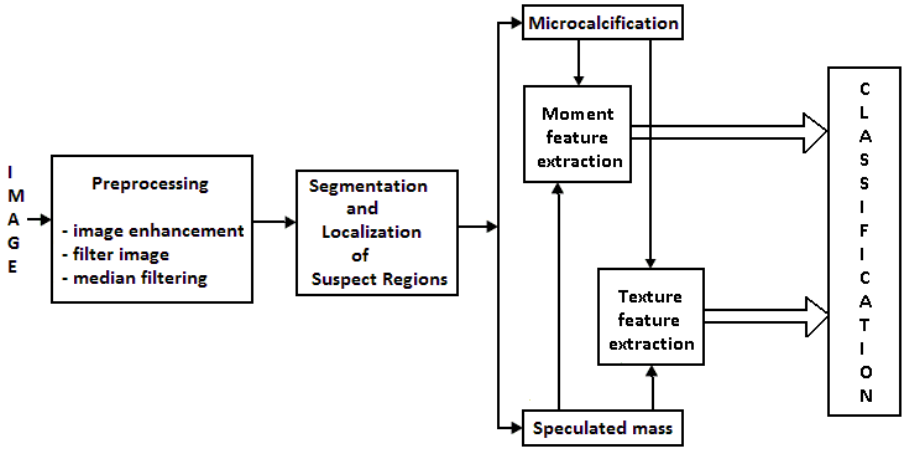

Fig. 2. Flow chart of the mammograms processing.

2. PREPROCESSING

Before performing feature extraction, the original mammogram images are subjected to some image processing operations, such as:

1. Image stretched. Contrast stretching is the image enhancement technique that commonly used for medical images. Contrast stretching process plays an important role in enhancing the quality and contrast of medical images [4].
   The contrast level is stretched according to

$$f_{out}(x,y) = 255 \times \left(\frac{f_{in}(x,y) - min}{max - min} \right)^{\gamma} \quad (1)$$

   f_{out} is the color level for the output pixel (x, y) after the contrast stretching process. $f_{in}(x,y)$ is the color level input for data the pixel (x, y). *max* - is the maximum intensity value in the input image. *min* – is the minimum intensity value in the input image, γ - constant that defines the shape of the stretching curve.
2. Median filtering.
3. Localized suspicious area and created of the ROI's (Region of Interest). This is not a trivial task, due to the ill-defined borders, which make difficult their discrimination from the parenchyma's structures. Suspicious area is brighter than its neighborhood region, has uniform density inside the

area and area edges are fuzzy. In every area is defined central points of ROI and ROI's were taken with $M \times N = 128 \times 128$ pixel resolution. The goal of this step is to extract one or more regions of interest (ROIs) from the background. The consecutive steps of the center pixel of ROI area search algorithm are presented in [3].

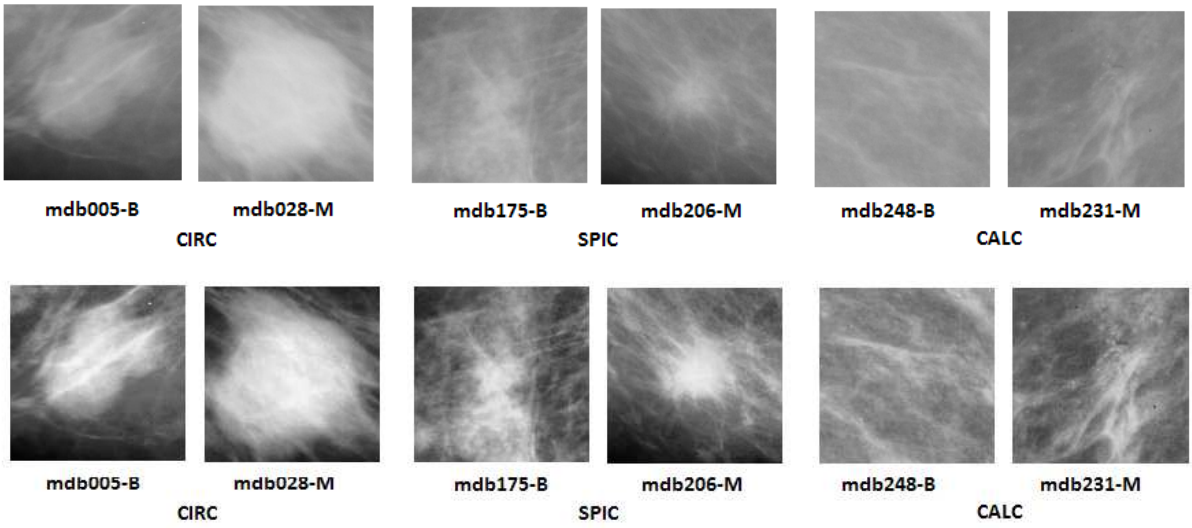

Fig. 3. Examples original and stretched (from top to below) of abnormal regions (from left to right): Circumscribed mass (image mdb005 – benign and image mdb028 – malignant); Spiculated mass (image mdb175 – benign and image mdb206 - malignant); Microcalcifications (image mdb248 – benign and image mdb231 - malignant).

3. MAMMOGRAM FEATURE EXTRACTION

3.1. STEERABLE FILTERS

To find the characteristic lines in mammograms ROI different edge filters can be defined. In this paper, the steerable filters are used for detection of characteristic mammogram lines. Steerable filters are a class oriented filters in which any filter is represented by a linear combination of set of basis filters. The concept of steerable filters was proposed by Freeman and Aldeson [2].
The idea of steerable filters is brief overview below. Let a 2*D* Gaussian function $g(x, y)$ is defined as

$$g(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{2}$$

The first derivative with respect to *x* is defined as the following equation:

$$g_x^{'}(x,y) = \frac{\partial}{\partial x}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} = -\frac{1}{\sqrt{2\pi}\sigma^3} x e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{3}$$

Function with eq.(3) rotated by angle $\theta = 90$, it is the same as the derivation of $g(x, y)$ with respect to *y*

$$g_y^{'}(x,y) = g_{x(\theta=90)}^{'}(x,y) = \frac{\partial}{\partial y}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} = -\frac{1}{\sqrt{2\pi}\sigma^3} y e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{4}$$

The first derivative of $g(x,y)$ rotated by any angle $\theta = 90$ can be obtained by the linear combination of eqs. (3) and (4):

$$g'_\theta(x,y) = \frac{\partial}{\partial\theta} g(r,\theta) = -\frac{1}{\sqrt{2\pi}\sigma^3}(x\cos\theta + y\sin\theta)e^{-\frac{(x^2+y^2)}{2\sigma^2}} = \cos\theta g'_x(x,y) + \sin\theta g'_{x(\theta=90)} \tag{5}$$

The convolution of the filters with image $f(x,y)$ is defined as:

$$Con'_\theta = g'_\theta(x,y) \otimes f(x,y) = \cos\theta(g'_x(x,y) \otimes f(x,y)) + \sin\theta(g'_y(x,y) \otimes f(x,y)) \tag{6}$$

In generally the steerable filters along of any orientation θ is consider as the form

$$g^\theta(x,y) = \sum_{i=1}^{M} k_i(\theta) g^{\theta_i}(x,y) \tag{7}$$

where M is the number of basis functions to steer a function $g\theta(x, y)$ and interpolation function are following [3]

$$k_i(\theta) = (-1)^i \begin{pmatrix} M \\ i \end{pmatrix} \cos^{M-i}(\theta)\sin^i(\theta) \tag{8}$$

We assume that the mammogram features are oriented in some direction for their detection we design of steerable filters. According to [1, 2] we used to edge detection third and fifth order edge template and to ridge (line) detection second and fourth ridge template (Fig. 4).

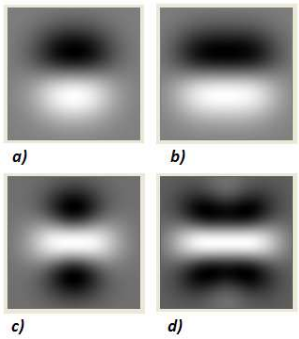

Fig. 4. Example of steerable filters. a) and b) edge filters for orders M=3 and M=5 ; c) and d) ridge filters for orders M=2 and M=4.

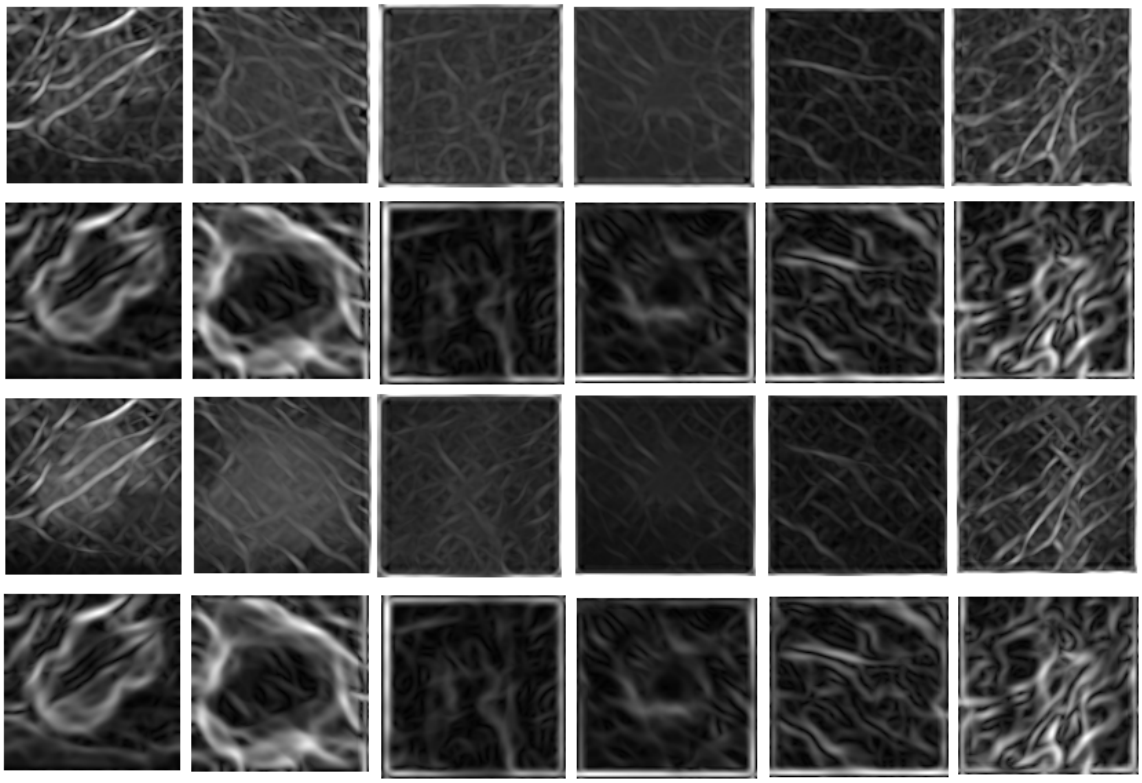

Fig. 5. Examples of convolution steerable filters with stretched images from Fig.3. From top to below respectively M = 2, M = 3, M = 4 and M = 5.

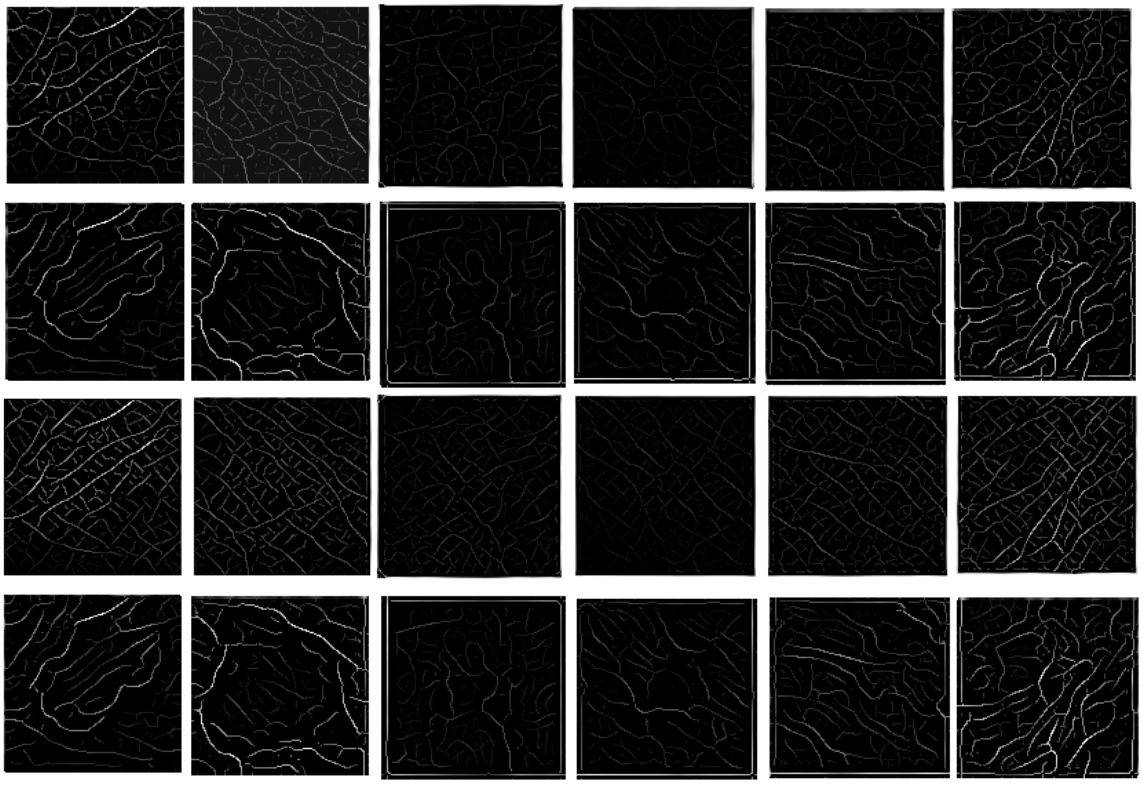

Fig. 6. Examples of line detection and thinning for images from Fig.3.

3.2. FEATURE EXTRACTION

The important step is feature selection and calculation. The features of the mammograms ROI image are represented by

$$E(x.y) = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(Con_{\theta}^{'})^2 \tag{9}$$

Additionally from each ROI with Fig. 5 and Fig. 6 were extracted the moment features and texture features, which are explained below.

i. Moment features

2D moments are defined as

$$m_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} i^p j^q f(i,j) \tag{9}$$

and central moments of order *(p+q)* are defined as

$$\mu_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N}(i-\bar{i})^p (j-\bar{j})^q f(i,j) \tag{10}$$

where $f(i,j)$ is an intensity of a digital image with $M \times N$ size, and $\bar{i} = \frac{m_{10}}{m_{00}}$, $\bar{j} = \frac{m_{01}}{m_{00}}$.

We used as a moment features three Zernike moments given as follows [5]:

$$Z_1 = \frac{3\left(2\left(\mu_{20}+\mu_{02}\right)-1\right)}{\pi} \tag{11}$$

$$Z_2 = \frac{9\left(\left(\mu_{20}-\mu_{02}\right)^2+4\mu_{11}^2\right)}{\pi^2} \tag{12}$$

$$Z_3 = \frac{16\left(\left(\mu_{03}-3\mu_{21}\right)^2+\left(\mu_{30}-\mu_{12}\right)^2\right)}{\pi^2} \tag{13}$$

Table 1. Zernike moments Z_1, Z_2, Z_3.

	mdb 005- B	mdb028-M	mdb175-B	mdb206-M	mdb248-B	mdb231-M
Z_1	50.5398	42.19968	103.5749	102.24305	77.7258	45.07815
Z_2	5.73011	0.51861	12.2330	4.194677	14.78254	1.781632
Z_3	27.88688	0.222264	36.64614	5940.753	37.17811	58.939085

ii. Texture features

Texture features are based on the co-occurrence matrixes $P_{\delta,\phi}(x,y)$, they are bi-dimensional representations showing the spatial occurrence organization of the gray levels in a block image. They

represent a bi-dimensional histogram of the gray levels, where fixed spatial relation separates couples of pixels, defining the direction and distance (δ, ϕ) from a referenced pixel to its neighbor.
The features are:

- Second Angular Moment

$$SAM = \sum_{x=1}^{M}\sum_{y=1}^{N}\left[P_{\delta,\phi}(x,y)\right]^2 \tag{14}$$

- Contrast

$$Con = \sum_{x=1}^{M}\sum_{y=1}^{N}(x-y)^2 P_{\delta,\phi}(x,y) \tag{15}$$

- Correlation

$$Corr = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left[xyP_{\delta,\phi}(x,y)\right]-\mu_x\mu_y}{\sigma_x\sigma_y} \tag{16}$$

- Inverse Differential Moment

$$IDM = \sum_{x=1}^{M}\sum_{y=1}^{N}\frac{P_{\delta,\phi}(x,y)}{1+(x-y)^2} \tag{17}$$

- Entropy

$$E = -\sum_{x=1}^{M}\sum_{y=1}^{N}P_{\delta,\phi}(x,y)\log P_{\delta,\phi}(x,y) \tag{18}$$

Table 2. Texture features.

	mdb 005- B		mdb028-M		mdb175-B		mdb206-M		mdb248-B		mdb231-M	
Parameter		δ =5	ϕ	δ =5	ϕ	δ =5	ϕ	δ =5	ϕ	δ =5	ϕ	δ =5
SAM	0	1.84E-4	0	9.95E-5	0	4.0E-4	0	3.845E-4	0	2.089E-4	0	8.99E-5
	90	1.51E-4	90	9.26E-5	90	4.24E-4	90	3.728E-4	90	1.923E-4	90	9.71E-5
	180	1.74E-4	180	1.01E-4	180	3.97E-4	180	3.889E-4	180	2.118E-4	180	9.06E-5
	270	1.64E-4	270	9.2E-5	270	4.226E-4	270	3.672E-4	270	1.9E-4	270	9.24E-5
Con	0	986.05	0	1922.8	0	1677.64	0	1077.68	0	890.91	0	2848.29
	90	1489.5	90	2085.2	90	1120.57	90	1268.44	90	1861.04	90	2022.82
	180	1064.17	180	1672.7	180	2108.43	180	726.1	180	839.58	180	2880.34
	270	1391.38	270	2087.8	270	1136.33	270	2269.5	270	2720.35	270	2450.85
Corr	0	4.49E-4	0	2.40E-4	0	2.8E-4	0	4.45E-4	0	4.9E-4	0	1.46E-4
	90	3.38E-4	90	2.29E-4	90	4.19E-4	90	3.69E-4	90	1.43E-4	90	2.47E-4
	180	4.32E-4	180	2.59E-4	180	2.00E-4	180	5.69E-4	180	5.06E-4	180	1.43E-4
	270	3.58E-4	270	2.29E-4	270	4.141E-4	270	1.23E-4	270	1.4E-5	270	1.96E-4
IDM	0	0.062	0	0.044	0	0.079	0	0.089	0	0.056	0	0.031
	90	0.044	90	0.041	90	0.093	90	0.072	90	0.047	90	0.032
	180	0.062	180	0.044	180	0.079	180	0.089	180	0.056	180	0.031
	270	0.045	270	0.041	270	0.093	270	0.072	270	0.047	270	0.032
E	0	9.026	0	9.465	0	8.349	0	8.282	0	8.803	0	9.523
	90	9.128	90	9.542	90	8.301	90	8.305	90	9.891	90	9.484
	180	9.051	180	9.459	180	8.370	180	8.263	180	8.792	180	9.520
	270	9.093	270	9.545	270	8,309	270	8.329	270	8,905	270	9.506

The classification method is based on equation

$$EucDist(Reference\_mamograms, Patient\_mamogram) = \sum_{i=1}^{L} \sqrt{\left(R\_m^j(i) - P\_m(i)\right)^2} \tag{19}$$

where $R\_m^j(i)$ is the feature vector of the jth reference mammogram images and $P\_m(i)$ is the feature vector of the patient mammogram. Each feature vectors have L coefficients. The minimum distance between feature vectors determine similarity, and describe patient mammogram.

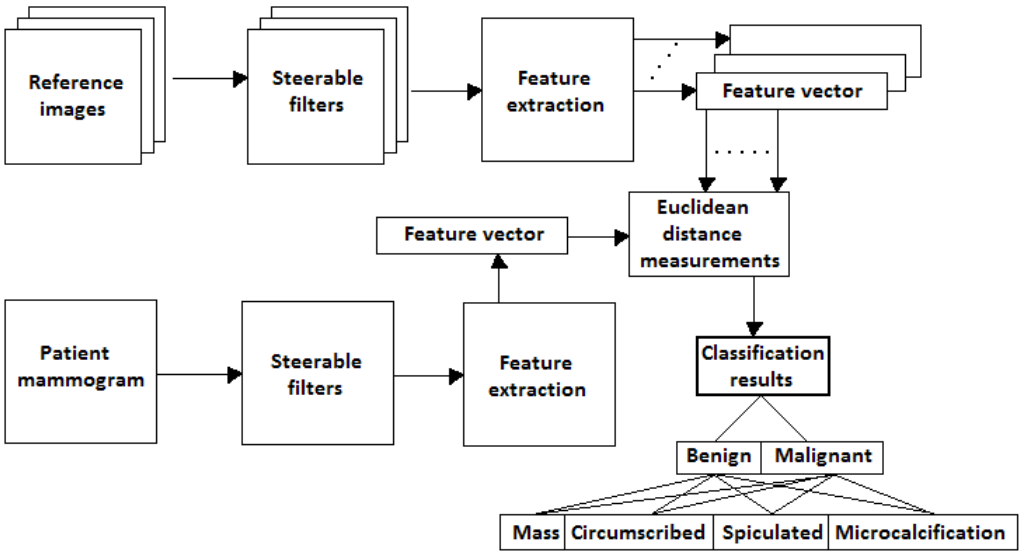

Fig. 7. Flow chart of the classification.

4. RESULTS AND CONCLUSIONS

In this paper the distinguish between the types of suspicious lesions in mammograms are considered. For proposed feature vector classification rate are 91.2 % for benign and 87.9% for malignant mammograms. Classification rate for the different types of tumors are following: 89.2% for speculated, 82.9% for circumscribed, 90.7% for microcalcification and 97.2% for mass tumors.

BIBLIOGRAPHY

[1] JACOB M., UNSER M., Design of steerable filters for feature detection using Canny-like criteria, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, No. 8, 2004, pp.1007-1019.
[2] FREEMAN W.T., ADELSON E.H., The design and use of steerable filters, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9, 1991, pp. 891-906.
[3] SINGH S., AL-MANSOORI R., Identification of regions of interest in digital mammograms, J. Intell. Systems, Vol. 10, 2000, pp. 183-210.
[4] LAINE A.F., SCHULER S., FAN J., HUDA W., Mammographic feature enhancement by multiscale analysis, IEEE Transactions on Medical Imaging, Vol. 13, No. 4, 1994, pp. 725–740.
[5] PEJNOVIC P., BUTUROVIC L., STOJILJKOVIC Z., Object recognition by invariants, Proc. 11th IAPR Int. Conf. On Pattern Recognition, 1992, pp. 434 – 437.