Similarity Analysis Based on the Weighted Moving Window for Tissue Characterization of Plaque in Coronary Arteries

Gancho Vachkov and Eiji Uchino

Abstract. This paper is dealing with the problem of tissue characterization of the plaque in the coronary arteries by processing the data from the intravascular ultrasound catheter. The similarity analysis method in the paper is applied in the frame of the moving window approach, which scans all cells in the matrix data from one cross section of the artery. The center-of-gravity model is used for evaluating the dissimilarity between any given pairs of data sets, belonging to pairs of windows. As a computational strategy, the use of weighted values of dissimilarity within the cells belonging to one window is proposed in the paper, rather than simply using an equal mean value for all cells in the window.

The similarity results from each cross section of the artery are displayed as gray scale image, where the darker areas denote the more similar areas to a predefined region of interest. The simulation results from the tissue characterization of a real data set show that the weighted moving window approach gives a sharper resolution of the similarity results that are closer to the real results, compared to the simple mean value approach. This suggests that the weighted moving window approach can be applied to real medical diagnosis.

Keywords: Similarity analysis, Weighted mowing window, Tissue characterization, Intravascular ultrasound, Classification.

1 Introduction

Health condition of the coronary arteries is vital for the normal functioning of the human heart since they supply a fresh blood to the muscular tissue of the heart.

Gancho Vachkov
Graduate School of Science and Engineering, Yamaguchi University, 1677-1 Yoshida, Yamaguchi 753-8512, Japan

Eiji Uchino
Fuzzy Logic Systems Institute (FLSI), 630-41 Kawazu, Iizuka, Fukuoka 820-0067, Japan
e-mail: {vachkov,uchino}@yamaguchi-u.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 1, SIST 15, pp. 297–307.
springerlink.com

When a gradual build-up of a *plaque* in the inner surface of the artery has occurred, it may cause under some circumstances severe heart diseases (acute coronary syndrome) such as myocardial infarction and angina.

The inner structure of the plaque tissue is directly related to the risk of a heart failure. Here the most important are two types of plaque structure, namely the *lipid* and the *fibrous* structure. A lipid plaque covered by a small and thin fibrous plaque is very likely to break and enter the blood stream, thus creating dangerous blood clots. Therefore it is of utmost importance to analyze the structure of the plaque and find out the so called lipid and fibrous *regions of interest*, abbreviated as Lipid ROI and Fibrous ROI.

The analysis and estimation of the size, shape and location of the Lipid ROI and Fibrous ROI is usually called *tissue characterization* in medical terms, which falls into the research area of pattern recognition and pattern classification.

One of the most frequently used techniques to get reliable information from the coronary artery for further visualization and tissue characterization is the *Intravascular Ultrasound* (IVUS) method [1]. The IVUS method uses a small rotating catheter with a probe inserted into the coronary artery that emits a high frequency ultra-sonic signal to the tissue. The reflected radio-frequency (RF) signal is measured and saved in the computer memory for any further analysis and visualization.

The IVUS is essentially a *tomographic imaging* technology, in which the reflected RF signal is preprocessed to produce a gray-scale image with a circular shape, called a *B-mode image* that is used by medical doctors for observation and analysis of the artery occlusion. One B-mode image corresponds to one *cross-section* of the coronary artery with a given depth-range in all *256* directions (angles) of rotation of the IVUS probe.

The main goal in this paper is not just the data visualization, but rather developing an appropriate method for tissue characterization. Therefore further on we represent the data and the respective results in a rectangular *X-Y* shape, instead of in circular shape. The axis *X* denotes the *angle* (direction) of the IVUS probe within the range of [*0, 255*], while the ordinate *Y* denotes the *depth* of the measurement. i.e. the *distance* between the probe and the current measured signal. The depth-of-interest in our investigations is within the range: [*0, 400*] since any lipid ROI found in the deeper inner areas of the coronary artery is considered as “not so risky”.

A graphical illustration of the matrix-type information obtained by the IVUS probe is presented in Fig. 1.The obtained large size matrix is called RF *matrix* and is further on saved in the computer memory It consists of every single measurement obtained for the IVUS probe for one cross section in the coronary artery.

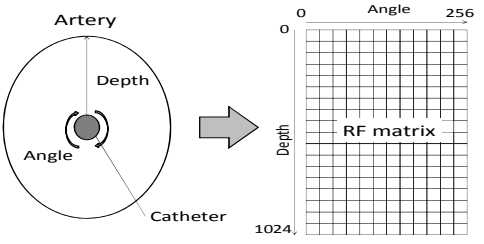

Fig. 1 Information obtained from the IVUS probe for one cross-section of the coronary artery. This is a rectangular X-Y matrix of data with *X* denoting the rotation angle and *Y* denoting the depth of the measurement.

Deep and long term research and data analysis have been done until now [1] – [4] to utilize the information obtained from the IVUS method for a proper tissue characterization of the plaque in the coronary artery. As a result, different classification techniques and algorithms have been proposed, developed and used for various simulations and comparisons. However, currently no "ideal" and easy-to-apply method still exists.

In this paper we propose the use of the moving window approach for *similarity analysis* of the data in the RF matrix, in order to find regions that are most similar to a given (pre-specified) Lipid ROI or Fibrous ROI. The work in this paper is a more detailed and advanced step of our previous work in [4] that also uses the concept of moving window and similarity analysis. In this paper we introduce a moving window with *maximal* overlapping ratio, a new method for similarity analysis and also a new way for calculating the similarity, based on the concept of the weighted moving window.

The rest of the paper is organized as follows. In Section 2 the standard moving window for similarity analysis is explained and in Section 3 the computational details of the new proposed weighted moving window are given. Section 4 explains the details of the center-of-gravity method used for similarity analysis. The experimental results are given in Section 5 and Section 6 concludes the paper.

2 The Standard Moving Window for Similarity Analysis

2.1 The Standard Moving Window Approach

The idea of the standard moving window approach is well known and used in many applications, especially in the research field of image processing and also in other areas that use a large *rectangular* data set as initial information.

First of all, a rectangular matrix of data to be analyzed should be available. In this research we use the RF data matrix, which contains all the reflected signal intensities from one *cross section* of the artery. This matrix consists of all the measured values of the reflected RF signal and has a size of *256* x *Dmax*, where *Dmax = 400* is considered as *sufficient maximal depth* for examination.

The next step is to select the *window size* $N_X \times N_Y$, with "reasonable" values for the *Angle_Range:* $2 \le N_X \le 60$ and for the *Depth_Range:* $5 \le N_Y \le 100$. Obviously these are *problem dependent* parameters that affect the final results from the tissue characterization. In many cases *reasonably* small window size gives better similarity results, while extremely small or extremely large sizes lead to deterioration.

The moving windows approach in fact performs a *scanning* procedure that starts with the first window being located at the *upper-left* corner of the RF matrix. Then this window gradually moves to the right with one-step (*one angle* position) only until the end of the line. After that the scanning is returned to the leftmost position, but shifted with one step (*one depth* position) downwards and continues on this new line. The scanning procedure continues until the last window reaches the bottom-right corner of the RF matrix.

Such scanning procedure ensures that every two neighboring windows have a *maximal overlapping* ratio, since they differ from each other by only one horizontal and one vertical position. This way of movement of the windows is different from our previous moving window approach in [4] where *no overlapping* between the neighboring windows was assumed.

The moving window process with maximal overlapping ratio leads to generating a large number of windows, thus increasing the overall computation time. The total number of the generated windows is: $N_W = 256 \times (D_{\max} - N_Y + 1)$. The total computation time depends not only on the number of the windows, but also on the complexity of the model that is calculated from the data in each window. In order to alleviate the computational burden, we have proposed in Section 4 a simple and easy to calculate representative model, called center-of-gravity model.

2.2 Similarity Analysis by Using the Moving Window

The similarity analysis used in frame of the moving windows approach is basically a method for calculating the difference (dissimilarity) between the structures of two data sets at each scanning step.

The first data set is fixed (constant) and is used as a *reference* data sets (an example) for comparison during the scanning process. These data should be available before the scanning process. They correspond to one typical (and correctly identified) Fibrous or Lipid ROI. It is obvious that the proper identification of such ROI depends on the medical doctor decision and experience.

The other (second) data set used for the similarity is different for each scanning step and is extracted from the respective window W_i, $i = 1, 2, ..., N_W$ for this step.

It becomes clear now that the proposed similarity analysis computation is essentially a *supervised procedure* for decision making, in which the *Dissimilarity Degree* DS at each step is calculated between one reference data set and one *unknown* set, belonging to the current window.

The value of DS is usually bounded: $DS_i \in [0,T]$ and shows *how close* is the data structure from the current window W_i, $i=1,2,...,N_W$ to the data structure in the predefined ROI. A value of dissimilarity, close to zero suggests that the two data sets are very similar while a bigger (closer to *T*) value stands for a bigger difference (bigger discrepancy) between the two data sets.

For the purpose of *fair* quantitative comparison between the data sets, the dissimilarity value is often *normalized* as: $DS_i \in [0,1]$.

It is clear that the calculated value of dissimilarity depends on the assumed model for describing the structure of the data set in each window. A simple and yet effective model for similarity analysis is presented in Section 4.

During the moving window process, the similarity value DS_i is calculated many times, namely for each current position of the window W_i. Then the main question is *where* (at which location) to assign the currently calculated value of DS_i? This problem arises because all $N_X \times N_Y$ data in the current window W_i have been used for calculation of the dissimilarity.

The simplest and reasonable decision is to assign the *same dissimilarity* degree DS_i to *all* coordinates (all *cells*) within the current window W_i that have been used in the calculations. In order to keep in a memory all these values, we create a *new* rectangular matrix called *Dissimilarity Matrix* DM, with the same dimension as the Data Matrix RF, i.e. *245* x *Dmax*.

It is easy to realize that each data item (each cell) in the original data matrix RF will be visited not only ones, but many times by the moving window. If *N* denotes the number of all visits of a given cell at location $\{i,j\}$; $i \in [0,255]$; $j \in [0,Dmax]$ by a moving window with size $N_X \times N_Y$, then this number is calculated as:

$$\begin{aligned} N &= (j+1)N_Y, \ \ if \ \ 0 \le j \le N_Y - 2; \\ N &= N_X \times N_Y, \quad if \ \ N_Y - 1 \le j \le D_{\max} - N_Y + 1; \\ N &= (D_{\max} - j + 1)N_Y, \ \ if \ \ D_{\max} - N_Y + 2 \le j \le D_{\max}; \end{aligned} \tag{1}$$

The DM matrix is actually an *additive matrix* that accumulates all the calculated values for the dissimilarity degrees at the same location $\{i,j\}$, as follows:

$$dm_{ij} \leftarrow dm_{ij} + DS_k, \ \ k=1,2,...,N \tag{2}$$

The *final dissimilarity* degree for each $\{i,j\}$ location can be taken in different ways, but in the in the simplest case of the standard moving window, we take it as a *mean value* of the accumulated similarity degrees in DM, namely:

$$dm_{ij} = dm_{ij} / N, \ \ i \in [0,255]; \ j \in [0, D_{\max}] \tag{3}$$

All final dissimilarity degrees (3) are saved in the DM matrix. Further on, they can be displayed in the following two ways for a final decision making:

- *Hard Decision* making with a user-defined *threshold Th* for separation of all the values into *2* crisp classes: *Similar* (with $dm_{ij} \le Th$) and *Not Similar* (with $dm_{ij} > Th$) to the pre-specified region of interest, such as *Lipid* ROI or *Fibrous* ROI. Then the *Similar* only class is displayed as a *crisp* dark image to the medical doctor for his final decision. Obviously the threshold selection is not a trivial task that could lead sometimes to ambiguous results.
- *Soft Decision* making. It is a kind of *fuzzy* way of displaying the results, in which all calculated values in (3) are visualized as a *gray scale* image with varying intensities. The similarity values closer to *zero*, corresponding to *very similar* areas, are displayed as *darker areas.* The values closer to *one*, corresponding to *less similar* areas are displayed as *brighter areas.* This gives additional information to the medical doctor in taking his final decision.

3 The Weighted Moving Window for Similarity Analysis

The idea of taking the dissimilarity degree in (3) as a *mean value* of all calculated dissimilarities in one window has a certain drawback. It is that the calculated *true* value in the center depends *equally* on all neighboring dissimilarities, even on those that are far from the center. As a result such assumption may lead to a serious deviation from the actual dissimilarity.

In order to make more precise calculation of the dissimilarities at each cell of the window, we propose here the idea of the *weighted moving window*. The essential point is that the dissimilarity in each cell $[i, j]$of the matrix DM is now *weighted* between *0* and *1* according to its *distance* to the center $[i_0, j_0]$of the window, as shown in the next equation.

$$w_{ij} = exp^{-\frac{|i-i_0|+|j-j_0|}{2\sigma^2}} \in [0,1];\ i=1,N_X;\ j=1,N_Y \tag{4}$$

As seen from (4), a Gaussian function with *Manhattan* (City Block) *distance* and a predetermined *spread* σ is assumed here to evaluate the amount of the weight for each cell in the window. For example, if the window size is 5 x 7, its center location will be [3,4].Then, the Manhattan distance between the center and a cell located at [4,2].will be 3.

The calculated weight by (4) is used to calculate the weighted dissimilarity that will be added to each cell of the matrix DM, in a similar way as in (2), namely:

$$dm_{ij} \leftarrow dm_{ij} + w_{ij} DS_k,\ k=1,2,...,N \tag{5}$$

The *final dissimilarity* degree for each {*i,j*} location will be calculated now as *weighted average* of the accumulated dissimilarities, taking into account the sum of all weights, namely:

$$dm_{ij} = dm_{ij} \Big/ \sum_{n=1}^{N_X} \sum_{k=1}^{N_Y} w_{nk}\,,\ i=1,N_X;\ j=1,N_Y \tag{6}$$

Thus all the final dissimilarities in (6) will be normalized between *0* and *1*.

An illustrative example of the Gaussian function from (4) is given in Fig.2.

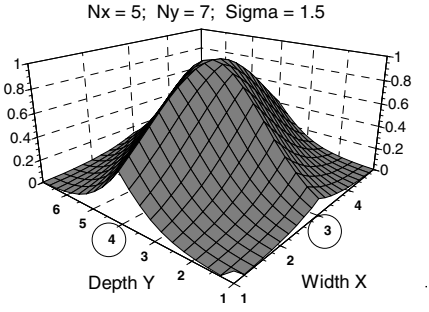

Fig. 2 Example of the Gaussian weight function from (4) used for calculating the weights of the dissimilarities in (5) in the case of a chosen window with size *5* x *7*.

4 The Center-of-Gravity Model Used for Similarity Analysis

In order to calculate the similarity degree between any two data sets, we need a *model* that describes appropriately the structure of the data set. Then, as an initial step, we calculate (once only) the two *Reference Models* RM_L and RM_F for the predefined *Lipid* ROI and *Fibrous* ROI., by using the respective available data: M_L and M_F. After that a *repetitive* model calculation is performed for each of the moving windows that contain the same number of $M_W = N_X \times N_Y$ data.

The Center-of-Gravity (COG) model, proposed and used in this paper produces a simple, but still representative estimation of the structure of the extracted data from each window. The COG model uses just two parameters with clear physical meaning, namely the *Center-of-Gravity* (CG) and the *Standard Deviation* (SD). The presumption is that data sets with different structures have different values of CG and SD, so the amount of the difference between the two parameters can be used to evaluate the similarity degree between them.

As initial information for constructing the RF matrix, we use the raw (actual) reflected RF signal with a simple *preprocessing*, as shown in Fig. 3, namely taking the absolute vale of the RF signal after subtracting the central (stationary) value of *2000* from it.

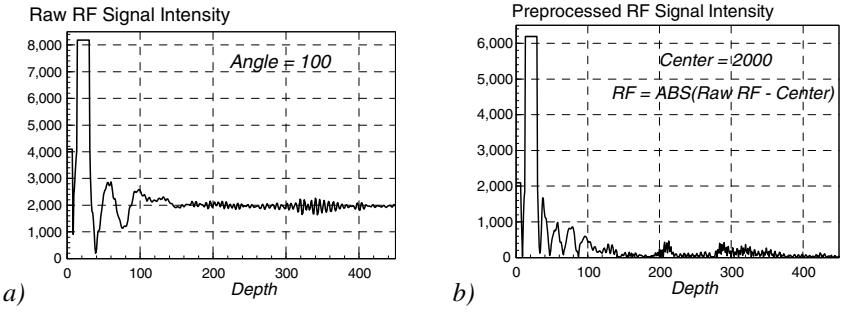

Fig. 3 The Intensity of the Raw (a) and the Preprocessed (b) reflected RF signal at one fixed angle (*100*) of the IVUS catheter and all depths, starting from *0* to *450*.

Let *M* denotes the number of data, i.e. the number of the preprocessed RF signal intensities R_i, $i=1,2,...,M$ from a given ROI or a given window W. Then the two model parameters CG and SD are calculated as follows:

- The *Center-of-Gravity* of the model is simply calculated as the *mean* value of the one-dimensional RF signal:

$$CG = \sum_{i=1}^{M} R_i / M \tag{7}$$

-The *Standard Deviation* is calculated as:

$$SD = \sqrt{\sum_{i=1}^{M} (R_i - CG)^2 \Big/ (M-1)} \tag{8}$$

For example, the calculated values of CG and SD for the Fibrous ROI and the Lipid ROI, extracted from the one experimental RF data matrix are as follows:

Fibrous ROI: CG = *47.565*; SD = *34.344*;
Lipid ROI : CG = *45.375*; SD = *17.045*.

For calculating the *normalized dissimilarity* DS between two COG models, namely the reference model *Mo* and the current window model *Mi*, we propose in this paper the following formula:

$$DS_i = DS(M_0, M_i) = 1 - exp^{-\frac{(CG_0 - CG_i)^2}{2\sigma_C^2}} exp^{-\frac{(SD_0 - SD_i)^2}{2\sigma_S^2}} \in [0,1] \tag{9}$$

Here σ_C and σ_S denote the predefined *width* for the *CG* and *width* for the *SD*, respectively. It is important to note that we are able to control the process of dissimilarity analysis by changing in appropriate way the tuning parameters σ_C and σ_S. For example, a smaller selected value of σ_S , i.e. a narrower width σ_S will give bigger importance to the standard deviation DS of the extracted data, than to

the center-of-gravity CG. Therefore the proper choice of these two tuning parameters (widths) is left to the user.

5 Simulation Results from Tissue Characterization

5.1 Simulation Details and Conditions

The above described moving window approach for similarity analysis was used for tissue characterization of several sets of real data, in the form of respective RF matrices, each of them with *X-Y* size: *256* x *400* (the maximal depth: *Dmax = 400*). For each matrix, the respective *Lipid* ROI and *Fibrous* ROI have been properly identified and marked by a medical doctor through a microscopic analysis. These ROI data were used for creating the *Reference* Lipid and Fibrous *models* for similarity analysis and also for testing and analyzing the correctness of the simulation results.

In the simulations, a moving window with size *21* x *31* was chosen, which generated *94976* windows in total, used for similarity analysis. Despite this large number, the calculations were fast, due to the simple structure of the proposed COG model. The CPU time for one RF data set was about 5 sec on a computer with 3.0 GHz *Intel 4* CPU unit. This suggests that the proposed tissue characterization method, if "accurate enough", could be applied in almost *real-time* mode.

5.2 Tissue Characterization Results from the Similarity Analysis

In the simulations we used one RF data set, corresponding to one cross section of the artery, for which the size, location and boundaries of the *Fibrous* ROI and the *Lipid* ROI were properly diagnosed by the doctor. The data extracted from these two ROIs were used to calculate the respective *Reference* COG models. Then the two reference models were used for similarity analysis by using the *Soft* decision between *0* and *1*, with pre-selected widths in (9), as follows: $\sigma_C = 8.0$ and $\sigma_S = 5.0$. Such settings put a bigger priority to the standard deviation (the *roughness* of the data) than to the CG (the *mean* value of the data). The user-defined value for the spread in (4) was: $\sigma = 2.5$.

The results from the similarity analysis, based on the proposed weighted moving window are shown in Fig. 4 for the case of Fibrous ROI and in Fig. 5 for the case of Lipid ROI. The results, based on the weighted moving window, with the dissimilarity calculated by (6), are depicted in Fig. 4*a*,4*b* and Fig. 5*a*,5*b,* while Fig. 4*c* and Fig. 5*c* show the results from the standard moving window, where the dissimilarity is taken simply as the mean value (3). It is easy to notice that the weighted moving window approach in Fig. 4*a* and Fig. 5*a* produces *sharper images* with higher resolution, which makes the final human decision easier.

Fig.4*b* and Fig. 5*b* are augmentation of Fig. 4*a* and Fig. 5*a*, where the locations of the pre-specified Fibrous ROI and Lipid ROI are added, in order to validate the correctness of the tissue characterization by our proposed method.

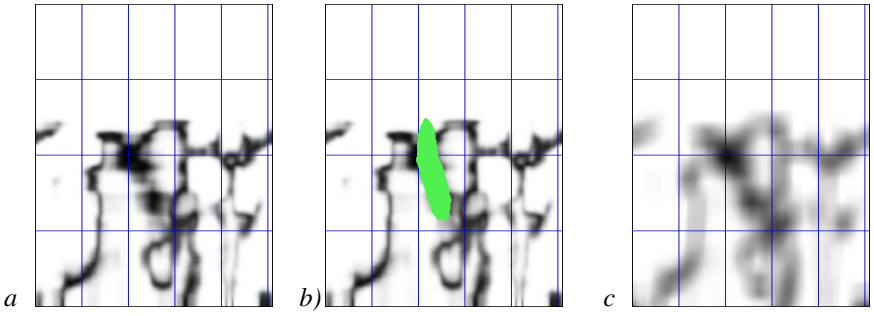

Fig. 4 Tissue characterization results for the case of Fibrous ROI; (a) and (b) are the weighted moving window results; (c) is the standard moving window result.

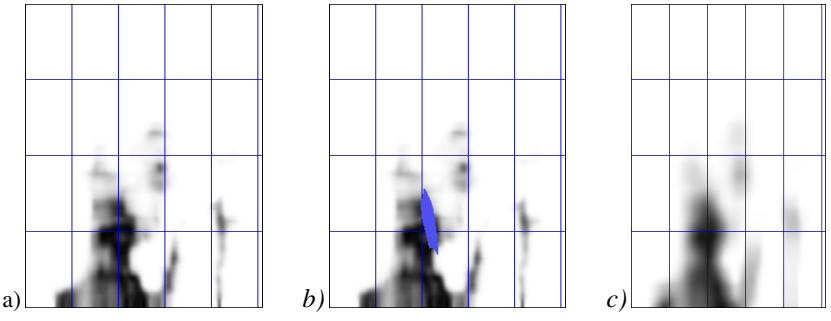

Fig. 5 Tissue characterization results for the case of Lipid ROI; (a) and (b) are the weighted moving window results; (c) is the standard moving window result.

The characterization results show that the proposed similarity analysis method has identified much larger areas as "belonging to" the *Fibrous* ROI and *Lipid* ROI, than the actually two ROIs, identified by the medical doctor. There are several different reasons for such discrepancy between the human and computer results.

One of them is that in reality there could be *several* (multiple) *Fibrous* and *Lipid* ROIs within the examined cross section, however the doctor has diagnosed and marked only one (a *typical*) example of a Fibrous ROI and a Lipid ROI. Another reason could be the *heuristic* (not optimal) selection of the window size $N_X \times N_Y$, as well as the widths σ_C and σ_S. all these are *problem dependent* parameters that have a strong influence on the final tissue characterization results.

6 Concluding Remarks

We proposed in this paper a general Moving Window computational approach with maximal overlapping ratio for tissue characterization of coronary arteries. This approach uses data obtained from the IVUS catheter and allows implementation of different methods and models for similarity analysis. One of them, the

Center-of-Gravity based model was proposed and used in the paper. The moving window approach has been applied in two versions, namely the standard and the weighted mowing window. They differ by the way of calculating the dissimilarity degree DS for the cells in each window. In the standard moving window the average is taken and applied to all cells. In the weighted moving window, a Gaussian function with Manhattan distance is used to assign weights between *0* and *1* of the calculated DS for all cells in the window.

A convenient and practical *soft decision* making for visualization of the similarity results is used in the paper. In this kind of decision all dissimilarities create a gray-scale image, where the *darker* areas represent the *more similar* areas to the predetermined region of interest ROI.

The simulation results by using a real set of data from the IVUS catheter show positive, but still "not perfect" results. They detect successfully large part of the actual Lipid and Fibrous ROI, but also show some other areas, as "very similar" to those ROI.

Possible further improvements of the proposed weighted moving window approach include optimization of the window size and the widths σ_C and σ_C used for decision making. Other ways for improvement of the tissue characterization accuracy are considered now by applying other, more precise models for similarity analysis and other calculation methods for the dissimilarity degree.

Acknowledgment. This work was supported by the Grant-in-Aid for Scientific Research (B) of the Japan Society for Promotion of Science (JSPS) under the Contract No. 23300086.

References

1. De Korte, C., Hansen, H.G., Van der Steen, A.F.: Vascular ultrasound for athero sclerosis imaging. In: Interface FOCUS, vol. 1, pp. 565–575 (2011)
2. Uchino, E., Suetake, N., Koga, T., Kubota, R., et al.: Fuzzy Inference-based plaque boundary extraction using image separability for IVUS image of coronary artery. Electronic Letters 45(9), 451–453 (2009)
3. Koga, T., Furukawa, S., Uchino, E., Suetake, N.: High-Speed Calculation for Tissue Characterization of Coronary Plaque by Employing Parallel Computing Techniques. Int. Journal of Circuits, Systems and Signal Processing 5(4), 435–442 (2011)
4. Vachkov, G., Uchino, E.: Similarity-based method for tissue characterization of coronary artery plaque. In: CD-ROM, Proc. 24th Annual Conference on Biomedical Fuzzy Systems Association (BMFSA), October 20-29, pp. 257–260 (2011)