Multiple k-Nearest Neighbor Classifier and Its Application to Tissue Characterization of Coronary Plaque

Eiji UCHINO†, Ryosuke KUBOTA††, Takanori KOGA†††, Hideaki MISAWA††††, and Noriaki SUETAKE††, Members

SUMMARY In this paper we propose a novel classification method for the multiple k-nearest neighbor (MkNN) classifier and show its practical application to medical image processing. The proposed method performs fine classification when a pair of the spatial coordinate of the observation data in the observation space and its corresponding feature vector in the feature space is provided. The proposed MkNN classifier uses the continuity of the distribution of features of the same class not only in the feature space but also in the observation space. In order to validate the performance of the present method, it is applied to the tissue characterization problem of coronary plaque. The quantitative and qualitative validity of the proposed MkNN classifier have been confirmed by actual experiments.

key words: acute coronary syndromes (ACS), coronary plaque tissue characterization, intravascular ultrasound (IVUS) method, multiple k-nearest neighbor (MkNN) classifier

1. Introduction

The coronary arteries supply oxygenated blood to the whole heart. If the coronary artery is occluded by thrombosis, anoxia-induced serious heart diseases such as acute coronary syndromes (ACS) are caused[1], [2]. It has been reported that most cases of ACS are caused by the rupture of coronary plaque. In order to judge whether a coronary plaque has a chance of rupture or not, and to find appropriate measures for ACS, characterizing the structural components of coronary plaque is very important for medical diagnosis[3].

The intravascular ultrasound (IVUS) method[4], [5], which is a tomographic imaging technology, is often used for the diagnosis of ACS [6]–[8]. The IVUS method gives a two-dimensional cross-sectional image of a coronary artery, which is called a B-mode image. A medical doctor makes a diagnosis of the coronary plaque by carefully observing this B-mode image. However, the diagnosis is very hard because this obtained B-mode image is very grainy.

The B-mode image is composed of the ultrasound signals reflected from various tissues of coronary artery. These reflected signals are called radiofrequency (RF) signals, which contain various noises and/or measuring errors. Many computer-aided methods[9]–[13] have been proposed so far to analyze directly those RF signals for the tissue characterization of coronary plaque.

The spectral analysis of the RF signal is a representative approach[10], [11], [14]. In this analysis, the Fourier spectra of the RF signal are used as a feature vector for the classification of the tissues of coronary plaque. In [11], the feature vectors are classified by using a k-nearest neighbor (kNN) classifier[15], [16]. This method, however, cannot perform a precise characterization of plaque because the distribution of the feature vectors of each tissue heavily overlaps each other in the feature space.

On the other hand, the integrated backscatter (IB) analysis[13] is another representative and popular tissue characterization method. In the IB analysis a locally averaged power of the RF signal, i.e., IB value, is used. The IB analysis can perform a precise tissue characterization in some cases but is not always effective because some types of tissues have nearly the same IB values.

In this paper we propose a modified k-nearest neighbor classifier, which we call a multiple k-nearest neighbor (MkNN) classifier. The proposed MkNN classifier considers the spatial continuity of data distribution, not only in the feature space but also in the observation space, and thus it performs a good classification even if the data distribution of each class overlaps with each other in the feature space. The performance of the proposed MkNN classifier is verified by applying it to the tissue characterization of coronary plaque of a patient in hospital.

In the next section we explain the IVUS method and conventional methods for the tissue characterization of coronary plaque by using the IVUS method. The algorithm of the proposed MkNN classifier is explained in Sect. 3. The experimental results and their quantitative and qualitative evaluations are shown in Sect. 4. Concluding remarks are given in the last section.

2. Tissue Characterization of Coronary Plaque

2.1 Intravascular Ultrasound (IVUS) Method

Intravascular ultrasound (IVUS) method[4], [5] is one of
the medical imaging techniques used for the diagnosis of ACS. In the IVUS method a specially designed thin catheter with an ultimately-miniaturized ultrasound probe attached to its distal end is used (see Fig. 1). A medical doctor steers a catheter with a very thin diameter from outside the body into the coronary artery to be visualized under angiography. The proximal end of the catheter is connected to the computerized ultrasound equipment. The ultrasound waves are emitted from the ultrasound probe, which are in the 40 MHz range in this study, and the probe also receives the reflected signal from the tissue. It is sampled at 400 MHz and stored in the computerized ultrasound equipment.

An example of a sampled radiofrequency (RF) signal is shown in Fig. 2 (a). An IVUS image is constructed by the amplitude information of the received RF signal. In order to visualize the inside of a coronary artery, the sampled RF signal shown in Fig. 2 (a) is firstly transformed into an 8-bit luminal intensity signal as shown in Fig. 2 (b) by taking the absolute value of the RF signal, then taking its envelope, and finally taking its logarithmic value. The luminal intensity signals in all radial directions are then used to obtain a tomographic cross sectional image of the coronary artery as shown in Fig. 2 (c). The IVUS image of Fig. 2 (c) is called a “B-mode image.” B-mode image shows a real time ultrasound cross-sectional image of a thin section of the blood vessel where the catheter probe is currently rotating.

In this study B-mode image is constructed with 2,048 pixels in the radial direction and 256 lines in the circumferential direction. Hence, the resolution of distance and angle are 1.95μm/pixel and 1.41°/line, respectively.

2.2 Conventional Methods for Tissue Characterization of Plaque

A variety of approaches to the tissue characterization of coronary plaque have been proposed and studied so far [9]–[13]. Especially, VH-IVUS [10] proposed by Volcano Corporation has been used by researchers. However, the performance of VH-IVUS is not good for a practical use. The integrated backscatter (IB) analysis [13] is another representative of IVUS-based conventional tissue characterization methods.

In the IB analysis tissues are classified by using a scalar feature called IB value, which is a locally averaged power of a backscattered RF signal. Each pixel of B-mode image, which corresponds to a constituent element of coronary plaque, is classified by this IB value.

Here let \( x(t) \) and \( x_0(t) \) be the amplitudes of the sampled backscattered RF signal and the smallest signal which can be sensed by the probe at time \( t \), respectively. The IB value at time \( t \) is given by:

\[
IBS(t) = 20\log\left( \frac{1}{T + 1} \sum_{i=-T/2}^{i=T/2} \frac{x(i)^2}{x_0(i)^2} \right),
\]

where \( T \) is a constant of a window size for the calculation of IB value. The time axis of \( x(t) \) or \( IBS(t) \) is easily transformed to distance by multiplying the speed of ultrasound in the blood vessel, that is, time \( t \) corresponds to the distance
Fig. 3  Distributions of the feature vectors. (a) B-mode image of fibrous tissue; (b) B-mode image of fatty tissue; (c) Distribution of the IB values for fibrous tissue of (a); (d) Distribution of the IB values for fatty tissue of (b); (e) Distribution of the Fourier power spectra for fibrous tissue of (a) compressed in 2D by linear discriminant analysis; (f) Distribution of the Fourier power spectra for fatty tissue of (b) compressed in 2D by linear discriminant analysis.

The IB analysis has been broadly used in the field of cardiology, and it can be regarded as a steady method in the tissue characterization of coronary plaque [13], [18]. In [18], the IB analysis was compared to VH-IVUS directly. The result of [18] showed that the IB analysis provided higher diagnostic accuracy than VH-IVUS.

However, the IB analysis is not always effective because some types of tissues have nearly the same IB values. Figures 3 (c) and (d) show the distributions of the IB values for fibrous and fatty tissues of Figs. 3 (a) and (b), respectively. It can easily be seen that the distributions of the IB values broadly overlap with each other, and thus it is impossible to classify the tissue precisely only by using those IB values.

Furthermore, the IB value is a locally averaged power of RF signal, and the power of RF signal is substantially affected by the distance between the plaque and the probe. That is, if the probe is near to the plaque, the IB value becomes large, and if the probe is far from the plaque, the IB value as a matter of course becomes small. The position of the probe in the coronary artery cannot be controlled nor measured. For those reasons, it is not appropriate to rely only on the IB values for the tissue characterization of coronary plaque.

The spectral analysis of RF signals is another representative approach for the tissue characterization [14], [19]. In the spectral analysis, the tissue characterization is performed by using the logarithmic Fourier power spectra of the local RF signal as a feature vector. The Fourier power spectra are calculated by the short-time fast Fourier transform (FFT) of the RF signal. The conventional tissue characterization methods based on the spectral analysis use the frequency components in various frequency bands. However, the frequency characteristics of some tissues are also nearly the same with each other as in the case of the IB analysis.

Figures 3(e) and (f) show the distributions of the Fourier power spectra for the plaques of Figs. 3(a) and (b), compressed into a two-dimensional linear sub-space by the linear discriminant analysis (LDA) [16]. LDA is one of the dimension reduction methods, which finds out the best separation of the two classes. Under those states of affairs, a new approach is desired for a precise tissue characterization of coronary plaque.

3. Proposed Tissue Characterization Method

Figure 4 shows the conceptual classification scheme of the proposed multiple k-nearest neighbor (MkNN) classifier for a two-class classification problem. In Fig. 4, the observation space is an observed image, and the feature vectors constituting the feature space are obtained at each pixel of the observed image. The feature vectors are obtained usually, e.g., by using PCA, Fourier transformation, and so on. In the experiments of tissue characterization, the observation space corresponds to B-mode image and the feature vectors are obtained by Fourier transformation of RF signals.
The feature vectors are fed to the MkNN classifier in the same manner as to the ordinary kNN classifier. Now, suppose that a set of \( N \) pairs \( \{(\mathbf{v}_n, i_n); n = 1, \ldots, N\} \) are given, where \( \mathbf{v}_n \) is a feature vector which belongs to class \( i_n \in \{1, \ldots, C\} \). \( C \) is the number of classes.

In the MkNN classifier, the classification of the pixel of interest \( p = (x_1, x_2) \) in the observation space is carried out as shown in Fig. 4. Note that \( k \) represents the number of the neighboring pixels \( \{\mathbf{p}_m; m = 1, \ldots, k\} \) of the pixel of interest (POI) \( p \) in the observation space, and \( k' \) represents the number of the nearest neighbors around the feature vectors \( \{\mathbf{u}_m; m = 1, \ldots, k\} \) in the feature space. The feature vector \( \mathbf{u}_m \) is the one for the pixel \( \mathbf{p}_m \) in the observation space.

The following are specific differences between the ordinary kNN classifier and the proposed MkNN classifier. The MkNN classifier uses not only the feature vector of POI \( p \) (the pixel to be classified) but also the feature vectors of the neighboring pixels \( \mathbf{p}_m \) for the classification. In the MkNN classifier, \((k \times k')\)-nearest feature vectors are selected for POI \( p \). Finally, the majority decision considering the class labels of all the selected feature vectors is performed in order to classify POI \( p \). On the other hand, in the majority decision process, the ordinary kNN classifier just uses the class labels of the \( k' \)-nearest feature vectors corresponding to POI \( p \).

The following is a brief description of the procedure of the proposed MkNN classifier.

**Step1:** The neighboring pixels \( \{\mathbf{p}_m; m = 1, \ldots, k\} \) of POI \( p \) in the observation space are selected. Note that \( \{\mathbf{p}_m\} \) includes \( p \).

**Step2:** The Euclidean distances \( \|\mathbf{v}_n - \mathbf{u}_m\| \) between each \( \mathbf{u}_m \) and all the feature vectors \( \{\mathbf{v}_n; n = 1, \ldots, N\} \) are calculated.

**Step3:** The feature vectors \( \mathbf{v}_n \) which satisfy the following condition:

\[
\|\mathbf{v}_n - \mathbf{u}_m\| \leq \|\mathbf{u}_m - \mathbf{v}_m^{(k')}\|,
\]

are selected for each \( \mathbf{u}_m \), where \( \mathbf{v}_m^{(k')} \) stands for the \( k' \)-th nearest feature vector around \( \mathbf{u}_m \).

**Step4:** The feature vector of POI \( p \) in the observation space is classified into class \( c \) as follows:

\[
c = \underset{j}{\arg \max} \sum_{m=1}^{k} \sum_{n=1}^{N} U(\mathbf{u}_m, \mathbf{u}_n, i_n, j),
\]

where \( j \) represents a class label. Equation (4) is a function for a majority decision making of a class label.

In the proposed MkNN classifier, the spatial continuities both in the observation and feature spaces are utilized in the selection process of **Step3**. That is, in the MkNN classifier, \((k \times k')\)-nearest feature vectors are selected for each feature vector corresponding to POI \( p \) based on the spatial relationship among \( p \) and \( \mathbf{p}_m \). This is summarized that the characteristic of the MkNN classifier is to utilize the information on the spatial continuities in both the observation and feature spaces. The MkNN classifier then realizes a fine classification even for the B-mode image with noises and/or measuring errors.

Parvin et al. [20] proposed the modified k-nearest neighbor (MKNN) method so far. This method has a similar idea to our proposed MkNN classifier on the point that it utilizes the information of the neighborhood of POI. However, that method only uses the information of the neighborhood in the feature space by employing a probability density estimation in the feature space.

The proposed MkNN classifier is related to image segmentation. Image segmentation methods utilizing spatial information have been proposed (e.g., [21], [22]). Although these methods take account into spatial information such as local smoothness and spatial continuity, the usage of spatial information in these methods is different from that of the MkNN classifier.

In our proposed MkNN classifier, we assume that the pixels with the same class label exist nearby in the observation space because of the continuity of tissue. Under this assumption, the proposed MkNN classifier utilizes not only the feature vector of POI \( p \) itself but also those of the neighboring pixels \( \mathbf{p}_m \) in the observation space.

The space complexity of the proposed MkNN classifier depends on the number of feature vectors stored for classification and vector quantization can be used to reduce the number of feature vectors. The proposed MkNN classifier takes more calculation time for classification than the ordinary kNN classifier. A high-speed calculation of the MkNN classifier was discussed in another paper [23].

4. Experimental Results

4.1 Quantitative Validation of Tissue Characterization of Coronary Plaque

Here we quantitatively validate the characterization performance of the proposed MkNN classifier. The tissues are
classified into the three tissues of fatty, fibrofatty and fibrous tissues. In the experiments, 20,000 data (pixels to be classified) were randomly selected from six IVUS B-mode images of the cross sections of the coronary artery of a patient. The data set contains nearly the same number of tissues for each class.

The classification results by the proposed MkNN classifier are compared to those by the conventional methods of the IB analysis, the ordinary kNN, and the Parvin’s MKNN classifiers by using 5-fold cross-validation.

In the IB analysis, the window size \( T \) was set to 32 and the threshold values for the classification were determined experimentally so that the classification performance became the best after cross-validation.

In the experiments of the ordinary kNN, the Parvin’s MKNN and the proposed MkNN classifiers, the feature vectors were power spectra calculated by the short-time discrete FFT of RF signal. The window size of the FFT was 32 pixels. The elements of the feature vector were normalized in the range of \([0, 1]\). The parameter \( k' \), which is the number of the nearest neighbors in the feature space, of the ordinary kNN, the Parvin’s MKNN and the proposed MkNN, was changed from 1 to 15. In the case of the Parvin’s MKNN classifier, the numbers of the neighborhood for the evaluation of validity in the feature space \( H \) were set to 9 and 25. In the case of the proposed MkNN classifier, the parameter \( k \), which is the number of the neighborhood in the observation space, were set to 9 (3x3) and 25 (5x5).

In this paper, C-support vector classification (C-SVC) \([24, 25]\) was adopted as a state-of-the-art classifier for further comparison with the proposed method. The C-SVC is a kind of soft-margin support vector machine. In the experiments, a radial basis function (RBF) was used as a kernel function and the one-versus-one method was used for the classification. The parameters \( C \) and \( \gamma \) of C-SVC were decided by using the grid search algorithm with cross-validation. The training data set was separated into several folds, and possible intervals of \( C \) and \( \gamma \) with grid space were provided. All the grid points of \((C, \gamma)\) were set to give the highest cross-validation accuracy. A fold was considered as a validation set and the rest were for training. For the 20,000 data, the parameters \( C \) and \( \gamma \) of C-SVC were decided to be \( C = 2.089 \) and \( \gamma = 0.201 \), respectively. The feature vectors were the same as those of the ordinary kNN, the Parvin’s MKNN and the proposed MkNN classifiers.

Figure 5 shows the classification accuracy rates of all the methods under comparison. The classification accuracy rate is the mean value of the true positive rate (TPR) of each class. The results of the IB analysis and C-SVC are shown by the flat dotted-lines with narrow pitch and broad pitch, respectively, because they are independent of \( k' \). With those results, it is observed that the proposed MkNN classifier gives a superior performance to the IB analysis, the ordinary kNN, and the Parvin’s MKNN classifiers.

Comparing to the C-SVC, the performance was evaluated by Welch’s statistical significance test \([26]\) in two cases: C-SVC versus the proposed MkNN classifier \((k = 9)\), and C-SVC versus the proposed MkNN classifier \((k = 25)\).

**Fig. 5** Classification accuracy rate by each method for the tissue characterization of coronary plaque. The horizontal axis represents the number \( k' \) for the ordinary kNN, the Parvin’s MKNN, and the proposed MkNN classifiers. The vertical axis represents the classification accuracy rate. A classification accuracy rate is the mean value of the true positive rate (TPR) of each class. The results of the IB analysis and the C-SVC are shown by the flat dotted-lines with narrow pitch and broad pitch, respectively.

With all those results, we conclude that the proposed MkNN classifier \((k = 25 (1 \leq k' \leq 11)), k = 9 (1 \leq k' \leq 5)\) has better classification performance than the other methods. Especially, the comparison between the Parvin’s MKNN and the proposed MkNN classifiers fully clarifies the effectiveness of utilizing the spatial continuities both in the observation and the feature spaces.

### 4.2 Qualitative Validation of Tissue Characterization of Coronary Plaque Superimposed on IVUS B-Mode Image

In order to evaluate qualitatively the performance of the proposed MkNN classifier for the tissue characterization of coronary plaque, the characterization was executed for three kinds of tissues using three test data which were taken from the same patient and were different from the six images in Sect. 4.1. The tissue characterization results are superimposed on the IVUS B-mode images.

Figures 6 (a)-(c) show the diagnosed compositions for each region of interest (ROI). The areas encircled by the yellow, blue, and green curves correspond to fibrous, fibrofatty, and fatty tissues, respectively. The diagnosis was done by a medical doctor observing the tissues stained with Masson’s trichrome staining through the microscope.
The diagnosed areas (ROIs) are mapped on the IVUS B-mode images in Figs. 6 (d)-(f). The composition of tissues is a little bit different between the corresponding images, i.e., (a)-(d), (b)-(e), and (c)-(f). This is because the stained thin slice of the coronary artery is slightly transformed and sometimes damaged when expanded on the slide prepared for viewing through a microscope.

The tissues (pixels) were classified into three classes: fibrous, fibrofatty, and fatty tissues. The feature vectors, whose classes were known in advance, were provided from the other IVUS B-mode images. The number of the feature vectors was 50 for each class, hence the number of all the feature vectors was 150. The feature vectors had been quantized by vector quantization of the 20,000 original feature vectors.
vectors in order to reduce the computational cost. The sizes of the neighborhood areas in the observation space and in the feature space for the proposed MkNN classifier were assigned to be $k = k' = 9$ by considering the experimental results of the pre-experiments, the computational cost, and so on. Although the condition $k' = 1$ gave the best classification performance in Sect. 4.1, the characterization result with $k' = 1$ is not good for visual inspection because it looks like having many impulse noises. We conducted the experiments to find a better setting such that the detail information such as a fibrous cap was kept as much as possible and noise-like characterization was reduced for visual inspection by medical doctors.

The classification performance of the proposed MkNN classifier was compared to that of the IB analysis because the IB analysis has been broadly used in the field of cardiology. In the IB analysis, the window size $T$ was set to 32, and the threshold values for the classification were determined experimentally so that the classification performance became the best.

Figures 6(g)-(i) show the characterization results by the IB analysis for the images shown in Figs. 6(d)-(f), respectively. The characterization results by the proposed MkNN classifier are shown in Figs. 6(j)-(l). It is seen that the proposed MkNN classifier gives better characterization results than the IB analysis. Many pixels far from the ultrasound probe are misclassified by the IB analysis.

A fibrous cap is a thin layer which covers the soft tissues such as fatty and fibrofatty tissues. With this structure of plaque, this may cause a serious rupture of plaque, and it must be treated at once, otherwise it goes to intravascular coagulation. The proposed MkNN classifier was able to visualize this dangerous fibrous cap as seen in Fig. 6(j), which however could not be correctly characterized by the IB analysis.

In Fig. 6(h), the upper part of ROI is misclassified by the IB analysis. This part should have been classified as the fibrous tissue. In addition, a part of the left-hand side of ROI in Fig. 6(h) had to be characterized as a layer of fibrofatty and fatty tissues. The proposed MkNN classifier characterizes this part sufficiently well as seen in Fig. 6(k).

In Fig. 6(i), in the left-hand side of ROI, the fibrofatty and fatty tissues are misclassified completely in reverse by the IB analysis. In this area, the plaque is layered with fibrous, fatty, and fibrofatty tissues in this order outward. With this composition, the fibrous tissue is thick enough to guard the soft fatty core (fatty and fibrofatty tissues) inside, and the urgent medical treatment is not necessary for the moment. This composition is adequately grasped by the proposed MkNN classifier as shown in Fig. 6(l).

From all those results, it is qualitatively confirmed as well that the proposed MkNN classifier is most effective among the methods compared.

5. Conclusions

In this paper we have proposed an enhanced k-nearest neighbor classifier named the multiple k-nearest neighbor (MkNN) classifier. The proposed MkNN classifier uses the information both in the feature space and in the observation space.

In order to verify the performance of the proposed MkNN classifier, it was applied to the tissue characterization problem of coronary plaque. In the experiments, not only the feature vector in the feature space corresponding to the pixel of interest (the tissue to be classified) but also the feature vectors corresponding to the pixels around the pixel of interest in the observation space were used. The experimental results suggest that the proposed MkNN classifier is promising for a practical tissue characterization of coronary plaque.

Future works are the improvement of classification accuracy by using the other feature vectors for coronary plaque tissue characterization and the applications to the other classification problems. The applications to the remote sensing images are also left for future works.

Acknowledgements

Many thanks are due to Dr. T. Hiro and Dr. H. Matsuzaki for providing the medical data and for their valuable discussions. Thanks are also due to Mr. T. Okamoto and Ms. M. Kunihiro for their helpful assistance.

References

[10] A. Nair, B.D. Kuban, E.M. Tuzcu, P. Schoenhagen, S.E. Nissen, and


