Classification Performance Comparison of Feature Vectors Based on Summation Scheme and Maximization Scheme

Haitao Yu

***Abstract*—The basis of object recognition is the feature extraction from the received signals. But the real signals have the three-non characteristics of nonlinear, non-stationary and non-Gaussian properties, which make it difficult to extract features of targets accurately. Hence, Higher-Order Statistics (HOS) and generalized diagonal slices spectra are introduced. Feature vectors E_1 and E_2 are extracted on the basis of generalized diagonal slices spectra using simple summation and maximization schemes for underwater targets, respectively. Summation and maximization of the horizontal or vertical slices spectra, assigned feature vector names as E_3 and E_4, respectively, are used as comparison. In order to compare the performance and reduce the dimension of feature vectors based on the generalized diagonal slices spectra and horizontal or vertical slices spectra, feature selection scheme based on Fisher's class separability is introduced. The classification accuracies of feature vectors E_1, E_2 E_3, and E_4 are testified by One-against-One (OAO) method of multi classification of Support Vector Machine (SVM) for different segment numbers of the training sets computing and different largest measure number M. The results show that the total performance of feature vectors based on diagonal slices spectra is better than that of horizontal or vertical slices spectra and that the performance of feature vectors based on summation scheme is better than that of maximization scheme.**

***Index Terms*—feature extraction, feature selection, Fisher's class separability, generalized diagonal slices**

I. Introduction

Automatic detection, classification and recognition technologies of underwater targets have become the bottleneck of the development of underwater technology due to three-non characteristics of nonlinear, non-stationary and non-Gaussian properties of underwater signals [1], [2]. Whereas, Higher-Order Statistics (HOS) is applied extensively for it can not only detect and represent the nonlinear characteristic in signals and systems, but also restrain the impact of Gaussian colored noises from non-Gaussian signals [3].

However, the large computation load and output dimensions are the main factors of hindering its realistic application. Many scholars have done quite a number of researches in feature extraction and dimension reduction based on HOS. One of the outstanding works is the diagonal slice and its spectrum proposed by Nagata [3]. But the diagonal slice is only one line of bispectrum. And it does not make full use of the information provided by bispectrum and has the risk of information lost. Reference [4] proposed a new feature extraction scheme based on horizontal slices spectra or vertical slices spectra in the low frequency domain and has good classification accuracy. From the point of view of the characteristics of underwater acoustic signals, it is feasible to choose the low frequency domain of bispectrum as the feature extraction region. But according to our simulations, there are still valid target features which can improve the classification accuracy in the middle and high frequency domain. And according to the habit of human eyes to classify the bispectrum, one may differentiate signals from the vice-diagonal direction of the bispectrum easily. Hence, a feature extraction scheme which considers the habit of human eyes and has no risk of information lost is expected. And lines parallel to the diagonal slice are attempted to be used for feature extraction in this paper. Thus, the generalized diagonal slices and their spectra are proposed in part II. And the feature extraction schemes based on generalized diagonal slices spectra are given in part III. Due to the use of the FFT in the generalized diagonal slices spectra estimation, feature vector dimension depends on the FFT points. If a large point of FFT is used, the feature vector dimension will also increase and classification will cost more time. Thus, dimension reduction of feature vector should be considered. Feature selection based on Fisher's class separability [5] is rewritten in part IV. The experiment results and conclusion will appear in Part V and Part VI, respectively.

II. Generalized Diagonal Slices and Their Spectra

Simply speaking, HOS [3] are the moments or cumulants higher than 2nd order and their spectra. Third-order spectrum, namely bispectrum, is the most common use in higher-order spectrum. It is the double FT of third order cumulant $c_{3x}(\tau_1, \tau_2)$ and its definition is as follows

$$B_x(\omega_1,\omega_2)=\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}c_{3x}(\tau_1,\tau_2)\,e^{-j(\omega_1\tau_1+\omega_2\tau_2)}, \qquad (1)$$

where

$$\begin{aligned}c_{3x}(\tau_1, \tau_2) &= cum\{\, x(n)\, x(n+\tau_1)\, x(n+\tau_2)\,\}\\ &=E\{[\,x(n)-m_{1x}][\,x(n+\tau_1)-m_{1x}][\,x(n+\tau_2)-m_{1x}]\}, \qquad (2)\end{aligned}$$

Manuscript received January 22, 2011. This work was supported in part by the Education Reform Program Funded by Xi'an University of Science and Technology (Program NO.JG10066).

Haitao YU is now with the School of Science, Xi'an University of Science and Technology, Xi'an 710054, P.R. China (phone: +86-29-85583133; fax: +86-29-85583131; e-mail: haitaoyu@xust.edu.cn). And he is also a Ph.D Candidate of the College of Marine, Northwestern Polytechnical University, Xi'an 710072, P.R. China.

and τ_1 and τ_2 are the time differences.

Nagata [3] proposed the concept of 1-D slices and their spectra in 1970. Many types of 1-D slices are possible, including radial, vertical, horizontal, diagonal, and offset-diagonal [6]. Tugnait and Friedlander et al. derived the generalized GM equation by setting $\tau_1 = \tau$ and $\tau_2 = \tau+m$ [6]. And they can be regarded as the rudiment of the generalized diagonal slices proposed next.

According to the habit of human eyes to classify the bispectrum, as shown in Fig. 1, one may differentiate signals from the vice-diagonal direction of the bispectrum easily. But only one line of bispectrum, for example, the diagonal slice, provides limited information for signals classification for both human eyes and classifiers. Hence, the slices parallel to the diagonal slice should be included. And the generalized diagonal slices of third order cumulant are proposed as

$$c_{3x}(\tau, \tau\pm q) = cum\{ x(n)\, x(n+\tau)\, x(n+\tau\pm q) \}, \quad (3)$$

where q is the amendment quantity of time differences τ and represents the distance between the original diagonal slice and other slices parallel to it in bifrequency plane. And its spectrum is

$$B_x(\omega,\omega\pm q) = \sum_{\tau=-\infty}^{\infty} c_{3x}(\tau,\tau\pm q)e^{-j(2\omega\tau\mp\omega q)}. \quad (4)$$

Let $X(\omega)$ be the Fourier Transformation of $x(n)$. Then the generalized diagonal slices spectra can be represented as

$$B_x(\omega, \omega\pm q)= X(\omega)\, X(\omega\pm q)\, X^*(2\omega\pm q)\,. \quad (5)$$

As we can see from (3), (4) and (5), the generalized diagonal slices and their spectra degrade into the diagonal slice and its spectrum, hereinafter, “the original diagonal slice” and “the original diagonal slice spectrum”, when amendment quantity q=0. Equation (4) and (5) provide two different implement methods of generalized diagonal slices spectra in programs. And the physical meaning of the generalized diagonal slices spectra depicts lines parallel to the original diagonal slice. It provides more information for signal classification than original diagonal slice.

III. Feature Extraction Schemes

Many scholars have proposed a large number of feature extraction schemes based on the original diagonal slice and its spectrum. The original diagonal slice spectrum is only one line in bifrequency domain. And feature extraction schemes based on it have the risk of information lost, which makes the classification accuracy not high enough.

Hence, feature extraction schemes based on the generalized diagonal slices spectra are expected. In other words, one can extract more features using lines parallel to vice-diagonal direction than the original diagonal slice spectrum which is only one line in bifrequency domain. And such features have little risk of information lost. But if more than one feature is extracted from one slice spectrum, there are too many features obtained by the generalized diagonal slices spectra. And if only one feature is acquired from one line, the dimension of feature vector will be reduced. Thus, two simple feature extraction schemes based on the generalized diagonal slices spectra are proposed as follows

$$E_1 = \{E_{1q}\} \quad (6)$$

and

$$E_2 = \{E_{2q}\}, \quad (7)$$

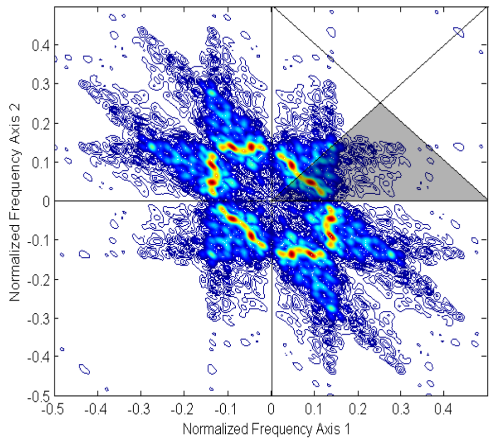

Figure 1. Shadow region based on the symmetric characteristics of bispectrum.

where

$$E_{1q}=\text{sum}\{B_x(\omega, \omega\pm q)\} \quad (8)$$

and

$$E_{2q}=\max\{B_x(\omega, \omega\pm q)\}, \quad (9)$$

q=0, 1, 2, … , N-1. And N is the FFT dot number.

The purpose of (6) and (7) is to obtain the summations and maxima of magnitudes of every slice spectrum, respectively. Hence, two feature vectors E_1 and E_2 are formed by summation and maximization, respectively. Maximization scheme is equivalent to selecting only the largest element. The summation scheme is equivalent to the energy accumulation. And it is expected that the summation scheme can effectively reduce the signal disturbance on the classification accuracy. In order to validate the separabilities of feature vectors E_1 and E_2, summation and maximization of the horizontal or vertical slices spectra, assigned feature vector names as E_3 and E_4, respectively, are used as comparison. Their separabilities are validated in the coming simulation.

According to the symmetric characteristics of bispectrum, the features should be extracted in the following triangle, $\omega_1\geq\omega_2\geq0$ and $\omega_1+\omega_2\leq\pi$, to cut down the computation load further, as shown in the shadow region of Fig. 1. It is noted that dimensions of feature vectors E_1, E_2 E_3 and E_4 depend on the FFT points because of the use of the FFT in the generalized diagonal slices spectra estimation. If a large point of FFT is used, the feature vector dimension will also increase and classification will cost more time. Thus, dimension reduction of feature vectors should be considered.

IV. Feature Selection Based on Fisher’s Class Separability

Xian-Da Zhang et al. [5] proposed using selected bispectra with the maximum interclass separability as feature vectors of signals. Its basic idea is to select only the bispectra at individual bifrequency points with the most discriminant power as feature vectors. And they used Fisher’s class separability as the discriminant measure. Now, we rewrite it for feature vectors mentioned above as follows.

Consider within-class or between-class separation of class i and class j using feature vectors, such as E_1, E_2 E_3 and E_4. Suppose the training set consists of feature vector samples

$\{E_k^i(q)\}_{k=1,2,...,N_i}$ and $\{E_k^j(q)\}_{k=1,2,...,N_j}$, where the subscript k stands for feature vectors computed from the kth set of observed data, the superscript i represents the ith class of the signal, and N_i and N_j are the set number of observed data of the ith and jth class signals, respectively. Therefore, the Fisher class separability measure [5], [7] between the ith and jth classes for feature vectors is rewritten as

$$F^{(i,j)}(q) = \frac{S_B(q)}{S_W(q)}, \tag{10}$$

where S_B is the between-class scatter matrix

$$S_B(q) = \sum_l [\mu_l(q) - \mu(q)][\mu_l(q) - \mu(q)]^T, \tag{11}$$

$$\mu_l(q) = \frac{1}{N_l} \sum_k E_k^l(q), \qquad l = i, j, \tag{12}$$

$$\mu(q) = \frac{1}{2}[\mu_i(q) + \mu_j(q)] \tag{13}$$

and S_W is the within-class scatter matrix defined as

$$S_W(q) = \sum_l S_l(q), \tag{14}$$

$$S_l(q) = \sum_k [E_k^l(q) - \mu_l(q)][E_k^l(q) - \mu_l(q)]^T, \quad l = i, j. \tag{15}$$

According to the definition of $F^{(i,j)}(q)$, smaller within-class scatter matrix S_W and larger between-class scatter matrix S_B will produce larger Fisher class separability measure. That is to say, feature vector elements with larger Fisher class separability measure have the more powerful separabilities. Therefore, use (10) to compute the Fisher class separability measure $F^{(i,j)}(q)$ for all class combinations (i, j) and requeue M largest measures, where M is sufficiently large and is assigned by operators. The corresponding amendment quantity is called the "effective" amendment quantity. And if the same amendment quantities for different combinations appear, only one of them is remained. Up to this point, the selected feature vector for both training sets and test sets can be obtained by the feature vector elements corresponding to the selected amendment quantities.

V. Experiment Results and Analysis

Data of 98,304 dots length for the radiated noise of underwater targets in three types, A, B and C, are obtained in various sea states respectively. And every class data are divided into 192 segments, the length of which is 512 dots. FFT of 256 dots for the short sample data of 512 dots is adopted to estimate the spectra of the generalized diagonal slices and vertical or horizontal slices spectra in this paper. And feature vectors extraction is limited in shadow region of Fig. 1 to reduce the computation load. Fig. 2 gives the normalized bispectra of the radiated noise of underwater targets in three types. And a certain similarities among them can be seen, which give rise to the inaccurate classification.

More and more scholars believe that Support Vector Machine (SVM) is becoming the new research focus after ANN and that it will promote the further development of theories and technologies in machine learning. Radial Basis Function (RBF) is efficient in most cases as the kernel function of SVM [8]. And the optimizing of the two parameters of RBF, C & γ, can be obtained by cross-validation and grid-search [8]. Hence, we choose RBF as the kernel function. And the optimizing for C & γ is conducted by 3-fold cross-validation and every 0.3 intervals grid-search in $[2^{-5}, 2^2]$. There are many multi-SVM methods. And classification accuracy of One-against-One (OAO) method is better for most data sets [9], [10]. Thus, OAO will be used for the classification of the radiated noise from underwater targets.

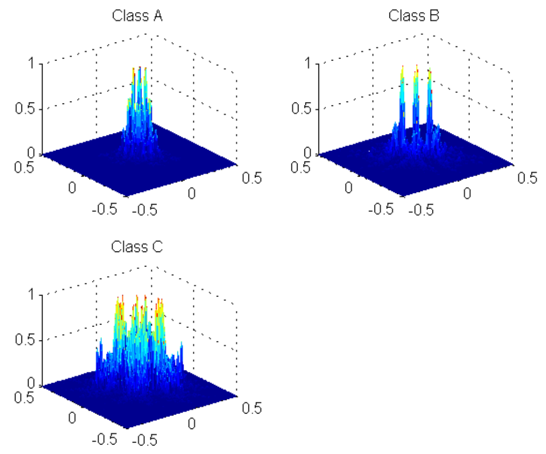

Figure 2. Normalized bispectra of the radiated noise of underwater targets.

TABLE I. Classification Results Based on Feature Vectors E_1, E_2, E_3 and E_4 by SVM When the Top 96th Segments of Every Class Is Used for the Computing of Training Sets

Feature Vectors	Class	A	B	C	Classification Accuracy (%)	Mean Classification Accuracy (%)
E_1	A	96	0	0	100	99.6528
	B	0	95	1	98.9583	
	C	0	0	96	100	
E_2	A	95	1	0	98.9583	99.3056
	B	0	95	1	98.9583	
	C	0	0	96	100	
E_3	A	96	0	0	100	100
	B	0	96	0	100	
	C	0	0	96	100	
E_4	A	95	0	1	98.9583	98.9583
	B	0	94	2	97.9167	
	C	0	0	96	100	

TABLE II. Classification Results Based on Feature Vectors E_1, E_2, E_3 and E_4 by SVM When the Top 96th Segments of Every Class Is Used for the Computing of Test Sets

Feature Vectors	Class	A	B	C	Classification Accuracy (%)	Mean Classification Accuracy (%)
E_1	A	96	0	0	100	100
	B	0	96	0	100	
	C	0	0	96	100	
E_2	A	96	0	0	100	98.9583
	B	0	94	2	97.9167	
	C	0	1	95	98.9583	
E_3	A	96	0	0	100	98.6111
	B	0	92	4	95.8333	
	C	0	0	96	100	
E_4	A	96	0	0	100	95.8333
	B	0	87	9	90.6250	
	C	0	3	93	96.8750	

In order to validate the separabilities of the feature vectors E_1, E_2, E_3 and E_4, the training sets are computed from the top 3rd, the top 9th, …, the top 96th segments for every class, respectively. And the left are used as the test sets. And the range of M mentioned in part IV is same as that of amendment quantity q.

A. From the View of Classification Results in Detail

The classification results of the top 96th segments of every class for training sets and test sets are shown in TABLE I and TABLE II, respectively.

As we can see in TABLE I, the classification accuracy of E_1 is higher than that of E_2 for training sets. Then classification accuracy of E_1 higher than that of E_2 for test sets is expected, as shown in TABLE II. The possible reason is that maximization scheme is equivalent to selecting only the largest element but the summation scheme is equivalent to the energy accumulation.

And it is expected that the separabilities of feature vector formed by summation scheme is higher than that by maximization scheme, even based on spectra of horizontal or vertical slices. And the simulation can verify the point. The classification results for both training sets and test sets are shown in TABLE I and TABLE II. It is apparent that the separabilities of feature vectors E_1 and E_3, which are formed by summation scheme, are better than feature vector E_2 and E_4, which are formed by maximization scheme for generalized diagonal slices spectra and spectra of horizontal or vertical slices, respectively.

The evident differences can be seen class A from class B and C in Fig. 2. Thus, the feature vector which can fully distinguish class A from class B and C should be gotten easily. And we can see that classification accuracies of feature vectors E_1, E_2, E_3 and E_4 for class A of test sets are all 100%.

And the evident similarities can be found in class B and class C in Fig. 2. The classification accuracy of E_4 for class B and C is much lower than that of E_1 and E_2 for both training sets and test sets. Especially, the classification accuracy of feature vector E_4 for class B of test sets is only about 90%. It is noted that the mean classification accuracy of feature vectors E_1 and E_2 based on the generalized diagonal slices spectra is higher than that of feature vectors E_3 and E_4 based on horizontal or vertical slices spectra. It indicates that feature extraction scheme based on the generalized diagonal slices according to the habit of human eyes to classify the bispectrum is feasible.

It is noted that the classification accuracy of E_1 for all class is 100% for test sets. It shows that feature vector extracted by the summation scheme based on the generalized diagonal slices according to the habit of human eyes to classify the bispectrum has the utmost classification accuracy.

B. From the View of Mean Classification Accuracy

As we can see from Fig. 3 ~ Fig. 12, the classification accuracies of feature vectors E_1, E_2, E_3 and E_4 is improved and become steady with the increasing of segments of every class used for the computing of training sets. And the classification accuracies of feature vector E_4 is the worst due to its low classification accuracy and its instability even the segments of every class used for the computing of training sets are the top 96th ones, half of the total segments for each class. The classification accuracies of E_3 is better than those of E_2 when the segment numbers for the training sets computing are less, but vice verse when the segment numbers for training sets are larger.

It is observed that the classification accuracies of the four vectors are almost the same when the largest measure number M is larger than 80 from Fig. 5. However, with the increasing of segment numbers for the training sets computing, the classification accuracies of feature vector E_1 are higher than those of the other three feature vectors. And several accuracies of 100% for feature vector E_1 are obtained for a few largest measure number M in Fig. 6. But the classification accuracies of 100% are not steady with the increasing of M. And we can see that the classification accuracies of 100% are steadier with the increasing of segment numbers for the training sets computing from Fig. 7 ~ Fig. 12. But the classification accuracies of feature vector E_1 is the steadiest one, compared to E_2, E_3 and E_4, when the training sets number equals the test sets number, as shown in Fig. 12.

It is noted that there is a little cut-down for the classification accuracies of E_1 when M is near 128 in Fig. 7,

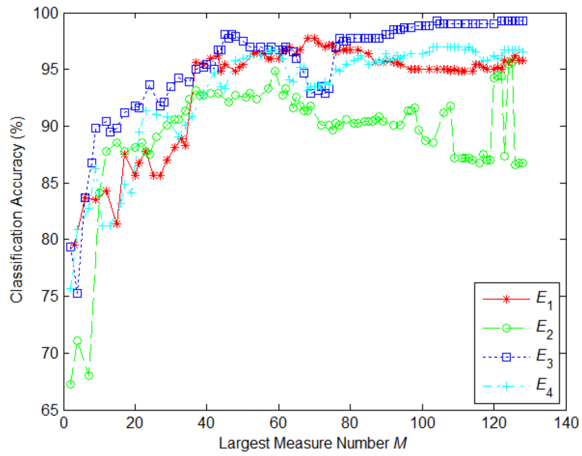

Figure 3. Classification accuracy versus largest measure number M when the top 18th segments of every class is used for the computing of training sets.

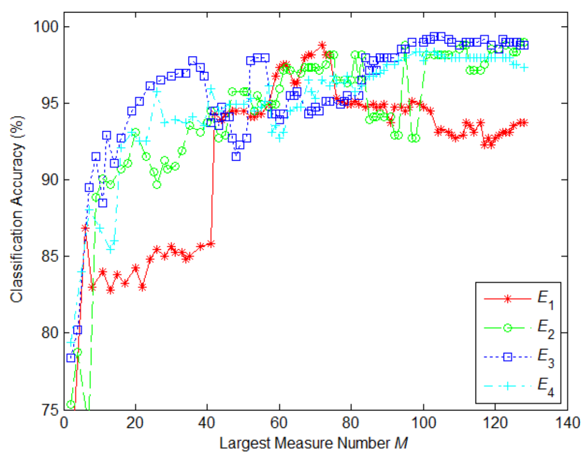

Figure 4. Classification accuracy versus largest measure number M when the top 27th segments of every class is used for the computing of training sets.

Fig. 10 and Fig. 11, but not has Fig. 12. And we can observe that the classification accuracies of E_1 reach 100% when the largest measure number M is about 100 in Fig. 7 and Fig. 10, but 60 in Fig. 11 and Fig. 12. It can be inferred that the less largest measure feature vector elements is needed when there are enough segments for training sets.

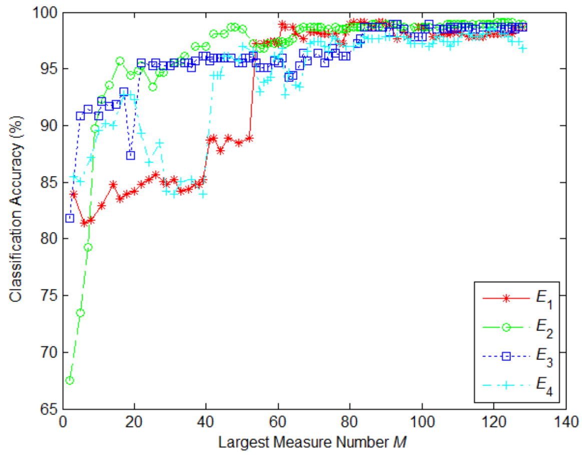

Figure 5. Classification accuracy versus largest measure number M when the top 36th segments of every class is used for the computing of training sets.

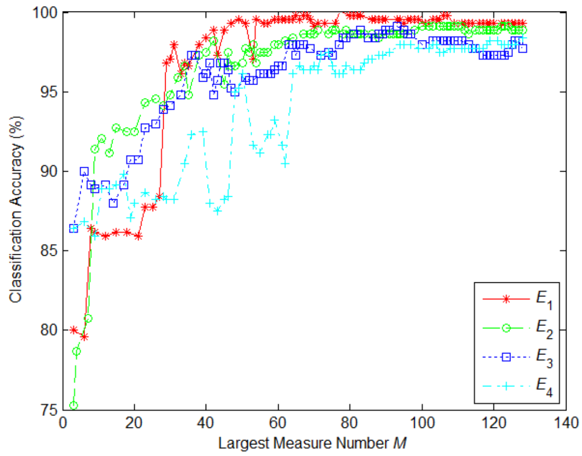

Figure 6. Classification accuracy versus largest measure number M when the top 45th segments of every class is used for the computing of training sets.

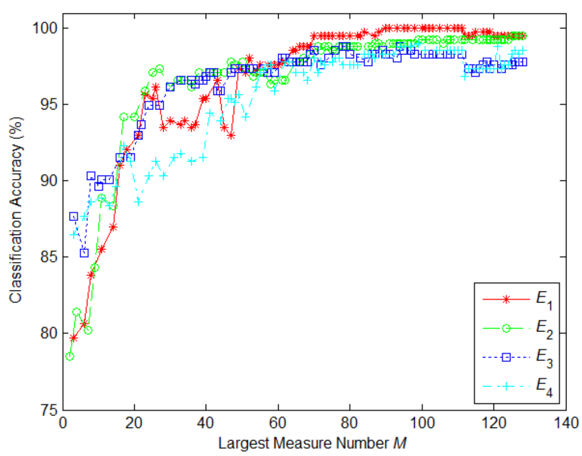

Figure 7. Classification accuracy versus largest measure number M when the top 54th segments of every class is used for the computing of training sets.

In general, the performance of feature vector E_1 is better than that of E_2 and E_3 is better than E_4. The common ground of E_1 and E_3 is that they are based on the summation scheme and E_2 and E_4 are on the maximization scheme. The possible reason is that maximization scheme is equivalent to selecting only the largest element but the summation scheme is

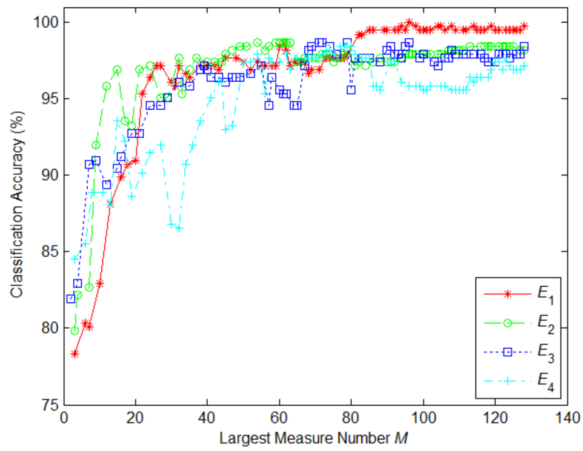

Figure 8. Classification accuracy versus largest measure number M when the top 63rd segments of every class is used for the computing of training sets.

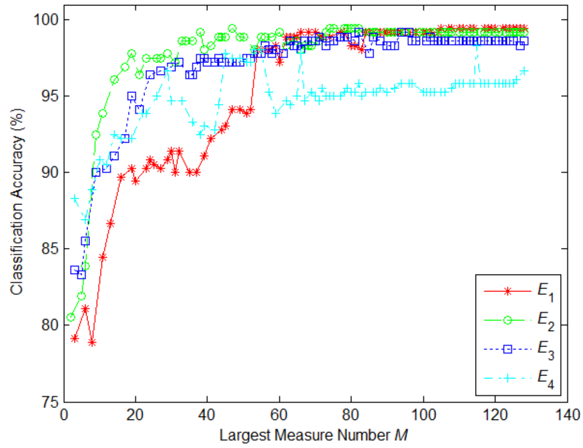

Figure 9. Classification accuracy versus largest measure number M when the top 72nd segments of every class is used for the computing of training sets.

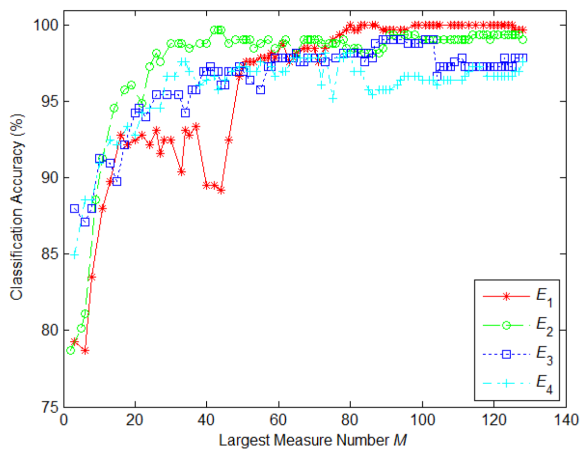

Figure 10. Classification accuracy versus largest measure number M when the top 81st segments of every class is used for the computing of training sets.

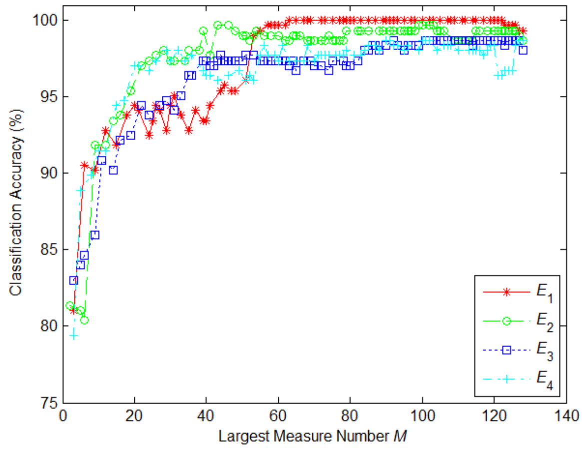

Figure 11. Classification accuracy versus largest measure number M when the top 90th segments of every class is used for the computing of training sets.

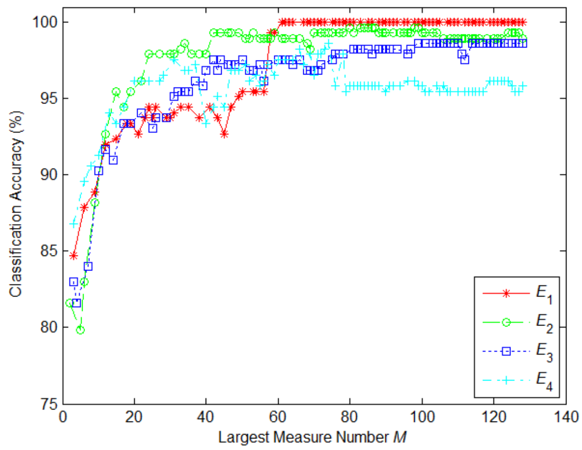

Figure 12. Classification accuracy versus largest measure number M when the top 96th segments of every class is used for the computing of training sets.

equivalent to the energy accumulation. And the summation scheme can effectively reduce the signal disturbance on the classification accuracy.

From another view, the total performance of feature vector E_1 and E_2 is better than that of E_3 and E_4. The common ground of E_1 and E_2 is that they are based on the generalized diagonal slices spectra and E_2 and E_4 is on the horizontal slices spectra. It shows that feature vectors based on the generalized diagonal slices spectra are consistent with the habit of humaneyes to classify the bispectrum. And they have the higher classification accuracies. Therefore, the feature vector extracted by the summation scheme based on the generalized diagonal slices has the utmost classification accuracy.

And one may find that there is a little classification accuracy cut-down for feature vectors E_1, E_2, E_3 and E_4 in Fig. 8 and Fig. 9. But the segment quantity of every class is used for training sets is the top 63rd and 72nd ones, respectively, whose numbers for training sets are not small. Hence, the cut-down reason is one of the next researches.

At the same time, the classification accuracies of the four vectors are almost the same when the largest measure number M is larger than 80 in Fig. 9. And this phenomenon also occurs in Fig. 5. The segment quantity of every class is used for training sets is the top 72nd and 36th ones, respectively. The number of the former is twice of the latter. Thus, one may think that this phenomenon may be conducted by classifier. And this could be another next research.

VI. Conclusion

The classification accuracies of feature vectors E_1, E_2 E_3, and E_4 are testified by OAO method of multi classification of SVM for different segment numbers of the training sets computing and different largest measure number M. The results show that the total performance of feature vectors based on diagonal slices spectra is better than that of horizontal or vertical slices spectra and that the performance of feature vectors based on summation scheme is better than that of maximization scheme. And the physical meaning of the selected features is the next research.

Acknowledgment

Special thanks to the organizing committee of the 3rd International Conference on Machine Learning and Computing 2011 (ICMLC 2011) for the recommendation to IJMLC. And an earlier version of this paper was presented at ICMLC 2011.

References

[1] Ma Yuanliang, “Challenges and potential of development for current signal processing on underwater acoustic,” the National Acoustics Conference 2002, Acoustical Society of China, pp. 5-8.

[2] Yu, Haitao; , " Feature Extraction and Selection Based on Generalized Diagonal Slices Spectra for Underwater Targets," Machine Learning and Computing, 2011., International Conference on , 26-28 Feb 2011, to be published.

[3] Zhang Xianda, Modern Signal Processing, 2nd ed., Tsinghua University Press: Beijing, 2002, Ch 5.

[4] Peng Yuang, Shen Liran, Li xueyao and Wang Kejun, “Bispectrum based feature extraction and classification of radiation noise from underwater targets,” Journal of Harbin Engineering University, vol. 24(4), Aug. 2003, pp. 390-394.

[5] Xian-Da Zhang; Yu Shi; Zheng Bao; , "A new feature vector using selected bispectra for signal classification with application in radar target recognition," Signal Processing, IEEE Transactions on , vol.49, no.9, Sep 2001, pp.1875-1885.

[6] Nikias, C.L.; Mendel, J.M.; , "Signal processing with higher-order spectra," Signal Processing Magazine, IEEE , vol.10, no.3, Jul 1993, pp.10-37.

[7] Vojtech Franc and Vaclav Hlavac, “Statistical Pattern Recognition Toolbox for Matlab User’s guide”, June 24, 2004, Available: ftp://cmp.felk.cvut.cz/pub/cmp/articles/Franc-TR-2004-08.pdf.

[8] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, “A practical guide to support vector classification,” Apr. 3, 2010, Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[9] Yu Hui and Zhao Hui, “New research on multi-classification based on Support Vector Machines,” Computer Engineering and Applications, vol. 44(7), 2008, pp. 185-189.

[10] Gou Bo and Huang Xianwu. “SVM multi-class classification,” Journal of Data Acquisition & Processing, vol. 21(3), Sep. 2006, pp. 334-339.

Haitao Yu was born in Changwu County, Shaanxi Province, P.R. China, on June 10, 1979. He received his master degree in System of Communication and Information from School of Communication and Information Engineering, Xi’an University of Science and Technology (XUST), Xi’an, P.R. China, in July, 2005. He is now a Ph.D Candidate of the College of Marine, Northwestern Polytechnical University, Xi’an, P.R. China. And his major field of study is information perception of underwater physical fields.

Mr. Yu has been positioned as a Lecturer since 2005 in XUST. And his current research interest areas include acoustic signal processing, image processing, physical field theory and teaching in physics.