USING PRINCIPAL COMPONENT ANALYSIS AND CANONICAL DISCRIMINANT ANALYSIS FOR MULTIBEAM SEAFLOOR CHARACTERISATION DATA

ZBIGNIEW ŁUBNIEWSKI, ANDRZEJ STEPNOWSKI

Gdańsk University of Technology, Department of Geoinformatics
Narutowicza 11/12, 80-233 Gdańsk, Poland
lubniew@eti.pg.gda.pl

The paper presents the seafloor characterisation based on multibeam sonar data. It relies on using the integrated model and description of three types of multibeam data obtained during seafloor sensing: 1) the grey-level sonar images (echograms) of seabed, 2) the 3^D model of the seabed surface which consists of bathymetric data, 3) the set of time domain bottom echo envelopes received in the consecutive sonar beams. The classification is performed by utilisation of several statistical methods applied for analysis of a set of seafloor descriptors derived from multibeam data. In the paper, the use of Principal Component Analysis (PCA), as well as Canonical Discriminant Analysis (CDA) for reduction of the seafloor parameter space dimension is presented along with the obtained results. In addition, the use of the open source World Wind Java SDK tool for implementation of imaging and mapping of seafloor multibeam data, integrated with other elements of a scene and overlaid on rich background data, is also shown.

INTRODUCTION

Multibeam sonars are widely used in applications like high resolution bathymetry measurements, underwater object detection and imaging, etc. Also, they are the promising tool in seafloor characterisation and classification, having several advantages over conventional single beam echosounders. The proposed approach to seafloor classification relies on the combined use of three different techniques. In each of them, a set of descriptors foreseen to be applied in seabed classification procedure, is calculated using a given type of data obtained from multibeam sonar system: 1) the grey-level sonar images of seabed, 2) the 3^D model of the seabed surface which consist of bathymetric (x, y, z) points, 3) the set of time domain echo envelopes received in the consecutive beams.

1. MATERIALS AND METHODS

The schematic concept of the applied approach was shown in Fig. 1. In the first technique used, i.e. Method 1 in the figure, the grey-level sonar echograms of seabed surface are utilised [1]. Usually, such images are generated by a multibeam sonar firmware. Next, a set of parameters describing the local region of sonar image is calculated for each bottom type. The parameters set include:

1. Basic statistical parameters describing the grey level distribution, i.e. local mean (*MEAN*) and standard deviation (*STD*).
2. Slope of the autocorrelation function of a grey level (in along track direction) approximated for a local region of the image (*SL_AUTC*).
3. Texture analysis parameters based on the Grey-Level CO-occurrence Matrix (GLCM) of a sonar image local region: entropy (*ENTR*) and local homogeneity (*HOMOG*). This technique description may be found in [1].

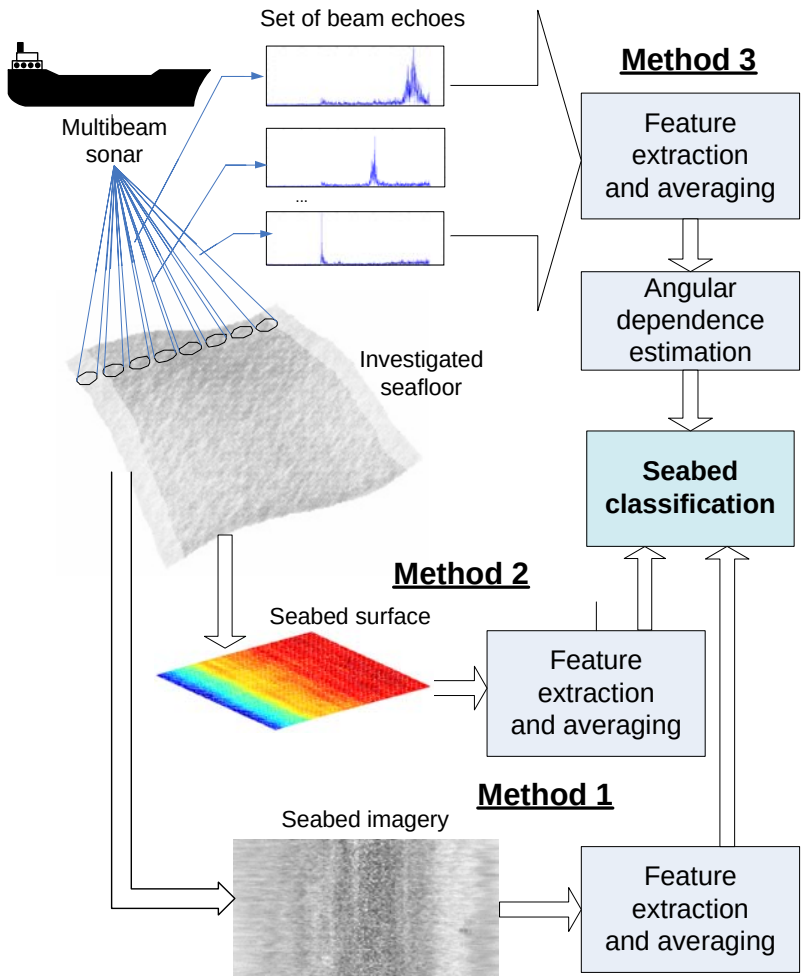


Fig.1. The concept of three combined methods of seafloor classification using multibeam sonar.

In the second technique of multibeam sonar data processing (Method 2 in Fig. 1), the 3^D "bathymetric" model of seabed surface is utilised [1]. It is constructed as a set of (x, y, z) points obtained from the detected bottom range for each beam, within the multibeam sonar seafloor imaging procedure. The examples of seabed surface model obtained for two bottom

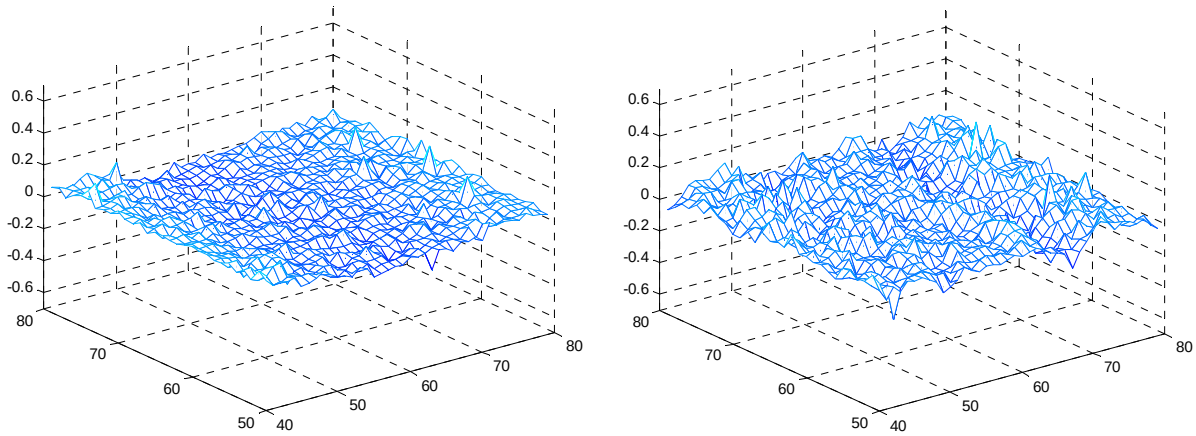
MOST WIEDZY Downloaded from mostwiedzy.pl

types, e.g. mud and coarse grained sand, are presented in Fig. 2. Next, for the local region of the constructed seabed surface, among some others, the following descriptors are calculated, viz.: rms height (*SURF_RMS*) and the slope of the seabed surface autocorrelation function (*SURF_AUTC*).

In the third technique of multibeam sonar data processing (Method 3 in Fig. 1), the set of echo signal envelopes received in the particular beams is analysed [2]. The data processing procedure in this method is more complex than in two previous ones. Firstly, after detection of a bottom echo in the received signal, the set of echo parameters is calculated for an appropriate part of each beam echo. The parameters include:

1. The normalised moment of inertia I of the echo envelope, with respect to the axis containing its gravity center [3].
2. Fractal dimension D of an echo envelope, interpreted as a measure of its shape irregularity. It is calculated as a box dimension approximation, as described in [2].

Next, for each seabed type, the dependence of I and D parameter values of the particular beam incident angle is estimated, and then, for the application in seafloor classification procedure, the following parameters are calculated for each sounding (swath): 1) the approximated slope of the angular dependence of the beam echo moment of inertia $I(\varphi)$, for the angle range of [2°, 17°] (*I_SLOPE*), and 2) the same approximated slope for the beam echo fractal dimension $D(\varphi)$, for the angle range of [4°, 19°] (*D_SLOPE*).



a) b)

Fig.2. The examples of seabed surface model obtained for two bottom types:
a) mud, b) coarse grained sand. Axes in meters.

Finally, using the results obtained by the techniques described above, the 2^D plots of calculated values for selected pairs of echo parameters were constructed. The obtained results were reported in [1], [2] and [4]. Sample result is presented in Fig. 3.

The field experiment summarizes as follows. The data used in the experimental verification of the proposed approach were acquired using Kongsberg EM 3002 sonar in the Gulf of Gdańsk region of the Southern Baltic from 2007 to 2009. Several sites of different seafloor types were investigated, but the results of the current investigation refer to 4 selected sites, characterised by the following true seabed types: mud, anthropogenic sand and mud, fine grained sand, and coarse grained sand.

The sonar operating frequency was 300 kHz, the beamwidth was 1.5° x 1.5°, the transmitted pulse length: 0.15 ms, the echo sampling rate: 14.3 kHz. The bottom depth was in

MOST WIEDZY Downloaded from mostwiedzy.pl

a range between 10 m and 100 m. Approximately, 1000 swaths from each of four seafloor types were processed. For each swath, 160 beams covered the angle sector from -65° to 65°. In the first – "imaging" technique, the seabed sonar image part corresponding to the beam angle sector between 15° and 30° was selected for further processing. In the estimation of mean, standard deviation, skewness and kurtosis of an image grey level, the size of a local image region was chosen as 11 x 11 pixels. The same local region size was used for entropy and local homogeneity calculation based on GLCM. For GLCM calculation, the image was quantised on 10 grey levels. In the estimation of the autocorrelation function slope, the used window size was 61 pixels and the maximum lag was 3 pixels. In the second technique of sonar data processing, the seabed surface part covering the beam angle sector between 15° and 50° was selected for processing. In the estimation of rms height, skewness and kurtosis, the size of a local image region was chosen as 11 x 21 pixels. In the estimation of seabed surface autocorrelation function slope, the used window size was 21 pixels and the maximum lag was 3 pixels. In the third technique, i.e. echo parameter angular dependence estimation, the beam echoes corresponding to the angular sector from -25° to 25° were selected for further processing and parameter calculation.

As a sample result, the 2^D plots of (*I_SLOPE*, *SL_AUTC*) calculated parameter values (e.g. one "echo" parameter combined with one "image" parameter) for 4 investigated seabed types are presented in Fig. 3. Using these two parameters allows for good separation of almost all seabed classes, with the only exception of fine grained sand mixed with coarse grained sand (however, these two bottom types are very similar to each other). It reveals the advantage of cross combined use of seabed characterising parameters derived by different techniques presented in Fig. 1.

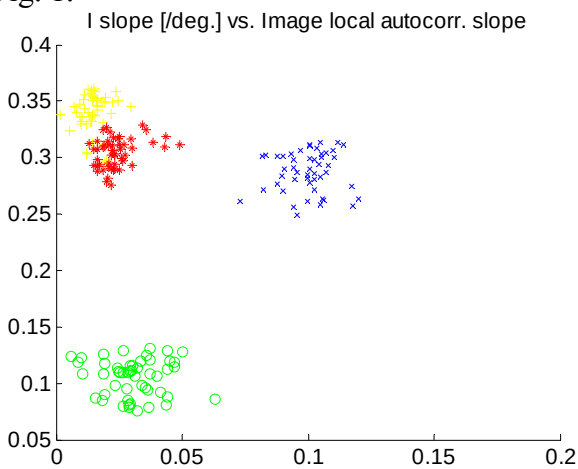


Fig.3. Plots of (*I_SLOPE*, *SL_AUTC*) calculated parameter values for 4 investigated seabed types: mud (x letters), anthropogenic sand and mud (circles), fine grained sand (crosses) and coarse grained sand (stars).

2. THE RESULTS USING PRINCIPAL COMPONENT ANALYSIS (PCA) AND CANONICAL DISCRIMINANT ANALYSIS (CDA)

Usually, the parameters (variables) used in classification as features describing the classified objects are quite highly correlated. In such a case, a reduction of the number of parameters may be applied using Principal Component Analysis method (PCA).

PCA [5] is a procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a (usually smaller) set of values of

MOST WIEDZY Downloaded from mostwiedzy.pl

uncorrelated variables called principal components (PCs). Having N variables $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_N$, which here will denote the sets of values of several calculated parameters as seabed descriptors, by PCA we are obtaining the a_{11}, a_{12}, ..., a_{1N} coefficients defining the first principal component (PC1) as a linear combination of the variables $\boldsymbol{X}$, the a_{21}, a_{22}, ..., a_{2N} coefficients defining the second principal component (PC2), and in the same way for next PCs. The criterion used in PCA is that PC1 should have the maximum possible variance from all linear combinations of X, PC2 – the maximum variance from all linear combinations ("directions") orthogonal to PC1, etc.

The PCA procedure was applied for the following set of variables: *I_SLOPE*, *D_SLOPE*, *STD*, *SL_AUTC*, *ENTR*, *HOMOG*, *SURF_RMS*, *SURF_AUTC*. Table 1 presents the obtained a_{ij} coefficients for PC1, PC2 and PC3 along with the percentage of the total variance incorporated in each of the first three PCs. Fig. 4 presents the 2^D plots of pairs of PCs' values for 4 investigated seabed types: the (PC1, PC2) pairs are presented in Fig. 4a and (PC1, PC3) pairs are presented in Fig. 4b.

Tab.1. a_{ij} coefficient values obtained for three first PCs of the analysed set of 8 parameters along with the variance percentage.

Parameter	PC1	PC2	PC3
I_SLOPE	-0.2720	0.5701	-0.0487
D_SLOPE	-0.1746	0.5310	-0.5043
STD	0.3976	-0.0713	-0.4331
SL_AUTC	0.3243	0.5114	0.1324
ENTR	0.4434	-0.0423	-0.3450
HOMOG	-0.4303	0.0920	0.2011
SURF_RMS	-0.3974	-0.2995	-0.2385
SURF_AUTC	-0.3043	-0.1629	-0.5674
% of total variance	53.7795	17.2761	11.0023

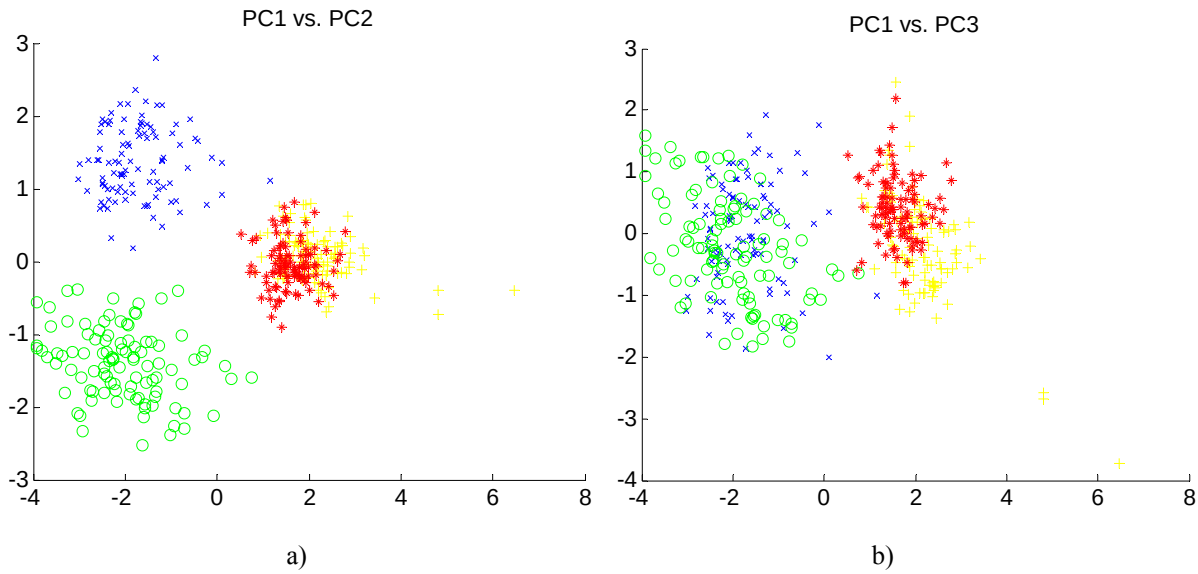


Fig.4. Plots of PC pairs: a) PC1 vs. PC2, b) PC1 vs. PC3 calculated for the analysed set of 8 parameters, for 4 investigated seabed types: mud (x letters), anthropogenic sand and mud (circles), fine grained sand (crosses) and coarse grained sand (stars).

MOST WIEDZY Downloaded from mostwiedzy.pl

It is visible that using the (PC1, PC2) pair (Fig. 4a) we may clearly distinguish the particular seabed types from the others, with one exception for fine grained sand and coarse grained sand. It confirms that we did not lose the significant information using PCA and suggests that we might simply use the (PC1, PC2) instead of the use of large number of the original parameters or instead the detailed investigation of the usefulness of each original parameter. It may also seen that PC3 seems to be not useful too much in classification (Fig. 4b), although it allows for a bit better distinction fine grained sand and coarse grained sand than in PC1 and PC2 cases.

However, it should be pointed out that in some other cases PCA may operate worse. It is due to PCA itself not use any information on the objects' true class assignment, from the training set for instance. So, it may work bad when the properties of the variables values distribution for the whole dataset differ much from those for particular classes. Then, other methods of data dimension reduction should be used, e.g. Canonical Discriminant Analysis (CDA) [6]. This method also relies on the linear, orthogonal transformation of the set of observation variables. The difference is that instead of maximalisation of the total variance, the criterion of the maximum between-classes variance and the minimum within-classes variance is used, under the assumption that the class membership is known at least for some subset of the data (i.e. the training dataset). Therefore, in the computational scheme of CDA, the $\boldsymbol{W}^{1}\boldsymbol{B}$ matrix (where $\boldsymbol{W}$ is the matrix of the within-classes sum of squares and cross-products of the variables values, and $\boldsymbol{B}$ is the matrix of the between-classes sum of squares and cross-products) and its eigenvaules are used instead of the use of covariance matrix in PCA.

The CDA procedure was applied for the same set of variables as PCA. Fig. 5 presents the 2^{D} plots of pairs of CDs' values for 4 investigated seabed types: the (CD1, CD2) pairs are presented in Fig. 5a and (CD1, CD3) pairs are presented in Fig. 5b.

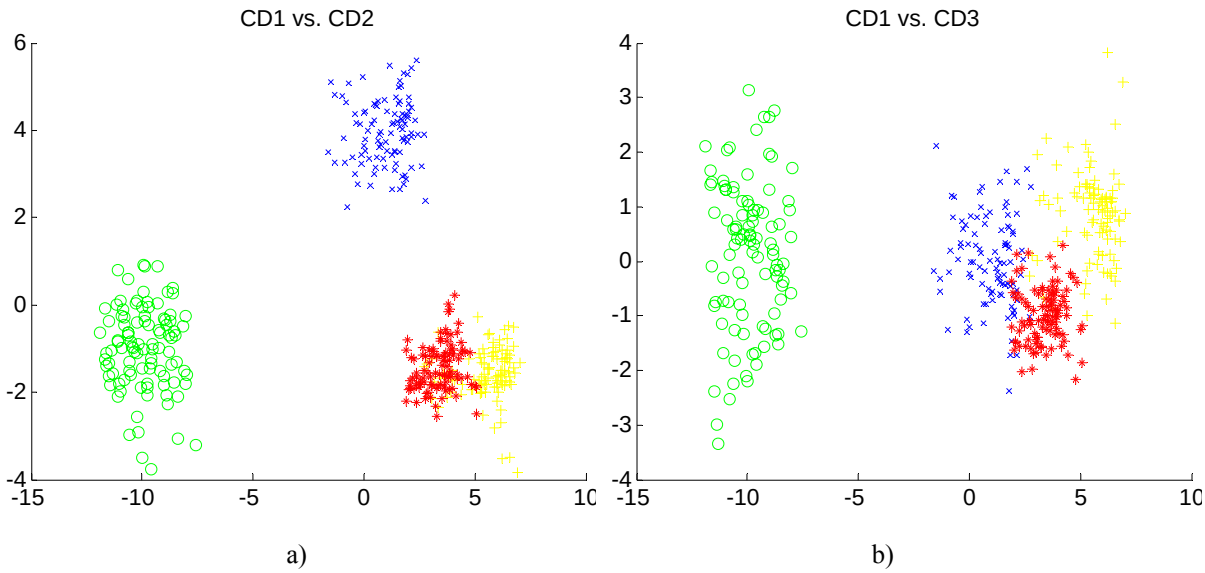


Fig.5. Plots of CD pairs: a) CD1 vs. CD2, b) CD1 vs. CD3 calculated for the analysed set of 8 parameters, for 4 investigated seabed types: mud (x letters), anthropogenic sand and mud (circles), fine grained sand (crosses) and coarse grained sand (stars).

While comparing the PCA and CDA results (Fig. 4 and Fig. 5), it is visible that, as it could be expected, CDs provide better separation of particular classes than PCs. What is

MOST WIEDZY Downloaded from mostwiedzy.pl

more, it may seen that all of three first CDs, on the contrary to the PCA case, are useful in seabed classification and provide good separation even between fine and coarse grained sand, which classes are quite similar to each other. It proofs the usefulness of the CDA application for the classification data reduction (the decreasing of the dimension of a parameter space) without the expected loss of the classification performance.

3. WORLD WIND IMAGING AND MAPPING OF SEAFLOOR MULTIBEAM DATA

In addition, the open source World Wind Java SDK tool for 3^D and 4^D implementation of imaging and mapping of seafloor multibeam data was used. The imaged data may be integrated with other elements of a scene and overlaid on rich background data. This software allows the user for easy implementation and customisation of a specific application, and taking advantage of an intuitive interface, the creation of 3^D images and 4^D animations, as well as easy integrating of the scene with additional, multidimensional environmental data from a huge number of sources. In the presented visualisation (Fig. 6) the multibeam bathymetric data are drawn using the color related to the pixel's depth (*z*-scale) value, but it is possible to easy switch to the mode where the color denotes the seabed class assignment.

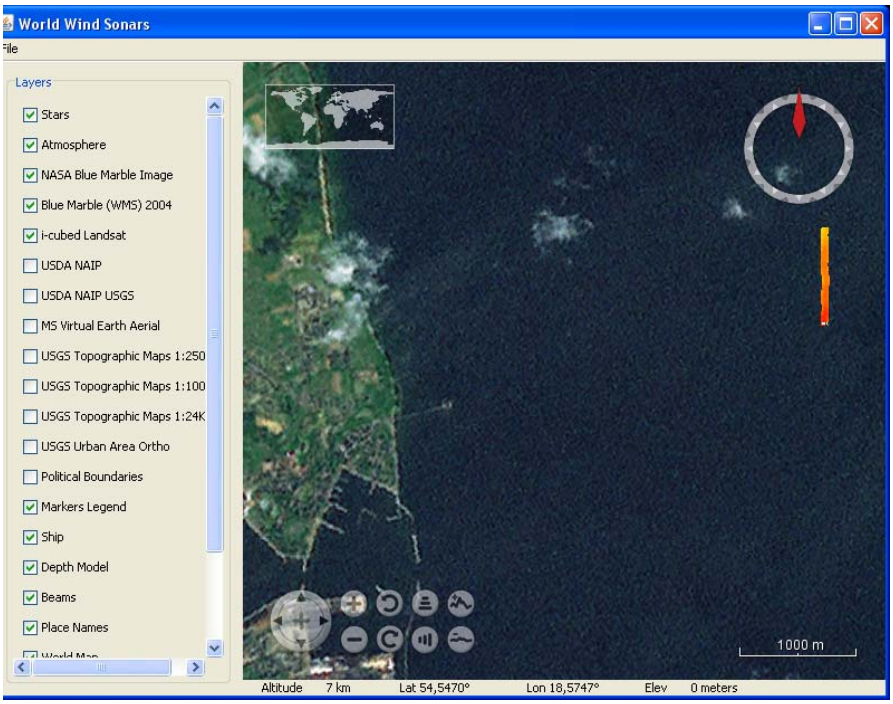


Fig.6. World Wind Java SDK sample visualisation of multibeam bathymetric data overlaid on other geographic data in Gdańsk Bay region.

MOST WIEDZY Downloaded from mostwiedzy.pl

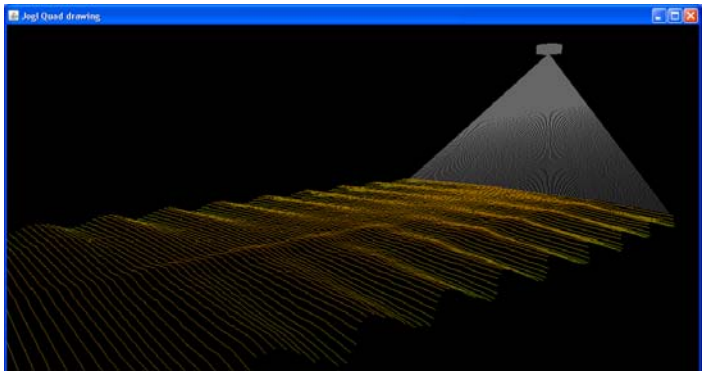


Fig.7. Screenshot of the high resolution 4^D visualisation and animation of seabed bathymetry or/and water column data, performed by the dedicated software module integrated with the Word Wind system.

Fig. 6 presents the World Wind visualisation of the multibeam bathymetric data overlaid on other geographic data, including land topography data, coarse bathymetry layer and atmosphere layer (clouds) etc. Fig. 7 presents the screenshot from the dedicated module (integrated with the Word Wind main application) for the high resolution 4^D visualisation and animation of seabed bathymetry or/and water column data.

4. CONCLUSION

The approach to seafloor characterisation, which relies on the combined, concurrent use of three different methods of multibeam sonar data processing, was presented. It has been confirmed that these techniques are useful in seafloor characterisation, and the fusion of them improves the classification performance. Two methods applied for the classification data space dimension reduction, namely, the Principal Component Analysis and the Canonical Discriminant Analysis, has been presented, and their influence on the classification performance has been preliminarily compared. In addition, the use of the open source World Wind Java SDK tool for implementation of imaging and mapping of seafloor multibeam data has been also shown.

REFERENCES

[1] Z. Łubniewski, A. Stepnowski, A. Chybicki, Seafloor characterisation combined approach using multibeam sonar echo signal processing and image analysis. Proceedings of the 10th European Conference on Underwater Acoustics, pp. 131-137, Istanbul, 2010, pp. 131-137.

[2] Z. Łubniewski, A. Chybicki, Using angular dependence of multibeam echo features in seabed classification, Proceedings of the 9th European Conference on Underwater Acoustics, pp. 717-722, Paris, 2008.

[3] J. Tęgowski, Z. Łubniewski, Application of some echo parameters to the seabed classification - methodological analysis, Hydroacoustics, 4, 237-240, 2001.

[4] A. Stepnowski, Z. Łubniewski, Combined Method of Multibeam Sonar Signal Processing and Image Analysis for Seafloor Classification, Proceedings of the 2011 Symposium on Ocean Elestronics, pp. 63-69, Kochi, 2011.

[5] W. W. Cooley, P. R. Lohnes, Multivariate Data Analysis, John Wiley & Sons, Inc., 1971.

[6] W. J. Krzanowski, Principles of Multivariate Analysis, Oxford University Press, 1988.

MOST WIEDZY Downloaded from mostwiedzy.pl