

Research on Pattern Recognition Applied for Volume Estimation of Blood Chamber with Matrix of Optical Sensors

P. MARCZYŃSKI*, G. KONIECZNY AND T. PUSTELNY

Department of Optoelectronics, Silesian University of Technology, Akademicka 2, 44-100 Gliwice, Poland

The following article describes research on possibility of using pattern recognition algorithms in the optical measurement system for estimation of the blood chamber volume in the Polish Ventricular Assist Device (POLVAD). The optical system is being developed at the Department of Optoelectronics, Silesian University of Technology, Poland. Data analysis methods include a feature subset selection algorithm involving principal components analysis and objective function as quality criterion. The analysis takes into account 17 patterns reflecting particular volumes. The k -nearest neighbours method is used as pattern classifier. The pattern recognition system was initially designed for an array of gas sensors and this paper describes its further development.

DOI: [10.12693/APhysPolA.124.498](https://doi.org/10.12693/APhysPolA.124.498)

PACS: 42.30.Sy, 47.54.-r, 43.60.Lq

1. Introduction

A pattern recognition mechanism generally consists of several subsequent stages: data acquisition and preprocessing, dimensionality reduction (also termed “feature extraction”) and classification [1]. Measurement samples are characterized with multidimensional feature vectors containing superfluous information about examined quantity and they are the input data for dimensionality reduction stage. Dimensionality reduction is responsible for extracting the most pattern discriminative features utilizing projection methods like principal components analysis (PCA) [2–4] and feature subset selection algorithms [5]. The last stage, classification, is about the assignment of a test measurement to particular pattern class.

Principal component analysis is a statistical technique commonly used for extracting useful information from multidimensional datasets. It involves a linear transformation of initial data to project existing variables onto new basis vectors, called the principal components (PC), with respect to maximum variance and minimum correlation. The succeeding principal components are sorted in order of decreasing variance, hence first several PCs contain the most of pattern discriminative information.

PCA does not examine the affiliation of data points to specified pattern classes; therefore, as a supervised learning method, needs to be aided with a feature subset selection algorithm [5]. The goal of feature selection is to find optimal feature subset providing the best separation and aggregation of pattern classes by means of feature subset quality criterion: the objective function [5].

In order to recognize a pattern in a test sample, the pattern recognition system first needs to be trained with

a set of known, labelled samples called the training dataset. Most commonly, for classifier validation purposes, a training and test datasets are obtained by slicing the measurement series into parts or by applying a cross-validation mechanism [1]. In the following research the k -nearest neighbours (kNN) algorithm [1, 6] is used as a classifier. The kNN method assigns the test sample to a specific pattern class basing on shortest Euclidean distance to labelled training dataset points in feature space. It is so-called “lazy learning algorithm”, as the training dataset is still required at the time of test sample classification.

The pattern recognition system described in this article was initially developed for the array of chemoresistive gas sensors, presented in [6]. This paper concerns the research on adaptation and development of the pattern recognition system for analysis of data acquired from the optical sensor matrix, aiming at estimation of the temporary blood volume in the POLVAD prosthesis prototype [7].

2. Data source and preprocessing

The datasets for numerical analysis were measurements of light reflected from the membrane in the blood chamber in all possible pair configurations of 12 light emitters and 32 light detectors. The details on experimental setup and data acquisition process are described in [7].

In a measuring series the blood chamber was alternately filled up and drained in 0–80 ml volume range. Samples were acquired in 1 ml steps (in case of static tests configuration) and in constant time intervals (semi-dynamic tests configuration). Every single measurement resulted in 384 dimensional feature vector (12 emitters \times 32 detectors). A total number of samples acquired during static tests was over 8000. Similar samples count was obtained during semi-dynamic tests.

Preliminary research involved analysis of 9 patterns (for multiples of 10 ml including 0) to verify capabili-

*corresponding author; e-mail:

Przemyslaw.Marczynski@gmail.com

ties of the dimensionality reduction mechanism. As the results were promising, a decision was made to define patterns for every multiple of 5 ml in 0–80 ml blood chamber volume range. That resulted in 17 volume patterns and hence was targeting maximal absolute deviation of ± 2 ml.

Data preprocessing included automated selection of samples for volumes that reflect defined patterns. In case of 17 patterns, the total training dataset counted around 2000 samples.

3. Dimensionality reduction

PCA method applied to the data array of all 384 features aggregates the measured samples in clusters in 3D reduced feature space. However it is not aware of defined pattern classes and as a result PCA creates two separate clusters from samples of the same volume but measured during filling up and draining the blood chamber. As Fig. 1 shows, in many cases these clusters are closer to points of another pattern class than to each other. A feature selection mechanism with an objective function solves the problem (see Fig. 2).

The pattern recognition system utilizes the plus- L minus- R selection (LRS) algorithm with an objective function as quality criterion (a filter approach) [5, 6]. The algorithm subsequently adds L and removes R features from active feature set, allowing further elimination of redundant and weak features which lower the feature subset rating. The L and R values were determined at level of $L = 20$, $R = 10$ as optimal for both selection quality and computation time.

The objective function is calculated in first 3 principal components' space. After applying PCA to every examined feature subset data, the whitening transformation [8] is performed (scaling to unit variance) in order to keep objective function value comparable for different feature subsets.

The objective function originally implemented minimization of the sum of pattern classes variances (1). For some measuring series problems emerged concerning unexpected values in several data array columns. Some detectors generated constant response and at the very beginning of feature selection, the algorithm failed at first three spoiled features indicating zero variance, hence the false best quality was being picked. A more robust objective function has been developed to solve the problem. The maximization criterion of cluster centres' variance, along with current criterion has been implemented. The new objective function also has been verified and further modified during legacy test on the dataset from the array of gas sensors [6]. Moreover, the final version of the objective function presented on (2) chooses higher quality feature subsets, leading to better classification results

$$\text{prev. obj. func.} = \sum_{k=1}^K \text{var}(X_k), \tag{1}$$

$$\text{new obj. func.} = \left(\sum_{k=1}^K \text{var}(X_k) \right)^2 - \text{var}(Y). \tag{2}$$

The optimization problem is to find minimum of the objective function, where K — number of pattern classes, k — pattern class index, X_k — coordinates of pattern class points, Y — coordinates of pattern clusters' centres.

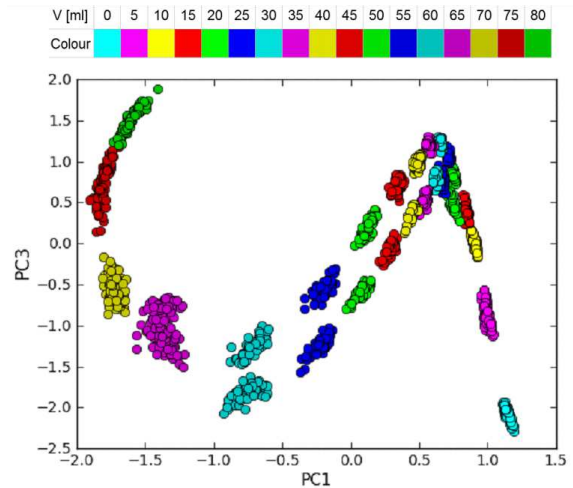


Fig. 1. All 384 features on PCA input. Separate clusters were created for the same volumes at filling up and draining the chamber.

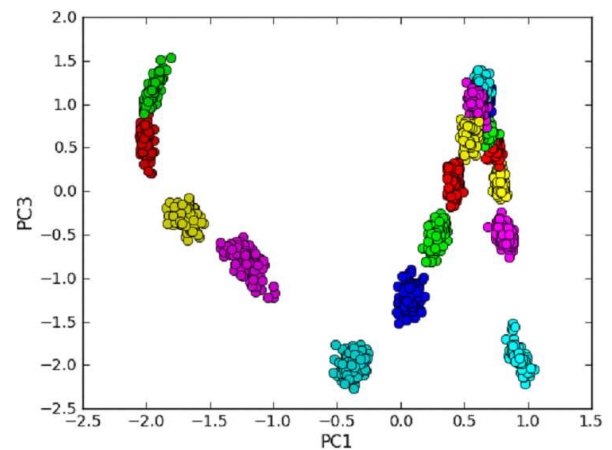


Fig. 2. Optimal feature set (100 features) discovered by feature selection algorithm. Pattern clusters are consistent for each volume.

For all above scatter plots (Fig. 3) feature space was reduced to 3D by means of PCA transformation. 2D figures show relations between two principal components: $PC3 = f(PC1)$.

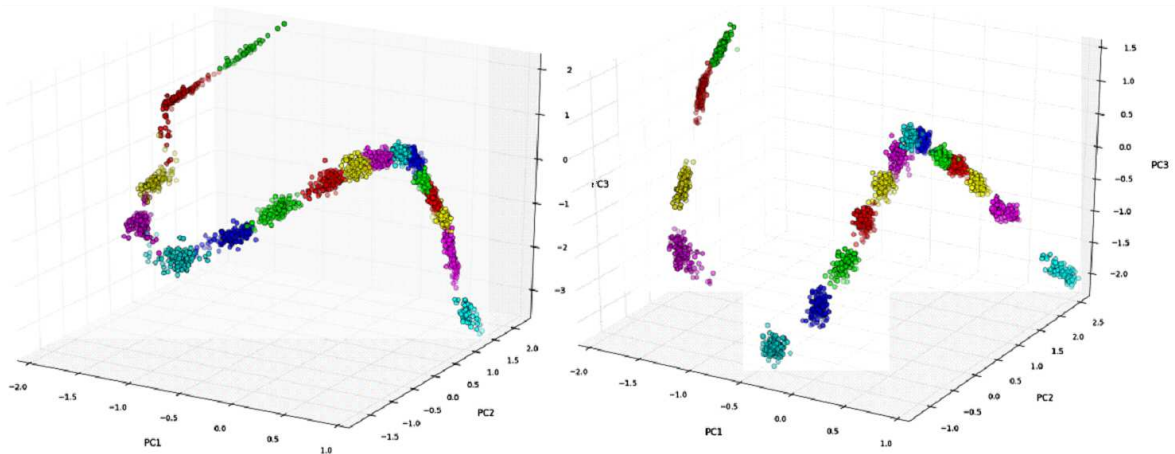


Fig. 3. Static (left) and semi-dynamic (right) method of measurement. 3D scatter plot for optimal feature set. (101 and 100 selected features out of 384, respectively).

4. Classification

For classification of the test samples to particular pattern class the kNN method is used [1, 6]. The classification tests were performed for $k = 5$ KNN in reduced, 3D feature space (the first 3 principal components).

The measuring series were divided equally into training and test datasets. Moreover, the test dataset included samples for all measured volumes, hence the classifier is expected to identify intermediate samples (in the ± 2 ml margin) as nearest known volume pattern (being a multiple of 5 ml). The classification results are presented in Fig. 4 and Fig. 5. The results show that the blood chamber volume is estimated with uncertainty of less than 5% of the measurement range.

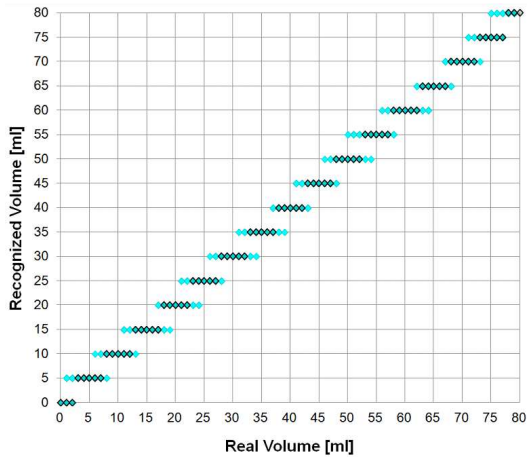


Fig. 4. Volume recognition results by means of kNN algorithm for semi-dynamic measurement method. All emitters included in initial feature set (selection resulted in 110 features out of 384).

The measuring series were divided in half into training and test datasets. Tested dataset counts over 3800

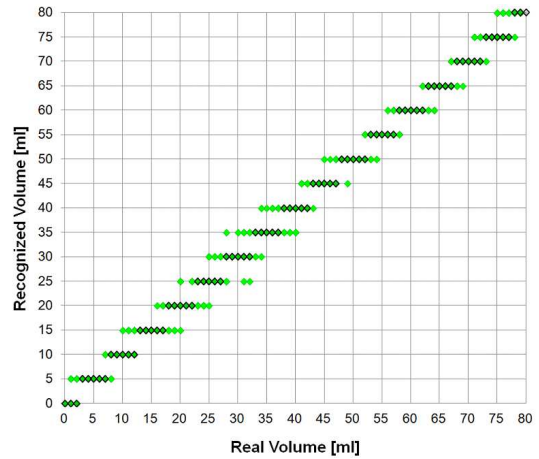


Fig. 5. Volume recognition results by means of kNN algorithm for semi-dynamic measurement method. Initial feature set reduced to 3 best emitters (selection resulted in 80 features out of 96).

samples. Each mark in the figures may represent more than one test sample. Darkened points on above plots denote ideal case of recognition (maximal deviation of ± 2 ml). The results represent $k = 5$ nearest neighbours classification.

5. Research on reducing the number of matrix optical elements

The other goal of feature selection in the research on blood chamber volume estimation is to find a possibly least numerous feature sets that preserve decent quality, so that the number of emitters and detectors could be reduced and data acquisition circuit [7] simplified. Especially reducing the number of emitters is important, as it results in shorter measurement time per sample.

An attempt was taken to reduce the number of emitters to 3 out of 12. All 220 combinations of 3 emitters were automatically examined by finding the best feature set for every combination and logging the objective function values. After finding the best three emitters, classification tests were performed. Figure 5 presents results of classification of test samples for training dataset limited to the best three emitters configuration.

For reduction of detectors count, a penalty factor for the objective function (3) has been developed. It modifies the current quality by adding a component proportional to the number of active features and to the current value of the objective function

$$Q_P = Q + |QFP|, \quad (3)$$

where F — number of features in active feature set, P — penalty factor (a constant within range of 10^{-4} – 10^{-1}), Q — feature set quality (objective function value), Q_P — penalized quality. Depending on the value of penalty factor (a constant), respectively less numerous feature sets are favoured.

However, less numerous feature sets (< 20 features) obtained with an assist of penalty factor lead to significantly lower pattern classes aggregation and unacceptable classification results. Further research in this area is planned.

6. Conclusions

The conducted researches show that the pattern recognition system has been successfully applied for estimation of the temporary blood volume in a pneumatic chamber of the POLVAD prosthesis.

The results of classification indicate that the blood chamber volume can be estimated with uncertainty of less than 5% by means of optical measurement method aided with pattern recognition.

The biggest difficulty at the classification stage is dealing with pattern classes, which are overlapping on scatter plots. The problem grows with the additional reduction of feature set count by means of penalty factor. Further

research will focus on improving the classification accuracy for very small feature sets (less than 20 features with only 2 or 3 emitters involved). As a preliminary concept, it is considered to implement a cascade classifier, that might identify the test sample in two steps: first analyzing all 17 patterns, then verifying the recognition result (e.g. 30 ml) with second classifier trained with relevant three “neighbouring” patterns only (e.g. 25, 30, 35 ml).

The other area of planned research is to explore different feature projection algorithms. The PCA method was used after assumption of Gaussian distribution of the datasets. Future research will examine applications of non-linear PCA (NLPCA) transformation [8] expecting to achieve better pattern classes separation and aggregation.

Acknowledgments

The work was partially financed by the Polish National Centre of Science NCN (Cracow, Poland), within the grant No. 4896/B/TO2/2011/40.

References

- [1] R. Gutierrez-Osuna, T. Nagle, *IEEE Trans. Syst., Man Cybernet. Part B, Cybernet.* **29**, 626 (1999).
- [2] A.H. Gómez, J. Wang, G. Hu, A.G. Pereira, *Sensors Actuat. B* **113**, 347 (2006).
- [3] Y. Yin, X. Tian, *Sensors Actuat. B* **124**, 393 (2007).
- [4] K. Gut, Bull. *Pol. Acad. Sci., Techn. Sci.* **59**, 395 (2011).
- [5] K.Z. Mao, *IEEE Trans. Syst., Man Cybernet. Part B, Cybernet.* **34**, 629 (2004).
- [6] P. Marczyński, A. Szpakowski, C. Tyszkiewicz, T. Pustelny, *Acta Phys. Pol. A* **122**, 847 (2012).
- [7] G. Konieczny, T. Pustelny, *Acta Phys. Pol. A* **122**, 962 (2012).
- [8] M. Scholz, R. Vigário, *Artificial Neural Networks*, Springer, Bruges 2002.