Unsupervised Fusion of Scattered Data Collected by a Multi-Sensor Robot on Concrete

Patricia COTIČ *, Ernst NIEDERLEITHINGER **, Markus STOPPEL **
* Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia
** BAM Bundesanstalt für Materialforschung und -prüfung, Berlin

Abstract. At BAM a multi-sensor robot system BetoScan is used for the investigation of reinforced concrete floors affected by corrosion in parking garages. Potential maps, as well as the distribution of concrete cover and moisture can be assessed simultaneously and data can be collected contactlessly. In order to evaluate the extent of degradation adequately and to divide the investigated structure into zones with defined damage classes, large data sets have to be collected and interpreted manually. Thus, to promote an efficient data evaluation framework, which could speed up and simplify the evaluation of large data sets, an unsupervised data fusion is of major interest. However, taking into account that collected data do not certainly coincide in space, a scattered data interpolation method should be applied prior data fusion.

In the paper, a case study involving a BetoScan data set acquired from a reinforced concrete floor of a parking garage in Germany is presented. The data set includes potential mapping, covermeter based on eddy current, as well as microwave moisture measurements. Among the examined methods for interpolation of scattered data, kriging shows to yield smooth interpolated data plots even in the case of very sparse data. In the post-processing step, the investigated structure is efficiently segmented into zones using clustering based data fusion methods, which prove to be robust enough also for handling noisy data. Based on the minimization of the XB validity index, an unsupervised selection of optimal segmentation into damage classes is derived.

1. Introduction

A large part of parking garages and bridge decks suffer from severe corrosion of the reinforcement, which is reflected in cracking, spalling and losses of the concrete cross section. In order to evaluate the extent of degradation adequately and to divide the investigated structure into zones with defined damage classes, investigation of surfaces of some thousand square meters with a dense grid is necessary. To solve this problem, a self-navigating mobile robot system has been developed within the BetoScan project (www.betoscan.de). With the BetoScan system, potential maps, as well as the distribution of concrete cover and moisture can be assessed simultaneously by an automated multi-sensor system, which is in detail described in [1].

On the contrary to the automated data acquisition system of BetoScan, data analysis is currently performed manually by direct comparison of the results from different NDT sensors. Such evaluation of the investigated structure into damage zones with defined probability of corrosion lacks both of speed and relevance. To tackle this issue, we propose using an unsupervised data fusion model based on clustering techniques that we have very recently derived for fusing images in NDT-CE [2]. Apart from NDT, clustering methods have already been applied in medical imaging for pattern recognition, e.g. for the detection of cancerous cells from magnetic resonance images [3] and mammograms [4].

Taking into account that BetoScan collects scattered data that do not coincide in space (due to practical limitations such as rough surface, the need to avoid obstacles and the sensor arrangement on the robot), an efficient scattered data interpolation method should be used in the pre-processing step. In the article, the results of several interpolation methods such as the nearest-neighbor and spline interpolation, inverse distance weighted method and kriging are presented and discussed.

2. Methods

2.1 The BetoScan measurement system

The data presented in this study were acquired with the BetoScan robot system using potential mapping, covermeter based on eddy current, as well as microwave moisture sensors (Fig. 1). For potential mapping, which measures the probability of active corrosion, a Canin+ instrument from Proceq was used in combination with a copper/copper sulphate reference wheel electrode. For eddy current, a Profometer 5+ from Proceq was used, which combines rebar detection and measurement of concrete cover. To measure the relative moisture distribution of the concrete, the microwave-based devices Moist RP (4 cm penetration) and Moist PP (20 cm penetration) from HF-Sensor were used.

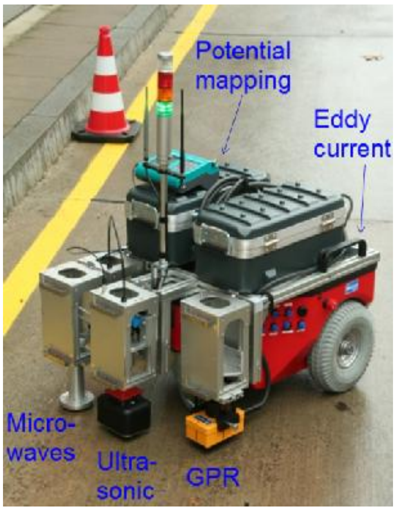

Figure 1. The BetoScan robot system with NDT sensors

2.2 Data evaluation methods

2.2.1 Scattered data interpolation methods

In the pre-processing step we took into account the following scattered data interpolation methods: Shepard's inverse distance weighted method, Wiener interpolation or kriging, the nearest-neighbor interpolation and multilevel B-splines. The aim of all interpolation methods is that an interpolation function F is obtained on the basis of measured data f_{ij} according to [5]

$$F(x,y) = \sum_{i,j} W_{ij}(x,y) \cdot f_{ij} \ , \tag{1}$$

where W_{ij} refers to a weight function. In the case of the inverse distance weighted method, the interpolation function at a particular position (x,y) is calculated as a weighted average of all measured data with respect to their inverse distances to (x,y) [5]. On the other hand, kriging is a statistically based method and tends to find an estimate, i.e. the most likely interpolation function, where the correlation function is used as the weight function [5]. In the case of the multilevel B-splines, the interpolation function at a particular position (x,y) is calculated from a linear combination of 16 most neighboring points Φ_{ij} in the control lattice [6],

$$F(x,y) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(s) B_l(t) \cdot \Phi_{(i+k)(j+l)} \ , \tag{2}$$

where $i = [x] - 1$, $j = [y] - 1$, $s = x - [x]$, $t = y - [y]$, whereas B_k and B_l are cubic B-spline functions [6]. The aim is to find a control lattice that minimizes the difference between interpolated $F(x_i,y_j)$ and measured values f_{ij}. On the contrary to the inverse distance weighted method and kriging, the multilevel B-splines is a very local method since it considers the contribution of only the most neighboring data points.

2.2.2 Clustering techniques

Clustering techniques tend to group a set of data into subsets (clusters or classes) according to their similarity to each other and do not require the knowledge about the statistical distribution of data. To partition an unlabeled data set $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}\}$ into c clusters, the target is to minimize an objective function J, which is the sum of distances of N data points to their respective cluster center $\mathbf{v}$ [7],

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^{m} \left\| \mathbf{x_j} - \mathbf{v_i} \right\|_{\mathbf{A}}^{2} \ , \tag{3}$$

where u_{ij} refers to the membership degree of datum $\mathbf{x_j}$ to cluster center $\mathbf{v_i}$ and m determines the fuzziness of the resulting partition. In our case we took into account fuzzy probabilistic clustering, in particular the Gustafson-Kessel (GK) clustering algorithm [8], where u_{ij} 0 [0,1] (data can belong to more than one cluster with a calculated probability) and the distance function is adopted with A for each cluster to account for their different shape and size.

To access the goodness of a clustering partition quantitatively, several validity indices have been proposed that tend to minimize the ratio between the compactness (the measure of data scattering within a cluster) and separation between clusters. In this study we refer to the Xie and Beni's (XB) index, given by [9]

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^{m} \left\| \mathbf{x_j} - \mathbf{v_i} \right\|_{\mathbf{A}}^{2}}{N \cdot \min_{i,j} \left\| \mathbf{x_j} - \mathbf{v_i} \right\|_{\mathbf{A}}^{2}} . \tag{4}$$

The GK algorithm and the XB validity index were chosen based on the results from our previous studies [2].

3. Results and discussion

For this study, a BetoScan data set that was acquired from a reinforced concrete floor of a parking garage in Germany was taken into account. The area inspected measured approximately 5×7 m^2. The spacing between tracks was around 20 cm, with about 1 cm between data points of each track.

The results of the inspected concrete floor with all used NDT methods are presented in Fig. 2. In the potential map (Fig. 2a), areas having electrochemical potentials lower than approximately -300 mV can be interpreted as critical areas with a high probability of corrosion. It can be seen that part of these areas coincide with the area of detected small concrete cover from eddy current (Fig. 2b), as well as with the area of high moisture content (more than 3 vol. %) from the sensor with deeper penetration capability – Moist PP (Fig. 2c). However, it can also be seen that at some areas with either low concrete cover or high humidity, active corrosion was not identified from the potential map. As both low concrete cover and high humidity may induce corrosion, it is evident that a reliable segmentation of the floor into damage classes could only be obtained by fusing the results.

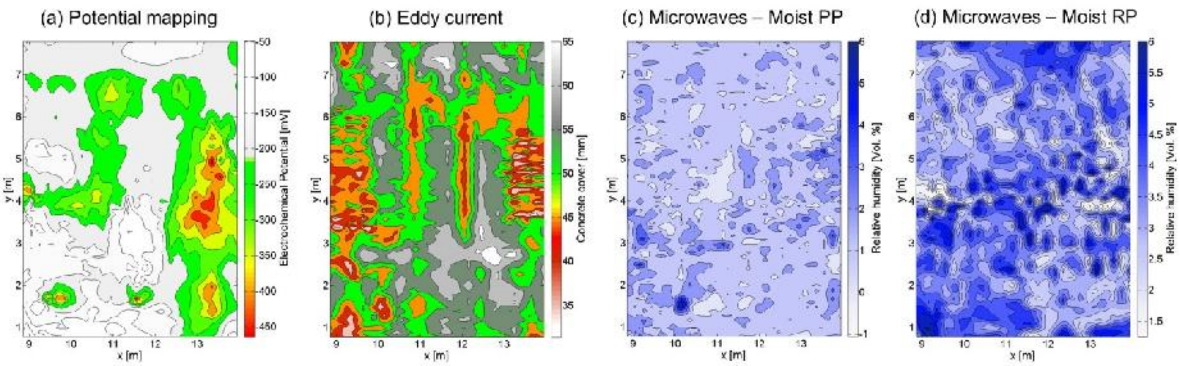

Figure 2. Results from the BetoScan measurement of a concrete parking deck using potential mapping (a), eddy current (b), microwaves with a Moist PP sensor (c) and microwaves with a Moist RP sensor (d)

3.1 Scattered data interpolation results

Before data fusion, a study of the optimal scattered data interpolation method was performed. Fig. 3 presents the results from different interpolation methods applied on the potential mapping data. The plots in the upper row refer to the original dense data, whereas the plots in the bottom row were obtained from data that were downscaled by 3 and 5 times in x and y direction, respectively, in order to study the effect of more sparse data on the interpolation result. The results show that the inverse distance weighted method and kriging yield very similar results in the case of dense and sparse data. However, part of the area with very high local changes in the electrochemical potentials (blue colored area) cannot be reconstructed properly in the case of sparse data. On the other hand, those areas could be identified when using B-splines since the method is very much affected by local changes in the data. Nevertheless, at areas where the spacing between data changes, the method's performance decreases largely as much higher electrochemical potentials are obtained than expected (top and bottom of the plot from sparse data). The results in Fig. 3 also show that the data cannot be reconstructed accurately when using the nearest-neighbor interpolation, which is especially pronounced in the case of the sparse data set.

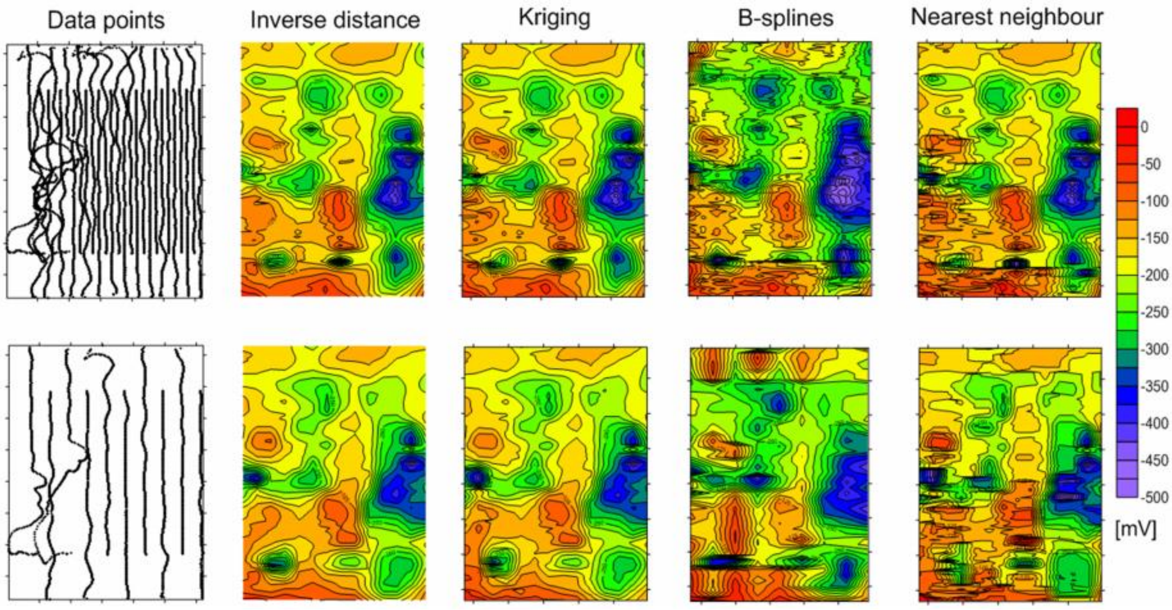

Figure 3. Results from different scattered data interpolation methods applied on the potential mapping data. The plots in the upper row refer to the original dense data, whereas the plots in the bottom row refer to the data that were downscaled by 3 and 5 times in x and y direction, respectively

The results from different interpolation methods applied on the eddy current data are presented in Fig. 4. The results obtained when using the nearest neighbor-interpolation again show that the method cannot be regarded as accurate enough for the interpolation of the BetoScan scattered data due to large spacing between tracks. However, contrary to the quite successful application of B-splines on the potential mapping data, the method reconstructs very high concrete cover results that are not in accordance with the results obtained from the inverse distance weighted method and kriging. An explanation for this could be that the eddy current data exhibits very high local changes to which B-splines are affected. To change the method's behavior to more global, the interpolation increment was increased slightly. This on the other hand produced an inaccurate distribution of the measured concrete cover.

The results in Fig. 3 and 4 show that the inverse distance weighted method and kriging are robust enough to handle both dense and sparse data. However, by evaluating more data sets it was observed that the inverse distance weighted method tends to produce flat regions around the measurement points as can also be seen from the plot in Fig. 4 obtained on sparse data. This favors kriging to be the most recommended interpolation method for the BetoScan data.

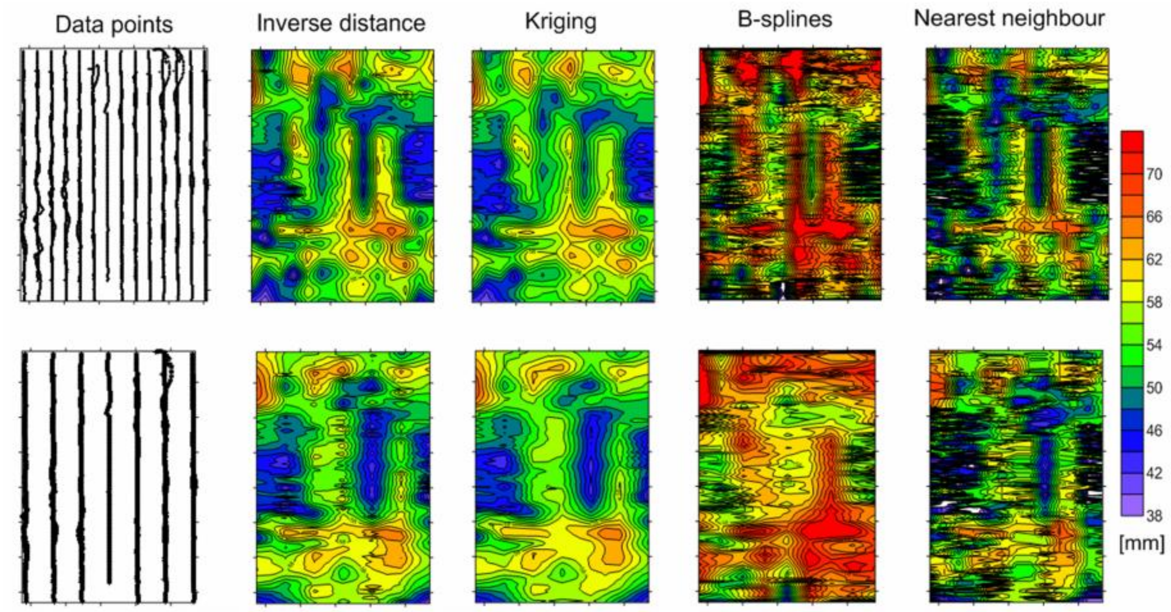

Figure 4. Results from different scattered data interpolation methods applied on the eddy current data. The plots in the upper row refer to the original dense data, whereas the plots in the bottom row refer to the data that were downscaled by 3 and 5 times in x and y direction, respectively

3.2 Clustering results

The efficiency of using clustering methods to estimate the damage zones with a different probability for corrosion is demonstrated in Fig. 5. Regarding the correlation between results in Fig. 2, the optimal clustering partition was expected for the combination of data from potential mapping, eddy current and the Moist PP sensor, which is shown in Fig. 5a. Here, the red cluster could be identified with the zone of active corrosion as the cluster coincides with the area having the lowest electrochemical potentials. The black colored cluster could be identified with the area of both higher electrochemical potentials and very low concrete cover. Thus it could be related to the area with highly probably corrosion. The white colored cluster could be characterized as undamaged (or minimally damaged) concrete. For the identification of the damage zone with the cluster color, see Fig. 5b, where the results of data correlation between electrochemical potentials and concrete covers are presented.

To evaluate the performance of clustering techniques in the case data are fused that are less correlated to corrosion; we clustered the whole BetoScan data set as given in Fig. 2. It can be seen that the inspected floor can be still well segmented and in particular, the most critical red colored cluster can be accurately detected. However, as data are less correlated to the detected corrosion, the membership degree to clusters decreases, which is in accordance with the transparency of the cluster color.

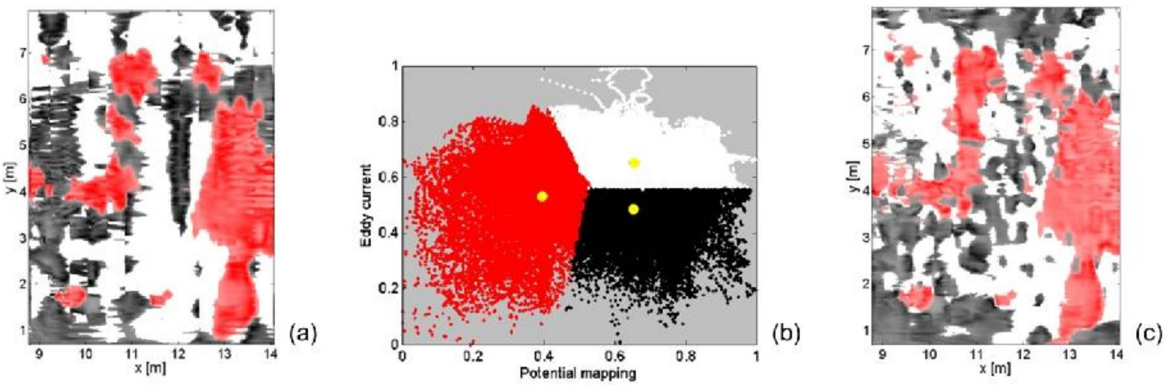

Figure 5. Resulting clustering partition when fusing data from potential mapping, eddy current and microwaves with a Moist PP sensor (a) and the corresponding correlation plot between potential mapping and eddy current normalized data (b); resulting clustering partition when fusing the whole BetoScan data set (c)

To provide an automatic selection of data that is best related to the identified corrosion, the XB validity index (Eq. 4) was evaluated for different combinations of used NDT sensors, aiming in properly evaluating the goodness of the clustering partition. According to Fig. 6, a relatively high index value is obtained when fusing only Fig. 2a and b (the results from potential mapping and eddy current). However, when adding Fig. 2c, the value drops remarkably, which is in accordance that a better partition derives a lower XB value. When changing Fig. 2c by Fig. 2d, the value of XB index increases as expected since less data prove a correlation with the identified corrosion. By means of selecting the combination of data yielding the minimal XB index, we could automatically obtain the optimal partition.

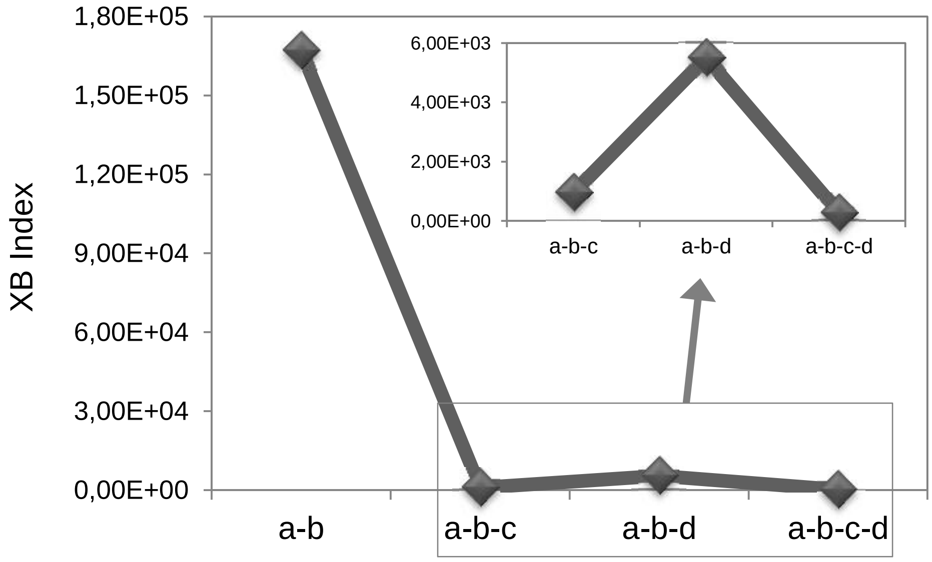

Figure 6. Results of the XB index value for different combinations of used NDT sensors. Here, a, b, c and d refer to the corresponding plots in Fig. 2

4. Conclusion

The use of clustering techniques for image segmentation of fused NDT data from a multi-sensor robot BetoScan has been tackled in combination with different studied interpolation methods for scattered data. The results of a case study show that kriging is the most recommended interpolation method for the BetoScan data and that an accurate zonation of corroded parking floors into damage classes is possible by using clustering. Moreover, the techniques prove to be robust enough to handle also noisy data that are poorly related to the identified corrosion. The presented results show that an automatic selection of data that produce the optimal segmentation of the inspected structure can be obtained using the XB validity index. This proves that clustering techniques could be incorporated into the BetoScan data evaluation framework in a fully unsupervised way. Hence, data could be easily and reliably evaluated also by untrained engineers.

Acknowledgement

The first author acknowledges the financial support of the Slovenian Research Agency through grant 1000-10-310156 and the Slovene Human Resources and Scholarship Fund through grant 11012-13/2012.

References

[1] Stoppel M, Taffe A, Wiggenhauser H, Kurz JH, Boller C (2011), Automated multi-sensor systems in civil engineering for condition assessment of concrete structures. In: Grantham M (ed.) Proceedings of the 4th International Conference on Concrete Repair. CRC Press, Boca Raton, pp 397–403.
[2] Cotič P, Jagličić Z, Niederleithinger E, Stoppel M, Bosiljkov V (2014), Image fusion for improved detection of near-surface defects in NDT-CE using unsupervised clustering methods. J Nondestruct Eval. Doi: 10.1007/s10921-014-0232-1.
[3] Bensaid AM, Hall LO, Bezdek JC, Clarke LP, Silbiger ML, Arrington JA et al (1996), Validity-guided (re)clustering with applications to image segmentation. IEEE T Fuzzy Syst 4(2): 112–123.
[4] Santoro M, Prevete R, Cavallo L, Catanzariti E (2006), Mass detection in mammograms using Gabor filters and fuzzy clustering. In: Bloch I, Petrosino A, Tettamanzi AGB (eds.) Fuzzy logic and applications. WILF 2005: Proceedings of the 6th international workshop. Springer, Berlin, pp 334–343.
[5] Lewis JP, Pighin F, Anjyo K (2010), Scattered data interpolation and approximation for computer graphics. In: Proceedings of the SA '10 ACM SIGGRAPH ASIA 2010 Courses. ACM, New York, pp 1–73.
[6] Lee S, Wolberg G, Shin SY (1997), Scattered data interpolation with multilevel B-splines. IEEE T Vis Comput Gr 3(3): 228–244.
[7] Kruse R, Döring C, Lesot MJ (2007), Fundamentals of fuzzy clustering. In: Oliveira JV, Pedrycz W (eds.) Advances in fuzzy clustering and its applications. John Wiley & Sons Ltd., Chichester, pp 3–30.
[8] Gustafson D, Kessel W (1979), Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of the 1978 IEEE conference on decision and control including the 17th Symposium on adaptive processes. IEEE Control Systems Society, San Diego, pp 761–766.
[9] Xie XL, Beni G (1991), A validity measure for fuzzy clustering. IEEE T. Pattern Anal. 13(8): 841–847.