

# **Hierarchical clustering of RGB surface water images based on MIA-LSI approach**

**Petr Praus<sup>1\*</sup> and Pavel Praks<sup>2</sup>**

<sup>1</sup>*Department of Analytical Chemistry and Material Testing, Faculty of Metallurgy and Material Engineering, VSB-Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic*

<sup>2</sup>*Department of Mathematics and Descriptive Geometry, Department of Applied Mathematics, Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava, Czech Republic*

## **Abstract**

Multivariate image analysis (MIA) combined with the latent semantic indexing (LSI) method was used for the retrieval of similar water-related images within a testing database of 126 RGB images. This database, compiled from digital photographs of the various water levels and similar images of surface areas and vegetation, was transferred into an image matrix, and reorganised by means of principal component analysis (PCA) based on singular value decomposition (SVD). The high dimensionality of original images given by their pixel numbers was reduced to 6 principal components.

Thus characterised images were partitioned into clusters of similar images using hierarchical clustering. The best defined clusters were obtained when the Ward's method was applied. Images were partitioned into the 2 main clusters in terms of similar colours of displayed objects. Each main cluster was further partitioned into sub-clusters according to similar shapes and sizes of the objects. The clustering results were verified by the visual comparison of selected images. It was found that the MIA-LSI approach complemented with a suitable clustering method is able to recognise the similar images of surface water according to the colour and shape of floating subjects. This finding can be utilised for the automatic computer-aided visual monitoring of surface water quality by means of digital images.

**Keywords:** multivariate image analysis (MIA), latent semantic indexing (LSI), RGB image, Ward's clustering, water quality