

Surface Classification Above Gas Pipeline Facilities

Dalibor Bartoněk, Irena Opatřilová and Jiří Bureš

Abstract The article discusses a surface classification above the stored gas facilities in the territory under the administration of the RWE Company in the Czech Republic (CR). The aim of the project was to determine reproductive values of these facilities and a reasonable estimate of the costs that would be spent on a construction of new pipelines. The classification uses a raster dataset of orthophoto with the resolution of 25 cm/1 pixel, a vector dataset of the Fundamental Base of Geographic Data (ZABAGED) and vector datasets of the route of underground utilities. As part of the assignment, it was necessary to classify a surface from two perspectives: from the object standpoint (the surface above gas facilities) and the spatial standpoint (administrative division of the territory of the CR). The authors have succeeded in creating a technological line (workflow) with a high degree of automation. The core of this workflow is a geographic information system (GIS). The technology has been implemented within a special application developed in Python language with a support of the ESRI libraries and ArcGIS 10.0. It is massive task given the territorial coverage of almost the whole of the CR with the area of 64,350 km² with a data volume of the order of 500 GB. The results demonstrated the effectiveness of proposed procedure to achieve low error rate in the range of 2–3 % across the modeled area.

Keywords Classification · Gas pipeline · GIS

D. Bartoněk (✉) · I. Opatřilová · J. Bureš
Institute of Geodesy Faculty of Civil Engineering, Brno University of Technology,
Veveří 331/95, 602 00 Brno, Czech Republic
e-mail: bartonek.d@fce.vutbr.cz

I. Opatřilová
e-mail: opatrilova.i@fce.vutbr.cz

J. Bureš
e-mail: bures.j@fce.vutbr.cz

D. Bartoněk
European Polytechnic Institute, Osvobození 899, 686 04 Kunovice, Czech Republic

1 Introduction

The aim of the project was:

- (a) The surface analysis of the RWE gas line in the Czech Republic (CR), i.e. processing of classification data of storage of gas facilities below certain types of terrain surfaces. This analysis was done in order to determine reproductive values of gas facilities (pipelines) and a valuation of costs which would be necessary to spend for building new networks.
- (b) The surface analysis in three variations for the route of gas line: a high-pressure gas line, a main line and service pipes. The requirement was in this subject-division to classify a surface in three separate groups:
 - 2 groups of a surface type: paved and unpaved
 - 5 groups of a surface type: asphalt, main road, local road, unbound, unknown
 - 10 groups of a surface type: forest, grassland, bare soil, asphalt, roof-tile, roof-straight, shadow, main road, local road, path.
- (c) The output was realized in a graphic form (*.shp format) and in a text form as a structured table in XLSx format with classified data of surface types above the route of gas line.
- (d) The extent of the analysis contained 12 regions from standpoint of the administration division in the CR. There are these regions: Karlovarský, Plzeňský, Středočeský, Ústecký, Liberecký, Královéhradecký, Pardubický, Vysočina, Moravskoslezský, Olomoucký, Zlínský and Jihomoravský. So it was processed total 188 municipalities with extended state administration (ORP). The outputs of classification were required in a graphic and in a text form to a territorial detail of the part of municipality within a territorial identification register of the CR.

The solution was based on a pilot project and a developed technology for this purpose called “Classification of data about storage of utilities and facilities below certain types of terrain surfaces” (Bureš et al. 2013).

2 Related Works

Works that deal with the results of image classification can be divided according to the application domain and according to the used methods. The field of remote sensing is represented most frequently in terms of application (Pupin et al. 2014; da Silva et al. 2014; Debes et al. 2014; Iames et al. 2013; Puertas et al. 2013; Mouelhi et al. 2013; Landgrebe 2003; Liang 2004). Another fields of image classification is used: medicine and biology (Gray and Song 2013), food industry (food quality testing) (Johnson 2013), agriculture (forecast of crop in the given field), ecology

(Makarau et al. 2013) and geodesy (Zhong et al. 2012) (classification of orthophoto).

There are a number of methods for improvement of the results of image classification and these methods are the most common:

- works dealing with the basic usage of image classification (Landgrebe 2003; Liang 2004),
- classical methods of image enhancement (filtering, transformation of brightness, discrete convolution) (Pupin et al. 2014),
- the use method of Support vector machine (SVM) (Iiames et al. 2013),
- a combination of different methods of image capturing of the Earth's surface (da Silva et al. 2014; Makarau et al. 2013),
- processing of time sequence of images (Debes et al. 2014),
- the use of graph theory (factors of graph) (Mouelhi et al. 2013),
- application of correlation analysis with a matrix of errors (Puertas et al. 2013),
- the use of metric spaces (Searcoid 2006) for evaluation of the results of image classification,
- the use of image classification in geodesy and surveying (Černota et al. 2011; Dandoš et al. 2013).

3 Theoretical Foundations

Let a non-empty set of $U \neq \emptyset$ is the universe of discourse and the X is a subset ($X \subseteq U$). The U set represents in our case the entire solved area and the X subset is part of the solved area. The elements of the X set are geographic objects of given locality. Then, let a system of relation of equivalence exists $\{R_i\}$ ($i = 1, 2, \dots, m$), where each R_i relation divides the X set into n subsets $X/R_i = \{X_{i1}, X_{i2}, \dots, X_{in}\}$ such that for every i, j holds:

$$\begin{aligned} X_{ij} \subseteq U, X_{ij} \neq \emptyset, \text{ for } (i = 1, 2, \dots, m), \\ (j = 1, 2, \dots, n) \text{ --- (all subsets are non-empty)} \end{aligned} \quad (1)$$

$$\begin{aligned} X_{ij1} \cap X_{ij2} \neq \emptyset, \text{ for } (i = 1, 2, \dots, m), (j_1, j_2 = 1, 2, \dots, n), \\ j_1 \neq j_2 \text{ --- (intersection of all subsets for each decomposition,} \\ \text{defined by the } R_i \text{ displays, is empty)} \end{aligned} \quad (2)$$

$$\begin{aligned} \bigcup_{i=1, j=1}^{i=m, j=n} X_{ij} = X \text{ (union of all subsets within each decomposition,} \\ \text{which is defined by } R_i \text{ displays, is just the whole } U \text{ set)} \end{aligned} \quad (3)$$

The R_i relation of equivalence can be considered as territorial division of the CR according to an administrative arrangement that divides this area into lower

administrative units (regions, districts, etc.). If the conditions in (1)–(3) are complied, each sub-area in terms of set theory has a character of the X_{ij} class.

Suppose there is another relation R_2 such that defines decomposition according to the subject (thematic) principle, i.e. according to geographical objects, which are elements of the X set. In our case, geographical objects are gas facilities stored in the ground.

Then the intersection of relations $S = R_1 \cap R_2$, $S \subseteq X$ defines the modeled area in which we will conduct spatial analysis. Because this territory is defaultly smaller than the original solved territory, this combination of decomposition leads to optimize the entire process of analysis and time saving of solution.

4 Method of Solution

The solution was carried out in the following stages:

1. Data preparation—generate data sets for each ORP for the entire territory of the CR (creation of binding database tables and dials, data sorting, repairing of topological errors in input data, building data sets, debugging of generating scripts)
2. Determination of parameters—measuring of the widths of roads on orthophoto for setting buffers of roads, creation of classification keys, data reclassification according to the classification keys
3. Data classification—classification of pipeline sections above various surface types based on image of orthophoto, classification keys and parameterized communication
4. Editing of errors—checking of completeness of image classification, semi-automated editing of errors of image classification, batch editing of changes to database tables
5. Final inspection—batch sorting outputs and files, checking completeness of the outputs and checksums
6. Completion—completion of data sets for a transfer, creation of XLSs file, creation SHP files for regions, a checking the completeness.

The scheme of technology of data analysis is illustrated in Fig. 1.

The data analysis was solved in the ESRI ArcGIS 10 environment. Total 3 powerful computers of the PC type and 5 notebooks were used for processing, which were connected in computer network (Fig. 2). The accessibility to the data was provided only for researchers so copyrights of data were protected. The processing of automated part was carry on computer machines of the PC type with parameters: CPU Intel Core i5 3.8 GHz, 16 GB RAM, NVIDIA GTX650 2 GB, HD of SDD type and VelociRaptor, OS of Microsoft Windows 7 (64 bit). Processing software was ArcGIS 10.0, Python 2.6 (both 32-bit).

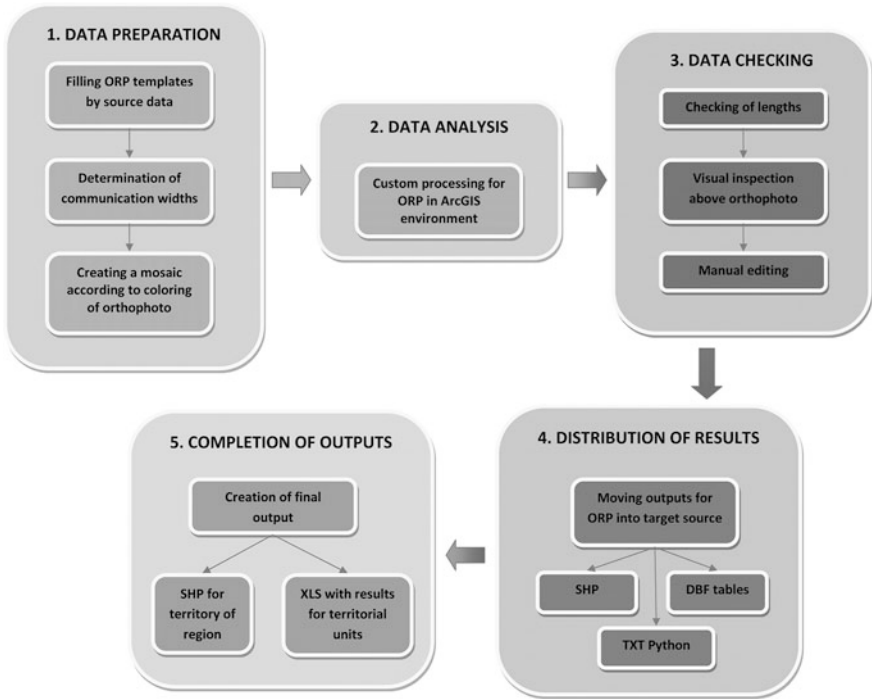


Fig. 1 The scheme of technology of data analysis

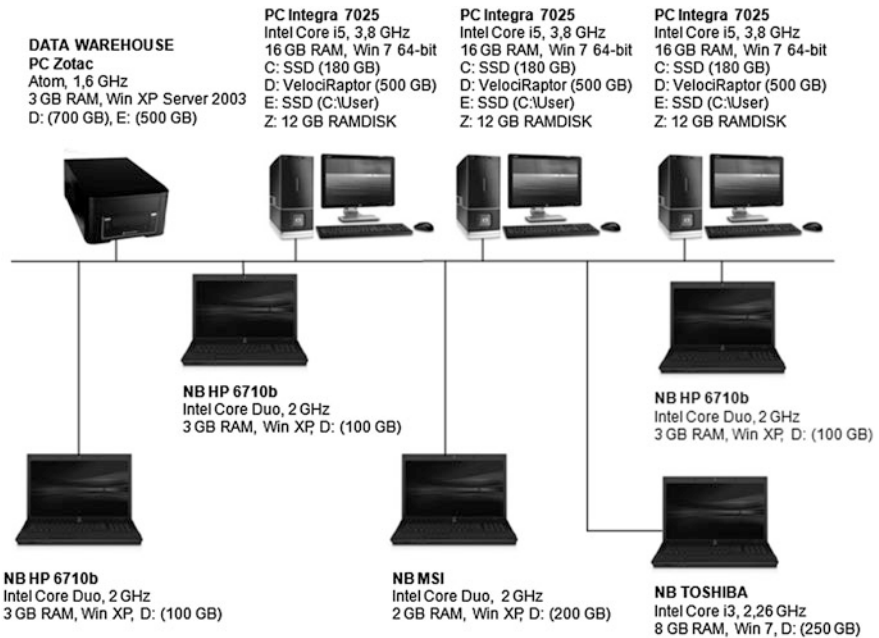


Fig. 2 The scheme of connection and parameters in computer network

5 Experimental Results

The project was implemented on a territory of the CR with the total area of 64,350 km². Input data for the classification were sheets of orthophotos with the resolution of 25 cm/1 pixel. The improvement of results of image classification was achieved by filtration from communication layers of the ZABAGED database. Processing was carried out according to the methodology described in the previous chapter. Results of the data analysis have been developed in text form (a pivot table in Excel, see Fig. 3) and graphical form (SHP format, see Fig. 4) in the dividing from standpoint:

1. Object standpoint, i.e. gas lines of types: high-pressure, main line and service pipes
2. Territorial standpoint, i.e. according to administrative division of the territory of the CR (regions, districts, ORP, municipality, parts of municipalities).

The structured table in Fig. 3 is the row structured list of about 13,000 rows (1 row contains data about a part of municipality). This list is arranged in the form of pivot table with options of sorting by common and widespread filter according to categories of a part of municipality, municipality, ORP, district and region (yellow parts in the figure). Surface analysis was carried out in three variants for the course of high-pressure gas pipeline, main lines and service pipes for 2 groups, 5 groups or 10 groups of surface type. The result is cumulative lengths of high pressure gas lines, main lines and service pipes under individual classes of surface for defined

The screenshot shows a Microsoft Excel spreadsheet with a pivot table. The pivot table is structured as follows:

Číslo	ČÍSLO ORCE	ORCE	ORP	OKRES	KRAJ	Klasifikace A			Klasifikace B			
						Data			Data			
						1 Způsob	2 Způsob	3 Způsob	1 Způsob	2 Způsob	3 Způsob	
6	Číslo ORC	Název ORC	Číslo ORP	Název ORP	Číslo Okres	Název Okres	Číslo Kraj	Název Kraj	Součet metrů A	Součet metrů B	Součet metrů C	Součet metrů D
7	000019	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	4065,33	2424,45	593,18	972,46
8	000027	Mláčava	0540	Olšovec	04036	Karlovy Vary	3051	Karlovarský kraj	1239,49	678,01	64,87	150,67
9	000051	Nová Ves nad Sázavou	0303	Brno-město	04054	Brno	3055	Jihomoravský kraj	0,00	0,00	0,00	0,00
10	000080	Nová Ves nad Sázavou	0303	Brno-město	04054	Brno	3055	Jihomoravský kraj	0,00	0,00	0,00	0,00
11	000108	Jeseník	067175	Jeseník	04095	Moravskoslezský kraj	3069	Moravskoslezský kraj	0,00	0,00	0,00	0,00
12	000116	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	4,24	3,49
13	000124	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	20382,29	11225,12	2644,85	3324,08
14	000131	Nová Ves nad Sázavou	0303	Brno-město	04054	Brno	3055	Jihomoravský kraj	4197,44	3032,41	1111,00	503,45
15	000135	Čáslav	057659	Čáslav	04052	Středočeský kraj	3055	Středočeský kraj	0,00	0,00	0,00	0,00
16	000281	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	2075,09	1215,38	342,67	108,35
17	000299	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	4679,74	698,18	939,36	659,34
18	000302	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
19	000311	Nová Ves nad Sázavou	0303	Brno-město	04054	Brno	3055	Jihomoravský kraj	3095,31	3446,70	395,70	414,93
20	000329	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	1041,43	1788,07	67,88	118,64
21	000337	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
22	000345	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
23	000353	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
24	000381	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
25	000370	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	74,46	167,17	10,62	20,48
26	000388	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
27	000396	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	3740,36	149,85	318,11	11,74
28	000400	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	5887,76	2622,25	745,10	1114,13
29	000418	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
30	000426	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
31	000434	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
32	000442	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00
33	000451	Albertyny	0540	Albertyny	04036	Karlovy Vary	3051	Karlovarský kraj	0,00	0,00	0,00	0,00

Fig. 3 The example of a pivot table of output in XLSx format

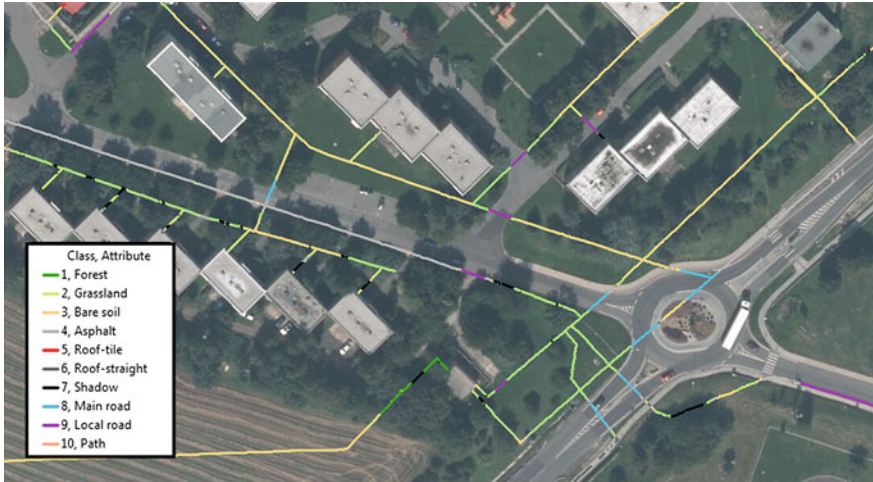


Fig. 4 Graphical output of a structured classified data in SHP format of the route of gas line

selection by a filter. In terms of volume data, it is a very detailed division of extensive territory-bound geo-database. Results include a total of $3 \times (2 + 5 + 10) \times 13,000 = 663,000$ items.

The graphical output of structured classified data in SHP format for 3 types of gas lines is in Fig. 4. This is a segmented course of gas line with an attribute table for 10 classes, which is georeferenced and viewable over any graphic base (orthophotos, cadastral map, etc.) in GIS of RWE Company.

GIS tools can conduct the function evaluation of the sum of the lengths of types of gas line under individual classes of surface in given categories (according to the definition by alphanumeric attribute) above any area in graphics, which is defined by e.g. fence, closed polygon, etc.

6 Analysis of Error Rate of Processing

The error rate of the F was evaluated as the ratio of the total length erroneously classified sections of the gas line to the total length of the gas line in percentage. The error rate was determined on a sample of 20 % of the territory (40 ORP) using locating and measuring lengths of erroneous sections of the S in a classified image. The orthophoto image was admitted as a reference fact. The error rate was assessed by the human factor by the following equation:

$$F = \frac{A}{C} * 100[\%] \tag{4}$$

Table 1 Error rate of processing

Error rate	High pressure[%]	Main line [%]	Service pipes [%]
Minimum	0.4	1.4	1.1
Maximum	2.6	5.4	4.6
Average error rate	0.5	1.9	1.5
Standard deviation	0.4	1.4	1.1

Average error rate contains statistical value of errors of surface classification through the whole territory of the Czech Republic

where

A is the length of incorrectly classified sections of the gas line

C is the total length of the gas line

The overview of the error rate for the high pressure gas line, the main series and service pipes is given in Table 1. The maximum error rate of sample does not exceed 5.4 %. The largest average error rate shows the main line.

7 Conclusions

These following conclusions are from our solution:

- (a) The technology enables with high information ability (error rate less than 5.5 %) and relatively rapidly (in the period to 16 weeks) processing data analysis of lengths of gas line sections that lie under the surface of various types:
 - 2 groups of a surface type: paved and unpaved
 - 5 groups of a surface type: asphalt, main road, local road, unbound, unknown
 - 10 groups of a surface type: forest, grassland, bare soil, asphalt, roof-tile, roof-straight, shadow, main road, local road, path.
- (b) The results of data analysis for 2 groups (paved and unpaved surface) have the largest information ability.
- (c) The technology of data analysis enables repeatability practically without influence of the human factor.
- (d) Data analysis shows very high information ability in terms a low error rate and it can be used repeatedly. The error rate is due mainly a limit quality of the default underlying data.
- (e) The technology of data analysis has qualitative potential of further refinement with possibilities of future use of new, qualitatively higher, source datasets.
- (f) The entire technological process requires about 54 % of skilled manual work, especially in terms of data preparation and qualified decision between processes—see Fig. 5

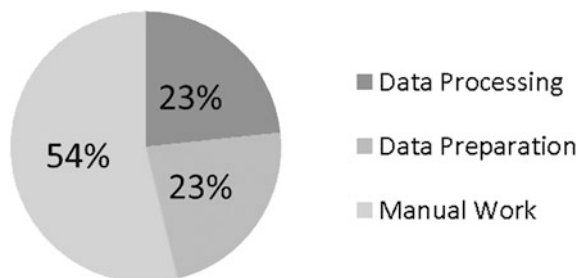


Fig. 5 Portion of time-demand of automated part of data analysis and manual work

The results showed that only one additional dataset for filtering of sub-results suffices for the effective achievement of the required quality classification for the given purpose and this layer refines results of the automated process. Subsequent visual inspection associated with editing then slightly improves the result of classification so that the absolute error rate of the fact, that is represented by the image orthophoto, does not exceed 2–3 %.

The process described in this paper has general character and it can be used for classification of the surface above utilities such as water supply, sewerage, electric power distribution, media distribution, etc.

Acknowledgments The work was solved within the project marked FAST-S-15-2723 and called “Methods development of qualitative appreciation of data of progressive measuring techniques via GIS tools on base of operational analysis” and FAST-S-14-2464 called “Research of the behavior of building structures under operating conditions via geodetic methods” and within research project of MŠMT AdMaS ED2.1.00/03.0097 Nr. HS12357021212200 “Data analysis of surfaces above RWE gas lines in the Czech republic”.

References

- Bureš J, Bartoněk D, Opatřilová I (2013) Data analysis of surfaces above RWE gas pipelines in the Czech Republic. Final report. Project AdMaS ED2.1.00/03.0097 Nr. HS12357021212200
- Černota P, Staňková H, Gašincová S (2011) Indirect distance measuring as applied upon both connecting surveys and orientation one. *Acta Montanist Slovaca* 16(4):270–275. ISSN 1335-1778
- da Silva AQ, Paradella WR, Freitas CC, Oliveira CG (2014) Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. *Remote Sens Environ* 155–167
- Dandoš R, Staňková H, Černota P, Subiková M (2013) Spatial visualization of the infantry blockhouse OP—S 25 “U trigonometru”. *Adv Mil Technol* 8(1):73–84. ISSN 1802-2308
- Debes C, Weiss C, Zoubir AM (2014) Development and assessment of a 250 m spatial resolution MODIS annual land cover time series (2000–2011) for the forest region of Canada derived from change-based updating. *Remote Sens Environ* 140:731–743
- Gray J, Song C (2013) A new automatic image analysis method for assessing estrogen receptors’ status in breast tissue specimens. *Comput Biol Med* 43(12):2263–2277

- Liames JS, Congalton RG, Lunetta RS, Brenning A (2013) Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975–2010). *Remote Sens Environ* 137:112–123
- Johnson DM (2013) Comparison between pixel- and object-based image classification of a tropical landscape using Systeme Pour l' Observation de la Terre-5 imagery. *J Appl Remote Sens* 7 (1):073512, pp 14
- Landgrebe DA (2003) Signal theory methods in multispectral remote sensing, vol XI. Wiley, Hoboken, p 508. ISBN 0-471-42028-X
- Liang S (2004) Quantitative remote sensing of land surfaces, vol XXXVI. Wiley, Hoboken, p 534. ISBN 0-471-28166-2
- Makarau A, Palubinskas G, Reinartz P (2013) Modelling habitat suitability for alpine rock ptarmigan (*Lagopus muta helvetica*) combining object-based classification of IKONOS imagery and Habitat Suitability Index modeling. *Ecol Model* 254:22–32
- Mouelhi A, Sayadi M, Fnaiech F, Mrad K, Ben R, Khaled P (2013) Alphabet-based multisensory data fusion and classification using factor graphs. *IEEE J Sel Top Appl Earth Obs Remote Sens* 6(2):969–990
- Puertas L, Brenning O, Meza AJ (2013) Estimating area from an accuracy assessment error matrix. *Remote Sens Environ* 132:202–211
- Pupin AM, Alves SD, Friedrich TRB, Ferreira SA (2014) Area-based and location-based validation of classified image objects. *Int J Appl Earth Obs Geoinf* 28:117–130
- Searcóid MO (2006) Metric spaces. Springer undergraduate mathematics series, Berlin. ISBN 1-84628-369-8
- Zhong L, Gong P, Biging GS (2012) Segmentation and thematic classification of color orthophotos over non-compressed and JPEG 2000 compressed images. *Int J Appl Earth Obs Geoinf* 15(SI):92–104