PFG 2011 / 1, 0017–0033
Stuttgart, February 2011

Article

A New Binary Encoding Algorithm for the Simultaneous Region-based Classification of Hyperspectral Data and Digital Surface Models

Huan Xie, Shanghai, China; Christian Heipke, Peter Lohmann, Uwe Soergel, Hannover; Xiaohua Tong, Shanghai, China & Wenzhong Shi, Hong Kong, China

Keywords: Hyperspectral images, DSM, binary encoding, region-based classification, integration

Summary: In this paper, an approach is proposed to integrate hyperspectral image data, object and height information into a new region-based binary encoding algorithm for automatically deriving land cover information. After georeferencing the different data sets and deriving a normalized digital surface model (nDSM), connected regions are extracted from the hyperspectral data by applying an edge-based segmentation algorithm. The mean spectrum per region is considered representative for the region. Five parameters are defined to describe the size and shape of the region, namely area, asymmetry, rectangular fit, ratio of length to width, and compactness. Together with the spectral information these parameters and the corresponding height values per region from the nDSM are converted into a binary code. This code is then matched to that of a training data set for classification.

In order to evaluate the suggested approach we applied it to a test area in Oberpfaffenhofen, Germany. A manually generated classification served as reference. We also compare our result with the well known support vector machine (SVM) classifier. Based on our test data, we could show that the inclusion of size, shape, and height improves the classification accuracy of binary encoding. We could also show that the new method obtained more accurate and more efficient results when compared to the SVM classification.

Zusammenfassung: *Ein neuer Binärcodierungsalgorithmus zur gemeinsamen regionenbasierten Klassifikation von Hyperspektraldaten und Digitalen Oberflächenmodellen.* In diesem Artikel schlagen wir einen neuen Ansatz zur automatischen Klassifikation der Landbedeckung durch Integration hyperspektraler Bilddaten mit Höheninformation vor. Der Ansatz setzt auf einer Segmentierung der Bilder auf und beruht auf der Binärkodierung. Nach der Georeferenzierung der verschiedenen Datensätze und der Ableitung eines normalisierten Digitalen Oberflächenmodels (nDOM) werden die Hyperspektraldaten mit Hilfe eines kantenbasierten Verfahrens segmentiert. Das mittlere Spektrum wird jeweils als repräsentativ für das gesamte Segment betrachtet. Fünf Parameter beschreiben die Größe und Form jedes Segments: Fläche, Asymmetrie, Rechteckigkeit, Verhältnis von Länge zu Breite und Kompaktheit. Diese Parameter werden zusammen mit der spektralen Information und der Höhe aus dem nDOM in einen Binärcode transformiert, der dann dem entsprechenden Binärcode von Trainingsgebieten zugeordnet wird. Zur Evaluierung des neuen Ansatzes wurden Tests mit einem Datensatz aus Oberpfaffenhofen durchgeführt, eine per Hand erstellte Klassifikation diente dabei als Referenz. Wir haben unsere Ergebnisse auch mit denen einer Support Vector Machine (SVM) verglichen. Für unser Testgebiet konnten wir zeigen, dass die Berücksichtigung von Größe, Form und Höhe die Ergebnisse im Vergleich zur standardmäßigen Binärklassifikation verbessert, dabei hat die neue Methode auch im Vergleich zu dem SVM Ansatz genauere Ergebnisse geliefert und war darüber hinaus deutlich schneller.

www.schweizerbart.de
DOI: 10.1127/1432-8364/2011/0072
1432-8364/11/0072 $ 4.25

1 Introduction

Hyperspectral imaging, also known as imaging spectroscopy, is a relatively new technique in remote sensing that generates hundreds of images by spectral filtering and binning of the received radiance into narrow bands (also called channels) of different center wavelength for the same area on the surface of the earth (Chang 2003). In principle, standard classification algorithms designed for multispectral imagery can be directly applied to hyperspectral data, because no theoretical limitations on the number of bands exist, provided that enough training samples are available. Otherwise, the Hughes phenomenon is likely to be encountered, see Hughes (1968). However, in practice algorithms such as Maximum-Likelihood (ML), even with improvements (Bolstad & Lillesand 1991, Lee & Landgrebe 1991), tend to perform poorly when applied to hyperspectral imagery. Furthermore, in case of ML the class mean vectors and the covariance matrices have to be estimated from the training data assuming a Gaussian distribution. Therefore, a number of alternative classification tools have been developed tailored to the processing of hyperspectral imagery. Their structure is not only driven by a need for efficiency, but also by different types of pattern recognition approaches, e. g., object-based analysis, made possible by the high-resolution spectral data.

The huge amount of information in hyperspectral imaging has always been a great challenge for scientific understanding and data exploration. Recently research focused on the retrieval of land surface variables (Verhoef & Bach 2007, Darvishzadeh et al. 2008, Martin et al. 2008), image classification and feature extraction (Bazi & Melgani 2006, Ball & Bruce 2007, Heiden et al. 2007, Hsu 2007, Duarte-Carvajalino et al. 2008, Prasad & Bruce 2008, Rajan et al. 2008), and preprocessing methods for classification, e. g., band selection and data reduction (Martinez et al. 2006, Martinez-Uso et al. 2007, Rogge et al. 2007, Serpico & Moser 2007, Demir & Erturk 2008). Since the first hyperspectral sensor became available in 1983 (Chiou 1984) a number of studies (e. g., Bruce & Li 2001, Nascimento & Dias 2005) were also concerned with reducing computational cost and improving algorithmic efficiency and such studies continue until today. These investigations mainly focused on the spectral information contained in the data and achieved great successes in their research fields.

The advent of advanced processing techniques and high speed computers have led to the possibility of supplementing hyperspectral image data with information about different kinds of objects during classification, for example, size and shape. Other data sources, e. g., digital surface models, provide height information and can also be integrated into the classification. Instead of focusing exclusively on the spectral information, several studies began to focus on methods for the representation of spatial information in hyperspectral data (Huang et al. 2007), the spectral-spatial classification methods (Zhang et al. 2006, Camps-Valls et al. 2006, Marconcini et al. 2009, Jia & Richards 2008), and the use of multisource data for hyperspectral classification (Hepner et al. 1998, Gamba & Houshmand 2000, Greiwe & Ehlers 2005, Feingersh et al. 2010). These studies show that additional information (e. g., spatial characteristics) can be very helpful during the application of hyperspectral data, however some are computationally rather expensive.

Our research considers hyperspectral image analysis from a broader perspective than the individual methods listed above. We focus on the simultaneous analysis of spectrum, size, shape and height from hyperspectral images and data from digital surface models (DSM). The current paper integrates a region-based classification approach with a traditional hyperspectral processing method (binary spectral encoding) to enhance the information extraction from remote sensing data. Binary spectral encoding is well known as a simple, effective hyperspectral analysis method with very small computational load in classification, search of similar spectra and identification of mineral components (Mazer et al. 1988). In the proposed method the information present in the input data after segmentation, i. e., spectrum, size, shape, and height per region, is translated into binary codes. Based on training data land cover classes of interest to the user are transformed into binary codes,

too. Subsequently, an algorithm to evaluate the similarity between the binary codes of regions and land cover classes is applied. The suggested procedure can be considered a useful pre-processing method for further image analysis, as it has proven to be an efficient classifier in our experiments.

Following this introduction, we describe the employed binary encoding method in Section 2. Especially the representation and coding of the different kinds of features and the computation of the signature distance are explained. Practical tests obtained using a test area in Oberpfaffenhofen, Germany, illustrating the proposed methods are presented in Section 3, and finally some conclusions are drawn.

2 Methods

2.1 *Overview*

Fig. 1 shows the overall flowchart of the suggested method. Several pre-processing steps for the hyperspectral images (we use HyMap data in our work, see Cocks et al. 1998 for a description of this sensor) and the DSM data are performed, e. g., the geometric, radiometric and atmospheric correction of the HyMap images and the generation of the normalized DSM (nDSM) from the given height data by first generating and then subtracting a digital terrain model (DTM) from the DSM (note that a DSM represents height including objects such as buildings and vegetation, whereas a DTM does not and a nDSM only contains these objects). Afterwards, image regions are extracted from the HyMap image using an edge-based segmentation algorithm. Then, the mean spectrum per image region is computed, as it is considered as a representative spectral description of the region. Subsequently, five descriptors, defined to represent object size and shape, are calculated for each region: Area, asymmetry, compactness, rectangular fit, and ratio of length to width. While it can be argued that size and shape descriptors may be less suitable in the presence of segmentation deficiencies – e. g., area may not be an appro-

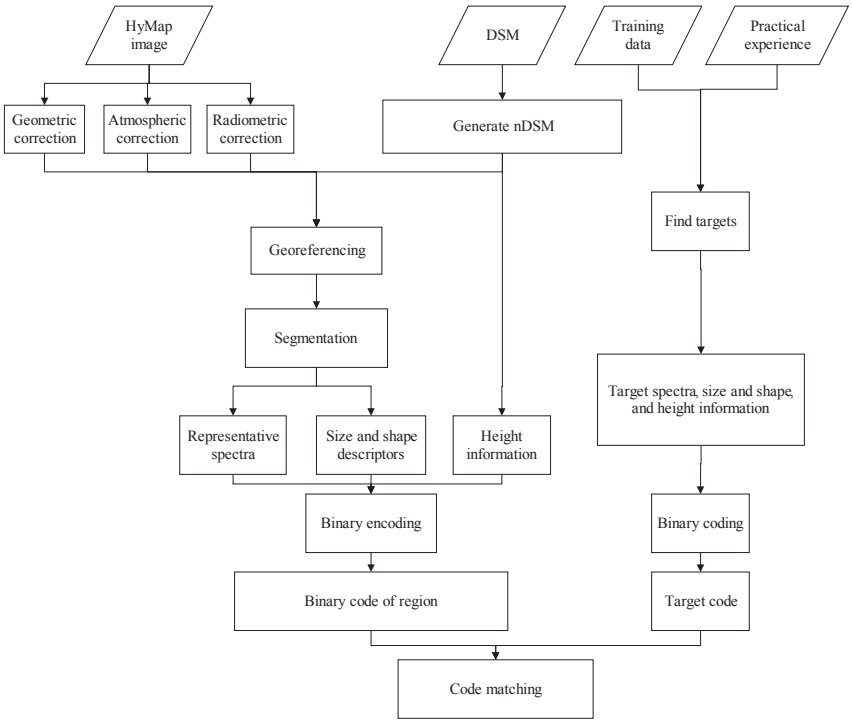

Fig. 1: Project flow chart.

priate descriptor in case of over- or under-segmentation – such descriptors have been successfully used in image classification in the past (Mehtre et al. 1997, Mohanty et al. 2005), and the descriptors selected here were carefully evaluated and have shown to be useful for our purposes.

After intersecting the nDSM and the region borders, for each region the average heights from the nDSM are grouped into three classes; all the information is finally converted into a binary code. This code is then matched to the target codes derived from training data for classification.

2.2 *Segmentation*

A segmentation algorithm is used with the expectation that it will divide the image into semantically significant regions useful for further processing. Several segmentation approaches have been proposed in the literature, e. g., Le Moigne & Tilton (1995), Kartikeyan et al. (1998), Acharyya et al. (2003), Duarte-Carvajalino et al. (2008), and comparisons were carried out to evaluate the quality of segmentation (Trias-Sanz et al. 2008). Since segmentation is only considered a pre-processing step in our research, we selected a simple edge-based segmentation and merging algorithm (Robinson et al. 2002), which according to visual inspection performed well on our test data. After subdividing the image into a large number of small segments based on the Mumford-Shah functional (Mumford & Shah 1985), the algorithm iteratively merges adjacent segments by combining spectral and spatial information. Merging proceeds if the algorithm finds a pair of adjacent regions, i and j, such that the merging cost $t_{i,j}$ is less than a predefined threshold λ:

$$t_{i,j} = \frac{\frac{|O_i||O_j|}{|O_i|+|O_j|}\left\|u_i - u_j\right\|^2}{length(\partial(O_i, O_j))} \qquad (1)$$

where O_i and O_j represent the two considered segments i and j, $|O_i|$ and $|O_j|$ describe the area of O_i and O_j, u_i and u_j are the vectors of pixel mean grey values (mean spectra) of O_i and O_j, $\|u_i - u_j\|$ is the Euclidean distance between the mean spectra, and $length(\partial(O_i,O_j))$ is the length of the common boundary of O_i and O_j.

2.3 *Binary Encoding for Image Regions*

Binary encoding is a standard technique in classifying hyperspectral images (e. g., Jia & Richards 1993). The basic idea is to reduce the large amount of data while preserving as much information as possible. Standard binary encoding reduces the information of a pixel (often represented as 8 bit per channel) into one or two bits per channel only. In our research we consider regions rather than individual pixels, and our code for an image region is $2L+28$ bits long, where L is the number of spectral channels of the hyperspectral image. The code consists of four parts, i. e., spectrum, size, shape, and height. The spectral amplitude and slope are represented by $2L$ bits, the size and shape of the segment is coded by 25 bits, and the relative height of a segment is represented by 3 bits (as will become clear below, size and shape as well as relative height could have been coded with less bits, but in this paper we are not concerned with minimum bit encoding; we have thus chosen a somewhat redundant but simple coding scheme).

2.3.1 *Spectrum*

According to Mazer et al. (1988), spectral mean values (mean grey values over all available channels) are calculated from the individual channel grey values in standard spectral binary encoding. A single spatial resolution element of the image (pixel) is denoted by an L-dimensional vector, where the indices (i, j) refer to the spatial location of the pixel within a given scene and $X(l)$ describes the grey value of channel l. Defining the scalar quantity v_{ij} as the spectral mean of pixel (i, j),

$$v_{ij} = \left[\frac{1}{L}\right]\sum_{l=1}^{L} X_{ij}(l) \qquad (2)$$

an L-bit binary code vector $\overrightarrow{Y_{ij}^{sa}}$ can be constructed from

$$\overrightarrow{Y_{ij}^{sa}} = \begin{cases} 0, \left[X_{ij}(l) - \nu_{ij}\right] < 0 \\ 1, \left[X_{ij}(l) - \nu_{ij}\right] \geq 0 \end{cases}, l = 1,2,\ldots,L \quad (3)$$

The constructed vector is a binary representation of spectral amplitude; however, considerable information is contained also in the local slope at each measured channel. Therefore, an additional L-bit code vector $\overrightarrow{Y_{ij}^{ss}}$ is constructed from

$$\overrightarrow{Y_{ij}^{ss}} = \begin{cases} 0, & [X_{ij}(l+1) - X_{ij}(l-1)] < 0 \\ 1, & [X_{ij}(l+1) - X_{ij}(l-1)] \geq 0 \end{cases}, l = 1,2,\ldots,L \quad (4)$$

Here $X_{ij}(0) = X_{ij}(L), X_{ij}(L+1) = X_{ij}(1)$, these two code vectors $\overrightarrow{Y_{ij}^{sa}}$ and $\overrightarrow{Y_{ij}^{ss}}$ are then concatenated to form a single, $2L$-bit code vector $\overrightarrow{Y_{ij}}$, which is taken to be the binary code word representing the spectrum of pixel (i, j).

In our region-based approach we use the average grey values per region for each channel in equations (2)–(4) rather than the grey values of the individual pixels, and then proceed in the way described above.

Size and Shape Parameters

The size and shape of a segment are represented by five descriptors: Area, asymmetry, compactness, rectangular fit, and ratio of length to width; they are explained in detail below; see also Benz et al. (2004). Although these descriptors are somewhat correlated as they encode similar information (e. g., asymmetry and length/width ratio) we have opted to use all five descriptors for our study based on the experimental results. Each descriptor has been encoded using 5 bits.

Area

The area of a region is measured as the number of pixels covering it.

Asymmetry

The lengthier a region, the more asymmetric it is. As we can see in Fig. 2, an ellipse is approximated to each region and the asymmetry is expressed as 1 minus the ratio of the length of minor axis n and the length of major axis m of this ellipse:

$$Asymmetry = 1 - \frac{n}{m} \quad (5)$$

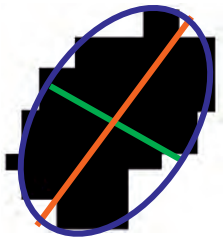

Fig. 2: The bounding box and ellipse of an image object (see also length/width ratio below).

The feature value increases with the asymmetry of the region, and the asymmetry value for a segment ranges from zero to one.

Compactness

Compactness is defined in our research as the ratio of the area A_p of a polygon representing the image region to the area of a circle with the same perimeter. The computation of polygons are obtained from the borders of each region and then generalized by means of the Douglas-Peucker algorithm (Douglas & Peucker 1973). The following formula is used to calculate the compactness of the selected polygon:

$$Compactness = \frac{4\pi A_p}{Perimeter^2} \quad (6)$$

Perimeter is the sum of the lengths of all edges which form the polygon. The compactness of a polygon ranges from zero to one and a circle has the highest compactness value.

Rectangular fit

A first step in the calculation of the rectangular fit is the creation of a rectangle with the same area as the considered region. In the creation of the rectangle the proportion of length and width of the bounding box of the region (see Fig. 2) is taken into account. Then, the area of the rectangle not covered by the image object A_O is compared with the area A covered by the image object. The better the fit the larger is the value, 1 stands for a perfect fit.

$$RectangularFit = 1 - \frac{A_o}{A} \quad (7)$$

Length to width ratio

The length to width ratio is approximated using the bounding box (see Fig. 2 and Russ 2006):

$$\gamma = \frac{a^2 + ((1-f)\cdot b)^2}{A} \tag{8}$$

Parameters a and b are length and width of the region's bounding box, respectively, f is the degree of filling, which is the area A covered by the region divided by the total area $a * b$ of the bounding box. The minimum value of the length/width is 1.

Encoding Size, Shape and Height

Encoding for the size and shape descriptors follows a similar way as for the spectra. The values of each descriptor are grouped into five bins. After constructing a histogram, boundaries are set at 20 %, 40 %, 60 % and 80 % of the total number of pixels (see Fig. 3 for an example of rectangular fit). In this example the peaks in the histogram are caused by different objects: The peaks around 0.1 mainly by connected roads, the peaks around 0.4 represent the majority of the image object, the peak around 0.55 is caused by regular cropland, while the peaks around 0.6 are mainly caused by buildings, and parking lots. The number and position of the bin boundaries was selected based on prior tests. Note that instead of using fixed boundaries, thresholds could have also been derived from the histogram minima. However, such an approach would have resulted in a more unstable procedure (consider various side minima in a local neighborhood) and a variable length code. Therefore, we have opted to work with fixed thresholds. Finally, a five digit code is constructed for each descriptor of a region: "1" is set for the bin the descriptor belongs to, "0" is used elsewhere. For example, 00100 means the descriptor lies in bin 3 of the histogram, i. e., between 40 % and 60 %.

The codes for the heights are determined from the average nDSM heights per image region. Heights are classified into three bins (again, this number was selected based on prior tests): Height less than 1.5 meters, height between 1.5 and 5 meters, and height larger than 5 meters. The codes are then generated in the same way as for size and shape, e. g., "001" means that the relative height of the considered region is larger than 5 meters.

The codes for the features size, shape, and height are concatenated in a vector of 28 bit length,

$$\overline{Z_i} = [Z_i(1), Z_i(2), \ldots, Z_i(28)]^T \tag{9}$$

The elements $Z_i(1)$ to $Z_i(25)$ encode size and shape in 5-digit codes each, $Z_i(26)$ to $Z_i(28)$ encode height.

2.4 *Encoding the Land Cover Classes*

The target land cover classes (or targets) need to be coded in a similar way as the input data. While in principle all necessary values could be learnt from training data, in our research we use a combination of empirical values derived from training data (e. g., for the spectra) and of general knowledge about the land cover classes under consideration (for size, shape, and height). The latter capture general statements such as 'buildings have a height of at least 3 meters' or 'forests are more likely to have lower compactness than grassland in a city'. While these statements are assumed to be valid for more than one test scene, adjustments may obviously be necessary if data showing different landscapes etc. are to be processed.

As an example consider the class 'Industrial buildings':

- The area of industrial buildings is not very large (compared to that of other classes); the buildings are likely to have different sizes;
- Industrial buildings have regular shapes; the compactness and rectangular fit are comparatively high;
- Industrial buildings are often higher than 5 meters.

Translated to binary codes and considering the existence of a considerable number of irregularly shaped buildings, we assume the area of industrial buildings to fall into bins 1 or 2, compactness to be mapped into bins 3, 4, or 5, rectangular fit to fall into bins 3, 4, or 5,

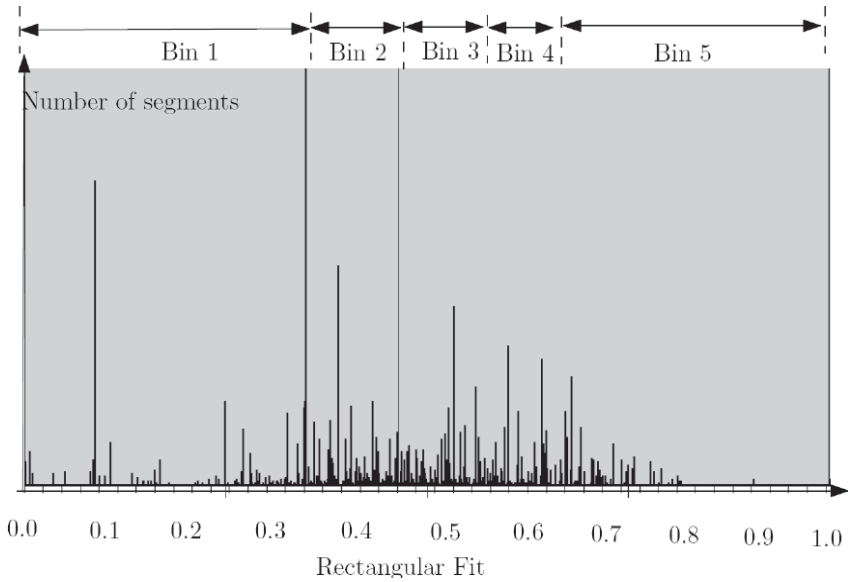

Fig. 3: The encoding rule for rectangular fit.

Tab. 1: Codes for the land cover classes used in our research.

Land cover class	Number of training samples	Bin(s) of histogram					
		Size and shape					Height
		Area	Asymmetry	Compactness	Rectangular fit	Length/width	
Stockyard or parking lot	4	1–2	1	4–5	4–5	1–2	1
Surface sealed with concrete	6	1–2	1–4	2–5	1–5	1–3	1
Street	8	1–2	4–5	1–2	1–3	5	1
Runway	3	4	5	1	1	5	1
Tennis ground	3	1	1–2	5	3–5	1–2	1
Farmhouse	8	1	1–4	2–5	1–5	1–3	1–3
Mixed trees*	6	1–4	1–4	1–5	1–5	1–5	1
Industrial buildings	14	1–2	1–5	3–5	3–5	1–5	2–3
Industrial area	9	1–3	1–5	1–5	1–5	1–5	1
Cropland	28	2–4	1–5	1–5	1–5	1–5	1
Grass	31	2–4	1–5	1–5	1–5	1–5	1

* trees were not captured as part of the DSM, therefore, the corresponding regions are assumed to feature a low height value (see Section 3.1 for more information).

and height to fall into bins 2 or 3. For the asymmetry and length/width ratio, we assume that all bins (bin 1 to bin 5) are possible.

In a similar way codes were constructed for all relevant land cover classes (see Tab. 1). Referring to the land use and land cover categories put forward by USGS (Anderson et al. 1976) and taking into account our test area (see Section 3), 11 classes were chosen: Stockyard or parking lot, Surface sealed with concrete, Street, Runway, Tennis ground, Farmhouse, Mixed trees, Industrial buildings, Industrial area, Cropland, and Grass. From the table it can be seen that while for some classes the additional size, shape, and height information should be of considerable value (e. g., for streets and runways), other classes will not be significantly improved by integrating the additional information (e. g., industrial area, cropland, grassland). This fact reflects the variable size and especially shape of these classes and should not been seen as a counterargument to region-based approaches as such.

2.5 *Code Matching*

The dissimilarity measure we use to determine spectral signature matches is the Hamming distance (Hamming 1950, Viterbi & Omura 1979):

$$D_s(\vec{Y}_i, \vec{Y}_m) = \sum_{l=1}^{2L} Y_i(l)(XOR)Y_m(l) \tag{10}$$

which is seen to be just a $2L$ sum of bit-wise exclusive-OR operations. Indices i and m refer to the considered region and the target under consideration, respectively. In the actual implementation of this algorithm, the Hamming distances D_s^{sa} and D_s^{ss} for the two components of the vectors $\vec{Y}_i$ and $\vec{Y}_m$ being compared (see equations (3) and (4)) are computed separately. This gives the user some additional flexibility in choosing weights for amplitude and slope information. This distance measures the dissimilarity between the two binary vectors. If the distance equals zero, then these two vectors are identical.

Different from the spectral data, the operator used in the similarity evaluation of size, shape, and height descriptors is the bit-wise AND operation, which is computed from

$$D_h(\vec{Z}_i, \vec{Z}_m) = 6 - \sum_{l=1}^{28} Z_i(l)(AND)Z_m(l) \tag{11}$$

The AND operation is more like a mask operation. For one descriptor (5 codes for shape, 3 codes for height), only 1 (matched) or 0 (unmatched) can result from the calculation. As a total of six descriptors (five size and shape descriptors, and one height descriptor, represented by 28 bits altogether) are used in our research, D_h equals zero if all descriptors are matched, while D_h equals 6 if all descriptors are unmatched. Thus, D_h measures dissimilarity, as does the Hamming distance. In the actual implementation of our algorithm, the separate distances D_h^h and D_h^s are computed for for height and for size and shape.

Using proper weights (see Section 3 for a discussion on how to select them), we now combine the different dissimilarity measures into the final measure D, which represents the dissimilarity between the image object and the target:

$$D = D_s^{sa} + D_s^{ss} + w_s D_h^s + w_h D_h^h \tag{12}$$

3 Results

In order to evaluate the proposed method, we tested it using data from an area in Oberpfaffenhofen, Germany. The area was mapped manually by an experienced human operator using the composite true color image (Bands 18, 8 and 1 of the HyMap image for red, green and blue channels, the center wavelengths of these bands are 735 nm, 570 nm, and 465 nm) to create reference data. This result was then compared to different versions of the new method and also to the results of a number of established classification algorithms. This section contains the results of these tests.

3.1 *Study Area and Data*

The study area Oberpfaffenhofen is located in the south of Germany. The available digital surface model of this region has a resolution of

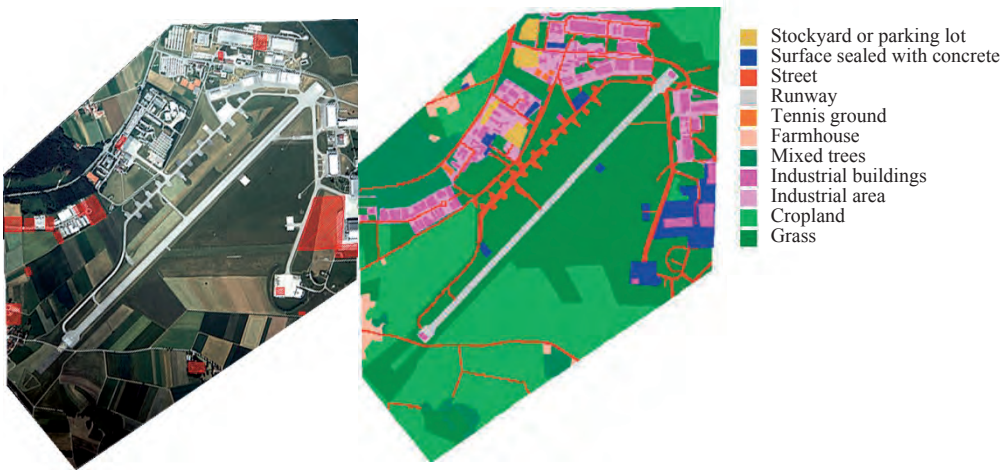

Fig. 4: The aerial photo of 1994 and the manual reference interpretation of the HyMap data of 2004.

0.5 meters and was generated by manual stereo measurements from a true color stereo pair of June 1994 at a scale of 1 : 13.000. While buildings were captured as part of the DSM, trees were not. HyMap data of this area was captured at 1:00 pm, June 7, 2004 (thus, 10 years later) at a flying altitude of 2580 m above sea level (corresponding to approximately 2000 m above the ground), and a flying direction from south to north (3.6°). These data have a ground resolution of 4 meters and 126 channels. The digital surface model (DSM) and the HyMap image are both available for an area of approximately 4.6 km^2. Technical details of the HyMap sensor can be found in Cocks et al. (1998). Fig. 4 shows the aerial photo of 1994 of the test area on the left and the manual reference interpretation of the HyMap data of 2004 on the right. As can be seen, the classes cover a significantly different amount of terrain, which somewhat reduces the value of the reference map, but conclusions as to the classification accuracy can still be drawn. Also, there have been some changes especially in the construction of new buildings within the 10 year difference of both data sets, part of the changes are marked by red regions on the aerial photo of 1994.

3.2 *Experimental Results*

Results of the New Method

We first tested the proposed new method in our test area with a number of training samples (and optimum weights, see Tab. 1 and below). In the segmentation step, a value of 88 was used for λ (see eq. (1)) based on a few tests, resulting in an average region size about 200 pixels. Note that the areas which changed during the time difference of 10 years between the two image acquisitions (depicted red in Fig. 4, left) are masked out and have no effect on the results.

Fig. 5 shows the HyMap RGB composite image and the classification result using the proposed method. Comparing the result to the reference classification (see Fig. 4, right) the first qualitative impression seems to be rather convincing.

The corresponding error matrix is shown in Tab. 2. From the table we can see that despite the good visual impression of the results the overall classification accuracy is actually not very high, i. e., 76.0 % with a kappa index of 0.677. Since about 2/3 of the pixels belong to either cropland or grass, we also report the overall accuracy for the remaining classes, it amounts to 69.9 %. Incorrect classifications can be found for Street, Farmhouse and Industrial area, these are the main reasons affecting

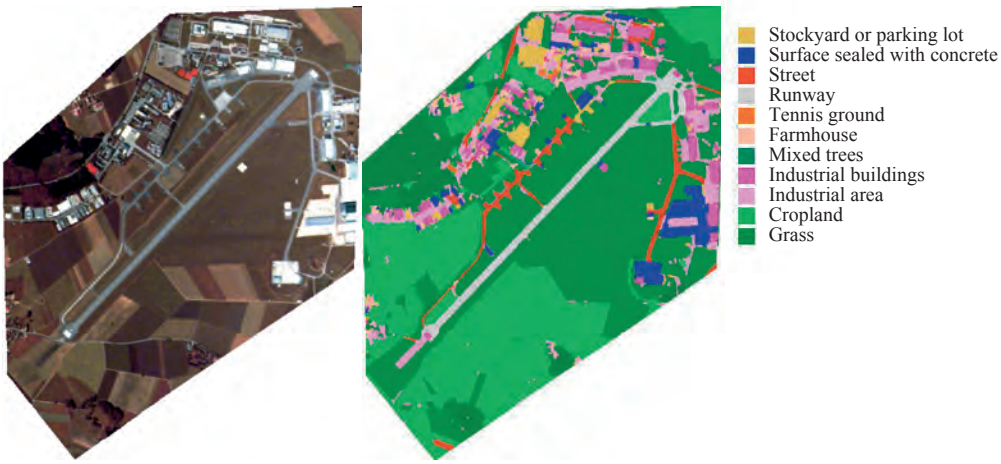

Fig. 5: The HyMap RGB composite image and the classification results derived using the proposed new method.

the overall accuracy. The ground resolution of only 4 meters of the HyMap image is seen as the main reason for the relatively low classification accuracy of Street and Farmhouse, the resolution is not sufficient to extract these relatively small objects. The low accuracy of Industrial area is probably due to the variability of this land cover class, which may contain industrial buildings, streets, small grass areas etc. Another potential reason for the relatively low classification accuracy is the time difference between image and height data: Several buildings present in the HyMap image do not exist in the DSM.

Setting the Weights

As was mentioned in Section 2.5, the spectral and the additional information (size, shape, and height) need to be properly weighted with respect to each other to obtain good results. Ideally, different weights should be chosen for each land cover class, e. g., based on the variability of the codes (see Tab. 1).

In order to investigate the impact of the weights w_s for size and shape and w_h for height on the classification performance, we systematically tested different set-ups. The results are shown in Fig. 6, the upper half shows the overall accuracy computed based on the training data to check the plausibility, while the lower half shows the overall accuracy with respect to data not used during the training phase. In Fig. 6, the case of $w_s = 0$ and $w_h = 0$ shows the classification result only using the spectral information. With the help of size, shape, and height the classification accuracy of the training data increases moderately from 93 % to 94 %, while for the test data the accuracy increases more significantly, namely from 69.0 % to 76.0 % (see Tab. 3). As can be seen, best results for the overall accuracy are reached for $w_s = 2$ and $w_h = 4$, these values are used in the remainder of this paper (they have also been used for the results presented in Fig. 5 and Tab. 2).

It should be noted that the relatively large difference between the results for the training and the reference data points to a potential over-fitting problem. This issue can be further investigated only by employing more and different data sets, which we do not have at hand. As a consequence, we are not able to fully assess the merits of the new method. However, we believe that our results can still be used to investigate its potential and to compare it with more traditional classification schemes.

Comparison with Standard Methods

In order to put the obtained results into context, we compared them to pixel- and region-based standard binary encoding and to a classification based on support vector machines (pixel- and region-based as well as including size, shape, and height features).

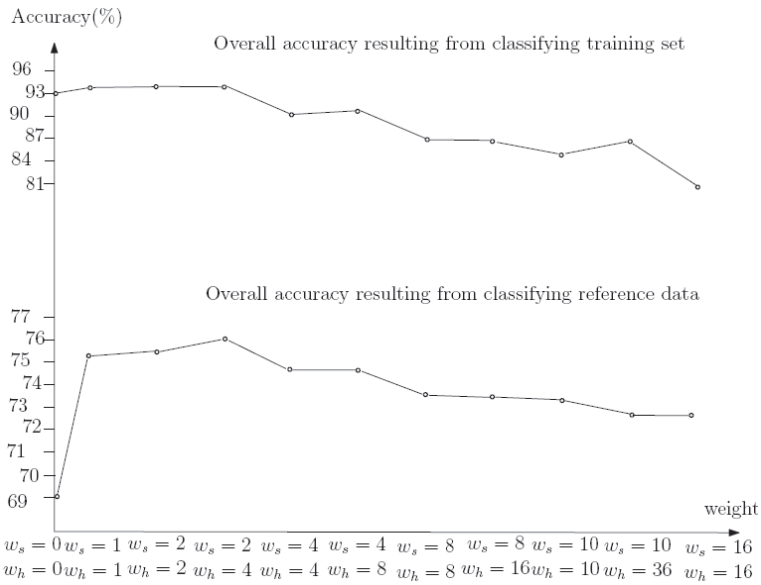

Fig. 6: The classification accuracy using different weights (see text for details).

The support vector machine (SVM) is a universal learning machine for solving classification or regression problems (Smola & Schoelkopf 1998, Vapnik 1998). In remote sensing applications the Gaussian radial basis function kernel $K(x,x_i) = \exp(-\gamma|x\text{-}xi|^2)$ proved to be effective with reasonable processing times, it's adopted in this paper to compare the results of the new method to those obtained with a standard classification scheme. In SVM the user needs to select two parameters: The parameter γ that defines the width of the Gaussian kernel function and a regularization parameter controlling the trade-off between the maximization of the margin between the training data vectors and the decision boundary plus the penalization of training errors (Burges 1998). In order to obtain suitable parameters for the SVM classification, a grid search (Rabe et al. 2009) is advisable, as default settings normally yield poor results.

SVM and binary encoded were both tested in three different versions: Pixel-based, region based using spectral information only, and region-based using spectral and spatial (size, shape, height) information. Of course, the last binary encoding run corresponds to the new improved method. In the SVM classifier we chose the radial basis function (RBF) kernel and all 126 channels. In all cases identical training areas were used, again the areas which had changed between image acquisiton were masked out, and all region-based methods are based on the same segmentation, as described in Section 2.2. All experiments were performed on a PC with a 2,53 GHz Intel Core 2 Duo CPU and 4 GB RAM.

The obtained results are visualized in Fig. 7. and Tab. 3, containing the user's and the producer's accuracy for each class, the overall accuracy, the kappa index, and the computing times for the different steps.

From Tab. 3, we can find that among all approaches, the proposed method has the highest overall classification accuracy and kappa index, 76.0 % and 0.667, respectively, while the pixel based binary encoding classification has the lowest overall accuracy and kappa index, 50.7 % and 0.348, respectively. The region-based approach significantly improves the binary encoding classification accuracy by 18.3 %, note however, the wrongly classified road depicted in red in the lower part of Fig. 7 (d). With the help of the additional features shape and height, the classification accuracy can be improved by a further 7.0 %. The same type of improvement cannot be reached using SVM. On the other hand, the pixel-based SVM

Tab. 2: Error matrix resulting from proposed method.

Reference Data[a]												
	SP	SC	S	R	TG	F	T	B	I	C	G	Row Total
Classification Data												
SP	3078	184	961	0	15	62	97	285	1281	7	210	6180
SC	0	3057	366	0	0	52	2	794	633	4	210	5118
S	2	80	5092	0	0	6	6	63	768	293	635	6945
R	0	0	618	6362	0	0	0	24	492	0	905	8401
TG	0	0	0	0	328	17	1	7	74	0	0	427
F	3	63	439	0	6	729	15	564	986	18	181	3004
T	68	134	796	0	0	285	16571	88	951	1581	1961	22435
B	0	322	209	23	0	132	3	7375	1323	0	322	9709
I	99	916	4056	49	28	849	1026	1181	7349	804	1777	18134
C	15	3	3261	13	34	585	1406	59	483	68431	8980	83270
G	78	327	4513	341	4	333	1840	333	2691	9633	88738	108831
Column Total	3343	5086	20311	6788	415	3050	20967	10773	17031	80771	103919	272454

	Producer's acc.[b]	User's acc.[c]		Producer's acc.	User's acc.
SP	92.07 %	49.81 %	T	79.03 %	73.86 %
SC	60.11 %	59.73 %	B	68.46 %	75.96 %
S	25.07 %	73.32 %	I	43.15 %	40.53 %
R	93.72 %	75.73 %	C	84.72 %	82.18 %
TG	79.04 %	76.81 %	G	85.39 %	81.54 %
F	23.90 %	24.27 %			

Overall accuracy = 76.0 % Accuracy in built-up areas[d] = 69.9 % Kappa = 0.677

[a] SP: Stockyard or parking lot; SC: Surface sealed with concrete; S: Street; R: Runway; TG: Tennis ground; F: Farmhouse; T: Mixed trees; B: Industrial buildings; I: Industrial area; C: Cropland; G: Grass

[b] Producer's accuracy: Relates to the probability that a reference sample will be correctly mapped and measures the errors of omission.

[c] User's accuracy: Indicates the probability that a sample from land cover map actually matches the reference data and measures the error of commission.

[d] Accuracy in built-up areas: Refers to the classification accuracy except Cropland and Grass.

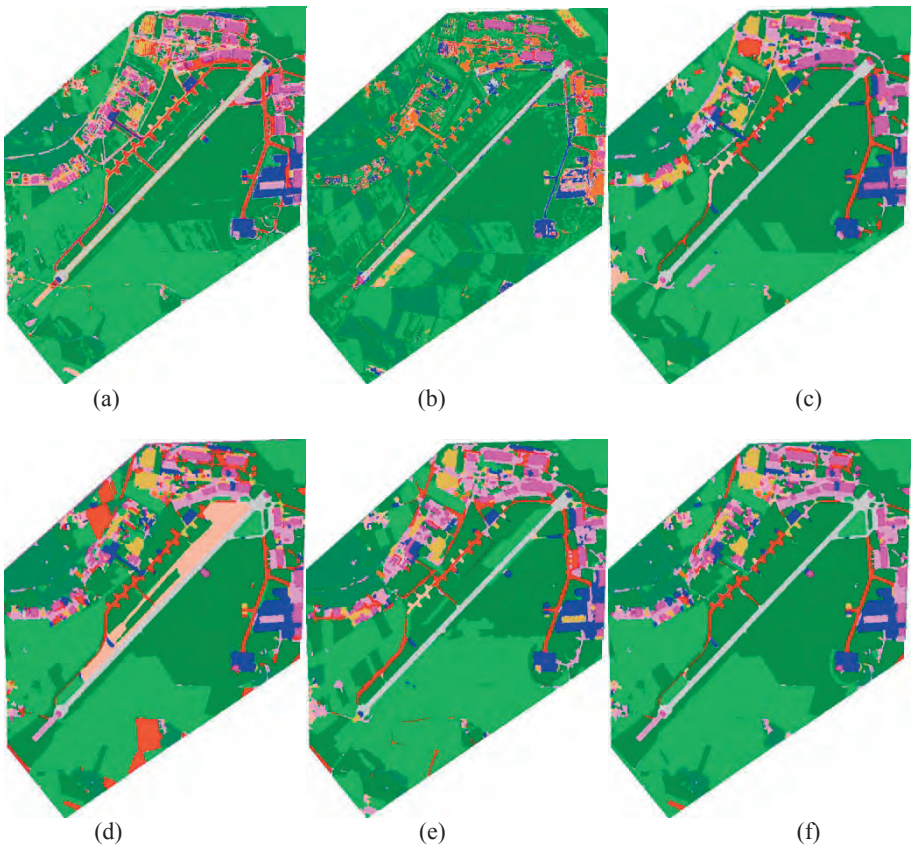

Fig. 7: Results of the classification using different approaches: (a) pixel-based SVM, (b) standard pixel-based binary encoding, (c) region based SVM using spectral information only, (d) region-based binary encoding using spectral information only, (e) region-based SVM using spectral and spatial information. (f) proposed new method.

classification already performed quite well, and shows almost the same classification accuracy (73.1 %) as the region-based SVM classification (73.2 %). The SVM classifier associated with shape and height information reached virtually the same classification accuracy (72.0 %). Similar trends can be observed for the results of the built-up areas, i. e., not considering the 2/3 of the area covered by cropland and grass. It should be noted that for all methods the accuracy figures as such are not extremely high, which could be a consequence of the Hughes phenomenon (Hughes 1968, see also Hsu 2007). Improvements are possible either by increasing the number of training samples or by reducing the number of employed channels.

With the exception of pixel-based binary encoding the reported accuracy figures are relatively similar, however, the computing times are not. The reason is the grid search to find suitable parameters for the SVM training (see Tab. 3). The grid search accounts for 70 to 90 % of the whole computing time and thus significantly slows down SVM classifiction. From the performance point of view, the newly developed binary encoding method thus has clear advantages.

Overall, the binary encoding result is consistent with our expectations: Ground objects with similar spectra do not always belong to the same class. For example, if the classification only considers the spectrum, a building with a concrete roof would probably be classi-

Tab. 3: Classification results.

Classification method		SVM		Binary encoding		SVM, only spectral		Binary encoding, only spectral		SVM, spectral and spatial		Proposed new method	
		Pixel based				Region based							
(see Fig. 7)		a		b		c		d		e		f	
Accuracy (%)		U[a]	P[b]	U[a]	P[b]	U[a]	P[b]	U[a]	P[b]	U[a]	P[b]	U[a]	P[b]
SP		18	64	7	14	23	60	36	92	46	89	50	92
SC		64	68	19	24	55	66	51	59	59	75	60	60
S		54	31	29	8	56	21	38	32	58	33	73	25
R		73	74	59	74	66	94	73	94	87	94	76	94
TG		75	77	4	71	69	73	78	79	75	77	77	79
F		26	25	2	4	34	38	5	22	28	30	24	24
T		87	73	32	77	81	78	74	79	84	75	74	79
B		86	67	79	27	61	55	69	60	83	67	76	68
I		25	32	24	8	33	27	42	35	42	49	41	43
C		83	85	58	40	84	82	80	77	72	86	82	85
G		82	81	66	74	78	87	81	75	80	72	82	85
Accumulated accuracy values	Overall [%]	73.1		50.7		73.2		69.0		72.0		76.0	
	kappa index	0.641		0.348		0.638		0.596		0.627		0.677	
	Overall[c] [%], built-up areas	67.0		36.1		68.2		67.5		70.7		69.9	
Computing time [min]	segmentation	-		-		0’58		0’58		0’58		0’58	
	grid search	18’12		-		7’18		-		7’24		-	
	classification	3’31		5’33		1’21		1’12		1’43		1’21	
	total	21’43		5’33		9’37		2’10		10’05		2’10	

[a] U: User’s accuracy [b] P: Producer’s accuracy

[c] Accuracy in build-up areas: Refers to the classification accuracy except Cropland and Grass

SP: Stockyard or parking lot; SC: Surface sealed with concrete; S: Street; R: Runway; TG: Tennis ground; F: Farmhouse; T: Mixed trees; B: Industrial buildings; I: Industrial area; C: Cropland; G: Grass

fied as Surface sealed with concrete, while a building covered by green plants would probably be classified as Grass or Cropland. In our method, with the help of height information, this kind of mistakes can be minimized. Taking the class Street as another example, the spectrum of road and the parking lot adjacent to the road are often very similar. If we do not consider the regular shape of the parking lot and the length of a road, these two types may readily be confused during classification, whereas shape information can help to improve the results. Obviously, this general line of argumentation should also hold for SVM classification; however we did not observe any related improvements in our experiments. At this point the reason for this observation is not clear, finding an answer will be part of our future research.

4 Conclusions

Based on the idea that integrating multi-source remote sensing data may improve automatically derived interpretation results, a new binary encoding classification method was presented. In this method regions rather than individual pixels are considered as basic units and size, shape, and height information for each segment are integrated into the classification. The information itself is encoded into a 280 bit code, which makes the method useful for hyperspectral image data.

The results obtained from our study area show that incorporating region-based information and also spatial additions significantly improve binary encoding, and that the new method, while somewhat suffering from a lack of representative training data, achieves higher accuracy and efficiency than the SVM classification. This is particularly interesting, as in binary encoding a portion of the available radiometric information is knowingly sacrificed in the coding step, whereas the SVM makes use of the full radiometric resolution of the image data.

In future, more analysis on refining the selection of shape descriptors and the similarity evaluation algorithm is necessary, as well as investigations into band selection and feature reduction to better cope with the Hughes phenomenon. We also plan a more comprehensive comparative experiment in a larger study area, and to introduce texture measures into our new method.

Acknowledgements

Special thanks go to the hyperspectral research group at DLR, Oberpfaffenhofen, especially Dr. Andreas Müller and Wieke Heldens, who enabled this work by supplying the hyperspectral data and the DSM. Thanks are also due to Dipl.-Ing. Ulla Wißmann and Dipl.-Ing. Adelheid Elmhorst, both from IPI, who generated the reference map.

This study was substantially supported by the China Scholarship Council, National Natural Science Foundation of China (Project No. 40771174), National High Technology Research and Development Program of China (863 Program No. 2007AA12Z178 and 2009AA12Z131), Key Laboratory of Advanced Engineering Surveying of SBSM (Project No. TJES0810)*, and grants from the Program for Young Excellent Talents in Tongji University (Project No. 2009KJ094).

References

Acharyya, M., De, R.K. & Kundu, M.K., 2003: Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework. – IEEE Transactions on Geoscience and Remote Sensing **41** (12): 2900–2905.

Anderson, J.R., Hardy, E.E., Roach, J.T. & Witmera, R.E., 1976: Land Use and Land Cover Classification System For Use With Remote Sensor Data. – U.S. Geological Survey Professional Paper **964**; USGS, Washington, D.C.

Ball, J.E. & Bruce, L.M., 2007: Level set hyperspectral image classification using best band analysis. – IEEE Transactions on Geoscience and Remote Sensing **45** (10): 3022–3027.

Bazi, Y. & Melgani, F., 2006: Toward an optimal SVM classification system for hyperspectral remote sensing images. – IEEE Transactions on Geoscience and Remote Sensing **44** (11): 3374–3385.

Benz, U., Hofmann, P., Willhauck, G., Lingenfelder, I. & Heynen, M., 2004: Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. – ISPRS Journal of Photogrammetry and Remote Sensing **58** (3–4): 239–258.

Bolstad, P.V. & Lillesand, T.M., 1991: Rapid maximum likelihood classification. – Photogrammetric Engineering & Remote Sensing **57** (1): 64–74.

Bruce, L.M. & Li, J., 2001: Wavelets for computationally efficient hyperspectral derivative analysis. – IEEE Transactions on Geoscience and Remote Sensing **39** (7): 1540–1546.

Burges, C.J.C., 1998: A tutorial on Support Vector Machines for pattern recognition. – Data Mining and Knowledge Discovery **2** (2): 121–167.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Frances, J. & Calpe-Maravilla, J.,

* Kwang-Hua Fund for College of Civil Engineering, Scientific Research Foundation of Key Laboratory for Land Environment and Disaster Monitoring of SBSM (Grant No. LEDM2010B02).

2006: Composite kernels for hyperspectral image classification. – IEEE Geoscience and Remote Sensing Letters **3** (1): 93–97.

Chang, C.I., 2003: Hyperspectral imaging: spectral detection and classification. – Kluwer Academic/Plenum Publishers, New York, US.

Chiou, W.C., 1984: Dynamic descriptors for contextual classification of remotely sensed hyperspectral image data analysis. – Applied Optics **23** (21): 3889–3892.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I. & Shields, T., 1998: The Hymap™ Airborne Hyperspectral Sensor: The System, Calibration and Performance. – 1st EARSeL Workshop on Imaging Spectroscopy, Paris, 37–42.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F. & Cho, M., 2008: LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. – ISPRS Journal of Photogrammetry and Remote Sensing **63** (4): 409–426.

Demir, B. & Erturk, S., 2008: Phase correlation based redundancy removal in feature weighting band selection for hyperspectral images. – International Journal of Remote Sensing **29** (6): 1801–1807.

Douglas, D.H. & Peucker, T.K., 1973: Algorithms for the reduction of the number of points required to represent a line or its caricature. – The Canadian Cartographer **10** (2): 112–122.

Duarte-Carvajalino, J.M., Sapiro, G., Velez-Reyes, M. & Castillo, P.E., 2008: Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods. – IEEE Transactions on Geoscience and Remote Sensing **46** (8): 2418–2434.

Feingersh, T., Ben-Dor, E. & Filin, S., 2010: Correction of reflectance anisotropy: a multi-sensor approach. – International Journal of Remote Sensing **31** (1): 49–74.

Gamba, P. & Houshmand, B., 2000: Hyperspectral and IFSAR data for 3D urban characterization. – IGARSS 2000, Honolulu, CD-ROM.

Greiwe, A. & Ehlers, M., 2005: Combined analysis of hyperspectral and high resolution image data in an object oriented classification approach. – 3rd International Symposium on Remote Sensing and Data Fusion over Urban Areas, 13–15.

Hamming, R.W., 1950: Error-detecting error-correcting codes. – Bell System Technical Journal **29:** 147–160.

Heiden, U., Segl, K., Roessner, S. & Kaufmann, H., 2007: Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. – Remote Sensing of Environment **111** (4): 537–552.

Hepner, G.F., Houshmand, B., Kulikov, I. & Bryant, N., 1998: Investigation of the Integration of AVIRIS and IFSAR for Urban Analysis. – Photogrammetric Engineering & Remote Sensing **64** (8): 813–820.

Huang, X., Zhang, L. & Li, P., 2007: Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. – IEEE Geoscience and Remote Sensing Letters **4** (2): 260–240.

Hughes, G.F., 1968: On the mean accuracy of statistical pattern recognizers. – IEEE Transactions on Information Theory **IT-14** (1): 55–63.

Hsu, P.H., 2007: Feature extraction of hyperspectral images using wavelet and matching pursuit. – ISPRS Journal of Photogrammetry and Remote Sensing **62** (2): 78–92.

Jia, X. & Richards, J.A., 1993: Binary coding of imaging spectrometer data for fast spectral matching and classification. – Remote Sensing of Environment **43:** 47–53.

Jia, X. & Richards, J.A., 2008: Managing the spectral-spatial mix in context classification using Markov random fields. – IEEE Geoscience and Remote Sensing Letters **5** (2): 311–314.

Kartikeyan, B., Sarkar, A. & Majumder, K.L., 1998: A segmentation approach to classification of remote sensing imagery. – International Journal of Remote Sensing **19** (9): 1695–1709.

Lee, C. & Landgrebe, D.A., 1991: Fast likelihood classification. – IEEE Transactions on Geoscience and Remote Sensing **29** (4): 509–517.

Le Moigne, J. & Tilton, J.C., 1995: Refining image segmentation by integration of edge and region data. – IEEE Transactions on Geoscience and Remote Sensing **33** (3): 605–615.

Marconcini, M., Camps-Valls, G. & Bruzzone, L., 2009: A composite semisupervised SVM for classification of hyperspectral images. – IEEE Geoscience and Remote Sensing Letters **6** (2): 234–238.

Martin, M.E., Plourde, L.C., Ollinge, S.V., Smith, M.-L. & McNeil, B.E., 2008: A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. – Remote Sensing of Environment **112** (9): 3511–3519.

Martinez, P.J., Perez, R.M., Plaza, A., Aguilar, P.L., Cantero, M.C. & Plaza, J., 2006: Endmember extraction algorithms from hyperspectral images. – Annals of Geophysics **49** (1): 93–101.

Martinez-Uso, A., Pla, F., Sotoca, J.M. & Garcia-Sevilla, P., 2007: Clustering-based hyperspectral band selection using information measures. – IEEE Transactions on Geoscience and Remote Sensing **45** (12): 4158–4171.

Mazer, A.S., Martin, M., Lee, M. & Solomon, J.E., 1988: Image processing software for imaging spectrometry data analysis. – Remote Sensing of Environment **24:** 201–210.

Mehtre, B.M., Kankanhalli, M.S. & Lee, W.F., 1997: Shape measures for content based image retrieval: A comparison. – Information Processing and Management **33** (3): 319–337.

Mohanty, N., Rath T.M., Lee, A. & Manmatha, R., 2005: Learning Shapes for Image Classification and Retrieval. – Lecture Notes in Computer Science **3568:** 589–598.

Mumford, D. & Shah, J., 1985: Boundary detection by minimizing functionals. – IEEE Conference on Computer Vision and Pattern Recognition **I:** 22–26.

Nascimento, J.M.P. & Dias, J.M.B., 2005: Vertex component analysis: a fast algorithm to unmix hyperspectral data. – IEEE Transactions on Geoscience and Remote Sensing **43** (4): 898–910.

Prasad, S. & Bruce, L.M., 2008: Decision fusion with confidence-based weight assignment for hyperspectral target recognition. – IEEE Transactions on Geoscience and Remote Sensing **46** (5): 1448–1456.

Rabe, A., van der Linden, S. & Hostert, P., 2009: imageSVM, Version 2.0, software available at www.hu-geomatics.de.

Rajan, S., Ghosh, J. & Crawford, M.M., 2008: An active learning approach to hyperspectral data classification. – IEEE Transactions on Geoscience and Remote Sensing **46** (4): 1231–1242.

Robinson, D.J., Redding, N.J. & Crisp, D.J., 2002: Implementation of a fast algorithm for segmenting SAR imagery. – Scientific and Technical Report, Australia: Defense Science and Technology Organization.

Rogge, D.M., Rivard, B., Zhang, J., Sanchez, A., Harris, J. & Feng, J., 2007: Integration of spatial-spectral information for the improved extraction of endmembers. – Remote Sensing of Environment **110** (3): 287–303.

Russ, J.C., 2006: The image Processing Handbook, Fifth edition. – Boca Raton, FL: CRC Press.

Serpico, S.B. & Moser, G., 2007: Extraction of spectral channels from hyperspectral images for classification purposes. – IEEE Transactions on Geoscience and Remote Sensing **45** (2): 484–495.

Smola, A. & Schoelkopf, B., 1998: A Tutorial on Support Vector Regression. – Statistics and Computing **14** (3): 199–222.

Trias-Sanz, R., Stamon G. & Louchet, J., 2008: Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. – ISPRS Journal of Photogrammetry and Remote Sensing **63** (2): 156–168.

Vapnik, V.N., 1998: Statistical Learning Theory. – Wiley, New York, NY.

Verhoef, W. & Bach, H., 2007: Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. – Remote Sensing of Environment **109** (2): 166–182.

Viterbi, A. J. & Omura, J.K., 1979: Principles of Digital Communication and Coding. – McGraw-Hill, New York, NY: 81.

Zhang, L., Huang, X., Huang, B. & Li, P., 2006: A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. – IEEE Transactions on Geoscience and Remote Sensing **44** (10): 2950–2961.

Addresses of the Authors:

Dr. Huan Xie, Prof. Dr. Xiaohua Tong, Tongji University, Department of Surveying and Geo-informatics, Siping Road No.1239, 200092, Shanghai, P.R. China. Tel.: +86-21-6598-3686, -8851, Fax: -1085, e-mail: huanxie@tongji.edu.cn, xhtong@tongji.edu.cn.

Prof. Dr.-Ing. Christian Heipke, Dr.-Ing. Peter Lohmann, Prof. Dr.-Ing. Uwe Soergel, Leibniz Universtät Hannover, Institut für Photogrammetrie und GeoInformation (IPI), Nienburger Str. 1, D-30167 Hannover, Germany, Tel: +49-511-762-2482, -2486, -2981, Fax: -2483, e-mail: heipke@ipi.uni-hannover.de, lohmann@ipi.uni-hannover.de, soergel@ipi.uni-hannover.de.

Prof. Dr. Wenzhong Shi, Hong Kong Polytechnic University, Department of Land Surveying and Geo-Informatics, Hung Hom, Kowloon, Hong Kong. Tel.: +852-2766-5975, Fax: +852-2330-2994, e-mail: lswzshi@polyu.edu.hk.

Manuskript eingereicht: Januar 2010
Angenommen: November 2010