Landscape similarity, retrieval, and machine mapping of physiographic units

Jaroslaw Jasiewicz[b,a], Pawel Netzel[c,a], Tomasz F. Stepinski[a]

[a] *Space Informatics Lab, Dept. of Geography, University of Cincinnati, Cincinnati, OH 45221-0131, USA*
[b] *Geoecology and Geoinformation Institute, Adam Mickiewicz University, Dziegielowa 27, 60-680 Poznan, Poland*
[c] *Dept. of Climatology and Atmospheric Protection, University of Wroclaw, Kosiby 6/8, 51-621, Wroclaw, Poland*

Abstract

We introduce landscape similarity - a numerical measure that assesses affinity between two landscapes on the basis of similarity between the patterns of their constituent landform elements. Such a similarity function provides core technology for a landscape search engine - an algorithm that parses the topography of a study area and finds all places with landscapes broadly similar to a landscape template. A landscape search can yield answers to a query in real time, enabling a highly effective means to explore large topographic datasets. In turn, a landscape search facilitates auto-mapping of physiographic units within a study area. The country of Poland serves as a test bed for these novel concepts. The topography of Poland is given by a 30 m resolution DEM. The geomorphons method is applied to this DEM to classify the topography into ten common types of landform elements. A local landscape is represented by a square tile cut out of a map of landform elements. A histogram of cell-pair features is used to succinctly encode the composition and texture of a pattern within a local landscape. The affinity between two local landscapes is assessed using the Wave-Hedges similarity function applied to the two corresponding histograms. For a landscape search the study area is organized into a lattice of local landscapes. During the search the algorithm calculates the similarity between each local landscape and a given query. Our landscape search for Poland is implemented as a GeoWeb application called TerraEx-Pl and is available at http://sil.uc.edu/. Given a sample, or a number of samples, from a target physiographic unit the landscape search delineates this unit using the principles of supervised machine learning. Repeating this procedure for all units yields a complete physiographic map. The application of this methodology to topographic data of Poland results in the delineation of nine physiographic units. The resultant map bears a close resemblance to a conventional physiographic map of Poland; differences can be attributed to geological and paleogeographical input used in drawing the conventional map but not utilized by the mapping algorithm.

Keywords: Landscape similarity, Landscape search, Physiographic mapping, Pattern recognition, Supervised classification, Web application

1. Introduction

Regionalization and mapping are the core elements of geomorphologic analysis. Traditionally, these tasks are carried out by analysts who rely on their visual perception of data and expert knowledge to delineate units of land surface within a given study area. Possible target units of mapping include – in order of increasing complexity – landform elements, landforms and landscapes (see Minar and Evans (2008) for a description of the hierarchical partitioning of land surfaces). With the increasing availability of medium-to-high resolution DEMs covering the entire land surface of the Earth as well as surfaces of other planets and because of the slowness, expense, and subjectivity of manual analysis, there is a significant interest in automating the process of geomorphologic mapping.

In this paper we present a novel methodology for the automated delineation of landscape types within a study area. To the best of our knowledge no previous work has addressed this issue directly by taking into account the complexity of landscape units as described, for example, by Minar and Evans (2008). Instead, previous work concentrated on the automatic classification of landforms – surface units of lesser complexity then landscapes. In practice, however, the methods employed in previous works tended to generalize the notion of "landform" to the point where the resultant maps (Iwahashi and Pike, 2007; Dragut and Eisank, 2012) delineate units that could be best described as physiographic units. Therefore, we will be able to compare the results of our mapping methodology with the results of previous auto-mapping techniques.

All previous methods share a common framework. They are classification schemes that assign a label to an areal unit on the basis of geomorphometric variables (Evans, 1972; Pike, 1988; MacMillan et al., 2004; Olaya, 2009) and/or their statistics calculated from DEM values at a given unit and/or from its immediate neighborhood. The first such classification scheme was devised by Hammond (1954) and was later implemented as a computer algorithm (Dikau et al., 1991; Gallant et al., 2005). Other landform classification schemes were proposed by Meybeck et al. (2001) and Iwahashi and Pike (2007) using different combinations of geomorphometric variables. Recently, Dragut and Eisank (2012) introduced the concept of Object-Based Image Analysis (OBIA) to classification of landforms. In their method a DEM is first segmented into multi-cell units which are homogeneous with respect to geomorphometric variables, and those units, rather then DEM cells, are the objects of classification.

The approach presented in this paper is based on different principles. We start with the concept of similarity between landscapes. Using this concept we design a computational framework for a landscape search and for auto-mapping of landscape types or physiographic units. According to the taxonomy of Minar and Evans (2008) landscapes are patterns of landforms which in turn are composites of landform elements. We skip the middle level of this hierarchy and consider landscape to be a pattern of landform elements over a site of interest. A similarity between two landscapes is defined as a single number that encapsulates all aspects of compositional and configurational alikeness between two patterns of landform elements.

Despite the great variability of local landscapes within a study area (a landscape at any specific site is unique in its details), there are a limited number of semantically different landscape types that can be discerned. We consider landscape types to be tantamount to physiographic units – regions of the study area having internal uniformity of landscape and clearly different from surrounding regions. A measure of similarity between landscapes enables the algorithmic identification of landscape types. The landscape search engine is an algorithm which, given a sample landscape (a query), parses the entire study area and retrieves sites having landscapes similar to that of the query. The set of all retrieved landscapes constitutes the landscape type exemplified by the query. An auto-mapper of physiographic units is an algorithm which delineates a study area into an exclusive and exhaustive set of physiographic regions.

Note that an auto-mapping algorithm that utilizes our framework could be based on the machine learning principles of either unsupervised learning (Duda et al., 2001) or supervised learning (Mehryar et al., 2012). An unsupervised learning algorithm delineates physiographic units without any guidance from an analyst by clustering similar landscapes. The number and character of these units emerge from the data and need to be interpreted afterward. An unsupervised learning mapping approach is most useful for the exploration of a study area with little prior knowledge about its physiography, like, for example, a planetary surface (Bue and Stepinski, 2006). A supervised learning algorithm delineates study area into an a priori known set of units on the basis of landscape samples provided by an analyst. A supervised approach is most useful when there is some prior knowledge about the physiography of a study area but objective delineation of units is desired. Note that the previous auto-mapping methods mentioned above are often referred to as "unsupervised" because they require no interaction between an algorithm and an analyst. However, they are not based on either supervised or unsupervised machine learning principles. They classify cells/segments into a priori defined landform types (a supervised aspect) but numerical criteria for belonging to a given type depends on the statistics of the data (an unsupervised aspect).

In this paper we focus on a supervised variant of our auto-mapping algorithm with the delineation of physiographic units achieved by repeated application of the landscape search algorithm. The methodology presented here is general and applies to any study area for which a DEM of sufficient quality is available. We illustrate the steps in our method using an entire territory of the country of Poland (represented by a 30 m resolution DEM) as a study area.

2. Analytical and computational framework

Because our methodology consists of several components, we start by describing its overall framework – a logical structure of several separate concepts and their computational implementations that together underpin our approach to landscape retrieval and mapping.

A schema of our analytical framework is shown in Fig. 1. The topography of a study area (Fig. 1A) is used as input data. Because we are concerned with the search for and mapping of spatially extensive areal units (of the size of physiographic units), a study area would typically cover a region which is very large in comparison to the resolution of a DEM. In this paper we consider

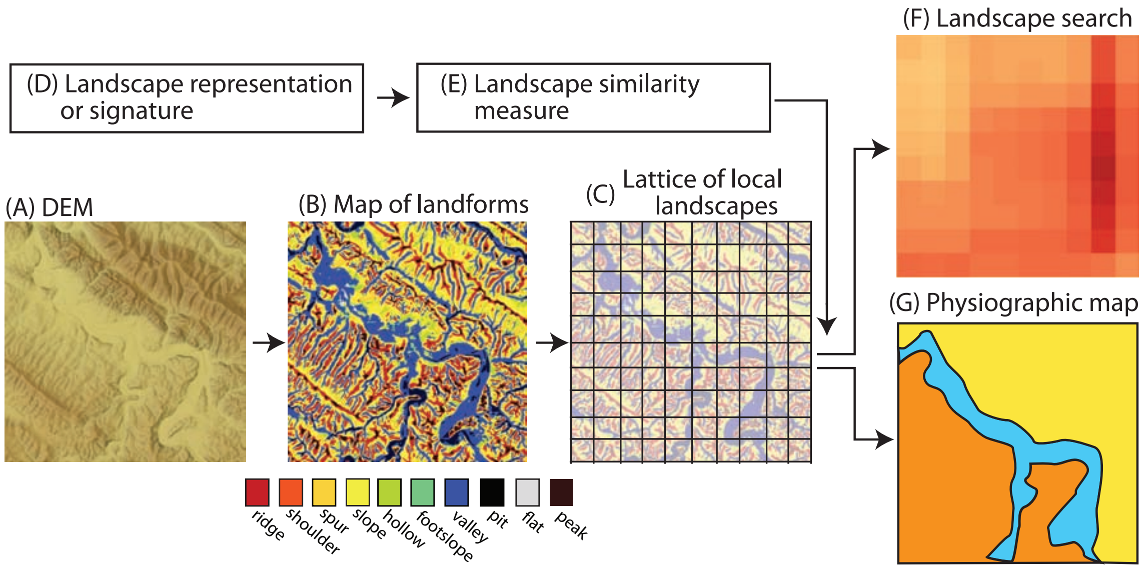

Figure 1: Schema showing an analytic framework of our methodology.

a study area containing the entire country of Poland at 30 m resolution (see Section 3 for details). The first element of our method is an automatic mapping of landform elements from a DEM (A→B transition on Fig. 1). This step could be achieved using several different methods (Wood, 1996; Dikau et al., 1995; Jasiewicz and Stepinski, 2013a) developed to classify DEM cells into a small number of categorical labels indicating an elementary form of a local surface. Extending our previous work we use the geomorphons method (Jasiewicz and Stepinski, 2013a) that allows for a direct, single-step classification of landform elements. The geomorphons method provides a fast and robust tool for achieving the A→B transition. It classifies DEM cells into the ten most common landform elements: flat, peak, pit, ridge, valley, shoulder, footslope, spur, and hollow (Fig. 1B). We have computed 30 m resolution maps of landform elements using the geomorphons method for Poland and, additionally for the United States. These maps can be explored and compared to a hillshade rendition of topography using our GeoWeb tools available at http://sil.uc.edu/.

The second element of our method is the conversion of a map of landform elements into a lattice of local landscapes (B→C transition on Fig. 1). We operationally define a local landscape as a square-shaped tile cut out of the map of landform elements. The size of a tile should be large enough so that local landscapes contain non-trivial mosaics of landform elements, but small enough to ensure a diversity of landscape types in the study area. The tiles are arranged in a lattice of local landscapes and together they cover the entire study area (Fig. 1C).

An overall, quantitative measure of similarity between two landscapes is the key concept of our methodology. To the best of our knowledge this concept has not been discussed with respect to its application to geomorphology. However, it has been studied in the context of landscape ecology (Wickham and Norton, 1994; Allen and Walsh, 1996; Cain et al., 1997) where the notion of landscape pertains to patterns of land use/land cover (LULC) categories rather than to the patterns of landform elements. There are two components of landscape similarity: (1) a concise numerical representation of landscape pattern hereafter referred to as a landscape signature (Fig. 1D) and (2) a similarity function (Fig. 1E) that uses this representation to calculate a number that encapsulates the overall degree of "alikeness" or affinity between two landscapes. In landscape ecology, a signature is a vector of landscape indices (O'Neill et al., 1988; Herzog and Lausch, 2001) and the Euclidean distance is used as the similarity function. Our choices for the landscape signature and similarity function are different from those used in the LULC context because the pattern characteristics of landform elements are different from those of LULC patterns (see details in Section 4).

The landscape search (Fig. 1F) utilizes a query-and retrieval technique to find all local landscapes similar to a sample landscape (also referred to as a "query").

A query doesn't have to be one of the local landscapes predefined by a lattice of tiles, and it doesn't have to be taken from the study area, however, in this paper all queries are samples from the study area. The search is performed by calculating the similarity between a query and each of the local landscapes. The result of this search is a "similarity map" (Fig. 1F) with locations colored in accordance with their similarity to a query. A landscape type exemplified by a query can be delineated as a set of all locations having a similarity to the query which is larger than a specified threshold.

Using several different templates and a repeated application of the landscape search we partition the study area into a set of physiographic units (Fig. 1G) corresponding to landscape types exemplified by respective templates. Using each template as a query we obtain a set of similarity maps, which, when reconciled, yield a single map of physiographic units (for details see Section 6).

3. Study area

Our study area is the country of Poland. Poland is a country in Central Europe located between latitudes 49° and 55° N and longitudes 14° and 25° E. Its surface area is 312,685 km^2. The territory of Poland exhibits a number of different landscape types including coastal plains in the northernmost part of the country, young–undenudated and old–denuded post-glacial lowlands of different ages in the northern and central parts, and uplands and mountains in the southern part. A color rendition of Poland's topography is shown in Fig. 2A.

A traditional physiographic map of Poland (hereafter referred to as a "reference" map) has been obtained by generalizing a physigraphic regionalisation of Poland proposed by Kondracki (2002). The original Kondracki map has been created manually on the basis of geomorphological as well as geological and paleogeographical information and includes many units, some of them delineated on the basis of regional position. The reference map shown in Fig. 1B delineates only 12 physiographic units as described by its legend (Fig. 1C). They include surfaces formed during the last glaciaction (young morainic hills, young moraininc plateaus and plains) and surfaces resulting from previous glaciation morphogenesis strongly modified by denudational process during the last glaciation (old morainic plateaus and hills, old morainic plains and old moraninc plains on older substratum). These units are frequently bundled together as "lowlands" on less specific maps or by global automatic classifiers based on geomorphometric variables like those by Iwahashi and Pike (2007) or Dragut and Eisank (2012). On the other hand, our landscape similarity-based methodology would be able to map these units quite well using only topographic data and without the benefit of additional information from geology and paleogeography.

The topographic data for Poland is a 1" integer-valued DEM (obtained from the Silesia University) which we reprojected to the PUWG92 coordinate system and converted by adaptive smoothing to a floating-point terrain model with a resolution of 30 m/cell. Our final DEM is 21,696 × 24,692 cells. The 30 m/cell categorical map of landform elements is calculated from this DEM using the geomorphons method (Jasiewicz and Stepinski, 2013a). This map was calculated using the following values for the two parameters required by the geomophons method: Search radius L =40 cells (1200 m), and Flatness threshold t =0.8 degree. The geomorphons code is available for download from http://sil.uc.edu/. The hillshade and the shaded relief of the DEM, as well as the map of landforms elements, are available for exploration using the GeoWeb tool TerraEx-Pl available from the same website.

4. Landscapes similarity

Being able to quantify the overall similarity between two landscapes using a single number is a key element of our methodology. In both landscape ecology and geomorphology landscapes can be considered as categorical spatial patterns, suggesting that we can apply similarity measures developed (Cain et al., 1997; Long et al., 2010; Kupfer et al., 2012) for LULC landscapes to patterns of landscape elements. However, this is not the case as these two types of patterns have different characteristics. While the most important discriminant between two LULC patterns is the presence or absence of specific LULC categories, most terrain patterns contain all categories of landscape elements. This follows from the fact that terrain consists of a series of landform elements, for example: valley, slope, ridge, slope, valley, etc. Thus, while LULC landscapes are predominantly distinguished from each other on the basis of land cover composition, terrain landscapes are predominantly distinguished from each other on the basis of their texture or the spatial configuration of their basic elements. In addition, the nature of land surfaces dictates that landscape patterns are dominated by just two elementary forms: "flats" in the lowlands and "slopes" in areas of higher relief. All other elementary forms, such as peaks, pits, ridges, valleys, footslopes, hollows, spurs, and shoulders are less abundant, but they are nevertheless crucial for characterizing terrain texture. In

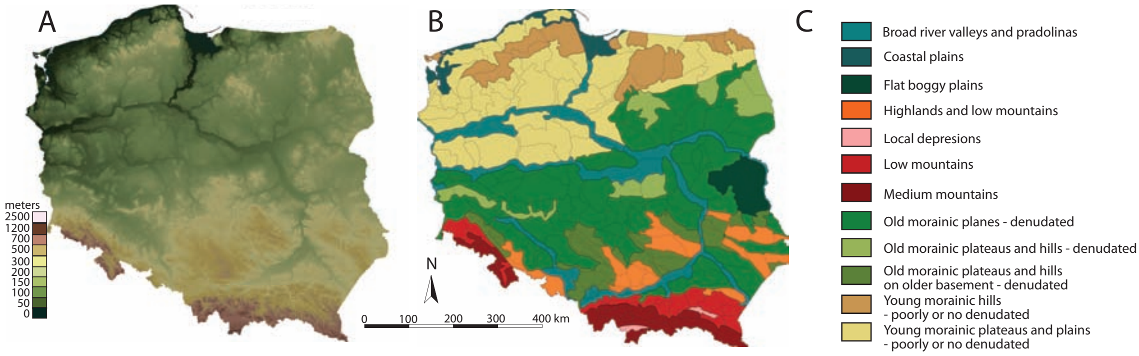

Figure 2: Topography and physiography of Poland. (A) Topographic map of Poland. (B) Reference physiographic map of Poland based on the Kondracki (2002) concept. (C) Legend for the reference physiographic map.

designing a quantitative measure of similarity between landscapes the contribution of these landform elements must be enhanced relative to their abundance.

In designing a landscape numerical signature we follow principles established in the field of Content-Based Image Retrieval (CBIR) (Gevers and Smeulders, 2004; Datta et al., 2008; Lew et al., 2006) – another domain where the issue of similarity between two rasters, in this case images, has been studied. In CBIR a pattern signature is calculated as a histogram of a pattern's "primitive features." A histogram is a good choice for pattern signature because of its rotational invariance; two patterns rotated with respect to each other will have identical histograms. Primitive features are simple measures designed to provide small, local pieces of information about a pattern. For example, a landform element at a given cell could be a primitive feature, and a histogram of all landform elements over the landscape could be a landscape signature. Such signature would, however, reflect only the abundance of landform elements in the landscape and would not characterize the landscape well enough (see discussion above) for effective comparison with other landscapes.

We use pairs of neighboring cells as primitive features as shown in Fig. 3A. This figure shows a small (8×8 cell) map of landform elements. Each cell in this map generates eight pairs which are shown by arrows in the case of three cells selected as examples (an 8-connected neighborhood is assumed). For example, the leftmost of the three example cells shown on Fig. 3A is labeled as "slope" and generates eight pairs: three slope-slope pairs, four slope-footslope pairs and one slope-channel pair. If a map of landform elements maps C different elements, there are $(C^2 + C)/2$ different possible types of neighboring cells. Because the geomorphons-generated map has 10 different elements, there are 55 different possible pairs of elements, examples include: flat-flat, flat-slope, slope-peak, etc.

The design of the landscape signature (first described by Barnsley and Barr (1996) in the context of land use reclassification) simultaneously encodes the composition and texture of the landscape in a simple histogram. (Note that this histogram contains the same information as a co-occurrence matrix (Haralick, 1986)). Fig. 3B illustrates how two different landscapes (depicted by their maps of landform elements) are represented by histograms of cell-pair features. The legend for Fig. 3A also applies to the maps shown in Fig. 3B. Histograms of the cell-pair features have 55 bins, with each bin height proportional to a number of cell-pairs belonging to a cell-pair category as indicated by color labels shown between the two histograms. The highest bins usually correspond to pairs of same category cell-pairs; they encode the composition of the landscape. For example, histogram of landscape 1 is dominated by flat-flat, slope-slope, and valley-valley bins, whereas the histogram of landscape 2 is dominated by slope-slope, valley-valley, and ridge-ridge bins. The bins corresponding to pairs of different cell categories encode the texture of the landscape.

The similarity between two landscapes is calculated using their signatures (histograms of features) and a similarity (or a distance) function. Note that a "distance" is a measure of dissimilarity between two landscapes and is thus the reverse of similarity. Choosing most appropriate similarity/distance function is largely an empirical decision. Cha (2007) provides a comprehensive review of possible functions to calculate the distance between two histograms. After extensive experimentation with different similarity measures we have observed that a modified Wave-Hedges similarity function (Cha, 2007) measures similarity between land-

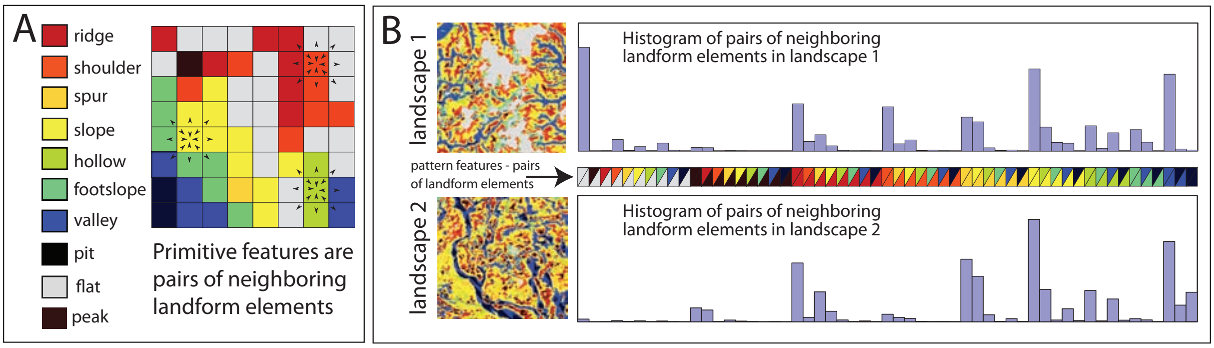

Figure 3: Landscape numerical signature. (A) Graph explaining pairs of neighboring landform elements as pattern primitive features. (B) Example of two landscapes and their corresponding signatures - histograms of primitive features.

scapes in a way that is most consistent with human perception. Let A and B denote two landscapes and A^h and B^h denote their corresponding signatures (histograms). The similarity between these two landscapes is given by the following formula:

$$\mathrm{sim}(A, B) = \mathrm{sim}(A^h, B^h) = \frac{1}{M} \sum_{i=1}^{M} \frac{\min(A_i^h, B_i^h)}{\max(A_i^h, B_i^h)} \quad (1)$$

where M is the number of positions in both histograms where at least one bin is non-zero and A_i^h and B_i^h are the values of bins in the i-th position. This measure takes into account only the cell-pair features that are present in at least one of the two landscapes. It compares each pair of corresponding bins separately by dividing a smaller bin value by its bigger counterpart. The result is a number between 0 and 1. Where the value of one of the bins is 0, there is no similarity with respect to this cell-pair feature between the two landscapes. If the values of both bins are identical, there is perfect similarity with respect to that cell-pair feature between the two landscapes. The overall similarity is an arithmetic average of all contributing similarities, its range is between 0 (no similarity between landscapes) and 1 (landscapes are identical).

Note that in Eqn. (1) the contributions of all cell-pair features to an overall similarity value are taken with the same weight regardless of each feature's abundance in the landscape. This ensures that composition-related features (pairs of same category cells) and texture-related features (pairs of different category cells) have the same chance to contribute to the overall similarity value despite the dominance of composition-related features in all realistic histograms (see Fig. 3B). It also ensures that landscape similarity will not be heavily skewed by the relative abundance of the most common landscape elements – flat and slope. Other potential similarity functions, like the Euclidean distance or the Jensen-Shannon distance, are dominated by similarity of the most abundant features. As a result, when those measures are applied to, for example, two landscapes both dominated by the flat element but with different secondary elements, they would yield a high value of similarity by focusing on the fact that both landscapes are basically flat. Frequently, this result will not correspond to the perception of an analyst for whom different departures from flat terrain are associated with significant dissimilarity between the two landscapes. However, the application of the similarity function given in Eqn. (1) to this example would result in lower value of similarity, more in agreement with how a human analyst would determine the similarity between those two landscapes.

5. Landscape search

The purpose of a landscape search engine is to enable the discovery of locations containing landscapes similar to a specified landscape of interest. Like other more familiar search engines it works on the principle of query-and-retrieval. However, the spatial aspect of topographic data calls for the presentation of search results in a manner that is different from those employed in non-spatial search engines. For example, an Internet search engine returns only several results (web pages, images) which represent the "best-fit" results to a given query. In contrast, the output of a landscape search engine is a similarity map which visually shows a degree of similarity to a query at all locations throughout the entire study area, thus providing geospatial context.

Fig. 4 illustrates the principle of the landscape search. Fig. 4A shows the topographic map of Poland with a green-to-brown color gradient illustrating low-to-high

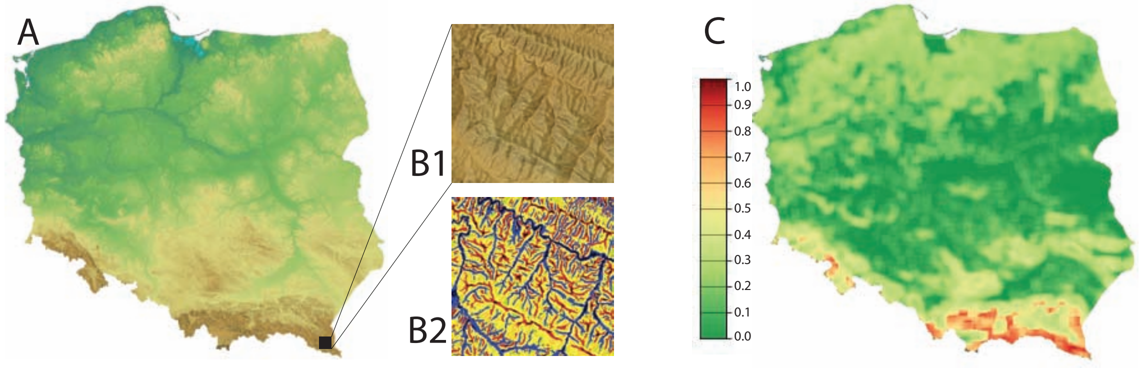

Figure 4: Concept of landscape search. (A) Topographic map of Poland. (B1) Shaded relief of query landscape. (B2) Map of landform elements of query landscape (see Fig. 5 for legend). (C) Similarity map. See main text for additional informations.

elevations. The small black square in the lower-right portion of the map indicates the location of a query. Fig. 4C shows the color-coded similarity values between that query and all local landscapes covering the entire territory of Poland. Locations on the similarity map which are shown in red are similar to the query; they are located along the southern border of Poland within a physiographic unit of medium mountains (see Fig. 2). Map locations shown in green are completely dissimilar to the query and correspond to lowlands. Map locations shown in various shades of yellow are also dissimilar to the query, but less so than the locations shown in green; they correspond to highlands and low mountains in the south and young morainic hills in the north.

Implementation of the landscape search in software follows our earlier design (Jasiewicz and Stepinski, 2013b; Stepinski et al., 2014) meant for searching the National Land Cover Database (NLCD) for U.S. locations having similar patterns of LULC. We use an overlapping sliding window approach. A square grid with a resolution of k raster cells is superimposed over the entire spatial extent of the study area. This grid forms a basis for the similarity map resulting from the query. Thus, the resultant similarity map has a resolution k times coarser than the map of landform elements. We will refer to cells of similarity map as "super-cells." The query is executed by means of exhaustive evaluation - the value of landscape similarity is calculated between the query tile (see section 2 to recall a definition of the tile) and all the local tiles assigned to the similarity grid.

The size of a tile is $N \times k$. If $N = 1$ the tiles are identical to super-cells and they don't overlap with neighboring tiles. However, in general, it is preferable to consider overlapping tiles ($N > 1$) to better accommodate the continuous character of landscape. In this paper we use $k = 50$ and $N = 10$ resulting in tile size of 15 km $\times$ 15 km and a similarity map with a resolution of 1.5 km. Fig. 5 illustrates the relationship between tiles and super-cells. Fig. 5A shows the 30 m resolution map of landform elements. The purple square denoted Q indicates a 15 km $\times$ 15 km tile containing the landscape to be used as a query. Fig. 5B shows the 1.5 km resolution similarity map. The clearly visible pixelation indicates super-cells; the color of each super-cell reflects the value of similarity between the query and the landscape contained in the 15 km $\times$ 15 km tile centered on this super-cell. The three examples of super-cell locations are labeled L1, L2, and L3 respectively. These super-cells are outlined by a thin dashed line and their corresponding tiles are outlined by a thick solid line. For example, a similarity value stored in the super-cell L1 indicates the similarity between the landscapes contained in tiles Q and L1.

In order to maximize the utility of the landscape search we have implemented it as a modern Internet application (called TerraEx-Pl) running in a web browser. TerraEx-Pl, available at http://sil.uc.edu/, is a computerized map application with all functionalities available through an active web page (as Google Maps). It enables real-time landscape searches to be performed over the country of Poland. The resultant similarity maps can be downloaded as GeoTiff files for offline analysis.

6. Delineation of physiographic units

We now demonstrate how to divide the territory of Poland into an exclusive and exhaustive set of physiographic units using a supervised learning approach that utilizes the landscape search method described in the previous section. Table 1 gives a description of

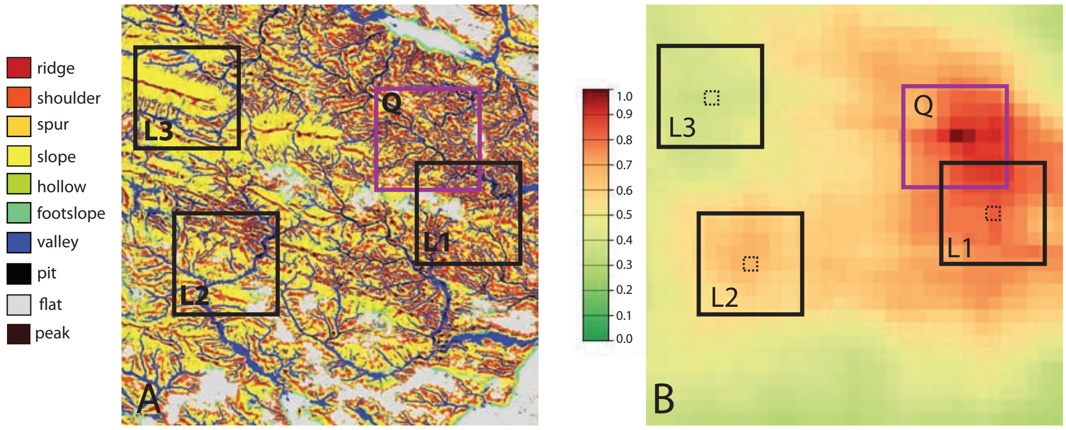

Figure 5: Relationship between tiles and super-cells. (A) Map of landform elements. (B) Similarity map to query Q. Purple square labeled Q indicates location and spatial extent of a query landscape. Black squares labeled L1, l2, and L3 indicate locations and spatial extents of sample local landscapes. See main text for more details.

nine physiographic units we have selected for mapping. They correspond to the 12 regions presented on our reference physiographic map (Fig. 2B). For auto-mapping we did not select regions from Fig. 2B which refer to their genesis rather than physiography as our method works (by our choice and design) only on the basis of landscape morphology alone. Thus, "broad river valleys and pradolinas", "costal plains", and "flat boggy plains" are combined into "flat plains", and "old morainic plains and hills" and "old morainic plains and hills on older basement" are combined into a single unit. Also "local depressions" are not selected as a unit to be mapped, but a new unit "inland dunes" is added.

Following the methodology of supervised learning we selected a number of template landscapes for each unit based on our expert knowledge. The third column in Table 1 shows a number of landscapes used as templates for a given unit. Altogether, 64 template landscapes where selected to represent the characteristic landscapes of various units. Locations of template landscapes and their samples are shown in Fig. 6.

We execute the landscape search with each template serving in turn as a query. The result is 64 similarity maps, each showing the spatial distribution of similarity to the given querying. It could be expected that similarity maps stemming from the set of queries representing a single unit should be very similar to each other. This is certainly the case for some sets of queries. For example, similarity maps for all 11 templates representing medium mountains (MM) are much the same. This means that landscape samples selected for medium mountains have all very similar terrain texture and that the entire unit of medium mountains has relatively uniform texture. On the other hand, similarity maps for 12 templates representing young morainic plateaus and plains (YMPP) show marked differences. This is because the set of YMPP landscape samples displays some variance in terrain texture. This follows from the fact the terrain texture across the YMPP unit is relatively less uniform than, for example, the terrain texture across the MM unit. The fact that some units display an internal variance of patterns is the reason for using a set of different landscape samples instead of a single sample to represent a "typical" landscape.

We average all similarity maps stemming from templates representing a single unit. Because the values on any similarity map range from 0 to 1, the average maps also have the same range of values and can be interpreted as a map showing the likelihood of local landscapes belonging to a given unit. The average similarity (likelihood) maps for all nine units are shown in Fig. 7A. The most distinct physiographic units are those for which likelihood maps are dominated by high (red) and low (green) values, thus clearly delineating a unit from the rest of the study area. Medium mountains (MM), low mountains (LM) and young morainic hills (YMH) are such units. A unit for which the likelihood map shows a lot of medium values marked by orange and yellow colors is less crisply defined. Our algorithm indicates that extended portions of the study area could be assigned to such unit but only with likelihood that is relatively small. Highlands (HL), young morainic plateaus and plains (YMPP), dunes (DN), old morainic plateaus and hills(OMPH), and old morainic plains (OMP) are examples of such units. The likelihood map for flat plains (FP) unit does not show high

Table 1: **Definitions of physiographic units**

Name	Abbreviation	# of samples	Description
Medium mountains	**MM**	11	Areas above 1000 m asl, medium dense texture and high, sharp relief, regular dendritic pattern, no flat areas.
Low mountains	**LM**	5	Areas between 500 and 1000 m asl, dense texture, medium and sharp relief, regular dendritic pattern, no flat areas, limited amount of slopes.
Highlands	**HL**	9	Areas between 300 and 500 m asl, dense texture, medium and high relief, regular dendritic pattern, limited flat areas.
Young morainic hills	**YMH**	10	Areas between 100 and 300 m asl, medium density texture, high relief and irregular pattern, limited flat areas.
Young morainic plateaus and plains	**YMPP**	12	Areas below 100-150 m asl, low and medium density texture, irregular pattern, significant amount of flat areas.
Inland dunes	**DN**	3	Elevation mostly below 100 m asl, very high density of texture, characteristic pattern, no flat areas in dune fields.
Old morainic plateaus and hills	**OMPH**	4	Areas between 100 and 300 m asl, low density texture, low and smooth relief, regular dendritic pattern, significant amount of flat areas.
Old morainic plains	**OMP**	5	Areas between 100 and 300 m asl, low density texture, very smooth or flat relief, large amount of flat areas.
Flat plains	**FLP**	7	Flat areas with very small addition of other forms.

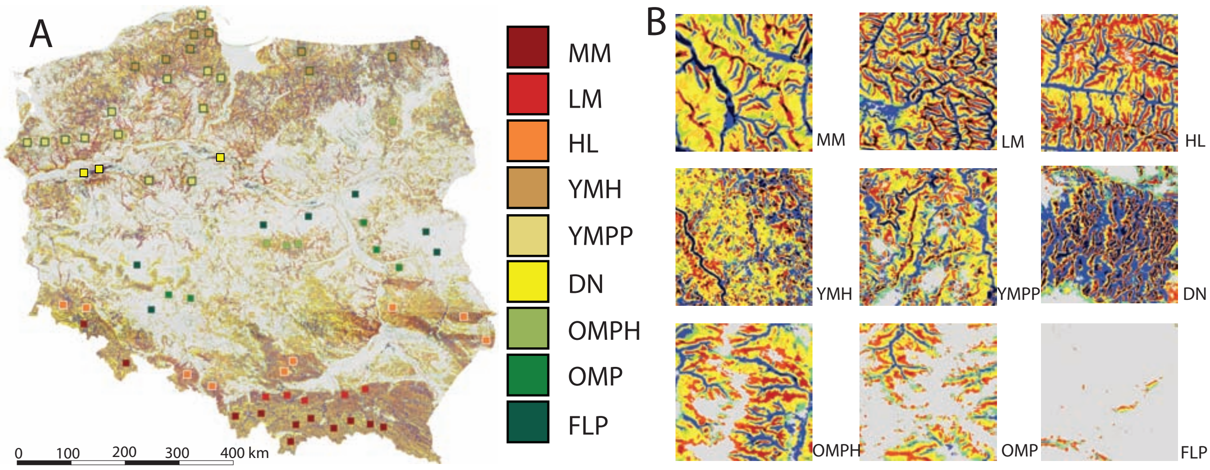

Figure 6: A set of template landscapes. (A) Landform elements map of Poland with locations of template landscapes indicated by squares having colors corresponding to units they exemplify (see legend to the right). (B) Nine sample landscapes (shown as maps of landform elements), each representing one of nine units.

values at all. This does not mean that local sites having flat landscapes are not similar to each other, but rather it reflects the nature of our similarity function (Eqn. (1)) that shows relatively smaller values of similarity between landscapes dominated by a single landform elements (flat) as it concentrates on differences between trace elements.

In the final step all nine likelihood maps are combined. This means that every super-cell temporarily stores nine values of likelihood, one for each unit. These values indicate the likelihood that a given super-cell belongs to each of the possible units. In order to produce a physiographic map of Poland we disambiguate these nine possibilities by assigning to a super-cell a unit label corresponding to the largest likelihood. This map is shown in Fig. 7B; it represents the final product of our calculations.

This generated map can be compared to the reference map (Fig. 2B) although we stress that reference map does not represent "ground truth" in the machine learning sense. Instead both maps are different models of reality made using different methodologies and different means. The reference map, like most manually created maps, delineates regions by delineating their boundaries. The homogeneity of landscape patterns within a boundary is implicit. Our algorithm delineates units by establishing regions with homogeneous landscape patterns. The boundaries between units are implicit. Nevertheless, the two maps tell the same story. The landscape types in Poland are arranged in a pat-

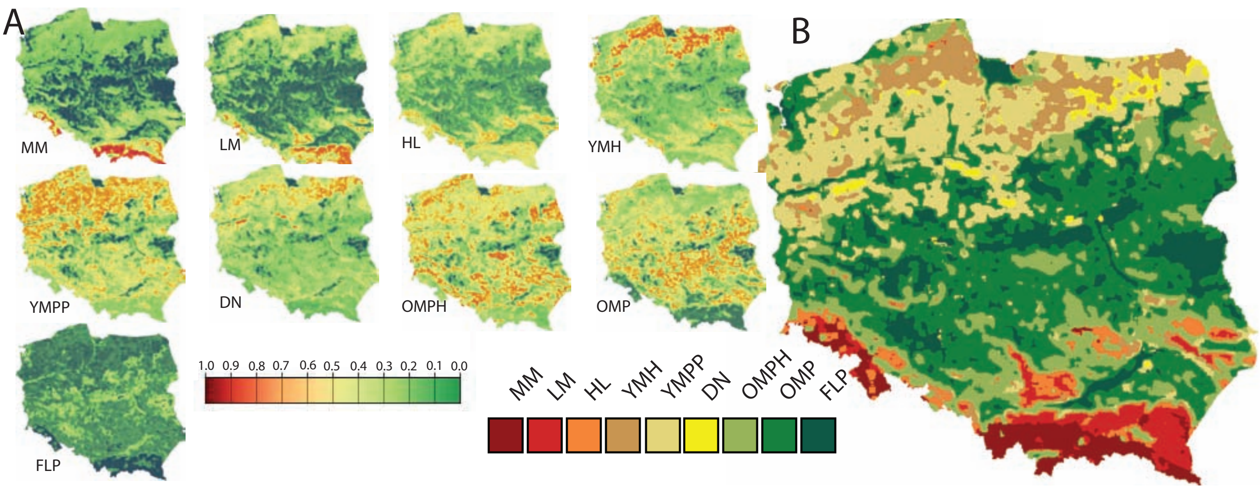

Figure 7: Delineation of physiographic units. (A) Spacial distributions of probability of belonging to a physiographic unit as indicated by a label. (B) The final, algorithm-delineated map of nine physiographic units across Poland.

tern of latitudinal belts (Lencewicz, 1937; Galon, 1972; Kondracki, 2002). These belts are the result of the general relief of Poland, with mountains and highlands in the south and lowlands in the middle and the north. Because the overall pattern of uplifted areas depends mostly on orography and its geological structure, the geomorphometry of lowlands are the result of the diminishing southward extent of successive Pleistocene glaciations (Marks, 2005). Every glacial epoch left several to hundred meters of new deposits, which after retreating revealed new, immature surfaces which became targets of denudation processes (Dylik, 1952, 1956). The overall differences between the two maps are minor and can be attributed to geological and paleogeographical inputs that went into the construction of reference map but were intentionally omitted from our algorithm whose goal was landscape comparison on the basis of terrain patterns alone. For example, the region located just south of the Notec pradolina in western Poland is classified exclusively as the YMPP unit on the reference map (Fig. 2B) up to the maximum extent of last glaciation (Marks, 2005). The same region is divided into several units (YMPP, YMH, DN, OMPH) on our map (Fig. 7B) because of differences in topography. Locally, the boundaries between corresponding units from the two maps are somewhat shifted, but the relative merits of specific delineations need to be discussed on a site-by-site case.

7. Discussion and Conclusions

In this paper we presented a new methodology enabling the quantitative comparison of landscapes, a search for landscapes similar to a given template, and, finally, the auto mapping of landscape types (or physiographic units). This methodology extends the field of geomorphometry - the science of quantitative land surface analysis - into the realm of content-based retrieval. Such an extension is significant because it opens up several new, practical possibilities.

First, our methodology provides for knowledge discovery through geomorphometry, as it makes possible a convenient exploration of very large topographic datasets (DEMs) in real time. Note that this is completely different from the capacity to search a DEM and its derivatives by attribute value using SQL queries built into most GIS systems. Whereas SQL queries retrieve individual DEM cells fulfilling predefined numerical conditions, our system retrieves entire landscapes on the basis of their similarity to a template. The only analog to our methodology is Content-Based Image Retrieval (CBIR) (Gevers and Smeulders, 2004; Datta et al., 2008; Lew et al., 2006) - the process of retrieving desired images from a large collection on the basis of features such as color, texture and shape that can be automatically extracted from the images themselves. In our method a local landscape or tile plays the role of an image and the set of all tiles covering the study area plays the role of a large image collection.

There are some important differences between the CBIR and our method. The most important difference is the expected outcome. The CBIR is expected to serve as an object-in-image retrieval system - a search is considered successful if retrieved images contain objects of interest. This expectation is very difficult to meet because a significant amount of high-level reasoning about semantic content of an image is required. However, available retrieval algorithms match images on the basis of primitive image features (much like in our method) that rarely, if ever, reflect the semantic meaning of an image. Thus, a general purpose CBIR often yields disappointing results (Hanjalic et al., 2008). On the other hand, our method serves as a pattern retrieval system - a search is considered successful if retrieved locations contain patterns of interest. This expectation is easier to meet because the relationship between primitive features and pattern is much closer than the relationship between primitive features and semantical objects. As a result our system provides a much higher level of user satisfaction and is ready to be used in practice. An additional reason our method performs quite well is because it is customized specifically for topographic patterns. A query is compared to scenes which are all landscapes. Because all landscape-derived patterns fulfill nature-imposed conditions our method avoids situations frequent in the domain of natural images, where scenes having patterns corresponding to very different landscapes have, nevertheless, very similar histograms of features.

Another difference between our method and the OBIR is the spatial character of landscape. Because of this character we are compelled to present the results of our search as a similarity map that returns not only the closest matches but also puts them in spatial context. This means that an execution of every query requires an exhaustive evaluation of similarities between the query and all other local landscapes. In contrast, the retrieval of similar images can be achieved by taking advantage of prior indexing. Despite the considerable computational cost of executing landscape search queries, our implementation (TerraEx-Pl) works in real time.

What are the potential uses for landscape search? The most obvious use is to identify locations having landscapes similar to a local landscape of interest. An example is provided by inland dune fields - a landscape that is quite rare in Poland and restricted to small patches of land. Using the TerraEx-Pl application a user can take a particular local dunes field (which happens to have a location known to the user) as a query and search for other potential dune fields across Poland. Such a search returns a similarity map that indeed identifies other dune fields. It also indicates a more extensive region showing elevated, but not high similarity to a dune query. The integrated environment of TerraEx-Pl allows for visual examination of this region. This underlines a human-computer interaction aspect of our landscape search application. The search algorithm application acts as a recommender, but a user has an ability to check these recommendations. Another potential use is the delineation of a region occupied by a certain landscape type. This applies to landscapes with relatively large spatial extent, like, for example, medium or low mountains discussed in Section 6. Both of these uses are quite powerful in application to Poland, but they would be even more powerful in application to larger datasets, such as the entire world, represented by the SRTM-derived DEM. The construction of such a world-wide landscape search engine is our ultimate goal.

Beyond landscape search, our method enables the auto-mapping of landscape types or physiographic units. Physiographic maps are important because they provide insight with regard to regional land-use planning, interpretation of landscape evolution, and the effects of physiography on other aspects of the surficial and ecological environment (Good et al., 1993; Martin-Duque et al., 2003; Daly et al., 2008; Fearer et al., 2008; Johnson and Fecko, 2008; Gawde et al., 2009). Because of their importance physiographic maps are developed by government-sponsored geological surveys at significant cost and effort. Our method is able to offer fast, custom physiographic mapping with minimum effort so that maps can be generated on demand by an end user.

In this paper we have demonstrated the process of physiographic map generation using a supervised approach. This choice was made primarily to demonstrate that our similarity measure is in agreement with human perceptions of landscape similarity. The performance of typical CBIR systems is tested using so-called ground truth set of images. These are images pre-labeled for content by an analyst. A system is considered to have a good performance if a query with label A retrieves predominantly images which are also labeled A. This standard performance test is not viable in the context of landscape similarity because it is difficult-to-impossible to label tiles of local landscapes with a small number of concise labels with clear semantic meaning. Instead, we test the design of our similarity measure implicitly by generating a physiographic map and comparing it to the manual map of the same set of units. Good agreement between the maps implies that landscapes similar according to our numerical measure are also similar according to human perception.

Auto-mapping of landscape types can also be performed using an unsupervised approach. Unsupervised approach uses the similarity measure but does not utilizes landscape search. Instead, regionalization of the study area with respect to landscape patterns is performed using either a clustering technique or a segmentation technique (Niesterowicz and Stepinski, 2013). As the aim of regionalization is to aggregate all local landscapes into a much smaller number of spatially contiguous regions – grouping landscapes having similar patterns – the output is tantamount to a physiographic map. In future work auto-mapping of Poland (or other regions) using an unsupervised approach will be performed and the results wiil be compared to those obtained using other auto-mapping techniques (Hammond, 1954; Iwahashi and Pike, 2007; Dragut and Eisank, 2012).

Finally, it needs to be pointed out that local landscapes can be compared at different characteristic length scales. In our method a single scale is used, but the value of the scale is a free parameter that can be changed to observe the influence of the scale on the results. In the TerraEx-Pl application only a single scale of 15 km is used so the computation is short enough to give a real time answer to a query.

Acknowledgments. This work was supported in part by the National Science Foundation under Grant BCS-1147702, the Polish National Science Centre under grant DEC-2012/07/B/ST6/01206, and by the University of Cincinnati Space Exploration Institute.

References

Allen, T. R., Walsh, S. J., 1996. Spatial and compositional pattern of alpine treeline, Glacier National Park, Montana. Photogrammetric Engineering & Remote Sensing 62(11), 1261–1268.

Barnsley, M. J., Barr, S. L., 1996. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. Photogrammetric Engineering and Remote Sensing 62(8), 949–958.

Bue, B. D., Stepinski, T., 2006. Automated classification of landforms on Mars. Computers and Geoscience 32, 604–614.

Cain, D. H., Riitters, K., Orvis, K., 1997. A multi-scale analysis of landscape statistics. Landscape Ecology 12, 199212.

Cha, S., 2007. Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods in Applied Sciences 1(4), 300–307.

Daly, C., Halbleib, M., Smith, J. I., P.Gibson, V., Doggett, M. K., Taylor, G. H., Curtis, J., Pasteris, P. P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28, 28, 2031–2064.

Datta, R., Joshi, D., Li, J., Wang, J. Z., 2008. Image Retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40, 1–60.

Dikau, R., Brabb, E., Mark, R., 1995. Morphometric landform analysis of New Mexico. Zeitschrift für Geomorphologie Supplement 101, 109–126.

Dikau, R., Brabb, E. E., Mark, R. M., 1991. Landform classification of New Mexico by computer. Tech. rep., US Department of the Interior, US Geological Survey.

Dragut, L., Eisank, C., 2012. Automated object-based classification of topography from SRTM data. Geomorphology 141-142, 21–23.

Duda, R. O., Hart, P. E., Stork, D. G., 2001. Unsupervised Learning and Clustering. In: Pattern classification (2nd edition). New York, NY: Wiley, p. 571.

Dylik, J., 1952. The concept of the periglacial cycle in middle poland. Bull. Soc. Sci. Letters 3, 5–29.

Dylik, J., 1956. Coup d'oeil sur la Pologne periglaciare. Biuletyn Peryglacjalny 4, 195–238.

Evans, I., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley, R. J. (Ed.), Spatial analysis in geomorphology. Methuen, pp. 17–90.

Fearer, T. M., Norman, G. W., Pack, J. C., Bittner, S., Healy, W. M., 2008. Influence of physiographic and climatic factors on spatial patterns of acorn production in Maryland and Virginia, USA. Journal of Biogeography 35, 2012–2025.

Gallant, A. L., Brown, D. D., Hoffer, R. M., 2005. Automated mapping of Hammond's landforms. IEEE Geoscience and Remote Sensing Letters 2, 384–288.

Galon, R., 1972. Geomorfologia Polski: Niż Polski. Vol. 2. Państwowe Wydawnictwo Naukowe.

Gawde, A. J., Zheljazkov, V. D., Maddox, V., Cantrell, C. L., 2009. Bioprospection of eastern red cedar from nine physiographic regions in Mississippi. Industrial Crops and Products 30, 59–64.

Gevers, T., Smeulders, A. W., 2004. Content-based image retrieval: An overview. In: Kang, G. M. S. B. (Ed.), Emerging Topics in Computer Vision. Upper Saddle River, NJ: Prentice-Hall, Ch. 8, pp. 333–384.

Good, J. E., Wallace, H. L., Williams, T. G. W., 1993. Management units and their use for identification, mapping and management of major vegetation types in upland conifer forest. Forestry 66, 261–290.

Hammond, E. H., 1954. Small-scale continental landform maps. Annals of the Association of American Geographers 44, 33–42.

Hanjalic, A., Lienhart, R., Ma, W. Y., Smith, J. R., 2008. The holy grail of multimedia information retrieval: So close or yet so far away? Proceedings of the IEEE 96(4), 541–547.

Haralick, R. M., 1986. Statistical image texture analysis. In: Young, T. Y., Fu, K. S. (Eds.), Handbook of Pattern Recognition and Image Processing. Academic Press, Orlando.

Herzog, F., Lausch, A., 2001. Supplementing land-use statistics with landscape metrics: Some methodological considerations. Environmental Monitoring and Assessment 72(1), 37–50.

Iwahashi, J., Pike, R. J., 2007. Automated classifications of topography from dems by an unsupervised nested-means algorithm and a three part geometric signature. Geomorphology 86, 409–440.

Jasiewicz, J., Stepinski, T. F., 2013a. Geomorphons - a pattern recognition approach to classification and mapping of lanforms. Geomorphology 182, pp. 147-156 182, 147–156.

Jasiewicz, J., Stepinski, T. F., 2013b. Example-based retrieval of alike land-cover scenes from NLCD2006 database. Geoscience and Remote Sensing Letters 10, 155–159.

Johnson, P. A., Fecko, B. J., 2008. Regional channel geometry equations: A statistical comparison for physiographic provinces in the eastern US. River Research and Applications, 24, 24, 823–834.

Kondracki, J., 2002. Geografia regionalna Polski. 3rd Edition. Wydawnictwo Naukowe PWN, Warszawa.

Kupfer, J. A., Gao, P., Guo, D., 2012. Regionalization of forest pattern metrics for the continental United States using contiguity constrained clustering and partitioning. Ecological Informatics 9, 11–18.

Lencewicz, S., 1937. Polska. Vol. 7. Nakl. Trzaski, Everta i Michal-

skiego.
Lew, M., Sebe, N., Lifi, C., Jain, R., 2006. Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput., Commun., Appl., 2(1), 1–19.
Long, J., Nelson, T., Wulder, W., 2010. Regionalization of landscape pattern indices using multivariate cluster analysis. Environmental Management 46(1), 134–142.
MacMillan, R., Jones, R., McNabb, D., 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). Computers, Environment and Urban Systems 28, 175–200.
Marks, L., 2005. Pleistocene glacial limits in the territory of Poland. Przeglad Geologiczny 53 (10), 988–993.
URL `http://www.pgi.gov.pl/pdf/pg_2005_10_2_25.pdf`
Martin-Duque, J. F., Pedraza, J., Sanz, M. A., Bodoque, J. M., Godfrey, A. E., Diez, A., Carrasco, R. M., 2003. Landform classification for land use planning in developed areas: An example in Segovia Province (Central Spain). Environmental Management, 32, 32, 488–498.
Mehryar, M., Rostamizadeh, A., Talwalkar, A., 2012. Foundations of Machine Learning. The MIT Press.
Meybeck, M., Green, P., Vorosmarty, C., 2001. A new typology for mountains and other relief classes: an application to global continental water resources and population distribution. Mountain Research and Development 21, 34–45.
Minar, J., Evans, I., 2008. Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. Geomorphology 95, 236–259.
Niesterowicz, J., Stepinski, T., 2013. Regionalization of multicategorical landscapes using machine vision methods. Applied Geography 45, 250–258.
Olaya, V., 2009. Basic land-surface parameters. In: Hengl, T., Reuter, H. (Eds.), Geomorphometry, Concepts, Software, Application. Elsevier, Ch. 6, pp. 141–169.
O'Neill, R., Krummel, J., Gardner, R., Sugihara, G., Jackson, B., DeAngelis, D., Milne, B., Turner, M., Zygmunt, B., Christensen, S., Dale, V., Graham, R., 1988. Indices of landscape pattern. Landscape Ecology 1(3), 153–162.
Pike, R., 1988. The geometric signature: quantifying landslide-terrain types from digital elevation models. Mathematical Geology 20, 491–511.
Stepinski, T. F., Netzel, P., Jasiewicz, J., 2014. LandEx - A GeoWeb tool for query and retrieval of spatial patterns in land cover datasets. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(1), 257–266.
Wickham, J. D., Norton, D. J., 1994. Mapping and analyzing landscape patterns. Landscape Ecology 9(1), 7–23.
Wood, J., 1996. The geomorphological characterisation of digital elevation models. Ph.D. thesis, PhD Thesis Department of Geography, University of Lancaster, UK.