Geomorphology 221 (2014) 104–112

Contents lists available at ScienceDirect

Geomorphology

journal homepage: www.elsevier.com/locate/geomorph

Landscape similarity, retrieval, and machine mapping of physiographic units

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

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

ARTICLE INFO

Article history:
Received 1 March 2014
Received in revised form 18 May 2014
Accepted 7 June 2014
Available online 17 June 2014

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

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.

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 than 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

* Corresponding author at: Space Informatics Lab, 215 Braunstein Hall, Cincinnati, OH 45221, USA.
E-mail address: stepintz@uc.edu (T.F. Stepinski).

http://dx.doi.org/10.1016/j.geomorph.2014.06.011
0169-555X/

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 the 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 than 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 depend 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 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 (Dikau et al., 1995; Wood, 1996; Jasiewicz and Stepinski, 2013b) 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, 2013b) 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 does not have to be one of the local landscapes predefined by a lattice of tiles, and it does not 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.