Content-based landscape retrieval using geomorphons

Jaroslaw Jasiewicz
Geoecology and Geoinformation Institute
Adam Mickiewicz University
Dziegielowa 27, 60-680 Poznan, Poland
jarekj@amu.edu.pl

Pawel Netzel
Department of Climatology and Atmospheric Protection
University of Wroclaw
Kosiby 6/8, 51-621, Wroclaw, Poland
netzelpl@ucmail.uc.edu

Tomasz F. Stepinski
Space Informatics Lab
University of Cincinnati
Cincinnati, OH 45221-0131, USA
stepintz@uc.edu

Abstract—**We introduce a concept of content-based landscape retrieval (CBLR). Our CBLR system retrieves morphometrically similar landscapes from a large DEM thus acting as a search engine for landscapes. The system works on the principle of query-by-example, a reference landscape is given and the system outputs a map showing degree of similarity between a reference and all the other local landscapes across the spatial extent of the DEM. Landscape is defined as pattern of landform elements. A DEM is converted into a map of landform elements using the geomorphons method. The core of the CBLR is the similarity function between two landscapes that encapsulates a degree to which their patterns of landform elements correspond to each other. The search relies on exhaustive evaluation of similarity using an overlapping sliding window approach. In the featured case study we use our method to delineate a spatial extent of the characteristic landscape formed by the end moraine associated with the latest glaciations across the country of Poland. Good agreement is found between the region delineated by our method and a range of end moraine manually delineated from geomorphic, geological, and paleogeographical information. The CBLR can be implemented as a GeoWeb application and serve as a rapid and convenient tool for exploration of very large DEM datasets.**

I. INTRODUCTION

Large archives containing medium resolution (10-100 m/cell) digital elevation models (DEM) datasets of continental or global extent are now readily available (see, for example, DEM Explorer http://ws.csiss.gmu.edu/DEMExplorer/). Popular datasets include The Shuttle Radar Topography Mission (SRTM) and Aster Digital Global Elevation Map (GDEM). However, at present, these archives are predominantly used to access topographic information for sites for which a prior knowledge of their relevance already exists. The full informational potential of these archives is not fulfilled due to lack of intelligent methods aimed at their exploration and knowledge discovery. Possible approaches to the development of such methods include terrain classification and query-by-example. Previous work [1-3] has focused on development of algorithms capable of automatic delineation of physiographic units from a DEM. Resultant maps indeed utilize the entire dataset, but tend to be too generalized to allow for meaningful exploration and discovery.

In this paper we propose a content-based landscape retrieval (CBLR) system – a query-by-example method that identifies all locations in a DEM characterized by landscapes that are similar to a given example or a set of examples. The word “content” in the CBLR indicates that a query is based exclusively on quantitative values extracted from the DEM and not on any metadata. Proposed system is a modification of our earlier work [4] on query and retrieval of similar land cover scenes from the National Land Cover Dataset 2006 and is inspired by Content-Based Image Retrieval (CBIR) systems extensively studied [5] in the context of natural image retrieval. By not being restricted to a set of pre-defined classes, the CBLR system has much higher discriminating power than auto-generated physiographic maps. This makes it a great tool for exploration of large DEMs with capacity for meaningful discovery. Ultimately, the system is envisioned as a web-based real time “search” application for landscapes.

In our context, “landscape” means a pattern of landform elements over a scene of interest. In the rest of this paper we use term landscape in this specific morphometric meaning. Two scenes have similar landscapes if they have similar patterns of landform elements. Thus, the CBLR system must consist of a method of delineating landform elements, a similarity function capable of comparing spatial patterns of those elements, and an implementation of spatial query. For automatic delineation of landform elements we use the geomorphons method [6]. This robust and computationally efficient method converts DEM into a geomorphometric map (gmphMap) – a categorical raster indicating the most common landform elements. The gmphMap can be thought of as an interpreted map of topography and is ideally suited for our CBLR system. (For an example of a gmphMap and its comparison to a physiographic map see http://sil.uc.edu/dataeye/). We developed a pattern similarity function appropriate to morphometric landscapes and designed a spatial retrieval system to execute a query and display its results.

II. Methodology

A. Mapping landform elements

The input to the method is a DEM dataset over the region of interest (for example, a given country, continent, or the entire globe). The first step is to convert a DEM to the gmphMap. This is achieved using a public domain GRASS module *r.geomorphons* (http://sil.uc.edu/downloads.html). Geomorphons are pattern recognition-based method for delineation of landform elements. The method yields the gmphMap – a raster of the same size as the input DEM with each cell assigned one of ten common landform elements labels: flat, peak ridge, shoulder, slope, spur, hollow, footslope, valley and pit. The method has two parameters: the search radius L and the flatness threshold t. The search radius determines the maximum spatial scale at which any landform element is extracted and the flatness threshold is a angle of a slope below which a terrain is considered flat.

B. Similarity between two landscapes

A "tile" T is defined as a small subset of the entire gmphMap. For convenience we use square-shaped tiles with the of size n by n cells. Each cell is labeled by one of 10 landform elements labels. A spatial pattern of different labels (commonly visualized by different colors on the gmphMap) constitutes a landscape. Note that the size of the tile indicates a spatial extent over which landscape is captured. A query Q is a particular tile containing a landscape of interest.

Like in most CBIR methods, we don't calculate similarity between two tiles directly from the values of their cells, but rather from the histograms of their "primitive features." Histograms are widely used in the CBIR because of their rotational invariance. Primitive features are elements of pattern that are counted to form a histogram. For example, they could be just individual cells; counting cells of different labels produces a histogram reflecting "composition" of landscape. However, in this paper we use pairs of neighboring cells as primitive features (4-connected neighborhood is assumed). Because there are 10 different labels in the gmphMap, there are 55 different possible pairs, examples include: flat-flat, flat-slope, slope-peak, etc. Pairs are extracted for each tile and counted to form a histogram that encapsulates the pattern of the tile. By showing which connections between landform elements are most common such histogram emphasizes a texture of a given landscape.

Calculating similarity between two landscape tiles T and Q reduces to calculating similarity between the two histograms T^h and Q^h representing their patterns. It is customary to compare histograms using distance (dissimilarity) rather than a similarity metric. Choosing the best distance metric is largely empirical decision dependent on the actual retrieval system. In this paper we use normalized Wave Hedges distance metric [7] which takes values from 0 (identical histograms) to 1 (histograms do not share bins). The formula for Wave Hedges similarity (1-distance) is:

$$sim(T^h, Q^h) = \frac{1}{n}\sum_{i=1}^{n} \frac{\min(T_i^h, Q_i^h)}{\max(T_i^h, Q_i^h)},\ \text{if}\ \max(T_i^h, Q_i^h) > 0$$

The measure compares corresponding bins of two histograms and compute a fraction equal to (smaller bin)/(larger bin). If the two bins are both 0 the fraction is equal to 1. The bin comparison values are summed and the sum is divided by the number of bins. An overall similarity is built from similarities between individual features. Note that such similarity measure is not sensitive to absolute composition of landform elements in a landscape; for example, two landscapes dominated by the flat element may still be measured as significantly different if other, minor elements are different.

C. Query execution

A query over the entire dataset uses a square grid with the resolution of k cells superimposed on the entire spatial extent of the DEM. This grid forms a basis for a similarity map resulting from the query. The query is executed by means of exhaustive evaluation - the value of similarity is calculated between the query tile and all the local tiles assigned to a similarity grid. If $n > k$ the local tiles overlap. The resulting similarity raster (much coarser than an original DEM) can be displayed as a map showing how a degree of similarity to a query varies over the entire region.

III. Case Study

In order to demonstrate a practical application of our CBLR system we applied it to a DEM covering the country of Poland. The input is a 1" integer DEM which was converted by adaptive smoothing to the 30m resolution floating-point terrain model. The final DEM has the size of 21,696 by 24,692 cells. The gmpgMap was calculated using search radius L=40 cells (1200m) and flatness threshold t=0.8 degree. Fig. 1 shows resultant gmphMap; individual landform elements cannot be seen at this level of resolution, but different physiographic units can be recognized by different tones of colors resulting from different composition and patterns of landform elements. The territory of Poland exhibits a number of different landscapes. In particular, the northern parts of the country is occupied by postglacial landscape and the southern parts by uplands and mountains.

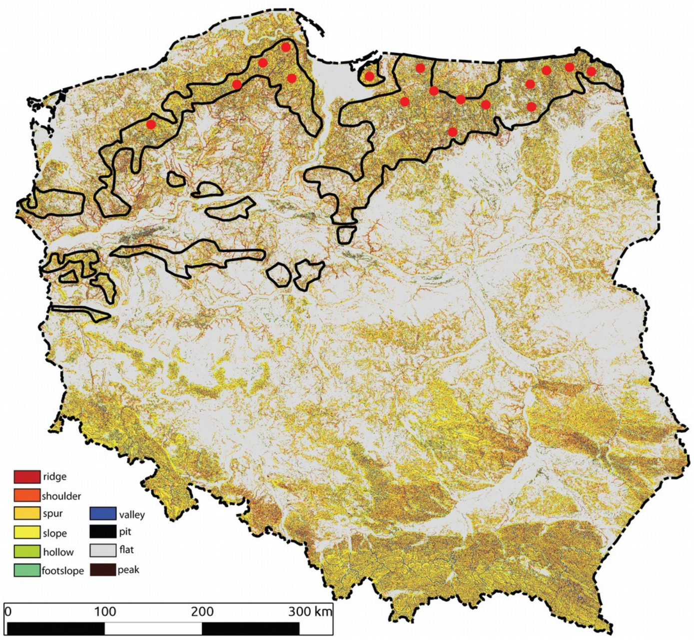

Figure 1. Geomorphometric map (gmphMap) of Poland calculated from the 30m resolution DTM using geomorphons method. Black contours indicate a range of end moraine according to Kondracki [8] and the red dots show location of our queries.

We demonstrate the ability of the CBLR system to delineate a spatial extent of a given landscape by applying it to calculate a region corresponding to end moraine resulting from glacial activity in the marginal zone of Pomeranian/Brandemburg phase. Such landscape, characterized by broad ridges and hills, is a very young surface not yet transformed by a denudation process. Therefore we expect that its morphometric landscape is sufficiently unique for the CBLR system to discriminate it from older surfaces which are also dominated by ridges and hills. On Fig. 1 the black contours indicate the range of end moraine landscape as manually delineated [8] on the basis of geomorphological as well as geological and paleogeographical information. This serves as a reference to our results.

The execution of a query was constructed using n=512 and k=64 cells. This means that we assess landscape on spatial scale of ≈ 15 km and we use similarity grid having size of ≈ 2 km. Thus, we sample local landscape with a high degree of overlapping. In its basic form a query-by-example works by choosing a single reference landscape and calculating a similarity map showing spatial distribution of similarity between this single example and the local landscapes. However, in a landscape formation, like the end moraine, not every location has exactly the same character of landscape as local variations are present. Therefore, it makes more sense to select a number of different queries, calculate similarity map for each one individually and integrate the results to obtain a region with local landscapes defined by similarities with all examples.

We have selected 17 examples (queries) from the end moraine range; their locations are shown by red dots on Fig. 1. The resultant similarity maps were combined in two different ways. In the first the minimum from amongst all 17 local similarity values was assigned as an overall local similarity. Such procedure delineates areas which are most similar to *all* examples. The resultant map is shown on the left panel of Fig. 2. In the second, the median of all 17 local similarity values was calculated and assigned as an overall local similarity. Such procedure delineates areas whose *expected* similarity to the queries is high. The resultant map is shown on the right panel of Fig. 2. Both similarity maps cover the entire area of Poland, areas with high values (red colors) delineate a range of end moraine landscape. We observe that our median method delineated a region with high overlap (>70%) with a manually delineated reference. Small areas in southern Poland also show high average similarity despite having no geological connection with end moraine. As expected, the minimum method is more restrictive, but its map does not show high values of similarity beyond the putative range of end moraine. Overall, the delineation of end moraine is good considering that our method takes into account only the DEM without any additional geologic information.

IV. CONCLUSIONS

The CBLR offers a search engine-like capabilities for landscapes embedded in large DEMs. In most applications a query is taken from the same DEM, but it may be taken from another source including a simulated landscape. In addition, a query may not correspond to any landscape at all, but rather be just a histogram of features designed to test a specific hypothesis. Unlike other, more familiar search tools, the CBLR does not output a short list of best-matching landscape locations, but rather a similarity map showing a degree of similarity between a query and every other landscape in the DEM. Such presentation of search results is more appropriate for spatial dataset where geographical context matters.

The CBLR is an ideal tool for rapid exploration of large DEMs. For such a tool to be practical it needs to be implemented as a GeoWeb application so it can be accessed by anyone. We plan on offering such application in the near future; it would work similarly to LandEx [9] (http://sil.uc.edu/landex/) - our existing tool for searching land cover across the United States.

The performance of the CBLR system depends predominantly on selection of features and a form of pattern similarity function. In addition to the features and the similarity function presented here we have also experimented with features described in [4] and the Jensen-Shannon similarity function. Such choice gives similar results in application to the case study presented here, but for other searches its performance was worse. Future research will evaluate other combinations of features and similarity functions. Note that search results will depend on the spatial scale of landscape; a scale used in this paper is about 50% larger than that advocated by Hammond [10] and within a range of macro landforms.

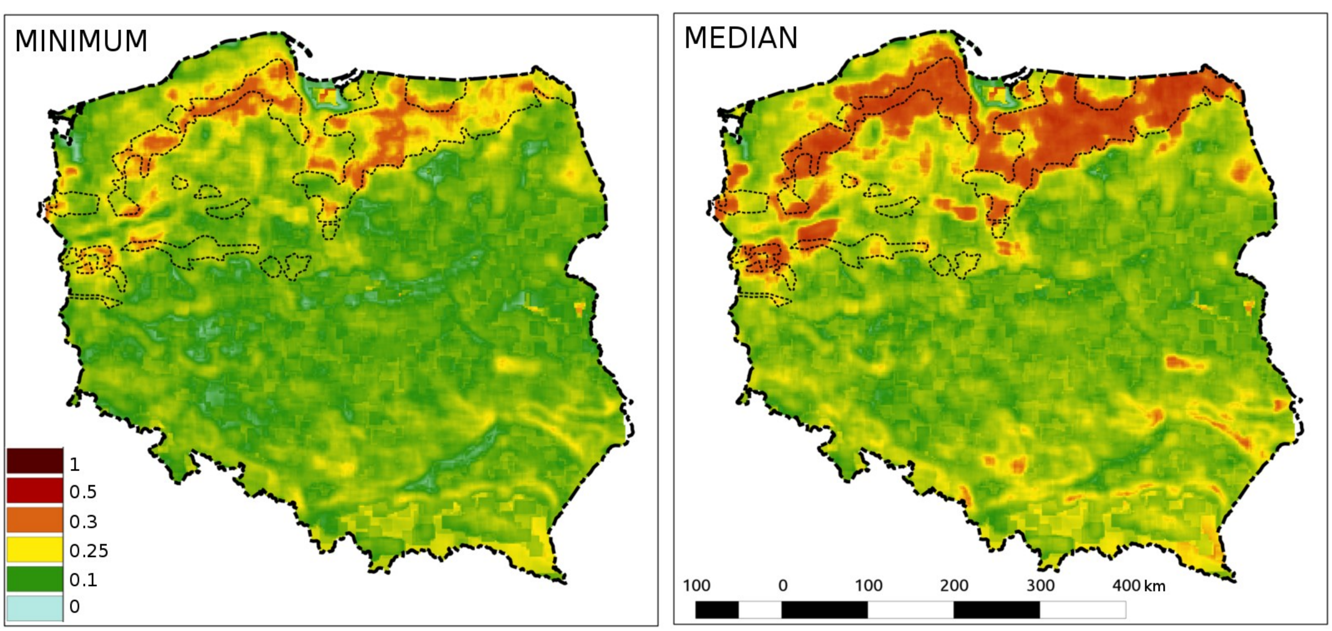

Figure 2. Landscape similarity maps constructed to delineate the range of end moraine landscape. (Left) Map obtained by taking a minimum similarity from all the queries. (Right) Map obtained by calculating a median similarity. Black contours indicate a range of end moraine according to Kondracki [8]

ACKNOWLEDGMENT

This work was supported partially by the National Science Foundation under Grant BCS-1147702 and National Science Centre DEC-2012/07/B/ST6/012206 and UC Space Exploration Found.

REFERENCES

[1] Dikau, R., Brabb, E.E, Mark, R.M., 1991.”Landform classification of New mexico by computer.” Open File Report 91-634. U.S Geological Survey. .

[2] Iwahashi, J., Pike, R., 2007. “Automated classification of topography from DEMs by an unsupervised nested-means algorithm and three-part geometric signature.” Geomorphology 86, 409-440.

[3] Dragut, L., Eisank, C., 2012. “Automated object-based classification of topography from SRTM data.”Geomorphology 141-142, 21--23.

[4] Jasiewicz, J., Stepinski, T.F. 2013. ”Example-Based Retrieval of Alike Land-Cover Scenes From NLCD2006 Database.” IEEE Geoscience and Remote Sensing Letters 10(1), pp. 155-159

[5] 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.

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

[7] Cha, S.-H. 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.

[8] Kondracki, J., 2002. ”Geografia regionalna Polski.” 3rd Edition. Wydawnictwo Naukowe PWN, Warszawa.

[9] Stepinski, T.F, Netzel, P., Jasiewicz, J., Niesterowicz,J. 2012. “LandEx - A GeoWeb-based Tool for Exploration of Patterns in Raster Maps, GIScience 2012 Conference. http://www.giscience.org/proceedings/

[10] Hammond, E.H. 1964. “Analysis of properties of land form geography: an application to broad-scale land form mapping.” Annals of the Association of American Geographers 54, 11-19.