LandEx – A GeoWeb-based Tool for Exploration of Patterns in Raster Maps

T. F. Stepinski[1], P. Netzel[1,2], J. Jasiewicz[3], J. Niesterowicz[1]

[1]Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131, USA
[1]Email: {stepintz; netzelpl}@ucmail.uc.edu niestejk@mail.uc.edu

[2]Department of Climatology and Atmospheric Protection, University of Wroclaw, Kosiby 6/8, 51-621, Wroclaw, Poland

[3]Geoecology and Geoinformation Institute Adam Mickiewicz University, Dziegielowa 27, 60-680 Poznan, Poland
[2]Email: jarekj@amu.edu.pl

1. Introduction

Advances in remote sensing and GIS make possible construction of high resolution categorical raster-based maps depicting spatial distribution of natural and/or anthropogenic features. The best know example is a land cover/land use (LCLU) map. GIS tools for performing queries for spatial extent of single map category are readily available. However, often, an analyst would like to query a map for an occurrence of a particular spatial pattern of categories rather than a single category; no tools exist to execute such query. We have developed the Landscape Explorer (LandEx) – an algorithm that performs query by pattern similarity (QBPS) on categorical rasters – to provide such functionality. In the context of land cover, a pattern of categories may be associated with semantic notions of "downtown", "typical small town", or "an irrigated agriculture in a desert environment" to give just a few examples. QBPS finds regions across the map having patterns of categories (and thus, presumably, semantic meanings) similar to a pattern of a reference region.

Fig.1 illustrates an idea of LandEx in a nutshell using the National Land Cover Dataset 2006 (NLCD2006) (Fry et al., 2011) map as an example. In this example LandEx takes as input a categorical raster map with K=16 land cover classes. A user selects a small region of interest (ROI) and wants to identify other regions across the conterminous United States having similar patterns (and thus similar semantic meanings). LandEx outputs map of similarity values thus enabling a user to put the ROI in a broader geospatial context.

Development of LandEx involves two separate tracks, one is a design of the core algorithm that executes a QBPS request and another is an implementation of this algorithm as a GeoWeb service. Our goal for LandEx is to offer a tool capable of intelligent knowledge discovery in large (or very large) spatial datasets in a convenient and intuitive manner, but also capable of delivering research-grade results. To fulfil this goal LandEx is accessible at http://sil.uc.edu/landex as a GeoWeb service that provides all expected functionalities.

2. Related research

To the best of our knowledge the topic of QBPS has not been previously studied in the context of raster maps. On the other hand, a somewhat similar topic of query by image content (QBIC) has been extensively studied (Datta et al., 2008), and the applications of QBIC for remotely sensed imagery data (a precursor to some of the raster maps) have also been considered (Daschiel and Datcu, 2005, Cerra et. al., 2010). The QBPS, as proposed here, can be considered as an unexplored niche of the QBIC. Our own previous relevant work includes a description of QBPS core algorithm for the land cover map (Jasiewicz and Stepinski, 2012).

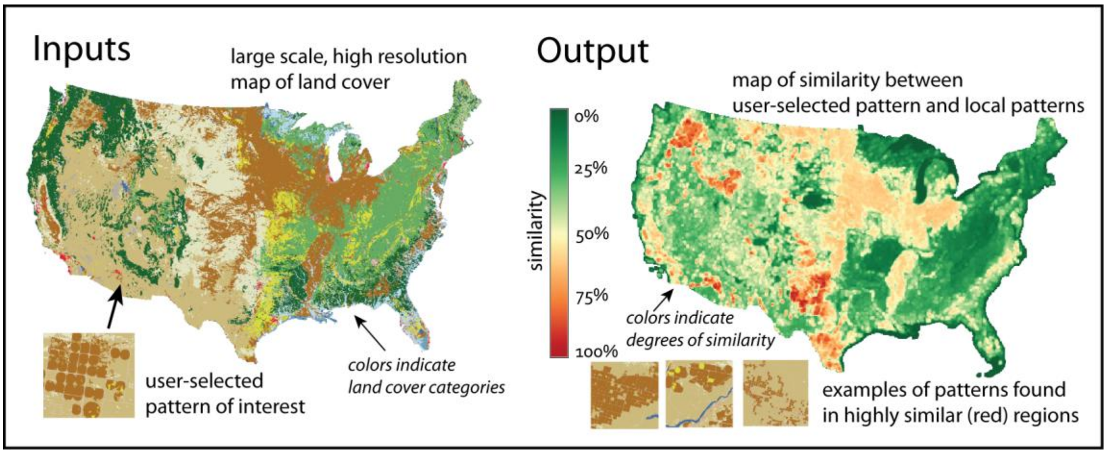

Fig.1. Idea of LandEx in a nutshell using the National Land Cover Dataset 2006 as an example.

3. LandEx core algorithm

The algorithm works on the principle of query-by-example. The purpose of the algorithm is to calculate a degree of similarity between a reference scene and all other scenes in a raster map. There are three major components of the algorithm: scene pattern signature, scene patterns similarity, and search. In our context "scene" refers to a spatial pattern of map categories (colors) over a local region, a scene signature is a compact mathematical description of such pattern, and scenes similarity is a function that assigns a degree of alikeness between two scenes on the basis of their respective scene signatures.

3.1 Scene pattern signature

Relative simplicity of a raster map (as compared to image) allows for a relatively simple design of pattern signature. We first segment the entire raster map into single color patches. After the segmentation each cell in the map is assigned two attributes, its color and the size of the patch to which it belongs. Scene signature is a two dimensional histogram of these attributes as collected over a local region. Incorporation of patch size attribute is crucial for describing a pattern rather than just a composition of colors in a local region. Note that color is a categorical variable; whereas patch size is a continuous variable and needs to be discretized for calculation of pattern signature. We discretize patch sizes (measured in units of cells) into bins with ranges based on the powers of two (e.g.1-2, 2-4, 4-8).

3.2 Scene patterns similarity

Pattern similarity in LandEx needs to be invariant to relative rotation, translation, or even some degree of deformation between the patterns. For this reason our patterns similarity measure is based on the notion of mutual information between two probability distribution functions (PDFs) related to the two local raster maps (Remmel and Csillag, 2006). The first PDF (called Y) is obtained by *combining* scene signatures of the two patterns into a single 2D histogram describing distribution of colors and patch sizes in the concatenated map. The second PDF (called X) assigns equal probabilities to selecting one of the two possible maps. A joint PDF assigns a probability to choosing one of the two maps and drawing a pair (color, patch size) from that map.

All PDFs are characterized using Shannon entropies values H. Specific conditional entropy $H(Y|X{=}x_i)$ is the entropy calculated only from the histogram of map x_i i = 1 or 2 and conditional entropy $H(Y|X)$ is an average of entropies calculated from histograms restricted to individual maps. Mutual information, $I(Y,X)= H(Y) - H(Y|X)$, measures an average reduction of unpredictability of Y if the specific map is set. The $I(Y,X)$ is a convenient measure of ``distance" between two maps, if by the distance we understand the increasing difference in patterns of colors and patch sizes in the two maps. Conversely, 1- $I(Y,X)$ is a convenient measure of similarity between the two maps.

3.3 Search

LandEx utilizes an overlapping sliding window approach to searching a large raster map for patterns similar to a reference. The search is executed by calculating similarities between a reference scene and all scenes contained in windows covering the entire dataset. The result of search is the similarity map indicating the degree of similarity between the local and the reference scenes.

4. LandEx as GeoWeb service

LandEx is implemented as a GeoWeb page and is available at http://sil.uc.edu/landex. LandEx user interface works in an internet browser environment and is based on JavaScript libraries: ExtJS with GeoExt and OpenLayers. Through this interface a user can access all functionalities expected from a GeoWeb page. Fig.2 summarizes major functionalities of LandEx user interface.

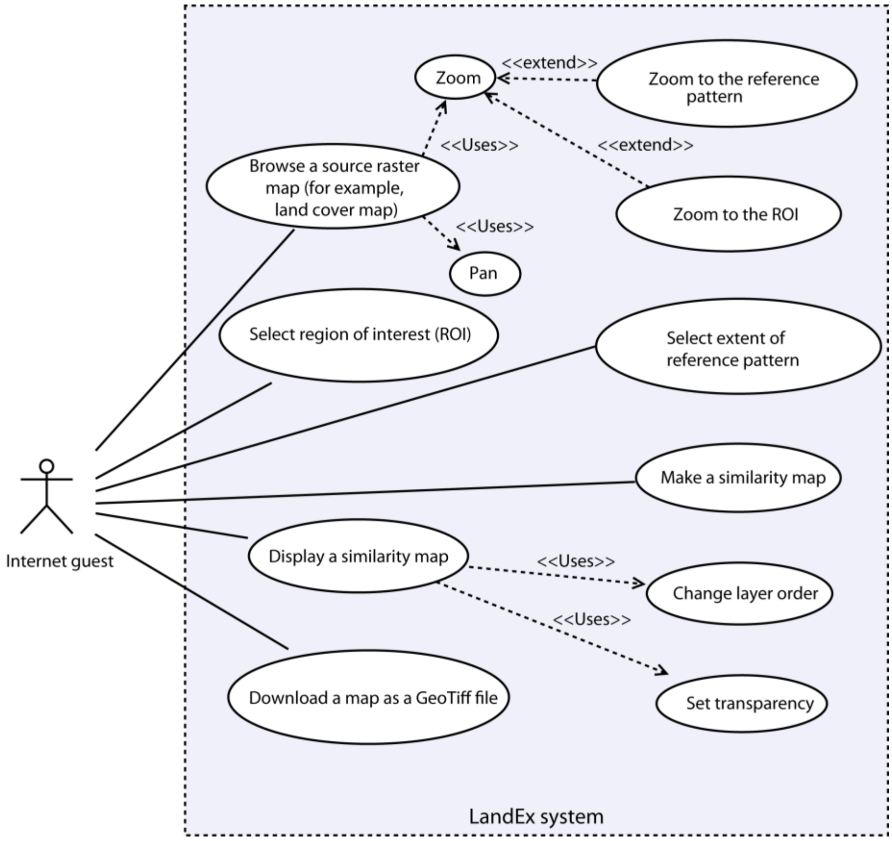

Fig. 2 LandEx use cases diagram.

The web browser client accesses a calculation engine using standards commonly used in geospatial web portals: WMS (Web Map Service), WCS (Web Coverage Service) and WPS (Web Processing Service). By using these standards we assure that LandEx adheres to the SOA (Service Oriented Architecture) architecture proposed by OGC (Open Geospatial

Consortium). By adhering to OGC standards we assure that LandEx can be accessed not only through our own internet browser interface but also by third party software packages compatible with WMS and WCS protocols. This provides an extra flexibility for utilizing the core idea of LandEx – generation of the map showing similarity between a local pattern and the reference.

Geospatial services are provided using the GeoServer software. Access to the calculation engine (which actually calculates a similarity map for a given reference pattern) is provided by the WPS available in the GeoServer as a plugin. The calculation engine itself is implemented in GRASS (Geographical Resources Analysis Support System) which uses its own geospatial database. The resultant similarity map calculated in GRASS is sent to GeoServer using FTP protocol and REST interface. It can be examined in the web browser interface (or third party software); it can also be downloaded as a GeoTiff file. All components of LandEx are Free Open Source Software (FOSS) software. The overall software architecture of LandEx is illustrated in Fig 3.

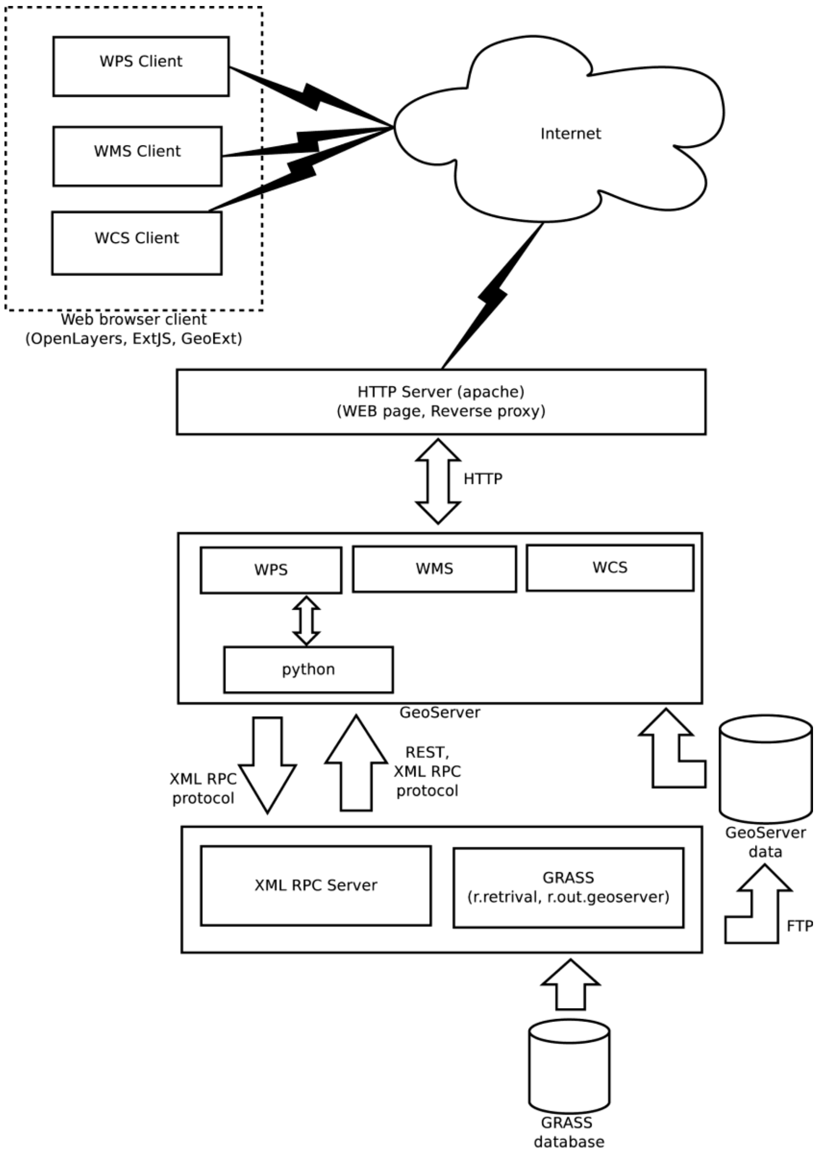

Fig. 3 LandEx software architecture

.

Acknowledgements

This work was supported by the National Science Foundation under Grant IIS-1103684.

References

Cerra, D., Mallet, A, Gueguen, L., and Datcu, M., 2010, Algorithmic Information Theory-Based Analysis of Earth Observation Images: An Assessment. *IEEE Geoscience and Remote Sensing Letters*, 7(1):8-12

Daschiel H and Datcu M, 2005, Information mining in remote sensing image archives: System evaluation. *IEEE Trans. Geosci. Remote Sens.*, 43(1):188–199.

Datta R, Joshi D, Li, J, and Wang JZ, 2008 Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40:5:1–5:60

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, *PE&RS*, Vol. 77(9):858-864.

Jasiewicz J and Stepinski TF, 2012, Example-based retrieval of alike land-cover scenes from NLCD2006 database. *IEEE Geoscience and Remote Sensing Letters* (accepted, available online http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6204321).

Remmel TK and Csillag F, 2006, Mutual information spectra for comparing categorical maps. *International Journal of Remote Sensing*, 27(7):1425–1452.