Pattern-based assessment of land cover change on continental scale with application to NLCD 2001-2006

Pawel Netzel and Tomasz F. Stepinski

***Abstract*—We present a method for assessing land cover change on continental scale and with high spatial resolution. This is a post-classification method, but instead of tracking transitions of land cover classes on cell-by-cell basis the method measures the change at a tile level by quantifying a difference between local patterns of land cover at two different time steps. Pattern-based change assessment is well suited for the large scale survey as it addresses landscape dynamics rather than just simple land class transitions. A tile is defined as a local area consisting of large enough number of land cover cells to sample a distribution of landscape but small enough to detect change with high spatial resolution; 4.5 km $\times$ 4.5 km square tiles are used. The level of change is measured as the dissimilarity between motifs of tile patterns at two time steps and is calculated using information-theoretic metric called the Jensen-Shannon similarity. The method is able to discriminate between different types of change including change in geometric pattern, change in class composition, and numerous class transitions without significant changes in either pattern or composition. The methodology is applied to the National Land Cover Dataset (NLCD) to obtain a 2001-2006 change map of the conterminous U.S. The resultant map shows (in a high resolution of 3 km/cell) a spatial distribution of the degree to which the landscape has changed in this time period. Both, large regions (southeastern and Gulf regions, Pacific Northwest region, and the state of Maine) of heightened landscape dynamics, as well as small regions of sudden change due to fires, urban growth etc. are clearly identifiable from the map. A fully-featured online application for fast and convenient exploration of the change map together with original land cover maps in their full resolutions is available at http://sil.uc.edu/dataeye/.**

***Index Terms*—land cover change, pattern recognition, NLCD 2001 and 2006**

I. INTRODUCTION

LAND use/land cover (LULC) composition and change are important factors affecting ecosystem condition and function. LULC change detection based on multi-temporal remote sensing data [33] has been established as a cost-effective means for providing information on the occurrence of change as well as its aerial extent and causes. Numerous techniques of LULC change detection have been developed (for reviews see [26], [6], [21]) as no single technique works equally well in all contexts. Different contexts of LULC change detection involve spatial scale (local, regional, continental, etc.), selection of thematic classes (general, vegetation-based, urban-based etc.) and purpose (high accuracy quantitative, medium accuracy qualitative, etc).

The majority of work on LULC change is conducted on the local-scale involving a specific study area; consequently, most change detection techniques are tuned to such applications [26]. Most change detection studies on a global-scale concentrate on utilizing Moderate Resolution Imaging Spectroradiometer (MODIS) products with a particular focus on the issue of deforestation. They rely on cell-by-cell comparison between data taken at two time steps [46], [47] or on examination of a time series of data at each cell [4], [28]. A recent study [15] assessed deforestation on global scale with resolution of 30 m using Landsat data. On a regional-scale, the MODIS data and the time series technique were applied [27] to the Albemarle-Pamlico estuary system to provide automated change detection. In the United States, the availability of the National Land Cover Database (NLCD) makes possible investigating regional-scale change detection in contexts other than deforestation. Xian et al. [44] used NLCD 2001 and 2006 data to derive a change matrix quantifying urban growth in the Gulf of Mexico region. Hollenhorst et al. [18] used integration of the NLCD and the Ontario (Canada) Provincial Land Cover data to derive a change matrix for the Lake Superior basin over the period of 1992 to 2001.

All existing large-scale LULC change detection methods are cell-based just like the vast majority of the methods developed for local-scale use. However, the purpose of assessing land cover change on a large-scale may differ from the purpose of assessing the change at a finer scale. Locally, change assessment is driven by a desire to accurately catalog all LULC transitions; the cell-based approach is very useful for such a purpose as it unambiguously indicates for each cell whether LULC transitions occurred or not. The aims of large-scale LULC change assessment can be different. The first aim is to search for all places where change has occurred [4], [28], [27]. Such an aim is well addressed by the cell-based approach, especially if coupled with segmentation technique [28]. The second aim is to provide a comprehensive overview of landscape dynamics in a geographical context, or, in other words, to produce a large-scale map showing the spatial distribution of the change rate and its character. Such an aim is poorly addressed by the cell-based approach (see section III). Thus, for example, the U.S. Geological Survey (USGS) Land Cover Trends project (http://landcovertrends.usgs.gov/) assessed LULC change over the U.S. at the level of ecoregions [25], [2]. This provides statistical synthesis of landscape

P. Netzel is with the Department of Climatology and Atmosphere Protection, University of Wroclaw, Kosiby 6/8, 51-621, Wroclaw, Poland.

T.F. Stepinski is with the Space Informatics Lab., University of Cincinnati, Cincinnati, OH 45221-0131, USA, e-mail: stepintz@uc.edu

Manuscript received xxx xx, 2014; revised xxx xx, 2014.

dynamics.

Our goal in this paper is to preset a pattern-based landscape change detection method that assesses landscape dynamics at spatial scale much finer than an ecosystem but much coarser than an individual cell. The method is intended for mapping landscape dynamics over continental or global extent. The method relies on a post-classification comparison; it uses LULC products, such as NLCD or CORINE as inputs. However, a basic spatial unit of comparison is not a cell characterized by its assigned LULC class but rather a locally-defined section (hereafter referred to as a tile) characterized by a pattern of LULC classes. LULC differences have often functional significance for landscape and pattern-based comparison addresses the question of landscape structural evolution rather than simple magnitude of change [42].

The importance of pattern analysis for characterization of LULC change (rather than its detection) has been recognized in previous works [16], [13], [45], [29], [38] where landscape indices [17] were used to characterize and compare LULC patterns at two different time steps. However, landscape indices are not well-suited for the single-valued comparison of multi-categorical patterns which is required for quantitative assessment of landscape change. This is because a large number of indices, all having different meanings, is necessary to describe such patterns without any guidance as to their relative contributions to the overall value of similarity. In the context of LULC, several methods for calculating the single-valued measure of similarity between patterns have been proposed including polygon-based fuzzy pattern matching [32], [42] and comparison of maps at multiple resolution [31]. These methods are best suited for assessing change on local scale as they are computationally expensive. They also lack rotational and translational invariance - a pattern and its rotated or shifted equivalent will be measured as being highly dissimilar.

Our methodology takes a different approach to quantification of similarity between patterns, one inspired by a domain of Content-Based Image Retrieval (CBIR). CBIR systems [12], [9], [22] are designed to query databases of natural images, but have also been applied to remotely sensed images [8], [7], [23], [39]. A method similar in spirit to a CBIR system was proposed by [3] for contextual reclassification of LULC maps to improve their accuracy in the urban environment. Because CBIR systems are designed for querying large datasets their similarity measures are computationally efficient. They are also invariant to rotation and translation. Recently, we have developed [20], [41] a Content-Based Map Retrieval (CBMR) system - a variant of CBIR especially designed for querying LULC maps (and other categorical rasters) rather than images. The GeoWeb implementation of our CBMR system for querying LULC patterns in NLCD 2006 is available online at http://sil.uc.edu/landex/. Here we have modified our CBMR system from its query and retrieval function (comparison of a reference pattern to all other local patterns at the same time step) to the change detection function (pair-wise comparison of spatially co-registered local patterns at two different time steps) and applied it to assess LULC changes over the conterminous United States using the NLCD 2001 and 2006. The result is a map showing the spatial distribution of the degree to which local LULC patterns have changed between 2001 and 2006. A zoomable web-based application showing the resultant map of change as well as the NLCD 2001 and 2006 maps in their full resolution is available at http://sil.uc.edu/dataeye/

The rest of this paper is organized as follows. Section II describes our pattern-based method of change detection. In Section III we demonstrate different modes of using our method and how they differ from a pixel-based methodology. Section IV presents and discusses the resultant U.S.-wide map of change. Discussion and future work directions are given in Section V.

II. Methodology

Our pattern-based LULC change assessment method relies on the post-classification comparison of local distributions of land cover class. As the focus is on large-scale assessment, potentially available input datasets are NLCD and CORINE. Another potential dataset is the MODIS MCD12Q1 product, however, this product lacks year-to-year consistency in category assignments and may not be feasible for change detection. In this paper we concentrate on the NLCD to assess landscape dynamics in 2001-2006 time period over the conterminous United States. The NLCD has a resolution of 30 m/cell. Each cell is labeled by one of $K = 16$ land cover classes. Note that we are using nominal NLCD class labels for our change assessment. The overall accuracy of class labels in NLCD 2001 and 2006 approaches 80% but it varies between different thematic classes and between various U.S. regions [43]. Most confusion occurs among the NLCD classes that contain grass in different contexts: "developed open space" (class 21), "grassland" (class 71), "pasture/hay" (class 81), "cropland" (class 82), and "emergent wetland" (class 95). In addition, the class "developed open space" is relatively highly confused with "developed low intensity" (class 22) and "shrub/scrub" (class 52). These inaccuracies need to be kept in mind when interpreting our results.

A spatial unit of change assessment is a local raster tile. A tile $\mathcal{A}$ is defined as a square-shaped local section of the NLCD having the size of $n \times n$ cells. A mosaic of land cover classes within a tile forms a local pattern. A change occurs if the local pattern alters from one time step to another (see Fig. 1); an overall degree of pattern dissimilarity between two time steps is taken as a degree of change. The tile's size determines the scale over which change is assessed. In this paper we use tiles with $n = 150$ (4.5 km $\times$ 4.5 km) but we have also run calculations using tile sizes of 15 km $\times$ 15 km and 30 km $\times$ 30 km. A tile size is an arbitrary choice but the change assessed over a larger area is statistically smaller than the change assessed over a smaller area, so maps created with larger tiles will show a smaller range of change. Using tiles smaller than 4.5 km $\times$ 4.5 km risks undersampling the landscape distribution.

We cover the entire U.S. with 1,684,540 4.5 km sized overlapping tiles. The purpose of the overlapping is to increase the spatial resolution of the change map resulting in its smoother appearance. Landscape changes in a continuous manner over

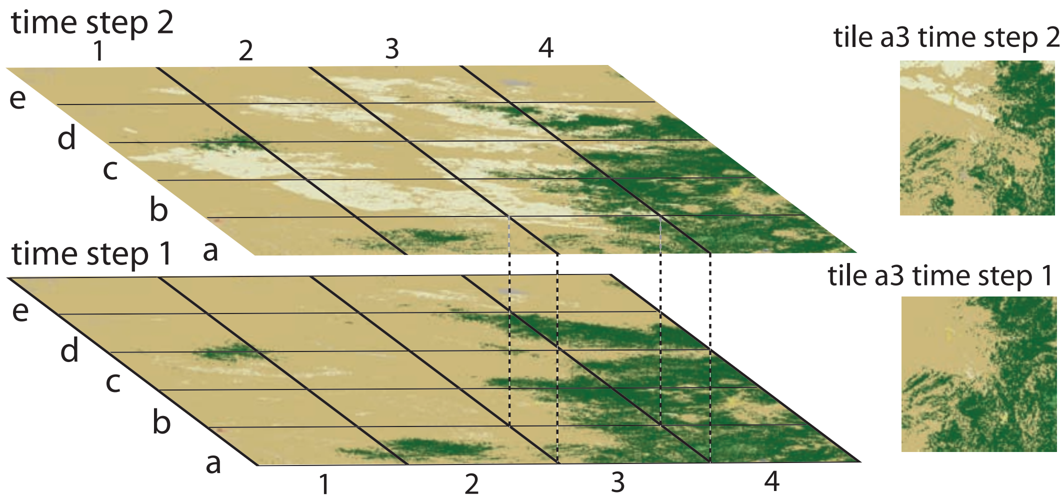

Fig. 1. The concept of pattern-based change assessment. LULC maps at two time steps are divided into spatially co-registered tiles. A dissimilarity between patterns within corresponding tiles is a measure of change. See Fig. 2 for the legend of land cover classes.

the U.S., so every point could be considered a center of a local landscape. We have chosen a grid of local landscape centers with grid spacing equal to 3 km, making neighboring tiles overlap each other. Change is determined by comparing 2001 vs. 2006 patterns at each local tile. The results of this comparison are stored in the 3 km/cell raster, which, upon visualization, yields a U.S.-wide map of change.

In order to calculate a dissimilarity value between a pair of local patterns at two time steps we use a method originally developed for querying the NLCD 2006 for locations having similar patterns [20], [41]. The method has two components, pattern signature and pattern dissimilarity. Pattern signature is a compact mathematical description of a pattern, and pattern dissimilarity is a function that assigns a numerical value to a pair of patterns on the basis of their respective signatures.

A. *Pattern signature*

For pattern signature we use a class/clump-size histogram constructed from the cells in the tile. A clump (also refereed to as a patch in context of landscape ecology) is a contiguous group of same-class cells. Segmentation of the NLCD into clumps is achieved using a connected components algorithm [34], [30]. A clump size is a number of cells in a clump. We quantize clump sizes by assigning them to bins with ranges based on powers of two (i.e. 1-2, 2-4, 4-8 etc). The number of bins, L, is determined by the size of the tile. In addition to its land cover class, each cell inherits a clump-size class from the clump to which it belongs. The 2D histogram of a tile‘s cells (with respect to LULC classes and clump-size classes) is a signature of a tile’s pattern. It estimates the joint distribution of land cover classes and clump sizes and is invariant to rotation and translation.

Fig.2 shows an example of constructing class/clump-size histograms from co-registered tiles showing an urban expansion between 2001 and 2006 in the suburbs of Las Vegas, Nevada. Fig.2A shows the two tiles with the size $n = 150$. Fig.2B shows the result of segmenting the two tiles into clumps. Fig.2C shows 2D class/clump-size histograms with 10 land cover classes arranged along the x-axis and 14 clump-size classes arranged along the y-axis. Note that only land cover classes present in at least one of the two tiles (in this case 10 out of possible 16 classes) are shown on histograms. The z-axis indicates a fraction of all the tile‘s cells that belong to a given class/clump-size bin. Because each histogram is normalized to unity it can be thought of as a probability density function (pdf) of a random variable X =(land cover class, clump-size class). Thus, in our method, calculating a dissimilarity between two local LULC patterns reduces to calculation of dissimilarity between two probability density functions. Note that a joint probability distribution (land cover class, clump-size class) can be marginalized with respect to the land cover by summing over clump-size bins yielding a 1D histogram reflecting just the bulk composition of a tile. For the location shown in Fig. 2 the most noticeable difference between 2001 and 2006 histograms is the reduction in cells belonging to the large clump-size “barren land class” bin and an increase in cells belonging to the medium clump-size“developed medium intensity” class.

B. *Similarity between two patterns*

The problem of calculating a dissimilarity value between two patterns has been well-researched in the CBIR domain. The challenge is to develop and use a dissimilarity measure that corresponds to how people perceive similarity. An additional constraint on the choice of dissimilarity measure is its computational cost; in our application we need to perform almost 1.7 million dissimilarity evaluations (see above) so a computationally efficient measure is needed. Given this constraint our choice of dissimilarity measure is restricted to those measures that do not take into consideration the non-overlapping parts of the two histograms. This means, for example, that semantic dissimilarities between land cover categories [1] cannot be taken into consideration in our present assessment. Within such constraints there is a large selection

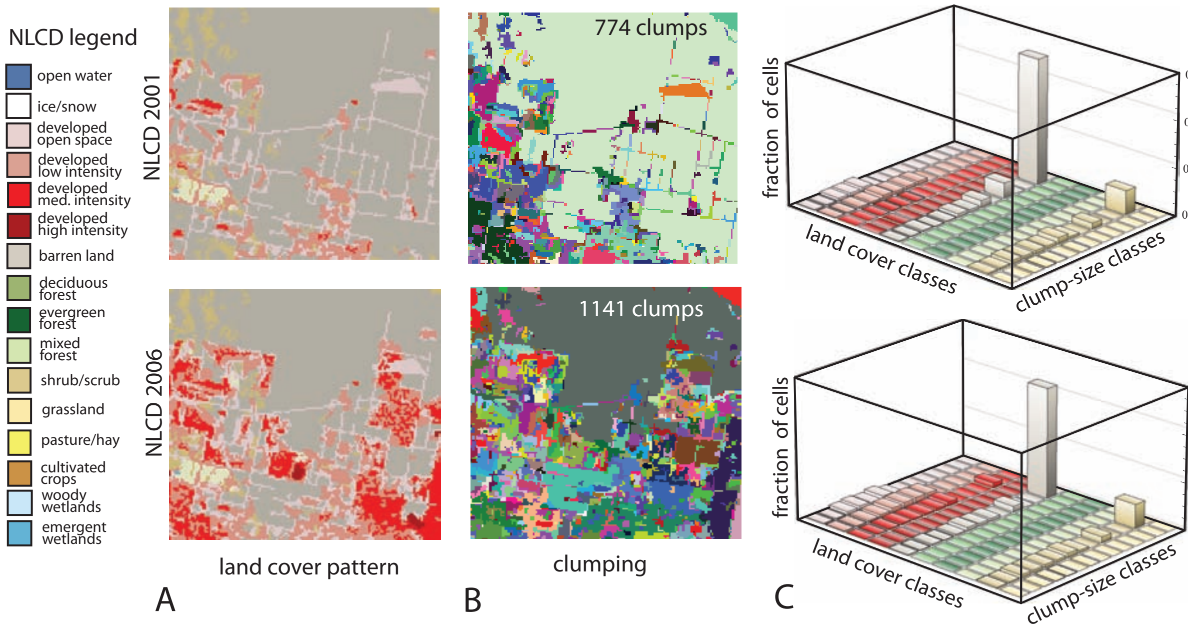

Fig. 2. Constructing pattern signatures. (A) Co-registered tiles of a NLCD at two time steps; different colors indicate different land cover classes as shown on the legend. (B) Tiles segmented into clumps; random (non-NLCD based) colors indicate individual clumps. (C) Class/size-clump histogram of the tiles.

of possible dissimilarity functions; for a comprehensive survey see [5]. We use the Jensen-Shannon divergence [24] to calculate dissimilarity between two histograms because of its robustness and good performance in the side-by-side comparison with other measures [35]. In this context "divergence" is synonymous with dissimilarity or distance - a quantitative degree of how far apart the two histograms are. For two histograms A and B (representing patterns at the same location at two different time steps) the Jensen-Shannon divergence (JSD) measures the deviation between the Shannon entropy [37] of the mixture of the two histograms $(A + B)/2$ (a histogram constructed by averaging corresponding bins of the two contributing histograms) and the mean of their individual entropies, and is given by

$$\mathrm{JSD}(A, B) = H\left(\frac{A+B}{2}\right) - \frac{1}{2}\left[H(A) + H(B)\right] \quad (1)$$

where $H(A)$ indicates a value of the Shannon entropy of the histogram A

$$H(A) = -\sum_{i=1}^{K}\sum_{j=1}^{L} A_{i,j} \log_2 A_{i,j}. \quad (2)$$

$A_{i,j}$ is a fraction of cells belonging to land cover class i and clump-size j. JSD is always defined, symmetric, bounded by 0 and 1, and equal to 0 only if $A = B$; $\sqrt{\mathrm{JSD}}$ has been proven [10] to be a metric and we use it to measure a "distance" between two patterns.

JSD can be interpreted as the mutual information between variable X having distribution $(A + B)/2$ and a binary indicator variable Z where $Z = 1$ if X is from pattern $\mathcal{A}$ and $Z = 0$ if X is from pattern $\mathcal{B}$. Mutual information gives an average reduction in unpredictability (entropy) of X if the pattern is set. The value of $H(A)$ reflects the distributional character of histogram A, a large value of $H(A)$ indicates A evenly spread between the bins (a scene where pixels are evenly distributed between patches of different sizes and LULC categories) whereas a small value of $H(A)$ indicates A concentrated in just few bins (a scene where pixels are concentrated in patches having just few sizes and/or LULC categories). JSD measures (in a single number) the difference between distributional characters of A and B. Note that if the two patterns, $\mathcal{A}$ and $\mathcal{B}$, have similar histograms the histogram of their mixture is similar to each of the two individual histograms and the value of JSD is small. If the two patterns have dissimilar histograms, the histogram of the mixture is more spread than each of the two original histograms and the value of JSD is large. A maximum difference, JSD=1, is assigned for two histograms where each is having only a single but different bin (two patterns each having a single but different land cover class).

We define the similarity between A and B (and consequently a similarity between the two patterns) as

$$\mathrm{JSS}(A, B) = \left[1 - \sqrt{\mathrm{JSD}(A, B)}\right] \quad (3)$$

where JSS stands for Jensen-Shannon Similarity. Both, JSD and JSS can be applied as well to the marginalized probability distributions (see previous sub-section) to assess changes in bulk composition of land cover classes. Small values of JSS indicate that tiles (their patterns or their bulk compositions) at two time steps are dissimilar so change has occurred. Large values of JSS indicate that the tiles are similar so no significant change has occurred.

An implementation of the pattern-based LULC change assessment methodology described in this paper is provided

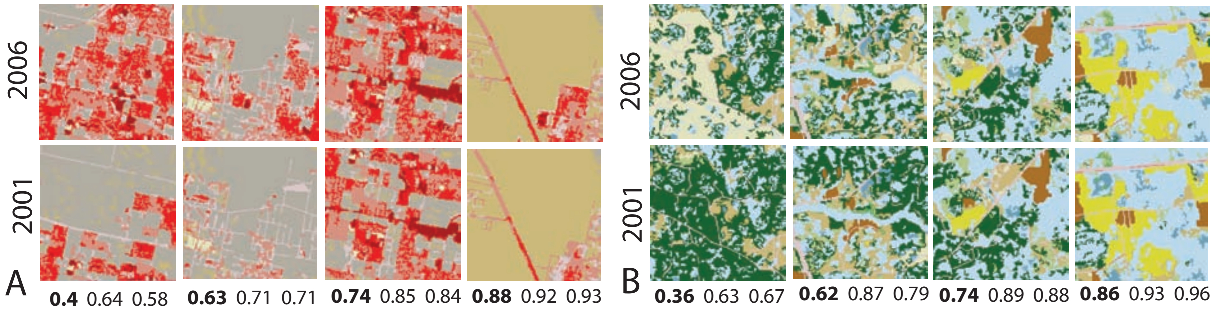

Fig. 3. Perceptual change in land cover and its numerical assessment. A) A series of four tiles in urban environments showing different perceptual degrees of change between 2001 and 2006. B) A series of four tiles in rural environments showing different perceptual degrees of change between 2001 and 2006. The numbers below each tile give the values of JSS (bold), JSS_1, and ρ, respectively. See Fig. 2 for the legend of land cover classes.

by the the Geospatial Pattern Analysis Toolbox (GeoPAT) - a collection of GRASS GIS modules that integrates various tools necessary for processing geospatial data using information derived from spatial patterns in categorical grids [19]. This software is freely available for download from http://sil.uc.edu/gitlist/geoPAT/.

III. Assessing change

Visual comparison of change maps with two layers of NLCD (for 2000 and 2006, respectively) using our online tool (see section 4.1) indicates that the JSS measure yields results that by large are in agreement with the perceptual notion of land cover change. Further verification by a diverse group of users is needed to confirm this conclusion. As a byproduct of our calculations we also compute two other potential measures of change, one is the JSS calculated for marginalized distributions (denoted by JSS_1) and the second is the fraction of unchanged cells in a tile (denoted by ρ). Whereas the JSS measures similarity with respect to pattern change, the JSS_1 measures similarity with respect to change in the bulk composition of land cover classes. The fraction of unchanged cells is calculated as the number of cells in a tile that keep the same land cover class at two time steps divided by the total number of cells in a tile. All three measures have a range between 0 (total change) to 1 (no change) but their meanings and values are different.

Fig. 3 gives examples of tiles that changed between 2001 and 2006 together with the values of JSS, JSS_1, and ρ. Fig. 3A pertains to urban environments and shows a series of four tiles characterized by progressively (from left to right) smaller perceptual change. The three numbers below each tile give (from left to right) the values of JSS (bold), JSS_1, and ρ. Fig. 3B shows similar comparison pertaining to rural environments.

In order to better understand the three different measures of change we constructed two scatter plots. Fig. 4A shows a scatter plot of ρ vs. JSS_1. The points on the scatter plot correspond to all the tiles covering the United States. The upper-right corner of the plot groups tiles that show very little change, whereas the lower-left corner of the plot groups tiles that show massive change. The tiles located between these two extremes show some degree of change. Note that the great majority of tiles are located near the upper-right corner as they experienced little change between 2001 and 2006. The tiles are centered around the diagonal of the plot indicating that, on average, both measures, ρ and JSS_1 have about the same sensitivity. In other words, maps of change constructed using these methods will have similar overall character. However, the scatter of tiles around the diagonal line indicates that at any given location ρ assesses change differently from JSS_1. These differences are especially pronounced for tiles located far from the diagonal. For example, in the tile denoted by the symbol P1 about 50% of cells have changed their land cover class between 2001 and 2006 but the bulk composition of land cover classes has not changed that much (as indicated by the high value of JSS_1). This tile, located in the state of North Carolina, is shown at two time steps in Fig. 5A together with corresponding histograms of the bulk composition of land cover classes. The changes in this tile are due to the cyclic nature of forest harvesting and regeneration resulting in a chess board-like pattern of forest, shrub, and grassland. The measure ρ corresponds to a gross change in a tile's landscape composition, whereas the measure JSS_1 encapsulates (in a single number) net changes of individual land cover classes [40].

Fig. 4B shows a scatter plot of JSS_1 vs. JSS. The values of JSS are always smaller or equal to the values of JSS_1 indicating that patterns are more susceptible to change than bulk compositions. This is expected as, in general, more conditions need to be met in order to match the geometry of a pattern than in order to match bulk composition. Thus, JSS is the most sensitive of the three change measures inasmuch as a change map calculated using the JSS will show more change than the maps calculated with ρ and JSS_1. The tiles located on the diagonal have the same degree of change in the bulk composition as in the overall geometric pattern. The tiles located increasingly below the diagonal are increasingly assessed as showing more change in pattern than in bulk composition. Tiles located toward the lower-right corner of the plot (like the tile denoted by the symbol P2) experienced a large change in the geometry of a pattern but a small change in the bulk composition. This tile, located in the state of Oregon, is shown at two time steps in Fig. 5B together with

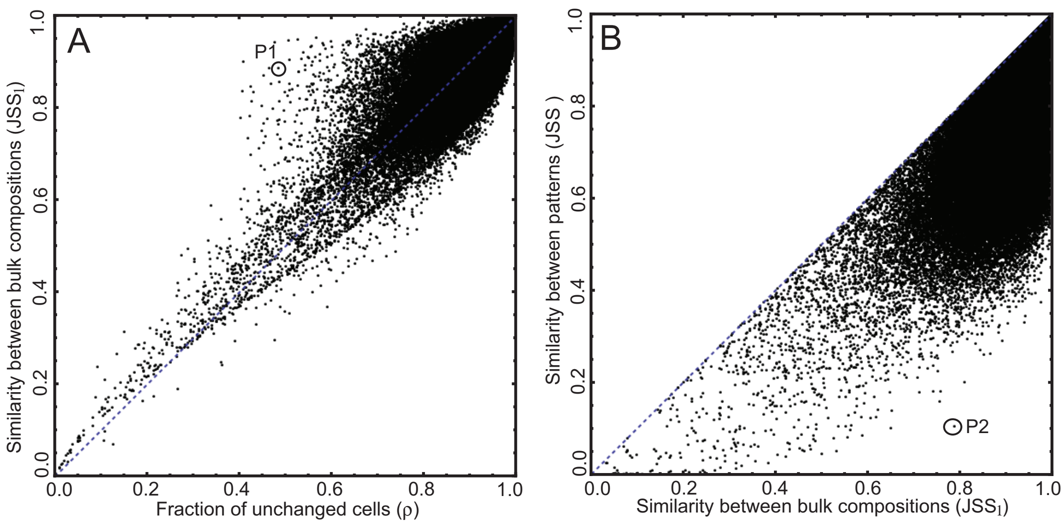

Fig. 4. Scatter plots comparing different measures of similarity between the tiles; all tiles covering the United States are plotted. A) The ρ vs. JSS_1 plot. B) The JSS_1 vs. JSS plot. Tiles denoted by symbols P1 and P2 are shown in Fig. 5 for in-depth comparison of change.

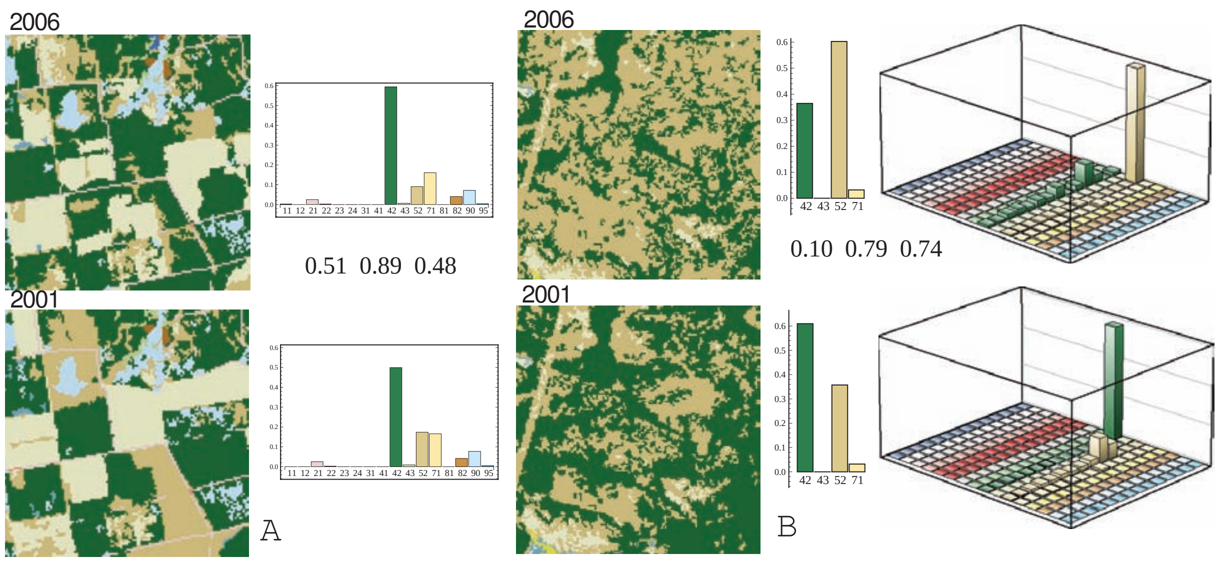

Fig. 5. Comparison of change in tiles P1 (A) and P2 (B). See main text for details and Fig. 2 for the legend of land cover classes. The numbers in the bulk composition histograms are standard labels of land cover classes. The numbers characterizing each tile give the values of JSS, JSS_1, and ρ, respectively.

corresponding histograms of the bulk composition and 2D histograms encapsulating the patterns. The changes in this tile are due to deforestation; in both years the site is dominated by forest and shrub, but the forest class is dominant in 2001 whereas the shrub class is dominant in 2006. Nevertheless, the histograms of bulk composition at these two time steps are not that different as their similarity is relatively high ($\text{JSS}_1 = 0.79$). Examining the 2D histograms reveals that in 2001 the forest was consolidated in a large clump while the shrub was scattered in smaller clumps, but in 2006 it is the shrub that is consolidated and the forest that is fragmented - hence large change in pattern has occurred (as indicated by the small value of JSS).

IV. Mapping Change over the Conterminous United States

For mapping the change over the entire conterminous United States we utilize a grid having dimensions of 1045 rows and 1612 columns. Each grid unit holds the JSS value calculated from NLCD 2001 and 2006 tiles centered on it. The unit is somewhat smaller than the tile because we want the tiles to overlap. Thus, whereas we use tiles having the size of $n = 150$ NLCD cells (4.5 km), we use grid units having the size of $n = 100$ NLCD cells or 3 km/unit. When interpreting the map it is important to remember that the value at the unit (pixel on the map) pertains to a change over somewhat larger area (150% $\times$ 150%) than the area of the pixel itself.

Fig. 6 shows the resultant map (see the next subsection on

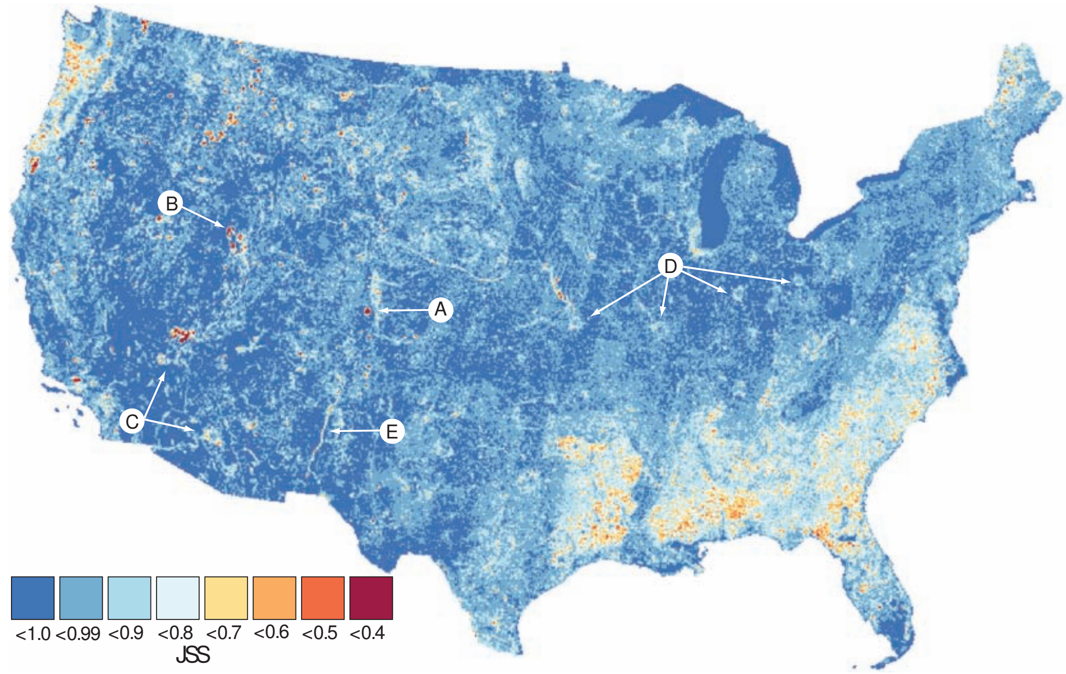

Fig. 6. Map of 2001-2006 land cover change over the conterminous U.S. constructed using the pattern-based approach and calculated using the JSS measure. See the main text for description of selected regions denoted by letters from A to E.

how to access our online tool designed to explore this map in detail). The color palette has been chosen to span from deep blue for large values of JSS (no change or insignificant change) to deep red for small values of JSS (significant change). The goal of the map is to show a spatial distribution of the degree to which the landscape has changed between 2001 and 2006 rather than to indicate specific transitions. The overall first impression is that most of the U.S. experienced little land cover change between 2001 and 2006 as the blue color dominates the map. However, there are notable regional exceptions including the southeastern and Gulf regions, the Pacific Northwest region, and the state of Maine. These regions show higher levels of landscape dynamic than the rest of the country. This is roughly consistent with the NLCD 2001-2006 change mapped [11] at the very coarse level of Landsat path/row scene (approximately 183km $\times$ 170km extent). This is also in agreement with an assessment of land cover change in the period of 1973-2000 [2] using sampling of ecoregions methodology. The ecoregions-based change map shows that the regions identified on our map as having relatively high levels of change in the period of 2001-2006, had also relatively high levels of change during the preceding period of 1973-2000. The ecoregions-based map also shows that additional regions including Central Great Plains, Southwestern Tablelands, Middle Rockies, Snake River Basin, and Columbia Plateau had somewhat elevated levels of change during the period of 1973-2000. Our map shows no significant change in these regions in the period of 2001-2006. This may be because the change in these regions has stopped or slowed down in the 2001-2006 period or because our method is not detecting change in regions composed mostly of NLCD classes with heightened levels of confusion between label assignment (see section II).

With the resolution of 3 km/cell our map can identify change on much finer spatial scale than an ecoregion or Landsat path/row scene. To fully appreciate the level of detail one has to use the online tool or download the full resolution map (http://sil.uc.edu/downloads.html#maps). However, even within the constraints imposed by the static medium of conventional figures, some individual specific locations of change can be identified on Fig. 6. Examples of such locations are highlighted on Fig. 6 by letters A to E and shown in details in Fig. 7.

Many red specks visible on the map in the western part of the U.S. are changes related to forest fires. In particular, a prominent red speck denoted by the letter A is the location of the 2002 Hyman fire [14]. Fig. 7A shows that this area has been covered by a well-consolidated forest in 2001, but the 2006 map shows a scar left by the fire. The location denoted by the letter B coincides with the Great Salt Lake in Utah. The close-up in Fig. 7B shows that the lake retreated from its 2001 levels to expose more “barren land” in 2006 - a dynamic that was registered as a ring of strong change on our map. Locations denoted by the letter C correspond to urban growth in the cities of Las Vegas, NV and Phoenix, AZ. Fig. 7C shows that in just five years the growth in Las Vegas has been significant enough to be noted even directly by comparing the two land cover maps, but much more clearly on the change map. Locations denoted by the letter D show (from left to right) cities of Kansas City, St. Louis, Indianapolis, and Columbus. In contrast to Las Vegas and Phoenix they show up on our map as rings of pale yellow instead of rings of orange and red indicating a small level of change. These cities did not experience growth to the same degree as Las Vegas or Phoenix did, but they had some level of urban development

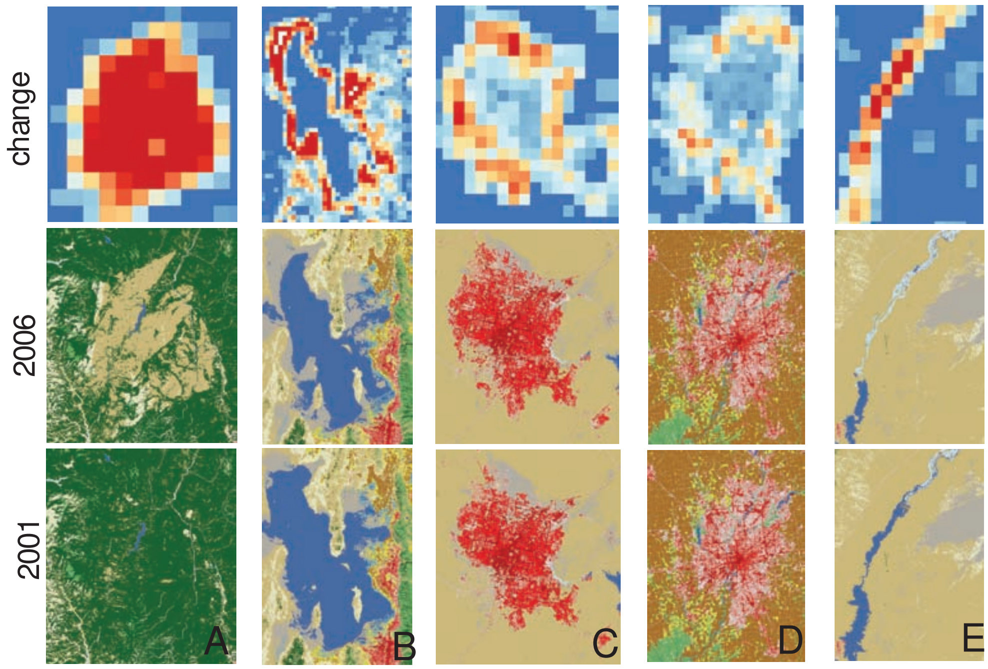

Fig. 7. Detailed view of selected regions highlighted on Fig. 6. The first row is the map of change, the other two rows are LULC maps in 2006 and 2001 respectively. See Fig. 2 for the legend of land cover classes and Fig. 6 for JSS color scheme. Regions are shown at different scales.

in their suburbia. Fig. 7D shows the city of Indianapolis; change cannot be detected easily from visual inspection of the land cover maps, but it shows clearly on the change map. Closer inspection using our online tool reveals details of development responsible for the change. Linear features on the map correspond to rivers and their surroundings. In particular, the letter E denotes the Rio Grande river in New Mexico. The portion of the Rio Grande river and the associated change in land cover between 2001 and 2006 are shown in Fig. 7E. The changes can be explained by fluctuating water levels.

A. *Online tool for exploring maps of landscape dynamics*

The best way to fully explore all three of our maps (based on JSS, JSS_1, and ρ) while being able to compare them with actual land cover in NLCD 2001 and 2006, is to use our GeoWeb application DataEye (http://sil.uc.edu/dataeye/). DataEye is a computerized map application with all expected functionalities. A user can navigate through the map, zoom to the region of interest, change layers (the three change maps and the two time steps of land cover map), set transparency, and download the portion of the map visible in the application window to a file in the GeoTiff format. The DataEye allows for rapid exploration of locations that experienced LULC change between 2001 and 2006 and for investigating their underlying causes.

V. Discussion and future directions

The methodology presented in this paper addresses the need for assessing land cover change on a continental scale and with high spatial resolution. Simultaneously fulfilling these two requirements calls for a new methodological approach as well as for a presentation of results in the form of interactive media.

Given that in our approach change is defined over the tile, we introduce two new measures of change encapsulating either a difference in land cover patterns within the tile between the two time steps (JSS), or a difference in tile land cover composition between the two time steps (JSS_1). Both of these measures are defined as similarities between probability distribution functions and thus are fundamentally different from more conventional measures based on the percentage of unchanged pixels (ρ). We argue that JSS is the best all-purpose measure of change, but for some specific aims one of the other two measures may be more appropriate. Moreover, new and interesting information can be revealed by comparing different measures. As demonstrated in Section III tiles characterized by a significant number of cell-based class transitions (relatively small values of ρ) and a small change in composition (large values of JSS_1) correspond to places where patterns of landscape alternate without much change in bulk composition. Tiles characterized by a small change in composition (large values of JSS_1) and a large change in pattern (small values of JSS) correspond to places where the geometry of the landscape changes without much change in composition. Identifying the spatial distribution of such places can be done by processing the maps of JSS, JSS_1, and ρ but this is beyond the scope of this paper.

We have also addressed the issue of presentation of our results by making available a fully-featured online application for fast and convenient exploration of our change maps. Using the online application one can find the places in the U.S. that have changed land cover (within the meaning of our three definitions) between 2001 and 2006 and to understand the reason for the change.

Our methodology has also room to grow. In particular, future research will explore the possibility of incorporating similarities between LULC classes to our similarity measure. Our formalism assumes that LULC classes are independent from each other. This is the standard assumption when dealing with categorical variables like LULC as calculating similarity between instances (LULC classes) must involve a significant dose of arbitrariness. However, it is clear that some LULC classes are more similar than others and [1] has quantified those dependencies using semantic similarities between the classes. Utilizing semantic similarities and the so-called Earth Mover's Distance (EMD) [36] – a measure of dissimilarity between two pattern signatures that can take into consideration dependencies between LULC classes – we would be able to recalculate the change maps to better reflect perceptual notions of change. Moreover, the EMD allows calculation of distance between two patterns having different LULC categories as long as a matrix of semantic similarities between two different legends is defined. Because [1] provides such a matrix for NLCD 2001 and NLCD 1992, we would be able to construct a 1992-2001 change map using the EMD. Finally, we will construct a 2006-2011 change map utilizing newly released NLCD 2011 data.

ACKNOWLEDGMENT

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

[1] O. Ahlqvist. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 US National Land Cover Database changes. *Remote Sensing of Environment*, 112(3):1226–1241, 2008.

[2] R. F. . Auch, M. A. . Drummond, K. L. . Sayler, A. L. . Gallant, and W. Acevedo. An Approach to Assess Land-Cover Trends in the Conterminous United States (1973-2000). In C. P. . Giri, editor, *Remote Sensing of Land Use and Land Cover Principles and Applications*, pages 351–368. CRC Press, 2012.

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

[4] S. Boriah, V. Mithal, A. Garg, M. Steinbach, V. Kumar, C. Potter, S. Klooster, and J. C. Castilla-Rubio. Automated detection of forest cover changes. In *Geoscience and Remote Sensing Symposium (IGARSS),*, pages 44–47. 2010 IEEE International, 2010.

[5] S. Cha. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307, 2007.

[6] P. Coppin, I. Jonckheere, K. Nackaerts, and B. Muys. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sensing*, 25(9):1565–1596, 2004.

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

[8] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia. Information mining in remote sensing image archives: System concepts. *IEEE Trans. Geosci. Remote Sens.*, 42(12):2923–2936, 2003.

[9] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 40:1–60, 2008.

[10] D. M. Endres and J. E. Schindelin. A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7):1858–1860, 2003.

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

[12] T. Gevers and A. W. Smeulders. Content-based image retrieval: An overview. In G. M. S. B. Kang, editor, *Emerging Topics in Computer Vision*, chapter 8, pages 333–384. Upper Saddle River, NJ: Prentice-Hall, 2004.

[13] F. Giordano and A. Marini. A landscape approach for detecting and assessing changes in an area prone to desertification in Sardinia (Italy). *International Journal of Navigation and Observation*, Article ID 549630, 2008.

[14] R. T. Graham. Hayman Fire case study: summary. Technical report, US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2003.

[15] M. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, and et al. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.

[16] M. Herold, J. Scepan, and K. C. Clarke. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A*, 34(8):1443–1458, 2002.

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

[18] T. P. Hollenhorst, L. B. Johnson, and J. Ciborowski. Monitoring land cover change in the Lake Superior basin. *Aquatic Ecosystem Health & Management*, 14(4):433–442, 2011.

[19] J. Jasiewicz, P. Netzel, and T. Stepinski. Retrieval of pattern-based information from giga-cells categorical rasters - concept and new software. In *IEEE International Geoscience and Remote Sensing Symposium 2014*, 2014.

[20] J. Jasiewicz and T. F. Stepinski. Example-Based Retrieval of Alike Land-Cover Scenes From NLCD2006 Database. *Geoscience and Remote Sensing Letters*, 10:155–159, 2013.

[21] G. Jianya, S. Haigang, M. Guorui, and Z. Qiming. A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37.B7:757–762, 2008.

[22] M. Lew, N. Sebe, C. Lifi, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput., Commun., Appl.,*, 2(1):1–19, 2006.

[23] J. Li and R. M. Narayanan. Integrated spectral and spatial information mining in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.*, 42(3):673–685, 2004.

[24] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 31(1):145–151, 1991.

[25] T. R. Loveland, T. Sohl, S. Stehman, A. Gallant, K. Sayler, and D. Napton. A strategy for estimating the rates of recent United States land-cover changes. *Photogrammetric Engineering and Remote Sensing*, 68(10):1091–1099, 2002.

[26] D. Lu, E. Mausel, E. Brondizio, and E. Moran. Change detection techniques. *Int. J. Remote Sensing*, 25(12):23652407, 2009.

[27] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, 105(2):142–154, 2006.

[28] V. Mithal, Z. O'Connor, K. Steinhaeuser, S. Boriah, V. Kumar, C. S. Potter, and S. A. Klooster. Time series change detection using segmentation: A case study for land cover monitoring. In *In Intelligent Data Understanding (CIDU)*, pages 63–70. IEEE, 2012.

[29] S. Narumalani, D. R. Mishra, and R. G. Rothwell. Change detection and landscape metrics for inferring anthropogenic processes in the greater EFMO area. *Remote Sensing of Environment*, 91(3):478–489, 2004.

[30] P. Netzel and T. F. Stepinski. Connected Components Labeling for Giga-Cell Multi-Categorical Rasters. *Computers & Geosciences*, 2013.

[31] R. G. Pontius. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering & Remote Sensing*, 68(10):1041–1049, 2002.

[32] C. Power, A. Simms, and R. White. Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *Int. J. of Geographical Information Science*, 15(1):77–100, 2001.

[33] J. Rogan and D. Chen. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, 61:301–325, 2004.

[34] A. Rosenfeld and J. L. Pfaltz. Sequential operations in digital processing. *J. ACM*, 13:471–494, 1966.
[35] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding*, 84:25–43, 2001.
[36] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
[37] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
[38] Y. Shi, J. Xiao, and Y. Shen. Landscape pattern change and associated environmental implications in the Haihe River Basin, China. *ISPRS J. Photogramm. Remote Sens.*, 37:569–574, 2008.
[39] C.-R. Shyu, M. Klaric, G. Scott, A. S. Barb, C. Davis, and K. Palaniappan. GeoIRIS: Geospatial information retrieval and indexing systemContent mining, semantics modeling, and complex queries. *IEEE Trans. Geosci. Remote Sens.*, 45(4):839–852, 2007.
[40] B. M. Sleeter, T. Sohl, T. Loveland, R. Auch, W. Acevedo, M. Drummond, K. Sayler, and S. Stehman. Land-cover change in the conterminous United States from 1973-2000. *Global Environmental Change*, 23(4):733–748, 2013.
[41] T. F. Stepinski, P. Netzel, and J. Jasiewicz. 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, 2014.
[42] R. White. Pattern based map comparisons. *Journal of Geographical Systems*, 8(2):145–164, 2006.
[43] J. D. Wickham, S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry, and T. G. Wade. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment 130 (2013): 294-304.*, 130:294–304, 2013.
[44] G. Xian, C. Homer, B. Bunde, P. Danielson, J. Dewitz, J. Fry, and R. Pu. Quantifying urban land cover change between 2001 and 2006 in the Gulf of Mexico region 27, no. 6 (2012): 479-497. *Geocarto International*, 27(6):479–497, 2012.
[45] X. Yang and Z. Liu. Quantifying landscape pattern and its change in an estuarine watershed using satellite imagery and landscape metrics. *Int. J. of Remote Sensing*, 26(23):5297–5323, 2005.
[46] X. Zhan, R. DeFries, J. R. Townshend, C. Dimiceli, M. Hansen, C. Huang, and R. Sohlberg. The 250 m global land cover change product from the Moderate Resolution Imaging Spectroradiometer of NASA's Earth Observing System. *Int. J. of Remote Sensing*, 21(6-7):1433–1460., 2000.
[47] X. Zhan, R. A. Sohlberg, J. R. Townshend, C. DiMiceli, M. L. Carroll, J. C. Eastman, M. C. Hansen, and R. S. DeFries. Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment*, 83(1):336–350, 2002.

Pawel Netzel is a mathematician, climatologist, and programmer. He works at the University of Wroclaw, Poland in the Dept. of Climatology and Atmosphere Protection. He promotes and develops Open Source Software for GIS. He is one of the founder-members of OSGeo chapter Poland. His interest includes spatial analysis, especially with applications of AI and adaptive systems, geographical web services, and tele-detection methods of atmospheric boundary layer. His E-mail address is netzelpl@ucmail.uc.edu

Tomasz Stepinski is the Thomas Jefferson Chair Professor of Space Exploration at the University of Cincinnati. He is a geoscientist with a PhD degree in Applied Mathematics from the University of Arizona. His recent area of research is automated identification, characterization, and classification of landforms and content-based query and retrieval of landscapes. He has established the Space Informatics Lab (http://sil.uc.edu/) at the University of Cincinnati. His E-mail address is stepintz@uc.edu