Article

Using Self Organizing Maps to Visualize Structural Change in Geographic Regions of Interest: Las Vegas across the Years

John M. Wandeto [1,2], Henry O. Nyongesa [2], Yves Remond [1] and Birgitta Dresp-Langley [1,*]

[1] ICube UMR 7357 CNRS - University of Strasbourg, France; birgitta.dresp@unistra.fr
[2] Dedan Kimathi University of Technology, Nyeri, Kenya; ndetos@yahoo.com
* Correspondence: birgitta.dresp@icube.unistra.fr

Abstract: Time-series of satellite images reveal important information about changes in environmental conditions and natural or urban landscape structures that are of potential interest to citizens, historians, or policymakers. We applied a fast method of image analysis using Self Organized Maps (SOM) and, more specifically, the quantization error (QE), for the visualization of critical changes in satellite images of Las Vegas, generated across the years 1984-2009, a period of major restructuration of the urban landscape. The satellite images were subdivided into geographic regions of interest. As shown in previous work by the authors, the QE from the SOM output is a reliable measure of variability and local changes in image contents. In the present work, we show how the QE from SOM run on satellite images of specific geographic regions of interest can be exploited for visualizing structural change across time at a glance, and facilitate the interpretation of related demographic data for a specific time period. The method is fast and reliable, and can be implemented for a rapid detection of critical changes in contents of time series of large bodies of satellite images.

Keywords: satellite images; image analysis; self organizing maps; quantization error; structural change; demographic data

1. Introduction

The analysis of time series of satellite images by computational methods or algorithms represents a highly complex challenge. The process of detection and characterization of changes in such images, reflecting potentially critical changes in public spaces of the natural or the built environment, is less than straightforward. It most often relies on human visual inspection and classification of expensive *ad-hoc* remote sensing imagery. The main problem is the lack of automatic techniques for discriminating between changed and unchanged contents in satellite images and discrimination is usually performed by using empirical strategies or manual trial-and-error procedures. These affect both the accuracy and the reliability of the whole change detection process. Given the cost, time, and organizational challenges involved, information from satellite images is rarely exploited further and not made useful to the general public. Yet, the automatic tracking and harvesting of landscape information could provide inexpensive, ready-to-use, and reliable information to the public, historians and scientists in general, administrators, and policy makers and could contribute substantially to improving both the quality and the timeliness of public resource management.

In this paper here, we introduce a fast, unsupervised change detection technique based on the functional principles of self-organizing maps (SOM), first introduced by Kohonen [1]. A whole satellite image, or specific sets of x,y image coordinates representing specific geographic regions of interest within a satellite image, were used as input to SOM. Image input is exploited directly without additional or intermediate procedures of analysis, bridging a gap between classic machine learning and some of the traditional methods of geographic image analysis [2, 3].

 doi:10.20944/preprints201710.0166.v1

Our previous work [4] had shown that a specific output variable of the SOM, the quantization error (*QE*), can be exploited as a diagnostic indicator for the presence of potentially critical local changes in medical image contents. In the present work, to further highlight the potential of this new computational approach, we used the *QE* output from SOM analyses of satellite images of Las Vegas, generated across the years 1984-2009. This period is of particular interest as a major time of structural changes in the urban landscape of Las Vegas. As a reliable indicator of density, variability and local change in image contents [4], the QE is used here to reflect structural changes that took place in Las Vegas across these critical years: in the town as a whole, and in five of its major subdivisions. We also show how the QE output may be used further to meaningfully highlight demographic data for the same critical time period.

2. Results

In a first analysis, four-by-four SOM with 16 neurons was run on each satellite image from a given year as a whole. The raw data in terms of QE output from this analysis are provided here below in Table 1. The principles of implementation of SOM are well-documented [1, 6]. The code used for the SOM analyses is provided in the Supplementary Materials section of this manuscript as a text file [S1 code.txt]. In a second analysis, the satellite images of Las Vegas and each of the corresponding Self-Organizing Maps were subdivided into 16 different image regions corresponding to 16 neurons in each SOM. The x-y coordinates of these image regions are used as image labels for the four main divisions of Las Vegas, as explained and illustrated further in the Materials and Methods section. The raw data, in terms of QE output, from these analyses are provided in the Supplementary Materials section here [S1 table1.xls].

2.1. Results from the global SOM on whole images as a function of time

The results of the global four-by-four SOM run on each of the six satellite images changes produced quantization errors (QE) that are consistent with the changes in building density across the whole of Las Vegas in the years of major restructuration (1984-2009). Photographic snapshots taken across this period of parts of the "Strip", the central artery of Las Vegas where most attractions are located, illustrate some of these changes (Fig 1).

Table 1: QE values from global SOM as a function of time.

Year	*QE*
1984	255.6264
1989	261.9832
1994	262.5629
1999	248.9702
2004	238.3870
2009	221.4607

These changes are consistently reflected by 1) an increase in the *QE* between 1984 and 1994 (increase in the variability of building structures due to partial destructions for reorganization), and a progressive decrease in the *QE* thereafter (progressive homogenization of building structures due to the new developments), as shown here above in Table 1.

 doi:10.20944/preprints201710.0166.v1

Figure 1: Photographic snapshots of parts of "The Strip" across the years 1982-2010. These photos, taken in 1982, 1995, 2005 and 2010 respectively, give some idea of the structural changes that took place in the time period when the satellite images retrieved for this study here were generated by NASA.

2.2. Results from SOM on different subsections of the satellite images

The results, in terms of QE output of the SOMs for the 16 different image regions corresponding to each of the 16 neurons in each SOM are shown here below in Figure 2 as a function of the year in which the image was taken. The variations in the QE shown here reflect varying pixel density in the images across time indicative of the major changes in structure of the urban landscape in the different regions across time.

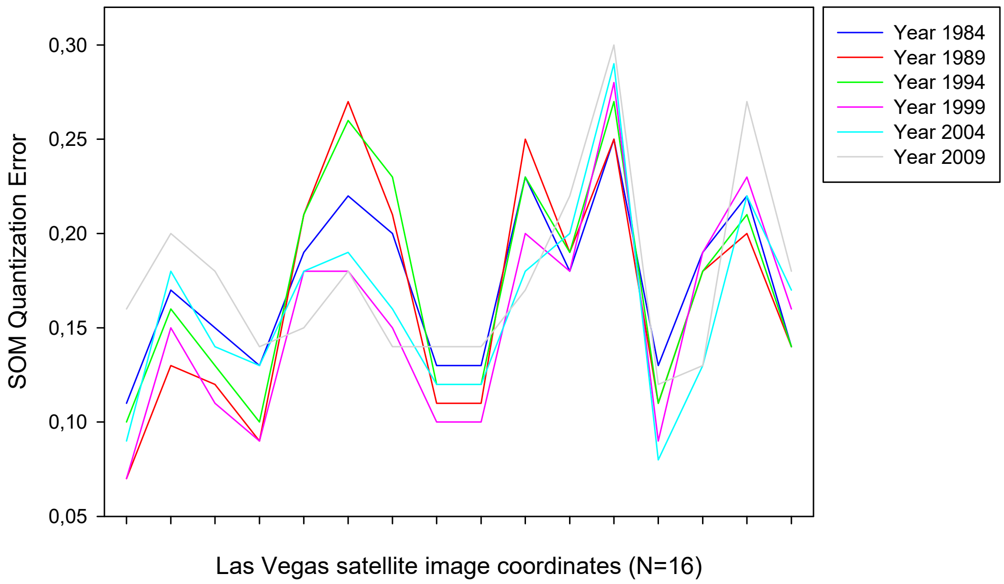

Figure 2: Variations in QE from SOM on 16 different image regions corresponding to different parts of Las Vegas are shown here as a function of the year when the image was taken.

In a next step, the QE values for each of the 16 image regions were plotted as a function of time. The results of this analysis are shown here below (Fig 3). In the SOM, each geographic region corresponds to specific x, y image coordinates in the mapping algorithm, as explained in greater detail in the Materials and Methods section. Here, data are shown as a function of the major geographic regions of Las Vegas corresponding to each of the 16 image regions. The results (Fig 3) clearly show different trends in QE as a function of time, reflecting meaningful differences in the way these regions have changed during the critical time period when the satellite images were taken. The geographic regions of interest that underwent the largest urban restructuration in the reference period between 1984 and 2009 were the North and South of the City Center around the Strip with its megastructures (casinos and hotel towers), the nearby East with shopping malls and apartment blocks in close proximity to the Strip and the airport, and the much sought after residential North where real estate prices were hitting the roof until the financial crisis in 2008. Reconstruction in these areas was booming in the reference period,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2017 doi:10.20944/preprints201710.0166.v1

resulting in densely built-up urban landscapes. This trend (Fig 3) is reflected here by the progressive decrease in QE from SOM on the corresponding satellite image regions for images taken between 1994 and 2009.

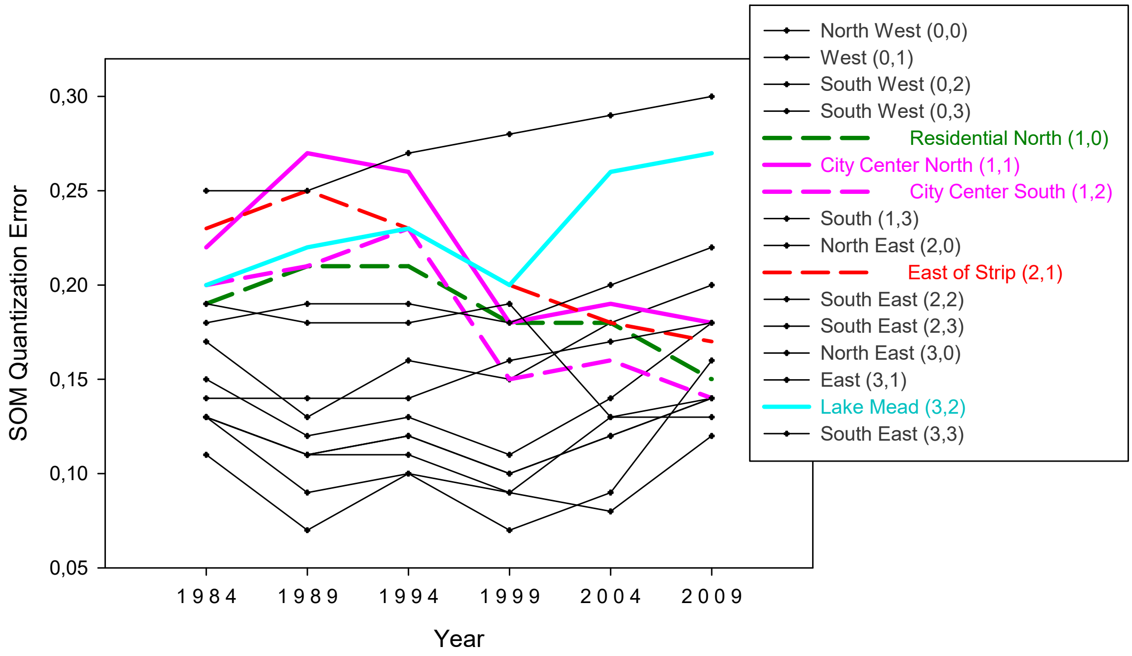

Figure 3: QE from SOM for each the 16 different image regions as a function of the years of the reference time period of this study.

Unlike the areas around the Strip, the central artery of Las Vegas, with increasing building density reflected by increasingly homogenous image regions captured by the steady decrease in QE from SOM on these image regions, other geographic regions of interest have evolved in a radically different way. This is shown in the results here (Fig) on the example of image region (3, 2), which corresponds to an area to the South East of Las Vegas City where Lake Mead pushes against the curved walls of Hoover Dam, built in 1935 to collect water from the Colorado River. The area around Lake Mead has become increasingly arid across the years of the reference period of this study here, and water levels have rather abruptly decreased between years 1999 and 2009. These changes are reflected by increasingly heterogeneous satellite image contents for this region, captured in the results here (Fig 3) by a sharp increase in QE from SOM on the corresponding image region (region with image coordinates (3, 2)).

2.3. Geographic regions of interest and statistics across time

Population statistics provided by the local government authorities as a function of the years when the satellite images were taken show that the estimated population total for Las Vegas increased (Fig 4) across these years of the reference period of this study. This trend can be directly linked to the increasing real estate developments in the East of Las Vegas nearby the Strip, and in the up-market residential North, reflected here by a gradual decrease in the QE

 doi:10.20944/preprints201710.0166.v1

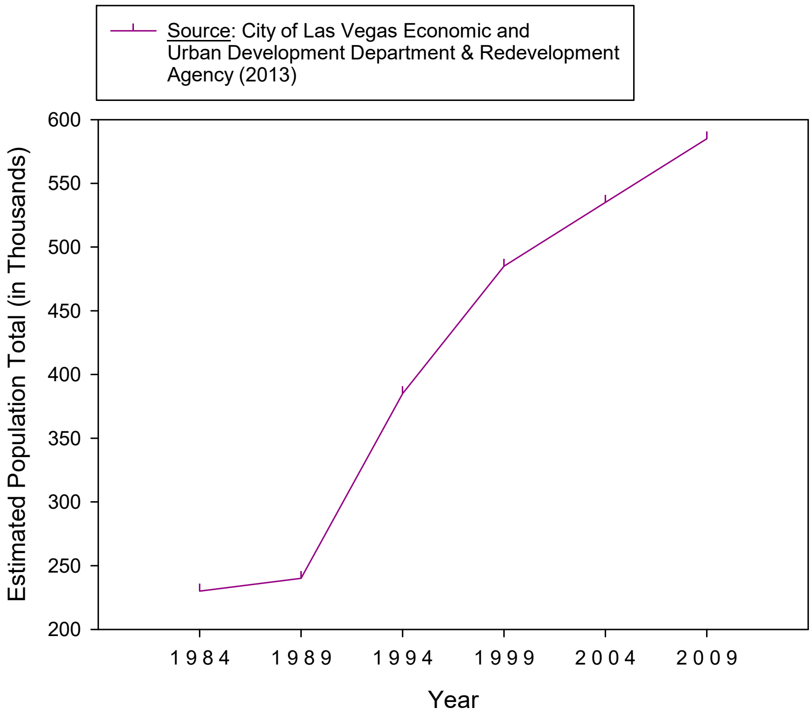

Figure 4: The estimated population total for Las Vegas provided by the local government authorities as a function of the years when the satellite images were taken.

from SOM on the corresponding image regions (2,1) and (1,0), respectively (Fig 3), across the years of the reference period. Other geographic regions of interest where changes detected in the satellite images by using the QE indicator can be meaningfully linked to government statistics are the City Center and the region around Lake Mead.

2.3.1. Las Vegas City Center

The average QE from SOM on image regions (1,1) and (1,2), corresponding to geographic regions of interest alongside the Strip, is shown to decrease across the years of the reference period (Fig 4), indicating increasing homogeneity of the satellite image regions as a result of increasing building developments for commercial purpose (new casinos, hotels, restaurants and other businesses) alongside the Strip during these years. These developments resulted in an increase in human mobility and local income, as demonstrated by a certain number of demographic facts and figures. For example, the estimated total number of visitors of Las Vegas provided by the local government authorities as a function of the years when the satellite images analyzed here in this study were taken was found to increase across these years (Fig 6). The gaming revenue total in millions of USD provided by the local government authorities also increased consistently (Fig 7) during the reference period, with a sudden drop between 2004 and 2009 coinciding with the time period of the economic crisis that culminated in the Wall Street banking scandal of 2008.

2.3.2. Lake Mead

The geographic region of interest around Lake Mead image and Hoover Dam corresponds to the image region with x, y coordinates (3, 2) in the satellite images analyzed here. The natural landscape around Lake Mead has become increasingly arid across the years of the reference period due to climate change. Water levels have dwindled away drastically between the years 1999 and 2009, causing the connection of Saddle Island to the shoreline.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2017 doi:10.20944/preprints201710.0166.v1

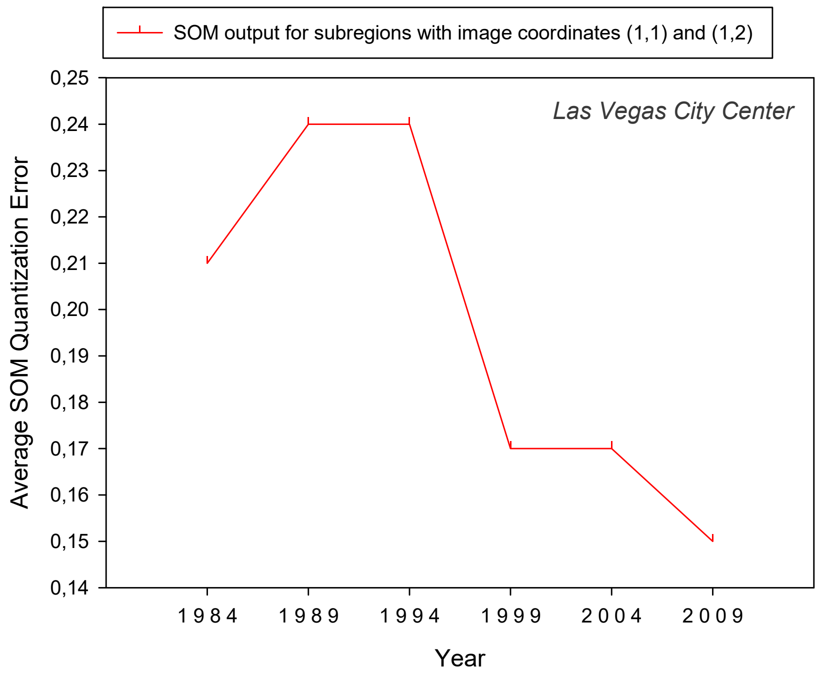

Figure 5: The steep and stepwise decrease of average QE from SOM on image regions (1,1) and (1,2), corresponding to the City Center with the Strip, between the years 1994 and 2009 of the reference period. The trend reflects the increasing homogeneity of these satellite image regions with time.

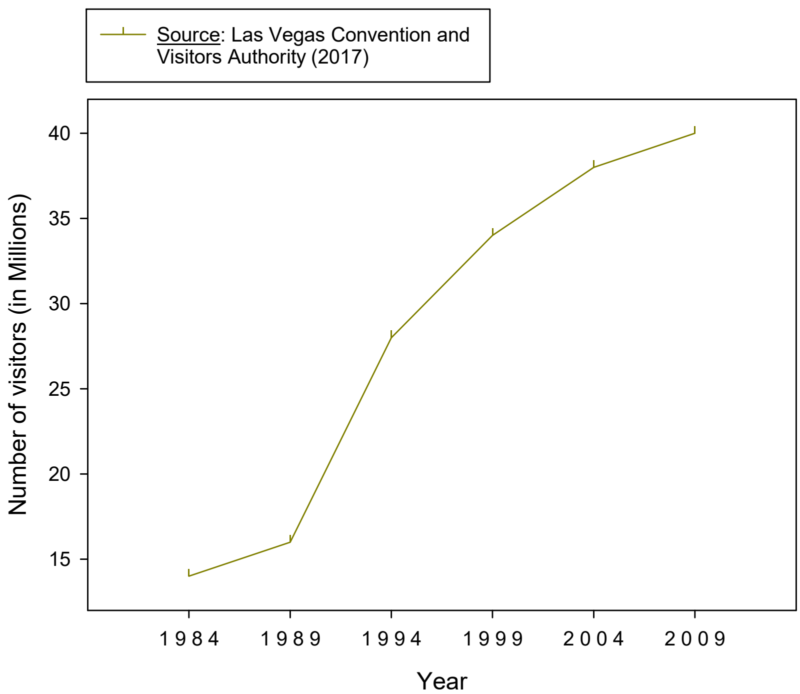

Figure 6: The estimated total number of visitors of Las Vegas provided by the local government authorities as a function of the years when the satellite images were taken.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2017 doi:10.20944/preprints201710.0166.v1

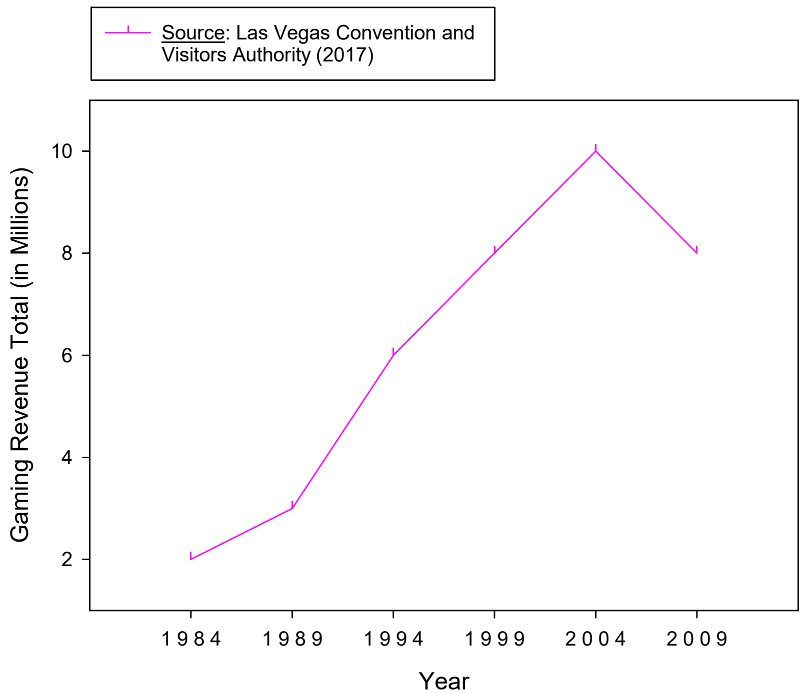

Figure 7: The gaming revenue total in millions of USD provided by the local government authorities as a function of the years when the satellite images were taken. The sudden drop between 2004 and 2009 coincides with the time period of the economic crisis that culminated in the Wall Street banking scandal of 2008.

The changes in the natural landscape around Hoover Dam and Lake Mead are reflected by increasingly heterogeneous satellite image contents in this specific geographic region of interest, captured in the results by a sharp increase in QE from SOM on the image region with the x, y coordinates (3, 2) between the years 1999 and 2009 (Fig7).

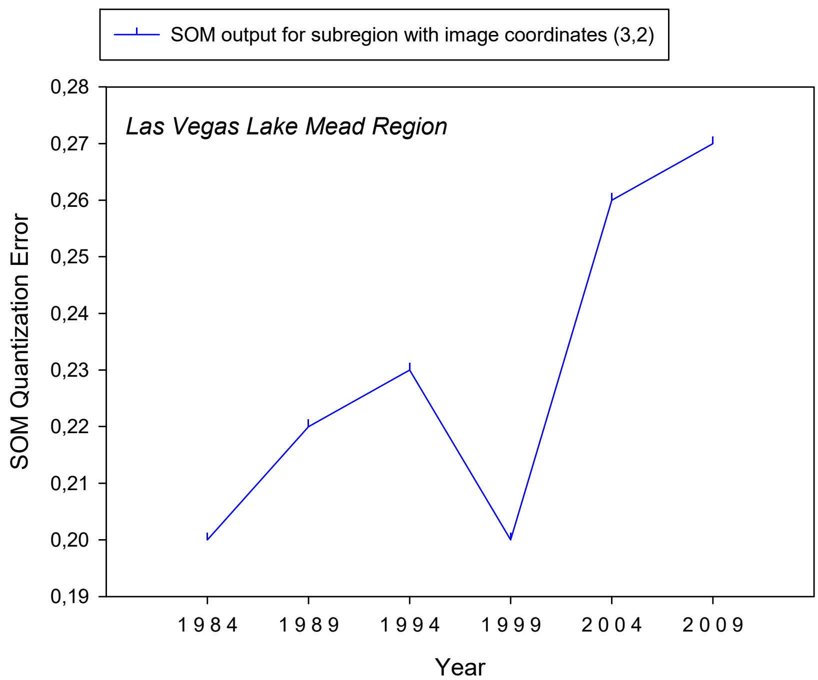

Figure 8: The sharp increase in QE from SOM on image region (3, 2) between the years 1999 and 2009 reflecting increasingly heterogeneous satellite image contents.

The US Department of Interior Bureau of Reclamation has issued statistics in 2010 [9] listing highs and lows of Lake Mead water levels over the years since Hoover Dam was built in 1935. Highs and lows for the reference time period of this study are plotted here below (Fig 9). The graphs show the steep decrease in water levels between the years 1999 and 2009, captured by the QE trend from the SOM (Fig 8) on the satellite image region corresponding to the area around Lake Mead in the reference period.

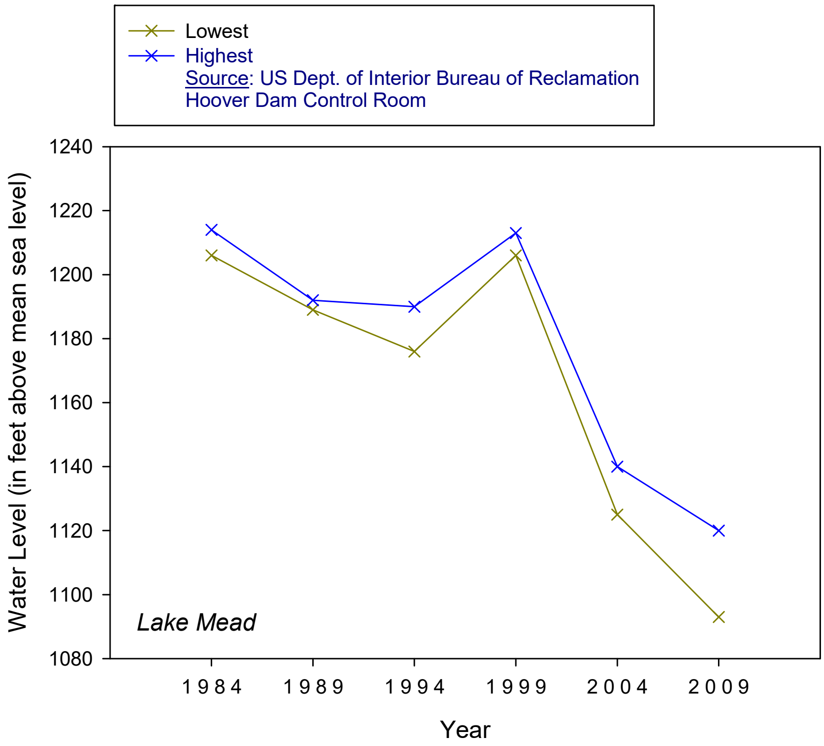

Figure 9: Statistics from the Hoover Dam control room showing the abrupt decrease in water levels of Lake Mead between the years 1999 and 2009..

3. Discussion

As shown in our previous work [4], the quantization error (QE) of the output of image analysis by Self Organized Mapping (SOM), an unsupervised neural learning algorithm, reliably reflects local pixel density variability in images and thereby provides an indicator of local changes in image contents in time. Here we show that the QE can be exploited effectively to signal changes in urban landscape the built environment of large cities, as demonstrated on the basis of SOM analyses of satellite images of Las Vegas taken at several moments of a critical time period of major restructuration of the city. Showing the variations in QE from SOM on different image parts relative to major regions of Las Vegas permits visualizing the extent to which these different areas have undergone structural changes, resulting in new building developments to a greater or lesser extent in different parts of the town. Greater increase in building density with time, especially in the City Center, went along as shown here with an increase in population estimates, number of visitors, and total sum of income from gaming revenue. The North, the sought-after residential part of Las Vegas further way from the Strip and close to leisure centers, golf course, and the nearby mountain ranges, progressively increased in building density over the years of the reference period here, as reflected by a consistently decrease in QE from SOM on the corresponding region of interest of the satellite images. In the East, closer to the heavily built up Strip and the airport, we find a similar evolution after a restructuration period between 1984 and 9994, also reflected by lower QE values from SOM on the corresponding image regions. A meaningful interpretation of demographic data, urban landscape characteristics such as water levels, as shown here on the example of Lake Mead and analyses of satellite image contents for the corresponding region of interest, can be provided in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2017 doi:10.20944/preprints201710.0166.v1

light of consistent variations in the QE. SOM analysis on time series of satellite images is fast (less than two minutes for a series of images), and represents a promising technique for the automatic tracking and harvesting of landscape information in large bodies of satellite images. SOM therefore proves a highly useful and easy to implement analytical tool for providing inexpensive, ready-to-use, and reliable information to the public, historians, administrators, or policy makers.

4. Materials and Methods

Satellite images of Las Vegas were retrieved from NASA (images provided by Jesse Allen and Robert Simon). They were acquired by NASA's Landsat 5 on October 22, 1984, May29, 1989, June 28, 1994, May 9, 1999, June 7, 2004 and January 12, 2009. The original images are available at https://earthobservatory.nasa.gov. Copies of the six satellite images are shown here for illustration (Fig 10). Images were of the same dimension and size, making them suitable for x/y coordinate-based comparison of their contents. A four-by-four SOM [1] was used for image learning, and the Quantization Error (QE) was computed from the output of two levels of analysis for each image. At a first level, SOM was run for each of the six images as a whole. In a second step the average QE value of all best-matched samples was calculated. These are the samples that best matched each of the sixteen neurons given that with the four-by-four SOM, there was a total of 16 neurons.

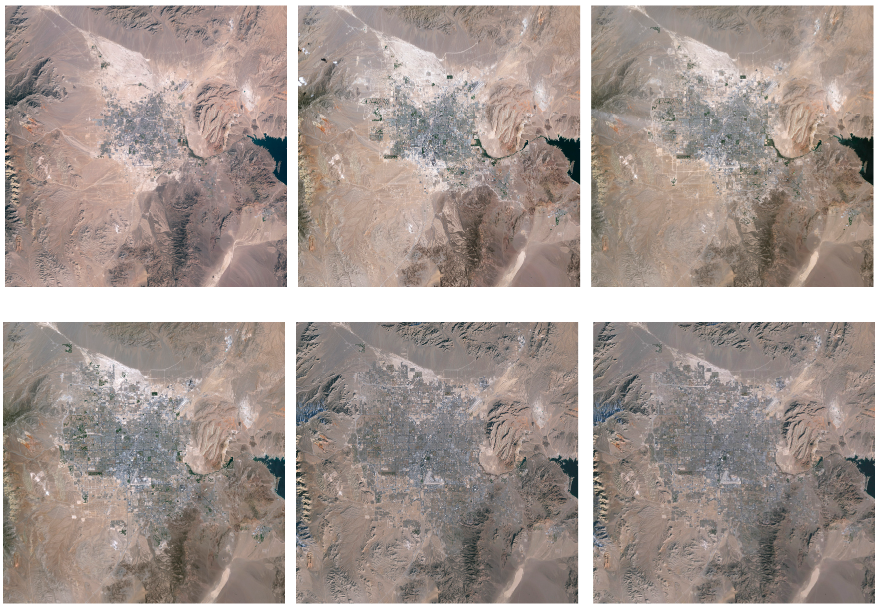

Figure 10: Satellite images of Las Vegas across the years 1984-2009. Copies of the original satellite images of Las Vegas are shown her for illustration. The images were generated in the years 1984, 1989, 1994, 1999, 2004 and 2009, shown here in their respective order.

Computer driven self-organized mapping [1] was used to analyze the images. A self-organizing map (SOM) is an unsupervised neural network learning technique that does not need target outputs required in error correction supervised learning, and is used to produce a lower-dimension representation of the input space. Thus, for each input vector, so called competitive learning is carried out to produce a lower-dimension visualization of the input data. SOMs are typically applied as feature classifiers of input data. From an initial randomization of a map, input data is iteratively applied to optimize the map into stable regions. Where the node weights match the input vector, that area of the lattice is selectively optimized to more closely resemble the data for the class the input vector is a member of. From an initial distribution of random weights and over multiple iterations the SOM eventually settles into a map of stable zones. Each region of the map becomes a feature class of the input space. Each zone is effectively a feature classifier, and the graphical output is a type of feature map of the input space.

The central idea behind the principles and mathematics of SOM is that every input data item shall be matched to the closest fitting region of the map, called the winner (as denoted by M_c in Fig. 11), and such subsets of regions will be modified for optimal matching of the entire data set [1]. On the other hand, since the spatial neighborhood around the winner in the map is modified at a time, a degree of local and differential ordering of the map occurs to provide a smoothing action. The local ordering actions will gradually be propagated over the entire SOM. The parameters of the SOM models are variable and are adjusted by learning algorithms such that the maps finally approximate or represent the similarity of the input data. The task of finding a suitable subset that describes and represents a larger set of data vectors is called vector quantization [5]. Vector quantization aims at reducing the number of sample vectors or at substituting them with representative centroids. The resulting centroids do not necessarily have to be from the set of samples but can also be an approximation of the vectors assigned to them, for example their average. Vector quantization is closely related to clustering, and SOM performs vector quantization since the sample vectors are mapped to a (smaller) number of prototype vectors, as demonstrated in [6]. The prototype vectors are called the best matching units (BMU) in SOM. As a property of SOM, the quantization error (*QE*) is used to evaluate the quality of SOM. The *QE* belongs to a type of measures that have been used to benchmark a series of SOMs trained from the same dataset. In the present study, the *QE* is used to perform a somewhat opposite measure, used to benchmark a series of datasets with SOM trained with the same parameters. In other words, the same SOM, same map size, feature size, learning rate and neighborhood radius is used to analyze series of image datasets with critical changes in contents across time.

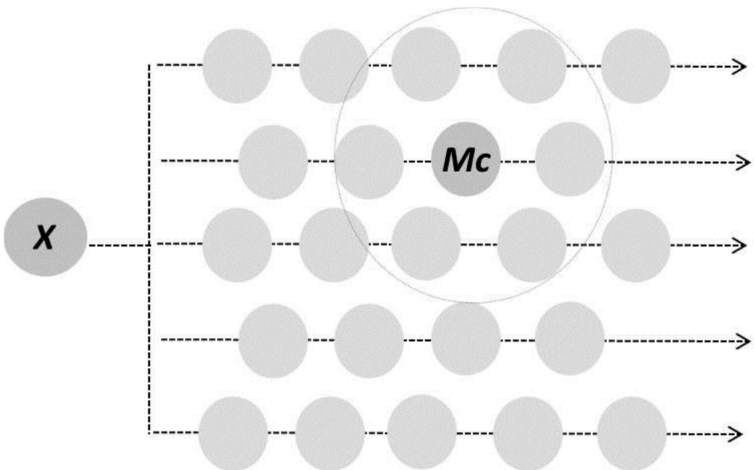

Figure 10: Schematic illustration of a self-organizing map (SOM). An input data item (*X*) is transmitted to an ensemble of models of which model *Mc* matches best with *X*. Models that lie in the neighborhood (indicated by the large circle here) of *Mc* in the map match better with X than all others (illustration adapted from [1]).

The *QE* is derived after subjecting an image to a self-organizing map algorithm analysis and by calculating the squared distance (usually, the standard Euclidean distance) between an input data, *x*, and its corresponding centroid, the so-called “best matching unit”, or BMU. This gives the average distance between each data vector (*X*) and its BMU and thus measures map resolution:

$$QE = \frac{1}{N}\sum_{i=1}^{N}\left\|X_i - \left(\mathrm{BMU}_{(i)}\right)\right\| \qquad (1)$$

where N is the number of sample vectors x in the image.

 doi:10.20944/preprints201710.0166.v1

This measure completely disregards map topology and alignment, as noted by [6], making it applicable for different kinds and shapes of SOM maps. Besides, the calculation does not rely on any user parameters as seen here above (Fig 11). A four by four SOM with 16 neurons and an initial neighborhood radius of 5 and learning rate of 0.2 was set up for the extraction of data from images. These initial values were obtained after testing several sizes of the SOM to check that the cluster structures were shown with sufficient resolution and statistical accuracy, as specified in [1]. The learning process was started with vectors picked randomly from the image array as the initial values of the model vectors. For analysis of the satellite images from the series here, all SOM parameters were kept constant.

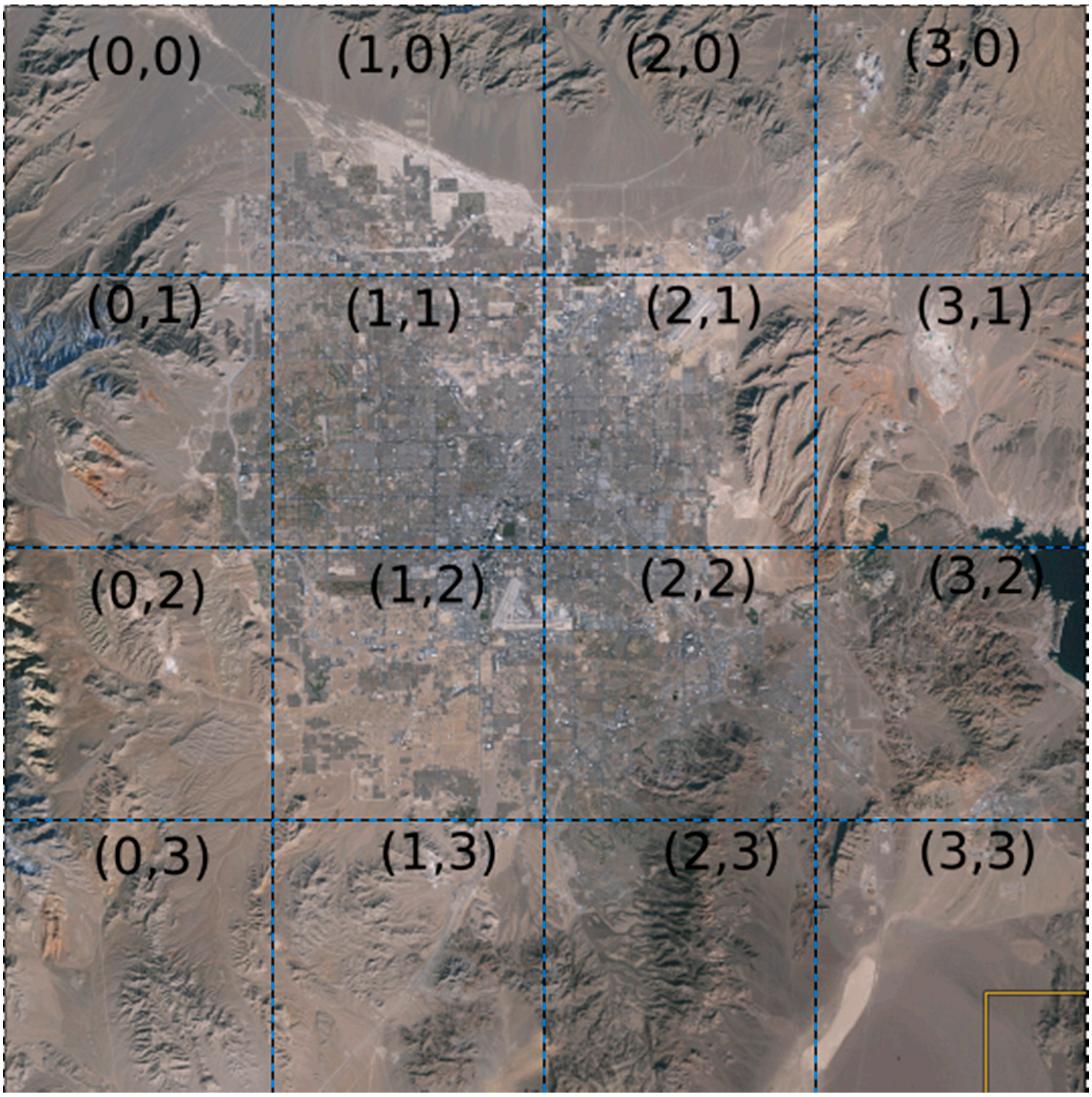

Figure 12: The satellite images of Las Vegas were subdivided into 16 different image regions corresponding to 16 neurons in each four-by-four SOM. The x-y coordinates of the different image regions are used as image labels for specific geographic regions of interest.

Each of the 16 neurons represents a particular section of the image (Fig 12). SOM preserves the topological location of samples as much as possible [1], and a neuron and its best matching samples can be seen as representing a specific region on the map. The satellite images of Las Vegas and each of the corresponding Self-Organizing Maps were subdivided into 16 different image regions corresponding to 16 neurons in each SOM. The x-y coordinates of these image regions (Fig 11) are used as image labels for specific geographic regions of Las Vegas. For instance, the North-West is

 doi:10.20944/preprints201710.0166.v1

labelled by the coordinates (0,0), the residential North by (1,0). The North-East is labelled by the coordinates (2,0), (3,0), the South-East by (2,2), (3,2), and (3,3), including the geographic region of interest around Lake Mead and Hoover Dam (3,2), which is located in the South-East of Las Vegas 30 kilometres away from the City Centre. The West and the East correspond to the labels (0,1) and (2,1), (3,1) respectively.

The South-West is tagged by image regions (0,2) and (0,3), the South by region (1,3). The Northern and Southern parts of the Strip constitute the City Centre and are labelled by the image coordinates (1,1) and (1,2) respectively. Major geographic regions of interest in this study here are the residential North (1,0), the City Centre (1,1), (1,2) and the area around Lake Mead (3,2), a man-made water reservoir pushing against the walls of Hoover Dam.

The code for the SOM that produced the QE raw data for each of the 16 image regions shown in Table S1 of the supplementary material section (QE data from SOM on the 16 image regions of satellite images taken by NASA's Lansat 5 across the years provided here in an excel file) can be accessed through the File S2 (pdf) in the supplementary material section. Python code was used for running the SOM (pdf file). Additional comments on the code are provided in File S3 (pdf file) in the supplementary material section.

Supplementary Materials: The following are available online at www.mdpi.com/link, Table S1: QE data from SOM on the 16 image regions of satellite images taken by NASA's Lansat 5 across the years (excel file). File S2: Python code used for running SOM (pdf file). File S3: Additional comments on the code (pdf file)

Acknowledgments: This research was supported by a study grant from the French Embassy in Nairobi to J. W.

Author Contributions: J.W., B.D-L., and H.N. conceived and designed the experiments; J.W. performed the experiments; J.W. and B.D-L. analyzed the data; Y.R. contributed analysis tools; J.W and B. D-L. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Kohonen, T. Automatic formation of topological maps of patterns in a self-organizing system, in 2nd Scand. Conf. on Image Analysis, Espoo, Finland, pp. 214–220, 1981.
2. Hays, J.; Efros, A. Im2GPS: estimating geographic information from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition, 2008.
3. Merril, K. R.; Jiajun, L. A comparison of four algorithms for change detection in an urban environment. Remote Sensing of Environment, 63, 95–100, 1998.
4. Wandeto, J. M.; Nyongesa, H.; Remond, Y.; Dresp-Langley, B. Detection of smallest changes in medical and random-dot images comparing self-organizing map performance to human detection. Informatics in Medecine Unlocked, 7, 39-45, 2017. https://doi.org/10.1016/j.imu.2017.03.001
5. Gray, R. Vector quantization. In: IEEE ASSP Magazine, vol. 1, no. 2, pp. 4–29, 1984. Available online: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1162229 (accessed on 02/10/2017).
6. Pölzlbauer, G. Survey and Comparison of Quality Measures for Self-Organizing Maps, Proceedings of the Fifth Workshop on Data Analysis WDA'04, pp. 67--82, 2004.
7. City of Las Vegas Economic and Urban Development Department and Redevelopment Agency. Demographics, 2013.
8. Las Vegas Convention and Visitors Authority. Statistics, 2017.
9. US Department of Interior Bureau of Reclamation, Hoover Dam Control Room. Statistics, 2010.