

Feature and information extraction for regions of Southeast Europe from Corona satellite images acquired in 1968

Dimitris Stratoulas^{*a,b}, M. Erdem Kabadayi^c

^aInformetrics Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam, dimitris.stratoulas@tdtu.edu.vn; ^bFaculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam; ^cKoç University, Rumelifeneri Yolu, Sariyer, 34450, Istanbul, Turkey, mkabadayi@ku.edu.tr

ABSTRACT

The Corona spy programme was a series of reconnaissance satellites which delivered more than 860000 images between 1960 and 1972. Since 1995, the data are declassified and a large historic earth observation archive is made accessible to the scientific community. Despite the large volume of information and the high spatial resolution of the satellite imagery, little has been done in the last 25 years in the context of image processing of this data source, a fact which perhaps can be attributed to the technical difficulties of these primitive images such as the lack of metadata, intense spatial and radiometric distortions, low Signal-to-Noise Ratio (SNR) and a single panchromatic band. Hence, the photogrammetric challenges to extract useful information are paramount. In this study, we present recent developments arising from our efforts to map settlements and agricultural parcels over the Plovdiv region, Bulgaria from a Corona image acquired in 1968. We, overall, present initial findings from the integration of earth observation into the ERC-StG project UrbanOccupationsOETR and evaluate the usability of such primitive images in feature extraction. We compare the areas corresponding to settlements and correlate them with concurrent population census. Based on the findings, we suggest that settlements and agricultural parcels can be mapped from a Corona KH-4B image with fine radiometric quality.

Keywords: Remote sensing, historical GIS, Corona, feature extraction, segmentation, settlements, agriculture.

*dimitris.stratoulas@tdtu.edu.vn

1. INTRODUCTION

The first constellation of reconnaissance spy satellites missions named Corona, Argon and Lanyard, operated between 1960 and 1972 and delivered a large satellite image archive; over 860000 film images depicting regions in eastern and southeastern Europe, Asia and South America were successfully physically retrieved and archived. Following its declassification in 1995 [1] as they were not anymore important for national security, this historic Earth Observation archive has been made available to scientists and the public alike. Since then, few efforts have been made by the scientific community to explore this unique dataset, a fact perhaps attributed to the primitive technical specifications of the images and the lack of coexisting validation data. Despite the large time span of the satellite mission (i.e. 12 years), the fine spatial resolution of the imagery and the 25 years of availability since the declassification, there have not been many studies attempting to associate information extracted from Corona images with field data during this historic time. This fact perhaps can be attributed to the lack of historic concurrent ground truth data from this past time and the technical peculiarities of the raw satellite data, the latter imposing challenges in information and feature extraction from the Corona imagery. Nevertheless, the very high spatial resolution (up to 2.5m) and the panchromatic nature of the imagery make it an unprecedented historic information source of land use and land cover during the cold war era.

The few studies having used Corona imagery in the scientific literature concentrate primarily on archaeological applications (e.g. [2-9]), land use / land cover studies (e.g. [10]) and terrain analysis (e.g. [11]) as reported by Awange and Kiema [12]. Digital Elevation Model (DEM) generation based on stereo-pairs of fore and aft Corona images has been another demonstrated application (e.g. [11, 13-14]) with an indicative vertical accuracy of 10m as reported from the first study conducted in Morocco. However, there is much more to anticipate from this historic archive; the 0.5–0.9 μ m spectral domain of the panchromatic layer has been shown to be the most important spectral region for vegetation-related studies (e.g. [14-17]) and the very high spatial resolution is suitable for small agricultural field parcels delineation [18-19] (if not necessary [20]) which appears to be the norm in Bulgaria of the 1960s.

During the era of the Corona missions, Bulgaria has observed unstable dynamics in its population of settlements in regard both to the total number and to the age structure of inhabitants. This was due to large scale social and economic engineering policies of a mature communist rule in Bulgaria with strong backing from the USSR at a border region with two NATO member states, i.e. Greece and Turkey both joined the military alliance in 1952. Given the availability of concurrent census population data, feature extraction from Corona images might be a useful technique for assessing settlement size information. While a high accuracy in classification is a consequence of a concurrent high spatial and spectral resolution of the image [21-23], which is not the case for Corona as it inherits a single panchromatic band, the very high spatial resolution of the dataset provides an opportunity to extract features through segmentation and object-based image analysis.

In this study, we attempt to distinguish generic features based on mean-shift segmentation of a Corona image acquired in 1968 over Plovdiv, Bulgaria. We present our initial efforts to extract features of interest, namely cities, town and villages (cumulatively referred to henceforth as settlements) and surrounding agricultural parcels. A concurrent national census with registered population, which is used as a proxy to settlement size, is used to showcase the population dynamics during the last decades. We reckon that feature extraction from a Corona image, after necessary and careful processing, is feasible and this unexplored archive can reveal historic land cover information in an era where Earth Observation data are scarce. All in all, we present initial findings from the project UrbanOccupationsOETR (Industrialisation and Urban Growth from the mid-nineteenth century Ottoman Empire to Contemporary Turkey in a Comparative Perspective, 1850-2000, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 679097)).

2. METHODOLOGY

2.1 Study area

The Plovdiv region, Bulgaria, Southeast Europe is selected as the study area (Figure 1). Bulgaria is a country that has suffered intensive depopulation from 1989 onwards nationally and as early as 1975 in the rural regions [24]. Moreover, an ageing population trend has been observed in the past [25] and future projections are discouraging. The UN Population Division estimates that Bulgaria will lose 23% of its current population by 2050, which will make it the fastest shrinking in the world [26]. In this context, settlement development and surrounding agricultural activity in this area are unique in the sense that they have largely remained stable for a very long period, even going back to the 19th century [27].

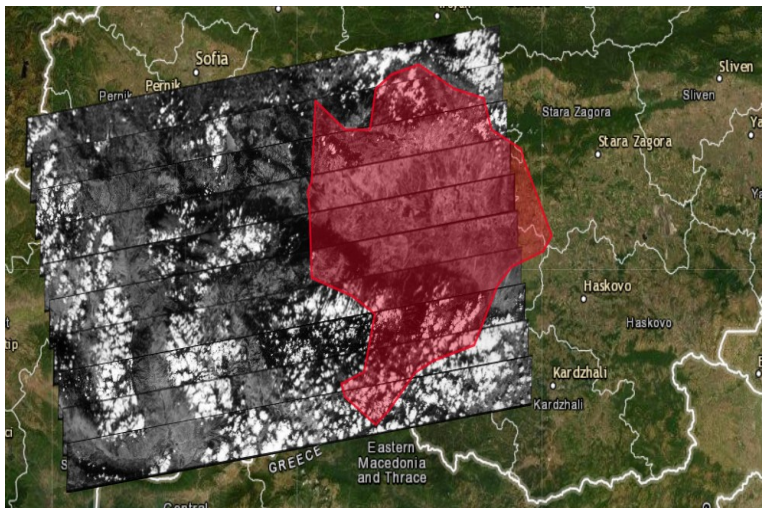


Figure 1. Study area in Plovdiv, Bulgaria overlaid with the Corona satellite image data set available and the location of the concave polygon containing all the available villages within Plovdiv region with population sampling.

2.2 Satellite data

The study area lies in the Southeast Europe and, as such, has been covered by the Corona mission; more specifically, the first available declassified image over Plovdiv dates back to 18th August 1960 and the last one to 24th July 1970. Ten adjacent Corona satellite images acquired during the KH-1104 mission on 17th August 1968 and covering 98% of the study area were used. This mission carried the Keyhole-4B (KH-4B) camera which delivered the best imagery to date on any KH-4 systems [28]. KH-4B recorded data from an altitude of 150km and with a focal length of 609.602 mm. Cloud coverage is the norm for Corona images as it is estimated by the data provider that “over 40 percent of the imagery is obscured by clouds”, however, based on visual assessment our data set has a cloud coverage of approximately 10%. The data were ordered from the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center; once ordered, the images are made freely available through the EarthExplorer interactive query system. The images inherit a single panchromatic layer in the 0.5–0.9 μ m spectral domain at 2.75m spatial resolution and, consequently, clear land cover structures (Figure 2) such as settlements and agricultural parcels can be visually distinguished. Unlike modern satellite systems which downstream data to receiving ground stations, Corona images were recorded on a film onboard which was subsequently returned to the earth for processing and analysis after each mission; hence, there is no associated metadata, neither ephemeris data, moreover a low SNR and intense spatial and radiometric distortions are observed. The images delivered are in essence film scans representing faithful copies of the source film. Potential distortions inherited during scanning have been compensated for. Each image is delivered as 4 high resolution complementary scans of the film in TIFF format (Figure 3).



Figure 2. Close-up of the scan depicting crisp land cover elements such as settlements, agricultural parcels and water bodies.

2.3 Census data

Census data reporting on the population per settlement and agricultural ownership from the 1970s are accessible. Population census of all individual populated places within Plovdiv region is available between 1965 and 2001 in a total of 6 intervals corresponding to census years of 1965, 1975, 1985, 1992, 2001, 2011 and the next census is anticipated in 2021. Moreover, agricultural data are available at aggregated level conducted in 1968, on the exact year of Corona imagery, and in 2010, the first one since Bulgaria joined the EU and compliant with EU regulations. Finally, the geographic locations of each settlement are known and with appended the Unified Classification of Administrative-Territorial and Territorial Units (UCATTU), or otherwise known as EKATTE, register number of places provided by the Bulgarian National Statistical Institute (NSI) [29].

Table 1. Sample subset of the ground truth data used as a proxy for settlement characterization. The settlements are georeferenced and an associated shapefile with the presented information as the attribute table was used.

EKAT TE	Name_en	Descr_en	Population 1965	Population 1975	Population 2011
56784	Plovdiv	city	229043	299638	338153
77373	Hrabrino	village	774	828	661
32408	Izvor	village	316	187	101
35095	Kadievo	village	1172	1266	1107
36244	Karadzhovo	village	1236	1149	1084
40004	Krumovo	village	3114	3779	3129
40467	Kuklen	village	5367	6327	5858
53833	Orizari	village	455	401	374
34028	Yoakim_Gruevo	village	2406	2478	2643
31036	Zlatitrap	village	1044	1388	1318

2.4 Data analysis

The census data were geo-spatially connected and geographic locations of each settlement were associated with the census data based on the EKATTE number; hence a vector file with the population data for whole Bulgaria was constructed. The population data from the 1965 census are a potential settlement proxy as this date is close enough to the actual acquisition of the Corona image and hence, we assume it represents the settlement population (and indirectly the settlement size) of this time period.

Each Corona image is delivered as 4 individual scans with 10% overlapping. In order to reconstruct the whole image, several stitching software were assessed including Orfeo ToolBox (OTB) (through homologous points with SURF), ISAM, Microsoft Image Composite Editor and Agisoft; eventually, the Adobe Photoshop 2020 was selected. Subsequently, the image was georeferenced on a high resolution recent map with Ground Control Points (GCPs) in QGIS [30] and projected on a map coordinate system (WGS 84/UTM zone 47N) with the Geospatial Data Abstraction Library (GDAL). The results from the stitching and georeferencing are depicted in Figure 3.

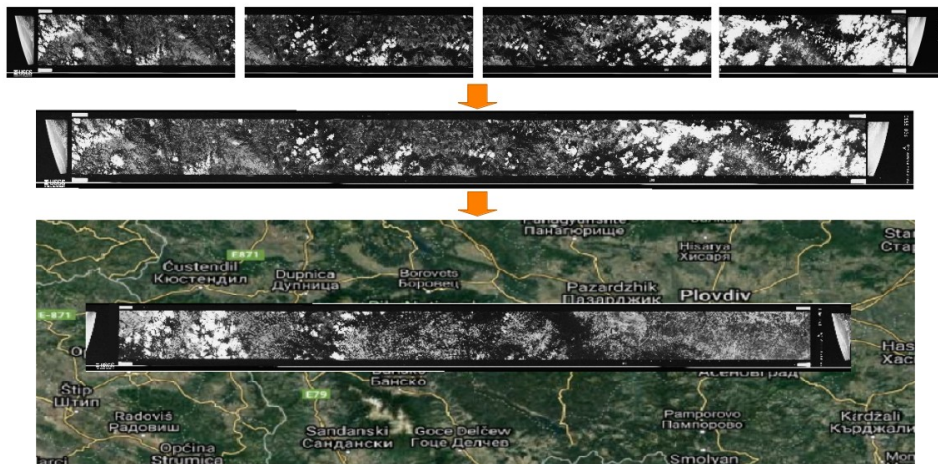


Figure 3. Individual film scans (tiles) with 10% overlap as delivered from the provider (top), the reconstruction of the whole image through stitching (middle) and the georeferenced product by using Ground Control Points (bottom).

We then segment the panchromatic Very High Spatial Resolution (VHSR) Corona image with the mean-shift algorithm with the aim to identify settlement structures, hence textural analysis is applied as in related studies (e.g. [31]). The segmentation was carried out on the OTB software plugin through QGIS. Three different minimum region sizes were selected, namely 0, 300 and 1000 pixels and the results are presented in Figure 4. When no minimum region size is selected (i.e. 0), the lines introduced from the scanning are taken into account as segments (central line in the middle of the left image in Figure 4); these lines disappear with higher order minimum region size. The 300 pixel value is selected eventually as it is large enough to eliminate the scanning lines and small enough to segment small agricultural parcels as indicated in the centre of the images in Figure 4. Thereafter, we identified 10 settlements around and including the town of Plovdiv and manually selected the polygons corresponding to the built-up areas in the Corona image (Figure 5). Adjacent polygons from the segmentation were merged to create a single polygon covering each settlement. Thereafter the vector file was reprojected to the BGS2005/UTM34N coordinate system to allow for metric estimation of ground surface areas; finally, the total coverage per settlement was calculated with the aim to be compared with the population of each settlement from the Census of 1965.

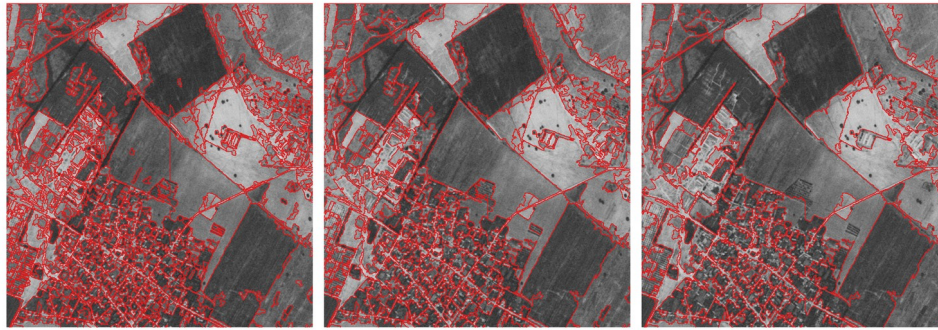


Figure 4. Close-up of the result of the mean-shift segmentation with 3 different minimum region sizes as of 0 (left), 300 (middle) and 1000 (right) pixels.

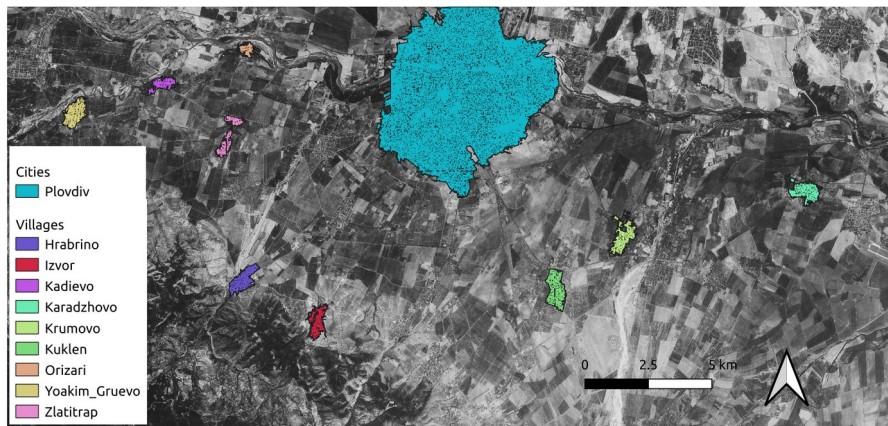


Figure 5. The 10 settlements (the city of Plovdiv and 9 surrounding villages) selected for this study and the area coverage resulting from the segmentation result.

3. RESULTS AND DISCUSSION

Census data analysis in Figure 6 present the decadal population dynamics and the 1965 population distribution for Bulgaria. It is worth noting the population peak around the 1980s and the sharp fall thereafter in this gaussian distribution. Figure 6b showcases a fragmented and large number of small settlements found in 1965 in Bulgaria, with only 11 settlements having a population larger than 50000 people and most settlements accounting for less than 1000 inhabitants.

The lack of metadata and image format impose difficulties (for instance Deroin et al. [32] in a study analyzing Corona KH-4A images avoid annotating geographical coordinates due to the presence of important conical distortion). These technical limitations have perhaps deterred scientists to attempt feature extraction but rather use Corona imagery for visual interpretation purposes. In our case, important displacement in the georeferenced product were observed in areas that are further from one of the GCPs; this denotes that perhaps polynomial of a higher order should be used when wrapping the image during georeferencing.

In figures 7 and 8 we present the segmentation in regard to settlements and agricultural parcels respectively. It is obvious, by visual assessment, that both settlements and agricultural parcels can be identified; moreover, the segment size resulting from this exercise seems to cluster settlement objects together, while splits individual agricultural parcels in parts; hence, a dual segmentation with different thresholds for settlements and agriculture could improve the results, however, it is worth noting that this will impose higher computational restrictions as a typical Corona image accounts for approximately 4GB. It is also worth noting that settlements and agricultural fields inherit a very different texture and inclusion of textural analysis is foreseeable. Last but not least, the low SNR and the single band nature of the Corona image, do not allow for a spectral analysis; comparing the Digital Number (DN) values of settlements, agriculture and other natural elements in the scene, there is no distinguishing pattern to discern land cover information based on the DN.

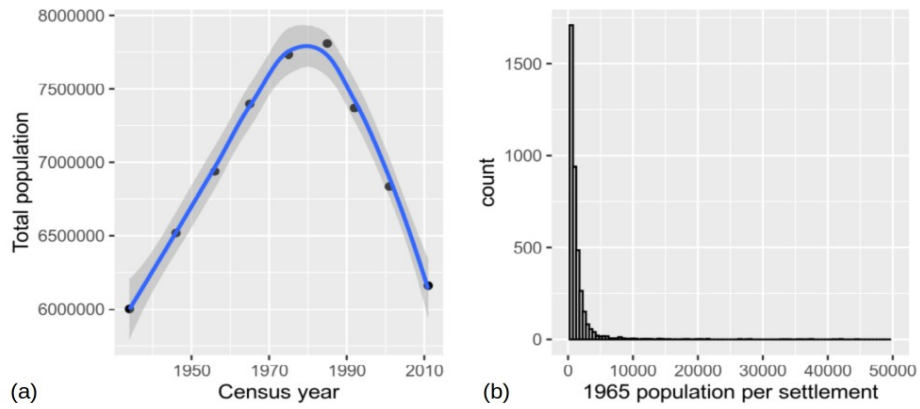


Figure 6. Temporal development of total population for Bulgaria based on Census decadal data (left) and population distribution for settlements in the year 1965, with settlements larger than 50000 accounting to only 11 cumulatively and having been omitted (right).



Figure 7. Segmentation of Corona image with the mean-shift algorithm. Segments denoting man-made structures in a settlement are depicted with the yellow polygons.

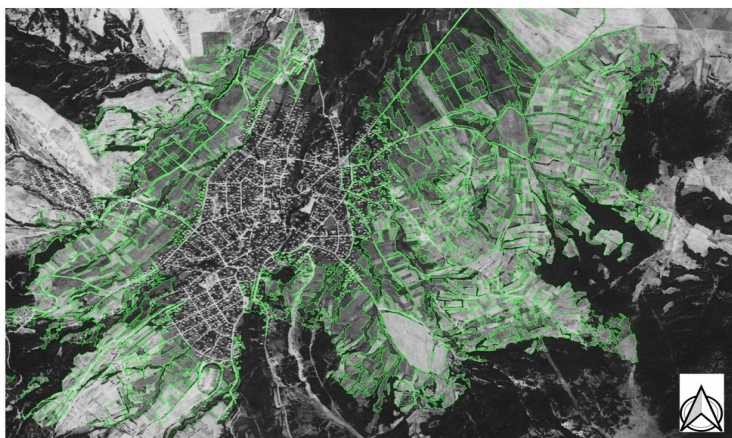


Figure 8. Segmentation of Corona image with the mean-shift algorithm. Segments denoting agricultural parcels are depicted with the green polygons.

In the last part of this study we attempt to examine the correlation between the population status of the year 1965 from the Census with the settlement coverage. In this exercise, several assumptions have been made; that the population between 1965 and 1968 has not changed, all the settlements have a proportional industrial activity (and hence industrial infrastructure within their settlement limits), the depopulation affecting the settlements is also proportional, the buildings populate similar-sized communities which consequently assumes the same population density for all rural settlements. In Figure 9 we present the findings from the linear correlation between the two attributes; the city of Plovdiv has been disregarded as it has a disproportional high population for a canonical data distribution in this dataset. We found that the linear regression has Intercept = -1902 and $R^2 = 0.5389$ (p-value = 0.02434). The relatively low intercept indicates that a hypothetically empty settlement would associate with a population of 1902 which is rather low and close to the null value that is anticipated. The R^2 value indicates a weak correlation, however in the context this exercise was designed, we interpret the results as encouraging; first and foremost, for all the reasons mentioned above, there are several assumptions made; moreover, the data sample we selected corresponds to very small villages and the sampling number is rather low; also, the de facto association between population density and settlement area is not direct and clear and a strong bias in the dataset distribution depending on the economic and social activity from each village is anticipated. Last but not least, in a similar study from Saleem et al. [33] for Iraqi Kurdistan from 1969 to 2014, they found that there is an exponential relationship between population increase and urban growth, and it is worth exploring in the future for our dataset. In this context, the regression between Census population and Corona-derived total area shows that there is a clear positive association and trend that is worth exploring with a larger dataset and more sophisticated texture analysis methods.

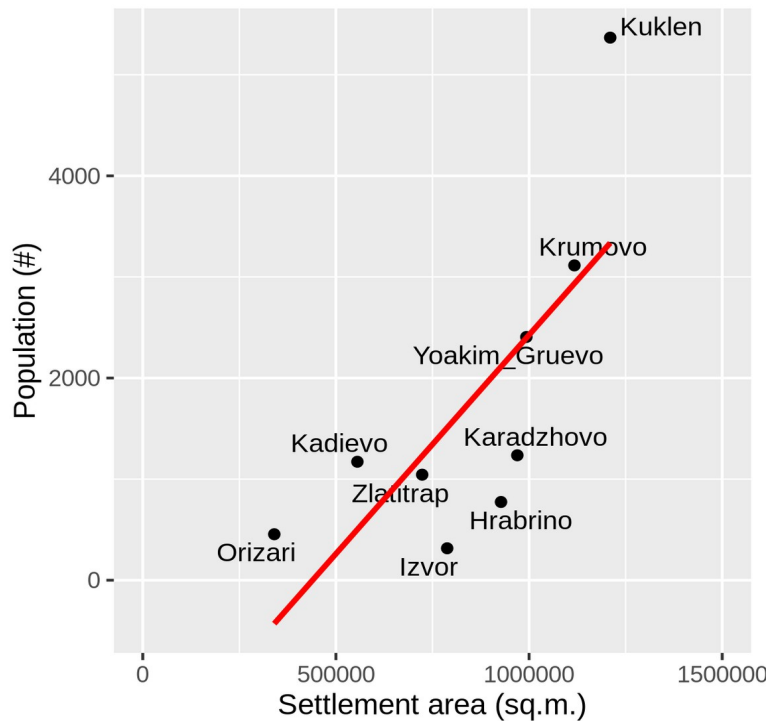


Figure 9. Linear regression of Population number (derived from Census data) and area coverage (derived from the Corona image) for 9 settlements around the city of Plovdiv. $R^2 = 0.5389$, p-value = 0.02434, Intercept = -1902.

4. CONCLUSIONS

In this study, we present initial findings arising from our efforts to extract features of interest from Corona images acquired over the Plovdiv region, Bulgaria in May 1968. We showcase that segmentation of a Corona image with the mean-shift algorithm can provide segments appropriate for settlement and agricultural parcels delineation. Moreover, we try to relate population statistics with the settlement area and we suggest that the weak correlation found is an indication that a positive association exists, and inclusion of a larger number of samples and further texture analysis will yield more accurate results.

Future efforts will focus on this direction, as well as feature extraction of earlier aerial photos acquired in the 1940s and later contemporary satellite images such as those from the Landsat series and Sentinel-2; combined with Corona images, a decadal time-scale archive can be built, such as the one presented by Tappan and McGahuey [34] who detected environmental and agricultural conditions from 1965 until 2001 in southern Mali; the results will showcase the correlations and differences between settlement development and agricultural practices for the period from 1968 until today.

REFERENCES

- [1] McDonald, R. A. (1995). Opening the Cold War sky to the public- Declassifying satellite reconnaissance imagery. *PE & RS- Photogrammetric Engineering and Remote Sensing*, 61(4), 385-390.
- [2] Ur, J. (2003). CORONA satellite photography and ancient road networks: A northern Mesopotamian case study. *Antiquity*, 77(295), 102-115.
- [3] Goossens, R., De Wulf, A., Bourgeois, J., Gheyle, W., & Willems, T. (2006). Satellite imagery and archaeology: the example of CORONA in the Altai Mountains. *Journal of Archaeological Science*, 33(6), 745-755.
- [4] Bitelli, G., & Girelli, V. A. (2009). Metrical use of declassified satellite imagery for an area of archaeological interest in Turkey. *Journal of Cultural Heritage*, 10, e35-e40.
- [5] Casana, J., Cothren, J., & Kalayci, T. (2012). Swords into ploughshares: Archaeological applications of CORONA satellite imagery in the Near East. *Internet Archaeology*, 32(2).
- [6] Min, L. (2013). Archaeological landscapes of China and the application of Corona images. In *Mapping Archaeological Landscapes from Space* (pp. 45-54). Springer, New York, NY.
- [7] Comer, D. C., & Harrower, M. J. (2013). *Mapping archaeological landscapes from space* (Vol. 5). Springer Science & Business Media.
- [8] McCoy, M. (2018). The race to document archaeological sites ahead of rising sea levels: Recent applications of geospatial technologies in the archaeology of Polynesia. *Sustainability*, 10(1), 185.
- [9] Lasaponara, R., Yang, R., Chen, F., Li, X., & Masini, N. (2018). Corona Satellite Pictures for Archaeological Studies: A Review and Application to the Lost Forbidden City of the Han-Wei Dynasties. *Surveys in Geophysics*, 39(6), 1303-1322.
- [10] Deroin, J. P., & Buslov, M. M. (2017). Geomorphic study of seismically active areas using remote sensing data. Case of the Gorny Altai (Siberia) affected by the 2003 Altai earthquake. *Bulletin de la Société géologique de France*, 188(1-2), 1-15.
- [11] Casana, J., & Cothren, J. (2008). Stereo analysis, DEM extraction and orthorectification of CORONA satellite imagery: archaeological applications from the Near East. *Antiquity*, 82(317), 732-749.
- [12] Awange, J., & Kiema, J. (2019). CORONA Historical De-classified Products. In *Environmental Geoinformatics* (pp. 191-199). Springer, Cham.
- [13] Schmidt, M., Goossens, R., & Menz, G. (2001). Processing techniques for CORONA satellite images in order to generate high-resolution digital elevation models (DEM). *Observing Our Environment from Space: New Solutions for a New Millennium*, 191-196.
- [14] Galiatsatos, N., Donoghue, D. N., & Philip, G. (2007). High resolution elevation data derived from stereoscopic CORONA imagery with minimal ground control. *Photogrammetric Engineering & Remote Sensing*, 73(9), 1093-1106.
- [15] Duveiller, G., & Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, 114(11), 2637-2650.
- [16] McCarty, J. L., Neigh, C. S. R., Carroll, M. L., & Wooten, M. R. (2017). Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery. *Remote Sensing of Environment*, 202, 142-151.

- [17] Stratoulas, D., Balzter, H., Zlinszky, A., & Tóth, V. R. (2018). A comparison of airborne hyperspectral-based classifications of emergent wetland vegetation at Lake Balaton, Hungary. *International Journal of Remote Sensing*, 39(17), 5689-5715.
- [18] Stratoulas, D., De By, R. A., Zurita-Milla, R., Bijker, W., Tolpekin, V., Schulthess, U., & Ahmed, Z. U. (2015, January). The potential of very high spatial resolution remote sensing in applications in smallholder agriculture. In *36th Asian Conference on Remote Sensing, ACRS 2015: Fostering Resilient Growth in Asia*.
- [19] Stratoulas, D., Tolpekin, V., De By, R. A., Zurita-Milla, R., Retsios, V., Bijker, W., Hasan, M. A., & Vermote, E. (2017). A workflow for automated satellite image processing: From raw VHSR data to object-based spectral information for smallholder agriculture. *Remote Sensing*, 9(10), 1048.
- [20] Pax-Lenney, M., & Woodcock, C. E. (1997). The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote Sensing of Environment*, 61(2), 210-220.
- [21] Welch, R. (1982). Spatial resolution requirements for urban studies. *International Journal of Remote Sensing*, 3(2), 139-146.
- [22] Herold, M., Gardner, M. E., & Roberts, D. A. (2003). Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and remote sensing*, 41(9), 1907-1919.
- [23] Stratoulas, D., Keramitsoglou, I., Burai, P., Csaba, L., Zlinszky, A., Tóth, V. R., & Balzter, H. (2017). A framework for lakeshore vegetation assessment using field spectroscopy and airborne hyperspectral imagery. *Earth Observation for Land and Emergency Monitoring*, 159.
- [24] Mladenov, C. (2002). Population distribution in Bulgaria. *Discussion papers, (Special)*, 20-26.
- [25] Rangelova, R. (2003). Bio-demographic change and socio-economic trends in Bulgaria. *Economics & Human Biology*, 1(3), 413-428.
- [26] UN Department of Economic and Social Affairs, Population Division. (2019). *World Population Prospects 2019 Highlights*, United Nations, New York.
- [27] Kabadayi, M. E., Gerrits, P., & Boykov, G. (2020). Bridging the Gap between Pre-census and Census-era Historical Data: Devising a Geo-sampling Model to Analyse Agricultural Production in the Long Run for Southeast Europe, 1840-1897. *International Journal of Humanities and Arts Computing*, 14(1-2), 46-63.
- [28] NASA 2020. <https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1968-065A> Accessed on 29 January 2020.
- [29] National Register of Populated Places (2020). National Statistical Institute, Republic of Bulgaria. <https://www.nsi.bg/nrm/index.php?ezik=en>. Accessed on 15 March 2020.
- [30] QGIS Development Team (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- [31] Alonso, M. C., Sanz, M. A., & Malpica, J. A. (2007, November). Classification of high resolution satellite images using texture from the panchromatic band. In *International symposium on visual computing* (pp. 499-508). Springer, Berlin, Heidelberg.
- [32] Deroin, J. P., Téreygeol, F., Cruz, P., Guillot, I., & Méaudre, J. C. (2012). Integrated non-invasive remote-sensing techniques and field survey for the geoarchaeological study of the Sud Lípez mining district, Bolivia. *Journal of Geophysics and Engineering*, 9(4), S40-S52.
- [33] Saleem, A., Corner, R., & Awange, J. (2018). On the possibility of using CORONA and Landsat data for evaluating and mapping long-term LULC: Case study of Iraqi Kurdistan. *Applied Geography*, 90, 145-154.
- [34] Tappan, G., & McGahuey, M. (2007). Tracking environmental dynamics and agricultural intensification in southern Mali. *Agricultural Systems*, 94(1), 38-51.