A multi-temporal classification approach to monitor a countertrend of urbanization in an era of economic crisis

Siachalou S.[1], Tsakiri-Strati M.[1], Papadopoulou M.[2], and Lafazani P.[2]

[1]*Laboratory of Photogrammetry and Remote Sensing, School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece,* ssiacha@topo.auth.gr, martsaki@topo.auth.gr
[2]*Laboratory of Cadastre and GIS, School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece,* papmar@topo.auth.gr, lafazani@topo.auth.gr

Abstract
In the majority of rapidly developing economies, significant land use changes are being witnessed. In such cases there is an increase in the rate of the migration from the countryside to the city, while higher income groups move to newer residential areas leaving the center of the cities to be populated by lower status residents and the newest immigrants. The newer residential areas are situated farther away from the center of the cities causing an expansion of the suburban areas. Urbanization and suburbanization imply the physical growth of a city and the expansion of the built-up suburban areas. In order to monitor the urban extent constantly, the mapping community has resolved into using Remote Sensing data which covers the surface of the Earth on a regular base. Very high spatial resolution imagery (e.g. Ikonos, Quickbird, Worlview) has made possible the track of changes on a higher level of detail that is necessary to urban planners. However classification of urban areas still remains challenging due to the high heterogeneity of materials and urban objects in terms of size, shape and texture. In the present work we propose a multi- temporal classification approach for monitoring and mapping urbanization based on Hidden Markov Models (HMMs). Multi-temporal satellite images enable the monitoring of the rate of urbanization by inferring thematic information from frequent image acquisitions. Our approach exploits the spectral and temporal information content on high resolution satellite data in order to minimize classification errors. It uses HMMs for representing the evolution of urban sprawl and determines the most probable sequence of classes over time by estimating class transition possibilities. The performance of our approach is tested by using a set of six very high resolution images over a suburban area of Greece covering a period of 13 years (2001-2013).

Keywords: Urban, change detection, multi-temporal, HMM, time series analysis

1. Introduction
Cities in developing countries worldwide are undergoing constant change due to the high concentration of population and the rapid expansion of urban settlements at the expense of open air spaces, agricultural land or even forests. Urban planning is necessary to take measures towards water and land resource management, transportation, human activities management, pollution, etc. Thus monitoring the urban extent is a critical matter considering the issues that may arise from a non-regulatory growth of urban areas and the direct impact that may have on the quality of life. To this end, urban authorities and policy makers need accurate and frequently updated maps on regional scales, for long-term planning studies, covering the expanding residential areas.

In this regard, Remote Sensing offers a powerful source of spatial information by monitoring the Earth's surface for the past 20 years with Medium Resolution satellites (e.g. Landsat, SPOT) and for the last decade with the advent of Very High Resolution satellite sensors (e.g. Ikonos, Quickbird, Worlview) allowing mapping in higher thematic and geometric scales (Taubenböck et al. 2011). VHR refers to 3 - 0.5 m pixel size, which in practice enables the identification not only of urban blocks but also of individual elements (houses, trees, roads, etc.) (Tran et al. 2011). Traditional methods of creating or updating urban maps are based on manual delineation of urban objects and are subject to the individual's interpretation capabilities of aerial or remote sensing images. Such approaches are time consuming and are based on the knowledge of the image analyst, which poses the question of the reliability and the lack of uniform results. The increasing availability of VHR remotely sensed observations requires the development of new automated classification methods which map the temporal changes with high degree of accuracy.

Although VHR data offers great potential towards the analysis of the urban fabric, some points need to be taken under consideration on the processing of multitemporal images. In general, errors may stem from the changing solar elevation, atmospheric conditions, and satellite sensor incidence angles. In the case of building detection, their complex structures and height discontinuities create occlusions and shadows which provide a false representation of their shape (Pacifici et al. 2007; Hermosilla et al. 2011). In such large scales, the presence of temporary objects, such as cars, trucks, or open markets reduce the accuracy of the classification or change detection results.

In order to discriminate land cover types various techniques are applied in the fields of urban mapping. Image classifiers may incorporate different type of information: spectral, spatial and temporal (Mota et al. 2007). Spectral classifiers are the most commonly used and rely only on the spectral features of a pixel or object. Spatial knowledge can be exploited only in the case of object- based classification and refers to the spatial relations between objects produced by segmentation algorithms. Segments are regions which are generated by criteria of homogeneity in one or more dimensions (of a feature space) respectively. Thus segments have additional spectral information compared to single pixels (e.g. mean values per band, minimum and maximum values, mean ratios, variance etc.), but of even greater advantage is the additional spatial information for objects (Blaschke, 2010). Recently many studies focus on object-based classification techniques by including information on the spectral response, the texture, or the shape of the objects. Other researchers (Bouziani et al. 2010; Doxani et al. 2014) employed multispectral images and vector information, regarding the geographic position of urban objects (buildings, roads, etc.), and they built a knowledge-base with the spectral, spatial and contextual features of the existing building objects. The new buildings are detected by defining proper rules to compare the image objects with the created knowledge-base. Moreover, some studies combine the information of imagery and three-dimensional data (Hermosilla et al. 2011; Haala 1999). They create image-objects by using automatic segmentation techniques and then apply a threshold value to the nDSM to discriminate objects with similar spectral features but different heights, such as buildings and roads. Despite the promising results of the object-based classification schemes and the excessive trend to applications ‘beyond pixels’, one must consider the critical role of the segmentation level. Under-segmentation directly results in classification errors and must therefore be avoided. Ideally the objects resulting from the segmentation process should correspond to real-world objects but this can be quite difficult in an environment of very complex and heterogeneous buildings (Roeck et al. 2009).

Temporal classifiers were introduced due to the disposal of multi-temporal remote sensing observations and the need to process more than two temporal images. These classifiers process the information content on various dates taking into account the evolution of a phenomenon over time and the dynamics of the classes of the target area. The temporal knowledge usually is combined with other knowledge forms to create an adequate multitemporal classifier (Mota et al. 2007). In 2002, Bruzzone and Prieto presented a partially unsupervised approach to the classification of multitemporal remote-sensing images. Such an approach allows the classification of a remote-sensing image in the cases where training data are not available by estimating the temporal correlation between successive scenes. Neural networks (NNs) have also been used to combine spectral, spatial and temporal information to produce classification and change detection maps of increased accuracy (Melgani 2003; Pacifici et al. 2007). Moreover the HMMs approach has been proposed as an effective tool to incorporate temporal knowledge to model the evolution of environmental phenomena such as the growth of crop types or wetland dynamics (Viovy et al. 1994; Aurdal et al. 2005; Leite et al. 2011; Siachalou et al. 2014).

This paper presents a multi- temporal classification approach for monitoring urbanization based on the potential of HMMs to process simultaneously multitemporal remotely sensed data. To this end, previous work has implemented HMMs (Mithal 2013) aiming to enhance the results of urban change detection by using a time series of Landsat-7 images (of 30 m resolution). The objective of the developed methodology is to face the challenges and explore the level of detail of multitemporal VHR satellite data (of about 1 m resolution) over a suburban area in Greece. The outcome of the multi- temporal classification scheme is the production of land cover classification maps on each date of interest. The performance of the methodology is evaluated against a traditional single-date spectral classification by using the same set of images and the same training samples. Finally some conclusions are drawn over the advantages and disadvantages of the methodology and possible improvements are discussed.

2. Materials and Study Area

Our proposed approach uses as input data a set of 6 VHR pan-sharpened satellite images acquired from Ikonos, Worldview 2 and Quickbird (Tab. 1). Before the processing of the images, they were all co-registered with sub-pixel accuracy and resampled to 1 m pixel size. The dates of acquisition covered the period from 2001 to 2013 and therefore made it possible to monitor the changes in the urban fabric before and after the beginning of the financial crisis of 2009.

Table 1. Description of the satellite data used in the study

Satellite	Date of acquisition	Spatial Resolution	Spectral Resolution
Ikonos	23/05/2001	1 m	Blue, Green, Red, NIR
Ikonos	19/08/2003	1 m	
Ikonos	01/09/2007	1 m	
Ikonos	05/04/2010	1 m	
Quickbird	23/07/2011	0,6 m	
Worldview 2	10/03/2013	0,5 m	

For this study we selected the suburban area of Triadi, located near the city of Thessaloniki, northern Greece. Thessaloniki is the second largest city of Greece and its suburbs had a high expansion rate before 2009. Three years prior to crisis, the construction activity in Greece was increasing with a mean annual percentage change of 8.5%, whereas during the next three consecutive years there was a big decline by -10.4% approximately (Eriotis et al. 2013). The study area is a typical example of a Greek suburban area with a residential form, occupied by detached and semi-detached houses, with rapidly changing land cover classes, which made it a proper example to evaluate our classification model.

3. Proposed multitemporal classification model

3.1. Definitions and observations

According to the theory of HMMs we observe a system assumed to change through a series of states (Aurdal et al. 2005). In each state the system will produce observations with certain probabilities and the transitions between states can be described statistically by probabilities. Therefore in an HHM, every state is not directly linked to an observation but to a probability distribution of observations and every state is “hidden” in the sense that it is not directly observable. In the present study, the set of observations $O = \{O_1, O_2 ... O_n\}$ is composed by a stack of the six multispectral images and the states $S = \{S_1, S_2 \dots S_j\}$, are the land cover types which may change over time. The aim of the proposed approach is to assign in every pixel of each temporal image a land cover type. To define an HMM we need to estimate the following parameters based on a representative set of training data:

1. the emission matrix, which corresponds to the probability that observation O_t is produced while being in state S_j, i.e., $b_j(O_t) = P[O_t \mid q_t = S_j]$,
2. the transition matrix, which describes the probability of a pixel belonging to state S_i at time t-1 to change to state S_j at time t, i.e., $a_{ij} = P[q_t = S_j \mid q_{t-1} = S_i]$ and,
3. the initial probability of being in state S_i at the initial time instant t_1, i.e., $\pi_i = P[q_{t_1} = S_i]$.

In order to estimate the emission matrix we assumed that the set of multispectral data being used as observations in our problem, have a multivariate Gaussian distribution. The covariance matrix and the mean vector of each state, in each image is computed from the training data. Through the transition matrix, the proposed classifier incorporates temporal knowledge about the statistical relationship between the states, derived both by the training data and prior knowledge on the area of interest. It was decided to use the concept of "soft" transitions, which means that the transition probabilities range between 0 and 1. The alternative of using a "crisp" transition would limit the classifier to either being possible or not ("1" or "0") and degrade the role of the emission matrix.

Before proceeding to the application of the methodology we need to clarify some key points that were taken into account on the design of the classification model:

I. A common denominator for the effectiveness of the classifier is using a correct training set to estimate the emission and the transition matrix. Training the classifier involves selecting a representative set of samples for which the state S_i at time t-1 and the state S_j at time t are known and all possible transitions are included (Mota et al. 2007).

II. Some states-classes have a higher risk of being misclassified due to their spectral similarity. For example in the study area, buildings have tile roofs with spectral characteristics common to bare soil (Fig. 1). Their evolution through time is what could assist their discrimination (the different possibility of changing from on state to another).

III. The transition probabilities between certain states are higher than others; such is the case of changing from bare soil to building, which is more likely than changing from building to vegetation, in the particular area. The majority of pixels tend to remain to the same state through time, which means that the transition probability to a different state is lower than self-transition. The transition probabilities are equal only when referring to shadows which may be present even if they were previously recognized as bare land, or building or vegetation.

The different cover types identified in the specific area formulate the following five state- classes: 1) buildings 2) man-made surfaces composed from concrete, asphalt or gravel 3) vegetation 4) bare soil (including dry vegetation) and 5) shadows.

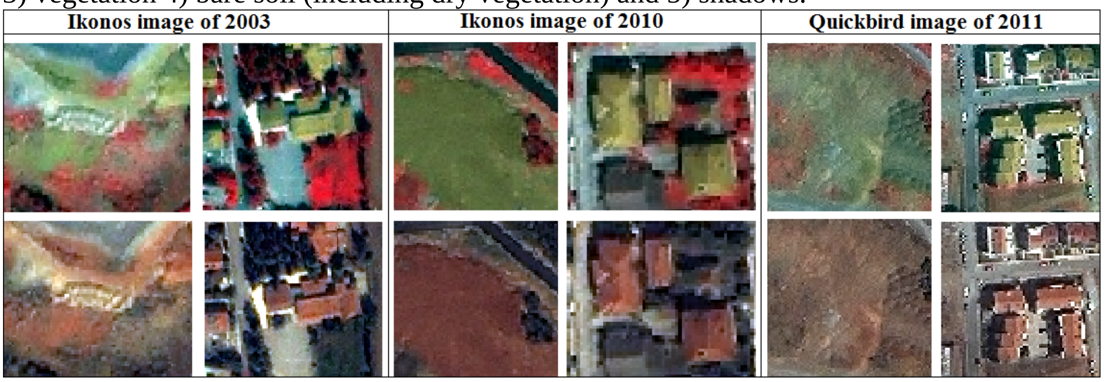

Figure 1. Examples of areas containing tile roofs and bare soil in the study area. Top row refers to false-color composite (R: IN Red, G: Red, B: Green) and bottom row refers to true-color (R: Red, G: Green, B: Blue).

3.2. HMMs classification algorithm

The basic steps of the developed methodology are illustrated in a flowchart in figure 2. In this paper, HMMs enable us to determine the state sequence that is most likely to have emitted the observed sequence of observations (i.e. the land cover class of each pixel on each time instant). The best state sequence $S_b = \{s_1, s_2 \dots s_n\}$ is defined by equation 1 which is computational intensive:

$$p(O_1, O_2, \dots, O_n | S_b) = \prod_{t=1}^{n} P(S_t|S_{t-1}) \cdot P(O_t|S_t) \quad (1)$$

In this equation, $P(S_t|S_{t-1})$ corresponds to the transition probability from state S_t to state S_{t-1} and $P(O_t|S_t)$ refers to the emission probability of observation O_t while being in state S_t. To reduce the computational complexity of the problem the Viterbi algorithm is most commonly used (Trier et al. 2011).

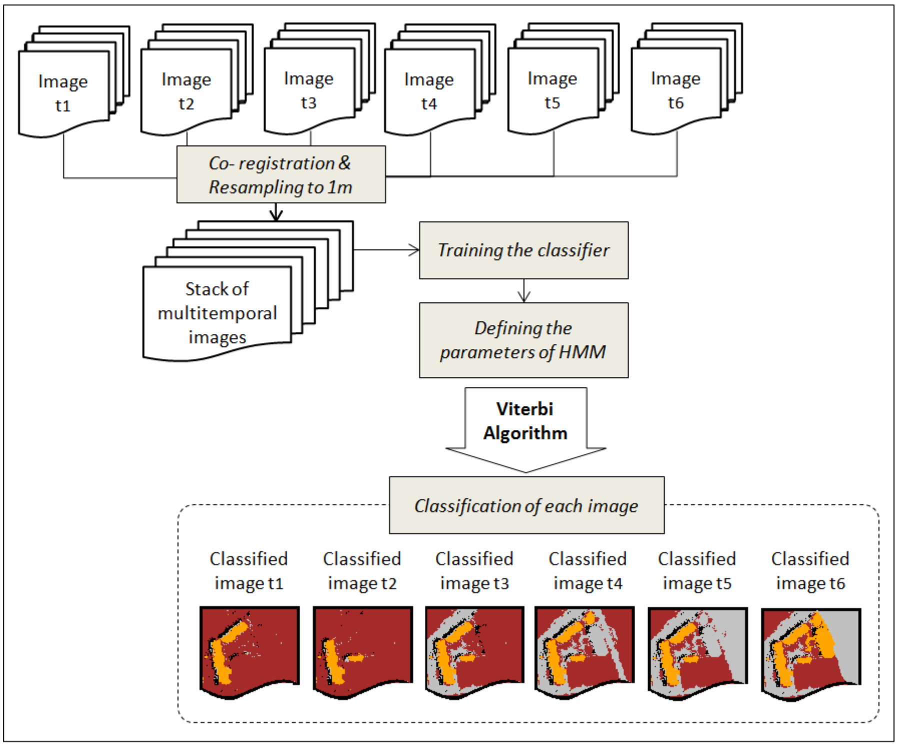

Figure 2. Schematic workflow of the developed classification methodology.

4. Results and Discussion

The method is applied on a set of 6 VHR images and 6 corresponding classified images are produced. Since the purpose of the study was to monitor the expansion of residential areas, the evaluation of our approach will focus on the validity of building detection. Hence, the classes of the produced classified images were grouped into two classes; “Buildings” and “Non-buildings”. To assess the contribution of our temporal classification scheme, the same set of images with the same training samples were also classified by the Maximum Likelihood (ML) algorithm.

The evaluation of the proposed methodology and the comparison with the ML algorithm was performed, with the help of appropriate ancillary ground truth data. The ground truth maps were produced by manual digitization of the boundaries of buildings in each image by an experienced image-interpreter. The evaluation measures of Completeness, Correctness and Overall accuracy were computed as they are usually used on building detection studies (Jin et al. 2005; Champion et al. 2009):

$$Completeness = 100 * \frac{TP}{TP + FN} \quad (2)$$

$$Correctness = 100 * \frac{TP}{TP+FP} \quad (3)$$

$$Overall\ Accuracy = \frac{TP+TN}{Total\ pixels} \quad (4)$$

In this context, True Positives (TP) are pixels that are correctly classified as "buildings", False Positives (FP) are the pixels that have been classified erroneously as "buildings" by the model and False Negatives (FN) are the pixels that have been classified as "non-buildings" while they were actually "buildings". The Completeness was computed by dividing the number of pixels that were True Positives by the correct number of pixels that were found as "buildings" from the image-interpretation. The Correctness was computed by dividing the True Positive pixels by the pixels that were classified as "buildings". The results of the accuracy assessment are illustrated in figures 3, 4 and 5.

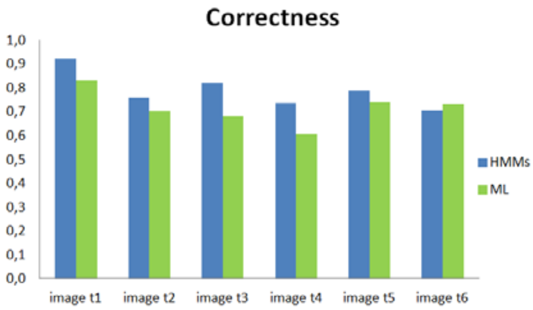

Figure 3. Diagram of the Correctness measure of each image classified by HMMs and ML.

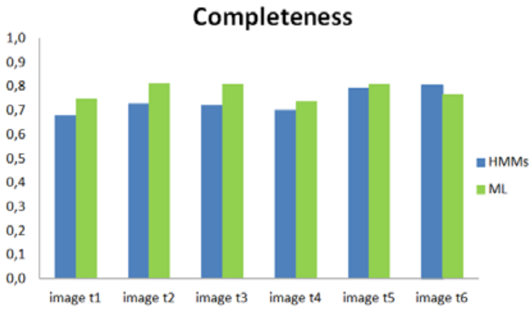

Figure 4. Diagram of the Completeness measure of each image classified by HMMs and ML.

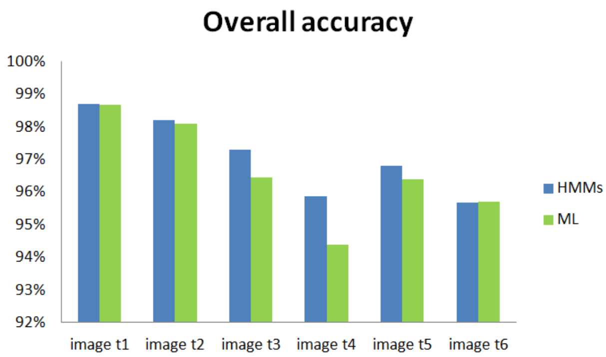

Figure 5. Diagram of the Overall accuracy of each image classified by HMMs and ML.

As we can observe the measure of Correctness in the case of the HMMs is higher than ML which means that in our method the pixels classified as buildings actually represented that class. This is actually the case in the examples of figure 6, where ML over-estimated the area of buildings in areas of bare soil. By a quick visual check on both classification maps some remarks can be made in favor of the temporal classifier. In cases of pixels which belong to class "Bare soil" at time t, but also have a spectral similarity to class "Buildings" the ML algorithm is likely to classify them as "Buildings". However, if at time t-1 those pixels belong to class "Bare soil" the temporal classifier is more likely to classify them as "Bare soil". This happens because the transition probability is higher from "Bare soil" to "Bare soil" than from "Bare soil" to "Buildings. Those pixels were correctly classified due to their increased transition probability of remaining at the same class.

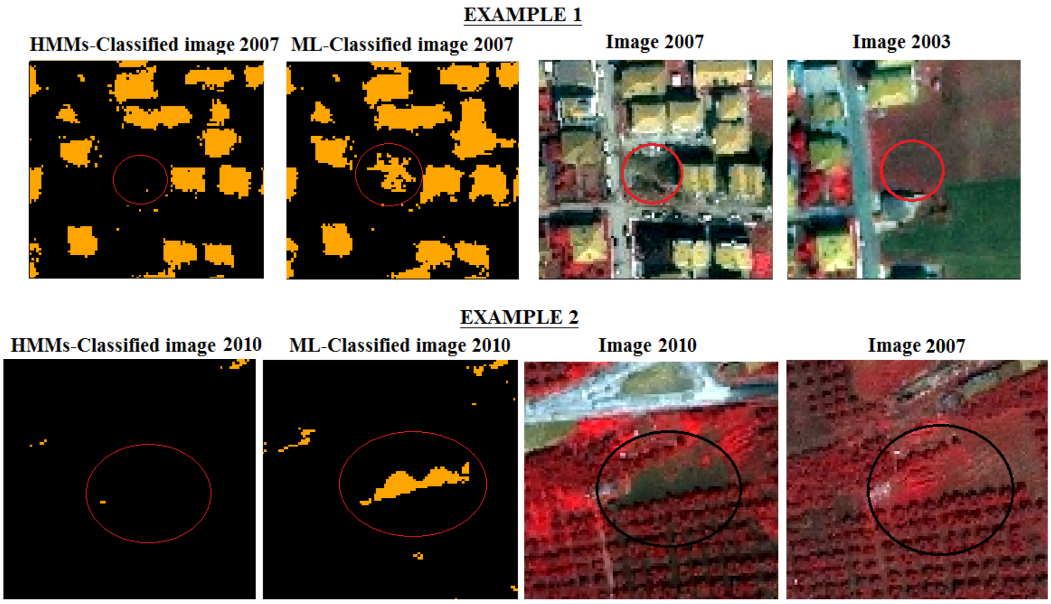

Figure 6. Two examples of the classified images produced by HMMs and ML algorithm, of areas whose true class was “Bare soil” in 2007 and 2010. In the classified images class “Buildings” is marked with yellow color and class “Non-buildings” is marked with black.

However the Completeness estimation proved that ML was superior in the sense that it did not omit to classify a pixel in its correct class compared to HMMs. This can be explained by the collection of multitemporal images; the look angle and Sun angle are critical factors affecting the reflectance of geographic objects. When the remote sensing images are not acquired at nadir both the building roofs and their sides are recorded and may be mapped at a different location in two images of different look angles (Chen et al. 2012) (Fig. 7). In figure 8 a building appears to be falsely underestimated by HMMs method compared to the ML algorithm. This is because those pixels in 2010 also resemble “bare soil” and were not “buildings” in the previous image of 2007 due to the different look angles of those images. To sum up, we conclude than in our approach the areas prone to errors are located in the perimeter of the true location of buildings. Hence, the use of multitemporal images can generate a number of erroneous “building” indications in those areas which can be either roofs or ground according to the look angle and are likely to be hidden by shadows in some images. Nevertheless the assessment of the Overall accuracy of each image (Fig. 5) showed that the HMMs approach prevails over ML in most images ranging between 95,7% and 98,7%.

Figure 7. Sub-scenes of images displaying roofs in acquisitions of 2010, 2011 and 2013 in different locations due to the different look angle of satellites.

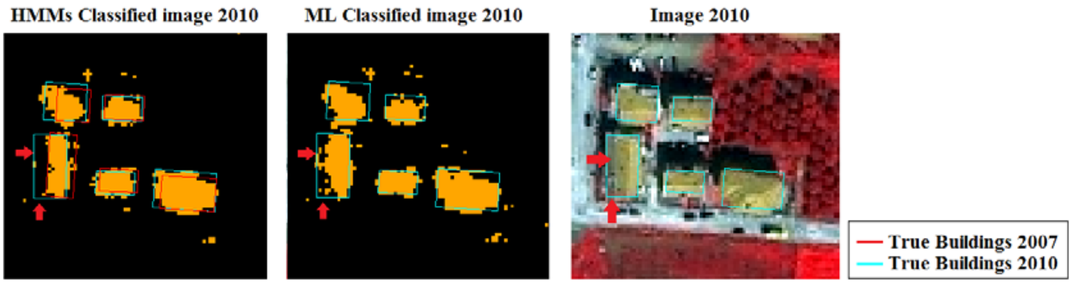

Figure 8. Sub-scenes of the classified images produced by HMMs and ML algorithm, and the corresponding image of 2010. Red color refers to the boundaries of buildings produced by image-interpretation of image 2007 and blue refers to buildings produced by image-interpretation of image 2010.

5. Conclusions

The goal of the present study was to develop a methodology to process multitemporal images simultaneously to produce improved classification maps. A HMMs classification approach was implemented on a time-series of 6 VHR images of a suburban area to monitor the expansion of urban sprawl and was compared to a ML single-image classification. The main advantage of the HMMs method is that it incorporates temporal knowledge apart from spectral. This benefit improves classification results when pixels resemble two classes but can be discriminated in previous images of the time-series set. Misclassification due to the resemblance of classes can be avoided if in the other images pixels can be correctly identified. The estimation of the overall accuracy of the proposed method has proven that the temporal classifier is superior to the single-date spectral classifier. However mapping buildings using multitemporal VHR imagery presents many difficulties due to imperfect image co-registration and to different image acquisition angles. The experimental results have shown that most of the errors refer to areas in the perimeter of buildings and occur frequently at the expense of the final multitemporal classification accuracy.

In future work many of the above limitations could be overcome by the use of true-ortho images. If the roofs of buildings are geometrically corrected and their location coincides in all images, multitemporal classification will not provide misleading information and will improve significantly the final classification maps.

References

Aurdal, L., Bang, H.R., Eikvil, L., Solberg, R., Vikhamar, D. & Solberg, A., 2005. Hidden Markov Models Applied to Vegetation Dynamics Analysis Using Satellite Remote Sensing. In Proc. of International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Biloxi, MS, USA, pp. 220-224.

Blaschke T., 2010. Object-based image analysis for remote sensing. ISPRS International Journal of Photogrammetry and Remote Sensing. 65(1):2–16.

Bouziani, M., Goita, K. , He, D.C., 2010. Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 65, Issue 1, 143-153.

Bruzzone L. & Prieto D. F., 2002. A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images Pattern Recognition Letters 23 (9), 1063-1071.

Champion N., Boldo D., Pierrot-Deseilligny M., Stamon G., 2009. 2D building change detection from high resolution satellite imagery: A two-step hierarchical method based on 3D invariant primitives. Pattern Recognition Letters, In Press, Corrected Proof, available online 22 October 2009.

Chen G., Hay G. J., Carvalho L. M. T. & Wulder M. A., 2012. Object-based change detection, International Journal of Remote Sensing 33(14): 4434-4457.

Doxani G., Karantzalos K., Tsakiri- Strati M., 2014. Knowledge-based building change detection from high resolution satellite imagery using ancillary geospatial information. In GEOgraphic Object Based Image Analysis (GEOBIA), Thessaloniki, 21-23 May 2014.

Doxani G., Siachalou S. and Tsakiri-Strati M., 2008. An object-oriented approach to urban land cover change detection. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37 (B7), p. 1655–1660.

Eriotis N., Poutos E., Retsinis E. and D. Vasiliou, 2013. The Impact of Greek Economic Crisis on Profit of the Greek Construction Industry, International Journal of Economy, Management and Social Sciences, Vol. 2, No.7, pp. 507-512.

Haala N., 1999. Combining multiple data sources for urban data acquisition. 47th Photogrammetric Week, Stuttgart, Wichmann, pp. 329-339.

Hermosilla T., Ruiz L. A., Recio J. A., and J. Estornell, 2011. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data, Remote Sensing, vol. 3, pp. 1188-1210.

Jin X. & Davis C. H., 2005. Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information. EURASIP Journal on Applied Signal Processing 14, 2196-2206.

Pacifici F., Frate F. D., Solimini C., and Emery W., 2007. An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery, IEEE Trans. Geosci.Remote Sens., 45(9), 2940–2952.

Leite, P., Feitosa, R., Formaggio, A., Costa, G., Pakzad, K. & Sanches, I., 2011. Hidden Markov models for crop recognition in remote sensing image sequences. Pattern Recognition Letters, 32, pp. 19–26.

Melgani F., Serpico S. B., Vernazza G., 2003. Fusion of Multitemporal Contextual Information by Neural Networks for Multisensor Remote Sensing Image Classification, Integrated Computer-Aided Engineering, vol. 10, no. 1, January 2003, pp. 81-90

Mithal V., Khandelwal A., Boriah S., Steinhaeuser K. and Kumar V., 2013. Change Detection from Temporal Sequences of Class Labels: Application to Land Cover Change Mapping In SDM 2013: Proceedings of the 13th SIAM International Conference on Data Mining, pages 650-658, 2013.

Mota G. L. A., Feitosa R. Q., Coutinho H. L. C., Liedtke C., Müller S., Pakzad K., and Meirelles M. S. P., 2007. Multitemporal fuzzy classification model based on class transition possibilities, International Journal of Photogrammetry and Remote Sensing, v. 62, iss. 3, p. 186-200.

Roeck, T.D., Der Voorde, T. V. and Canters, F., 2009. Full hierarchical versus non-hierarchical classification approaches for mapping sealed surfaces at the rural-urban fringe using high-resolution satellite data. Sensors, 9, pp. 22-45.

Siachalou S., Doxani G., Tsakiri-Strati M., 2014. Time-series Analysis of High Temporal Remote Sensing Data to Model Wetland Dynamics: a Hidden Markov Model Approach, SENTINEL-2 for Science Workshop, Frascati, Italy 2014.

Taubenböck H. and Esch T., 2011. Remote sensing – An effective data source for urban monitoring. In: Earthzine IEEE Magazine, URL:http://www.earthzine.org/2011/07/20/remote-sensing-%E2%80%93-an-effective-data-source-for-urban-monitoring/

Tran, T.-B., Puissant, A., Badariotti, D., Weber, C. 2011. Optimizing spatial resolution of imagery for urban form detection- The cases of France and Vietnam. Remote Sens. 2011, 3, pp. 2128–2147.

Trier, Ø.D., Salberg, A.-B., 2011. Time-series analysis of satellite images for forest cover change monitoring in Tanzania. 1st EARSeL Workshop on Operational Remote Sensing in Forest Management, Prague, Czech Republic, 2-3 June, 2011.

Viovy, N. & Saint G., 1994. Hidden Markov models applied to vegetation dynamics analysis using satellite remote sensing. Geoscience and Remote Sensing, IEEE Transactions on, 32, pp 906-917.