VillageFinder: Segmentation of Nucleated Villages in Satellite Imagery

Kashif Murtaza[1]
kashifm@lums.edu.pk
Sohaib Khan[1]
sohaib@lums.edu.pk
Nasir Rajpoot[2]
nasir@dcs.warwick.ac.uk

[1] Computer Vision Lab
LUMS School of Science and Engineering
Lahore, Pakistan

[2] Department of Computer Science
University of Warwick, UK

Abstract

Geo-spatial data on village locations, their size, population and other parameters is scarcely available to decision makers in many developing countries. In this paper, we demonstrate an automatic "crawler" which can segment nucleated villages from satellite imagery freely available in public domain geographic information systems such as Google Earth™ . Our approach is to use frequency and color features to generate a number of weak classifiers, which are then combined through Adaboost to produce the final classifier. We use a total of 69 features in the generation of the weak classifiers, including phase gradients, cornerness measures and color features. Our primary dataset consists of 60 images having more than 345 million pixels and covering more than 100 km^2 of area, containing nucleated villages in fifteen countries, spread over four continents and captured by different sensors. Using six manual annotations for ground-truth, we perform five-fold cross validation, using 25% of data for testing. Our results show an Equal Error Rate (ERR) of around 3.4%. Using the trained classifier, we detect villages on a 50 km^2 image (close to 184 million pixels) from a different site than the images used in training, and demonstrate highly accurate extraction of villages with 2.3% false positives and 0.01% false negatives.

1 Introduction

The importance of geo-spatial information for development planning cannot be over- emphasized. Remote Sensing technologies are extensively used in the West for land-use analysis, urban planning, crop mapping, planning of infrastructure, forest management and disaster assessment. A recent National Academies' report on the use of geographical information for sustainable development in Africa emphasizes the importance of geo-spatial information for developing countries, listing it as a key enabler of good governance [1]. Yet, in most developing countries, district-level administrators have limited, if any, access to geo-spatial information. This leaves them with no decision support mechanism and people have little information on their developmental rights, entitlements and comparative information of other similar communities.

While direct access to recent satellite imagery may be expensive for local authorities in developing countries, a wealth of image data is contained in online tools such as Google

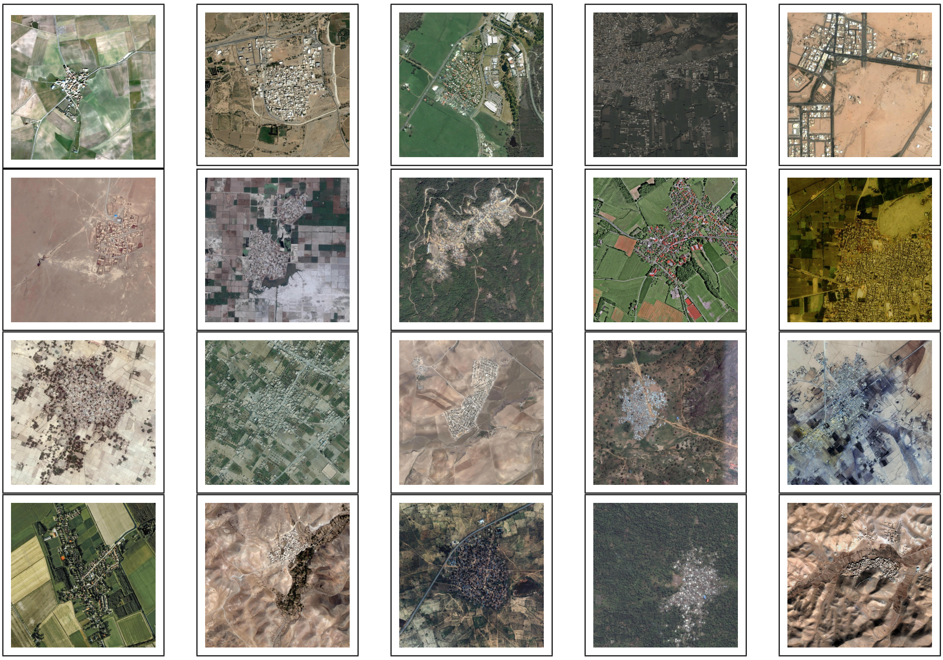

Figure 1: A selection of images containing nucleated villages. The terrain, type of construction, geographical location and imaging sensor generate high variability between images.

Earth™ and Microsoft's Live Search Maps™ . However, such images, while suitable for online viewing, do not yield the required geo-spatial information because of limited viewport and the fact that the relevant geographical features have to be extracted and processed before they are of use to the decision makers.

In this paper, we look at the problem of segmenting villages over a large area taken from Google Earth™ . The ability to accurately segment villages and measure their properties, such as their area, can be useful in a number of applications. For example, in disaster assessment, such data can used to quickly estimate of number of houses submerged in the case of a flood. We have developed a crawler to save high resolution image blocks of a large area from Google Earth™ . Once these images are saved, they are automatically stitched into a single image, which is then used for further processing. The extracted geo-spatial information may finally be presented in an appropriate manner, and may even be ported to GIS tools such as ArcGIS™ for further analysis and presentation. This provides a simple way for resource strapped users in developing countries to generate and access large scale geospatial datasets.

Robust segmentation of villages is a difficult problem because of the huge amount of variability in the shape, features and layout of a village. Two basic morphological forms of a village are *dispersed settlements* and *nucleated villages* [17]. Dispersed settlements, for example the common type of villages in the Himalayas, are actually a collection of homes scattered over a large area. This is in contrast to the concept of a nucleated village, the more common notion of a village, which is largely a contiguous collection of homesteads, built compactly around a village center. For the purposes of this paper, we concentrate on the segmentation of nucleated villages. Even within this category, variability can be huge; a mud-house village in Nigeria is very different in appearance from, say, a farmhouse

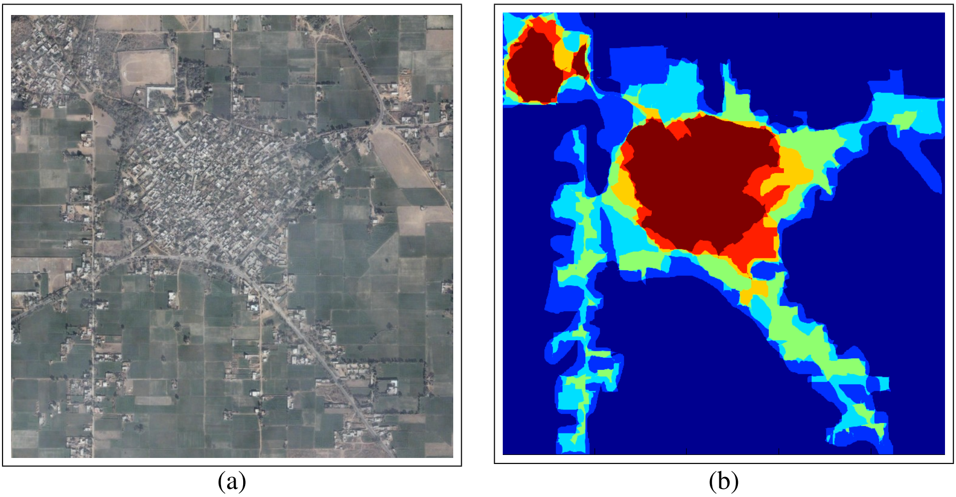

Figure 2: (a) Satellite image of a village near Faridabad, India. (b) Composite of ground truth marking by six annotators. The different colors indicate varying number of votes for each pixel (dark red - 6, orange - 5, yellow - 4, green - 3, cyan - 2, dull blue - 1 and dark blue - 0). Note the high variability in the understanding of village boundaries. Some observers mark the homes around roads as part of the village while others disagree.

community in the Netherlands. We tackle this variability by choosing examples of nucleated villages from a number of different countries in our training dataset, a portion of which is shown in Figure 1. Note the differences in sensor characteristics, terrain, type of construction and geographical layout in these examples.

Since we wish to accurately segment each village, the notion of what constitutes the boundary of a village becomes important. In most cases, the exact demarcation of where the village ends and the peripheral areas start is not clear. For the purposes of this paper, we consider the village extents to be delimited by the contiguousness of the houses in it; adjoining farmland or open spaces are not considered part of the village class. Even under this definition, the variability in what constitutes a village is very high between human observers. Consider the image of a nucleated village shown in Figure 2. Six annotators were asked to mark the ground truth of the village extents. Note the high variability in the marking, which is typical for most cases. Some annotators understood the village to extend to houses along the roads, while others disagreed. Similarly, an open field and clumps of trees north of the main cluster were marked by some as part of the village and not by others.

We collected a reasonably representative dataset of nucleated villages for training and testing our segmentation algorithm. We have created a dataset of 60 images, covering more than 100 km^2 of area. The approximate resolution of each image is 0.54 meters per pixel. The data consists of nucleated villages in four continents and 15 different countries. Manual annotations by six different persons are used as ground truth, and are fused together using a Kalman Filtering framework. Based on this, five fold cross validation, with 25% test data, is used during training. A total of 69 features, including phase gradient features [12], 'cornerness' features based on the eigenvalues of the gradient scatter matrix and a color measure are used. The features are given as input to Adaboost algorithm to yield the final classifier. Once the classifier is trained, we can now run it on a large image which is extracted from

Google Earth™ . We used an image spanning approximately 50 km^2, no portion of which is used in training the classifier. Our classifier extracted all villages reasonably accurately, with approximately 2.3% false positives at less than 0.01% false negatives when compared to the ground truth labeling.

1.1 Related Work

While land use mapping through satellite images has been in vogue in the remote sensing community since the sixties, most of the work till late eighties has concentrated on pixel based techinques, essentially learning the spectral signatures of various categories of land cover [4][7]. A number of more sophisticated approaches have been developed in recent years. A recent, excellent survey groups these methods into subpixel approaches, per-field approaches, contextual approaches, knowledge-based approaches and hybrid methods [9]. It has been only in the last few years, with the availability of high-resolution imagery, that segmentation and classification has been used together in land use analysis. The remote sensing community terms such methods as ‘object-oriented’ methods [9][14]. However, most of the papers concentrate on one type of sensor, such as QuickBird [18], IKONOS-2 [15]. While this paper is a limited application of land-use mapping with just one category, what distinguishes it is that we have not restricted our application to a single sensor, or to imagery captured at the same time. We have also not employed any radiometric or atmospheric calibration techniques, as is the norm in remote sensing literature. We have used a single spectral layer, with the view that this is what is available to end-users in the developing countries.

2 Features and Classification

Object segmentation can be achieved using shape, texture and color as cues [11, 19]. In case of our problem, however, nucleated villages can be of any arbitrary shape and size, which leaves us with texture and color as the other potentially useful cues. In this section, we present details of phase gradients as a texture descriptor. The phase gradient features combined with color and a measure of cornerness are used in our system to characterize nucleated villages.

2.1 Phase Gradient Features

Phase information, somewhat overlooked by researchers in texture analysis, can be used an effective cue for texture analysis since its gradient can both represent textural properties and point to the edge of textural regions. Let us model the visual texture in an image $v(\mathbf{x})$ as a linear combination of 2D sinusoidal wave functions and some additive white noise $w(\mathbf{x})$, as given below:

$$v(\mathbf{x}) = \sum_i a_i(\mathbf{x}) \cos \phi_i(\mathbf{x}) + w(\mathbf{x}) \tag{1}$$

where $\mathbf{x} = [x, y]^T$ and $\phi_i = [\phi_{ix}, \phi_{iy}]$ denotes the local phase.

In 1D, computing the gradient of local phase yields instantaneous or local frequency. Estimation of local frequency in 2D, however, is faced with a couple of problems. First, unlike signals in 1D, it is not entirely straightforward to compute the Hilbert transform of an image. Second, 2D signals such as images are usually multi-component, which begs the question: what exactly is local frequency? To overcome the former problem, solutions such as the analytic image [6] or the monogenic signal [11] have been proposed. The latter problem can be addressed by employing band-limited versions of the image, i.e., local frequency in 2D must be defined for each of the components, localized in frequency and orientation. We utilize log-Gabor filters [2, 8] to decompose a given image into its scale+orientation components, before computing the phase gradients for each of the components. These filters can be thought of as Gabor filters on a logarithm scale.

Let $v_i(\mathbf{x})$ denote the ith log-Gabor component of a given image $v(\mathbf{x})$. It can be represented as follows,

$$v_i(\mathbf{x}) = |v_i(\mathbf{x})| e^{j\phi_i(\mathbf{x})} \tag{2}$$

where $|v_i(\mathbf{x})|$ denotes the magnitude of v_i at a particular value of $\mathbf{x}$. Differentiating both sides of the above equation and re-arranging terms, we get the following expression for local phase gradient,

$$\phi_i'(\mathbf{x}) = j \left[\frac{|v_i(\mathbf{x})|'}{|v_i(\mathbf{x})|} - \frac{v_i'(\mathbf{x})}{v_i(\mathbf{x})} \right]. \tag{3}$$

The magnitude of $\phi_i'(\mathbf{x})$ as given below,

$$|\phi_i'(\mathbf{x})| = \sqrt{\frac{d\phi_i^2}{dx} + \frac{d\phi_i^2}{dy}} \tag{4}$$

gives local frequency in a particular direction, namely the direction perpendicular to the radial direction in the log-Gabor domain. To the best of our knowledge, ours is the first method to estimate the local frequency in this way. A vector of local frequency estimates for a range of scales and orientations yields a signature which is used to characterize the nucleated villages.

2.2 Additional Features

In addition to the phase gradient features, we also added "cornerness" measure as a feature. Motivated by [5] and [13], we compute the scatter matrix of the gradient vector over a 15 × 15 neighborhood and take the minimum eigenvalue as a feature. This captures the notion that we expect higher number of corner-type features in a village than in surrounding areas. This feature is computed at eight levels of the pyramid. We also added a color feature which is simply the green component of a pixel divided by its red plus blue components, to distinguish trees and foilage from the village structures.

2.3 Classification

For classification, we generate weak classifiers based on our features and combined them through Adaboost. We have used Gentle Adaboost, a variant of original Adaboost algorithm [3] that gives lesser weight to outlier datapoints. A publicly available implementation [16] was used in our experiments. This implementation uses CART trees as weak classifiers, but we set the depth of the tree to one (tree stump classifier), so that raw features are simply thresholded to generate the weak classifiers.

2.4 Fusion of Ground Truth

To fuse the multiple ground truth labelings of each image, marked by six annotators, we employed a Kalman Filtering framework as in [10]. To estimate the uncertainty of each

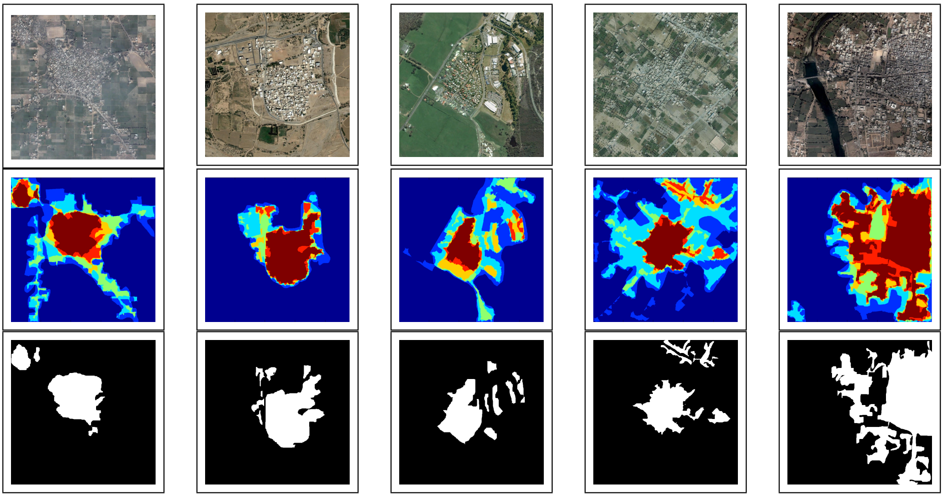

Figure 3: Fusion of ground truth. Top row shows the original images. Middle row shows a composite of six manual annotations (with the same coloring convention as used in Figure 2). Bottom row shows the fused ground truth.

annotator, we asked them to mark a couple of images thrice at different times. The variance of an annotator's markings, normalized by the sum of variances of all annotators, was used as a measure of uncertainty of that annotator, $\hat{\sigma}_i^2 = \sigma_i^2 / \sum_{j=1}^{6} \sigma_j^2$.

3 Experiments, Results and Discussion

In this section, we will describe our experimental setup and discuss the results which were obtained.

3.1 Generation of Datasets

We have written a crawler for Google Earth™ which has the capability of saving large satellite images. The user defines the resolution at which the image needs to be saved and the extents of the image, in terms of latitude and longitude coordinates of its bounding box. Since the entire image rarely fits on the screen, our program generates a sequence of requests for images at different camera locations, which are then screen-captured and saved. When the complete area has been captured, we mosaic the screen-captured images into a single large image. During this process, we keep the terrain layer turned off, so that perspective distortions are minimized, assuming that the original satellite data was reasonably orthorectified.

To select the locations of villages, a host of strategies were used. About 25 images were chosen by randomly browsing Google Earth™ . Another 25 were chosen by capturing 1000 random images from 10 different locations on the globe and manually identifying the ones which contained a village. An additional 22 images were captured by automatically picking random coordinates from settlement data available at `http://fallingrain.com`. However, because of registration errors, this strategy did not always yield a village at the captured location. Finally, all of the images were reviewed manually to see whether they contained a

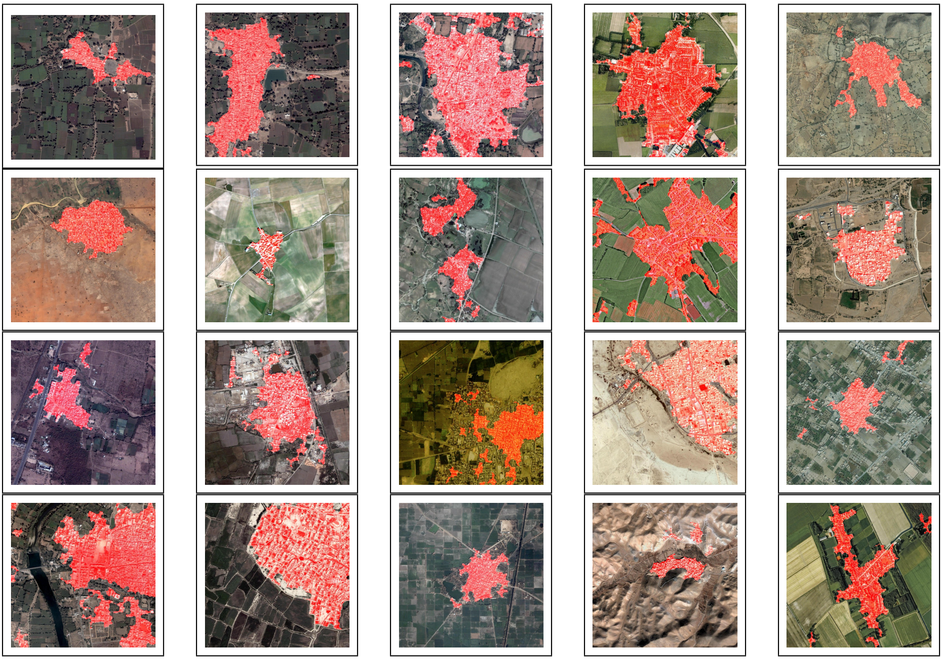

Figure 4: A selection of testing results obtained during cross-validation. Villages of differing characteristics are captured accurately, indicated by the red overlay.

portion of a nucleated village at a high spatial resolution. A total of 55 images containing nucleated villages and 5 images as negative examples were finally used in the dataset. Each image was captured at a resolution of 0.54 meters per pixel and is of size 2400 × 2400 pixels, covering approximately 1.7 km^2 on the ground. In this dataset, we have villages from fifteen countries spread over four continents, though around 80% of the dataset is derived from countries in Asia and Africa. Some of the images in this dataset are shown in Figure 1.

A secondary dataset for testing the trained classifier was generated by capturing a 50 km^2 image from a rural area with planar terrain. There is no overlap between this image and the original dataset of 60 images. This image was also captured at the same spatial resolution using similar methodology, and contains approximately 184 million pixels. The datasets used in this paper, along with the ground truths and the code will be made publicly available after the acceptance of the paper.

3.2 Marking of Ground Truth

We asked six volunteers to segment out the villages in each of the sixty images of the first dataset. The volunteers were given basic instructions in how to use Adobe Photoshop™ as a tool to mark the images. However, what constitutes a village was largely left to their judgement. The initial few markings of each volunteer were not used, and they marked those images again, presuming that this helped them learn the tool better. We fused these ground truths together using the Kalman Filtering framework described earlier.

The second dataset, consisting of the single large image, was marked only by one annotator. However, the annotator who had the most consistency in a test of marking the same

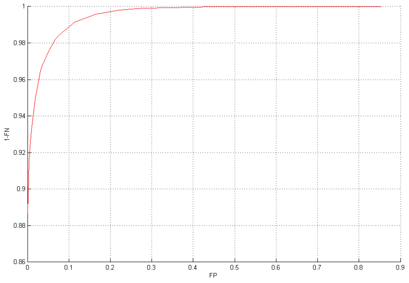

Figure 5: ROC curve of our classifier, obtained through five-fold crossvalidation on a dataset of 60 images, using 15 images in each fold for training. It shows an equal error rate of 3.4%

image at multiple times was asked to mark this image.

We have used the first dataset of 60 images for the training and testing of the classifier through cross-validation. A window size of 15 × 15 was used to compute phase gradient and cornerness features. The color feature was computed at each pixel independently. We used 6 scales and 10 orientation bins for the phase gradient features, 8 scales for cornerness features and one color feature. This resulted in a total of 69 features. For classification, we used an Adaboost implementation in tree stumps were used as weak classifiers [16].

3.3 Feature Extraction and Training

For training, we marked five 512 × 512 sub-images from each of the 60 images in the dataset. For each fold of cross-validation, we used sub-images from 45 of the 60 images for training. Testing was done on the rest of the 15 images using the entire image rather than the set of sub-images. We averaged features within a 16 × 16 window for training, and in a 8 × 8 window for testing. The final output of the classifier is passed through a set of morphological operations for cleanup. We first convolve the binary output of the classfier with a 32 × 32 box filter and threshold the result. Next we apply a closing operator and fill any interior holes. Finally we apply an area filter to remove small spurious detections. The same set of parameters for morphological operations were used in all our experiments.

3.4 Segmentation Results

Five folds of cross validation were used. We ensured that each image was used in testing at least once during cross validation. The results were averaged for each fold, which were then averaged to generate the final ROC curve, shown in Figure 5. The Equal Error Rate (ERR) on this experiment was observed to be approximately 3.4%. Figure 4 shows some of the testing results obtained during the cross-validation runs. Most of the results are highly accurate, even though we observed some false negatives, mainly in very low contrast images.

It should be appreciated here that the results are being compared against ground truth, the marking of which can have high variability, as illustrated in Figure 2. Moreover, our dataset is very dense in terms of the ratio of village pixels to non-village pixels. In an actual experiment over a large area, as the one described in the next subsection, the number of village pixels will be much smaller than the number of non-village pixels. This allows for

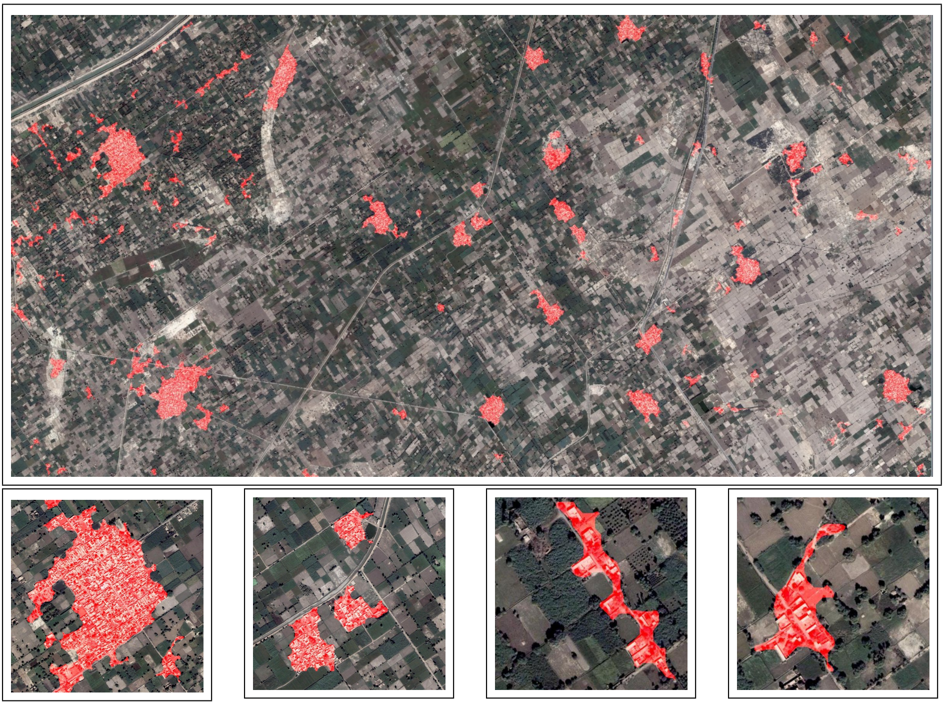

Figure 6: Top: Village segmentation of a 50 km^2 image of a rural area. The bottom row shows zoomed view of some of the segmented villages. The first two images are villages which were also marked in the ground truth. The last two images in this row show some of the “false positives” which actually contain housing clusters, but were not marked by the ground-truth annotator.

tuning of the classifier in favor of lesser false negatives, so that the impact of variability in marking of the villages can be reduced.

3.5 Results on a Large Image Obtained by the Crawler

The trained classifier was used to search for villages in a large contiguous 50 km^2 image obtained by our crawler. Even though each fold of cross validation returned a different trained classifier, we observed little difference in the final output between the different trainings. Thus, we simply used the first fold classifier for this experiment. The image was segmented using exactly the same parameters as before. The results are shown in Figure 6. Comparison with the ground truth yielded 2.3% false positives at 0.01% false negative rate. Most of the false positives were also observed to be small clusters of houses, which were not marked as a village by the ground truth annotator. Other false positives included small areas of shrubs which may accidently have similar frequency response. Almost no false negatives were observed, other than the slight variation in village boundaries between the ground truth and the output of our classifier.

4 Conclusions

We have demonstrated a system to accurately segment nucleated villages from freely available satellite imagery available in online geographic information systems such as Google Earth™ . The system employs phase gradient features along with color and cornerness measures to characterize the nucleated villages. To deal with the within-class variability of nucleated villages, we trained the system on images taken from varied locations and by different satellite sensors, and used six manual annotations to generate the ground truth, merged using a Kalman Filtering framework. Our approach shows excellent results when tested on a large image generated by a crawler working in Google Earth™ .

References

[1] Committee on Geographic Foundation of Agenda 21 Committee on Geography, Mapping Science Committee. *Down to Earth, Geographical Information for Sustainable Development in Africa*. National Academies Press, 2002. ISBN 978-0-309-08478-9.

[2] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394, 1987.

[3] Jerome Friedman, Trevor Hastie, , and Robert Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 38(2):337–374, April 2000.

[4] P.J. Hardin. Parametric and nearest-neighbor methods for hybrid classification: a comparison of pixel assignment accuracy. *Photogrammetric Engineering and Remote Sensing*, 60:1439–1448, 1994.

[5] C. Harris and M.J. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147 – 152.

[6] J.P. Havlicek, J.W. Havlicek, and A.C. Bovik. The analytic image. In *Proceedings International Conference on Image Processing*, volume 2, pages 446–449, 1997.

[7] J.R. Jensen. *Introduction to Digital Image Processing, A Remote Sensing Perspective*. Prentice Hall, 2nd edition edition, 1996.

[8] H. Knutsson. *Filtering and Reconstruction in Image Processing*. PhD thesis, Linköping University, Sweden, 1982. Diss. No. 88.

[9] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5): 823–870, March 2007.

[10] R.C. Luo, C.C. Yih, and K.L. Su. Multisensor fusion and integration: approaches, applications, andfuture research directions. *IEEE Sensors Journal*, 2(2):107–119.

[11] B.W. Mel. Seemore: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural computation*, 9(4):777–804, 1997.

[12] Nasir Rajpoot and RR Coifman. Phase gradients for texture analysis. In *MVA-AVA Symposium on Image Features and Statistics*, London, UK, October 2004.

[13] Jianbo Shi and Carlo Tomasi. Good features to track. In *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593 – 600.

[14] Gregor Willhauck Iris Lingenfelder Markus Heynen Ursula C. Benz, Peter Hofmann. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58:239–258, 2004.

[15] C. J. van der Sande, S. M. de Jong, and A. P. J. de Roo. A segmentation and classification approach of ikonos-2 imagery for land cover mapping to assist flood risk and flood damage assessment. *International Journal of Applied Earth Observation and Geoinformation*, 4(3):217 – 229, 2003. ISSN 0303-2434.

[16] Alexander Vezhnevets. GML Adaboost Matlab Toolbox. http://research.graphicon.ru/machine-learning/gml-adaboost-matlab-toolbox-3.html.

[17] David Waugh. *Geography, an integrated approach*. Nelson Thrones, 3rd edition edition, 2000. ISBN 978-0174447061.

[18] Wenbo Xu, Bingfang Wu, Jianxi Huang, Yong Zhang, and Yichen Tian. A segmentation and classification approach of land cover mapping using quick bird image. volume 5, pages 3368–3370 vol.5, Sept. 2004.

[19] Y. Zhong and A.K. Jain. Object localization using color, texture and shape. *Pattern Recognition*, 33(4):671–684, 2000.