THE EFFECTS OF VIEWING- AND ILLUMINATION GEOMETRY ON SETTLEMENT TYPE CLASSIFICATION OF QUICKBIRD IMAGES

F. van den Bergh

Remote Sensing Research Unit
CSIR Meraka Institute
Pretoria, South Africa
fvdbergh@csir.co.za

Index Terms**—** Image classification, Urban areas, Image texture analysis

ABSTRACT

Image texture features extracted from high-resolution remotely sensed images over urban areas have shown promise in their ability to distinguish different settlement classes. Without any explicit mechanism to counter the effects of variable illumination- and viewing geometries, these features may not generalize well in multi-date applications such as change detection. This paper presents the results of a small study of the effects of unwanted variability on low-income settlement classification performance in the Soweto residential area of the city of Johannesburg, South Africa.

Somewhat surprisingly, the Gray-Level Co-occurrence Matrix (GLCM) features were found to perform better than Local Binary Pattern (LBP) features on combined spatial and temporal generalization tasks, although the LBP features offered better performance on spatial-only generalization problems.

1. INTRODUCTION

Image texture features have been shown to be an effective means of describing the structural features that sets apart different urban settlement classes. Early work in this category includes that of Pesaresi [1], Benediktsson *et al.* [2]. These features can also be used to detect the presence of urban areas within satellite imagery, as recently demonstrated by May *et al.* [3].

Texture features have also been demonstrated to be effective in classifying low-income and informal settlement classes such as those found in South Africa [4]. The regular monitoring of such informal settlements is critical to the service provisioning planning process, thereby helping to bring services such as water and sanitation to these areas in a timely manner.

Although some texture features, such as those generated by the Local Binary Pattern (LBP) algorithm, have been demonstrated to be highly effective in the settlement classification task, subsequent experiments have demonstrated that the generalization performance of classifiers using these features is not ideal. The study presented here is an early investigation into identifying the cause of poor generalization performance. It is expected that viewing- and illumination geometry are the two dominant factors that hamper the generalization performance of classifiers based on such texture features, since the illumination geometry has a direct impact on both the amount and direction of shadowing within a scene.

2. METHODOLOGY

A good image feature is one that is designed to have a representation that is sensitive to change in the desired variables, e.g., settlement type, whilst being insensitive to other types of change that may be present in the image. If an image feature has these properties, then it can be expected that the feature will lead to good generalization performance in classification tasks. The high spatial resolution of QuickBird presents several hazards to good generalization in that many differences can be observed in co-registered image pairs, but that most of these differences are unrelated or irrelevant to a settlement classification problem.

The QuickBird satellite can acquire images at fairly large off-nadir angles. Imaging the same area at two different off-nadir angles produces an image pair that may contain a large number of spurious differences, even though no real change occurred on the ground. This effect is referred to as viewing geometry differences, and is the first source of unwanted change that is observed in repeated QuickBird images of the same area.

Images acquired at different times of the year also have different illumination geometries, which manifests as shadows of varying lengths and orientations. This effect could also be exacerbated by large off-nadir viewing angles, because the local time on the ground may be different compared to a nadir view of the same area. This also introduces spurious differences in image pairs of the same area, and is referred to as illumination geometry differences.

A third type of spurious change can be observed if two images from different seasons are compared. This affects mainly vegetation, but the peak of the spectral response function of the panchromatic sensor on board QuickBird is shifted towards red / near infrared, which implies that seasonal vegetation changes are prominent in the panchromatic band. It is expected that the effect of the seasonal differences will be smaller in urban environments, or that these seasonal differences will not dominate real structural differences.

The following method is proposed to evaluate the sensitivity of different texture features to these sources of spurious change:

1. Obtain two QuickBird images over the same area, preferably acquired not too far apart in time (minimizing seasonal differences), but with different viewing- and illumination geometries.

2. From each image, extract polygons containing examples of different settlement types. Multiple non-overlapping examples of each settlement type are extracted.

3. From within each polygon, extract square tiles (120 m $\times$ 120 m) from random locations entirely within the demarcated polygons. Tiles are paired, so that the same location is extracted from both dates (images).

4. Extract image feature vectors (patterns) using various texture feature algorithms. Partition these extracted patterns into four sets, namely A_{d_1}, A_{d_2}, B_{d_1} and B_{d_2} where A_{d_1} denotes the first set derived from the scene acquired on date d_1, and B_{d_1} denotes a second set derived from the same scene. Observe that the B sets do not overlap the A sets spatially.

5. Evaluate the generalization performance of the different texture feature algorithms by evaluating the performance of a Support Vector Machine (SVM) classifier over the six possible combinations of training and test sets, e.g., A_{d_1} used as training set, evaluated on B_{d_2} as test set, and so on.

It is expected that texture features that are less sensitive to viewing- and illumination geometry will produce temporal generalization classification accuracies (training and testing data from different dates) that are comparable to their spatial generalization performance (training and testing data from same date, but spatially non-overlapping), whereas less robust features will exhibit a decrease in performance.

3. DATA SET

To evaluate the effect of these possible sources of spurious differences, a familiar study area, namely Soweto (Gauteng, South Africa) was selected. This area contains a large variety of settlement types, ranging from formal suburbs, to very informal settlements consisting of owner-built shacks. Two QuickBird images over this area were obtained: one acquired on 2005-10-18 (early summer, rain season, called d_1 in the sequel), and another on 2006-05-30 (early winter, called d_2 in the sequel). Large differences in viewing- and illumination geometry between these images lead to a pronounced difference in both the amount of shadow and the orientation of shadows.

This study area has previously been used in settlement classification experiments [4]. Unfortunately, the area of overlap between the two QuickBird images was smaller than previous studies, and it no longer contained sufficient examples of the unstructured *informal settlement* class. To compensate for the slight loss of variety, it was decided to merge the classes leaving just four final classes: *formal suburbs (FS)*, *formal suburbs with backyard shacks (FSB)*, *ordered informal settlements (OIS)*, and a *non-built-up (NBU)* class to represent vegetation and bare areas. For each date a set of 6794 patterns were selected for training data, leaving 7560 patterns in the testing set.

Examples of three of the settlement classes are provided in Figure 1. Observe how the different settlement classes each have a unique distribution of object (building) sizes. The relative position of the buildings is also an important attribute.

4. RESULTS

A previous study investigated the performance of various texture features in a comparable settlement classification task [4]. The three best performing algorithms from that study were selected: Gray-Level Co-Occurrence Matrix (GLCM) features [5], Local Binary Patterns (LBP) features [6], and Granulometric features [7]. Reasonable effort was expended to ensure that the parameter choices of these algorithms were suitable for the intended task. A total of 10 attributes were extracted from each image using the GLCM algorithm, with LBP and Granulometry weighing in at 30 attributes and 14 attributes, respectively.

In order to obtain standard deviations on the various classification accuracy results, the following procedure was used to evaluate a given configuration using data sets X and Y (where $X = \mathrm{A}_{d_1}$, and $Y = \mathrm{B}_{d_1}$, for example):

1. Train a support vector machine (SVM) using the whole of set X.

2. Partition set Y into 10 folds using stratified sampling to preserve relative class frequency.

3. Evaluate the SVM (trained on X) on each of the 10 folds of Y, obtaining one accuracy figure for each fold.

4. Exchange X and Y, and repeat steps 1–3.

This process, denoted $X \rightleftharpoons Y$, produces 20 individual values for each accuracy metric, which were then used to calculate a mean and standard deviation for each metric (Table 1).

(a) formal settlement (b) formal settlement with backyard shacks (c) ordered informal settlement

Fig. 1. Examples of three of the settlement classes found in Soweto

Table 1. Overall classification accuracy obtained with various texture algorithms, and per-class true positive (TP) rate. All values represent sample means followed by corresponding standard deviations.

Texture algorithm	**Data set**	**Overall Accuracy (%)**	**FS TP (%)**	**FSB TP (%)**	**OIS TP (%)**	**NBU TP (%)**
GLCM	$A_{d_1} \rightleftharpoons B_{d_1}$	92.58 ± 0.60	57.62 ± 43.14	99.52 ± 0.22	94.28 ± 5.85	100.00 ± 0.00
	$A_{d_2} \rightleftharpoons B_{d_2}$	95.96 ± 0.60	83.15 ± 15.21	97.84 ± 1.89	96.63 ± 2.15	100.00 ± 0.00
	$A_{d_1} \rightleftharpoons B_{d_2}$	87.18 ± 1.15	57.51 ± 34.79	86.69 ± 1.33	92.67 ± 7.62	99.92 ± 0.24
	$A_{d_2} \rightleftharpoons B_{d_1}$	88.21 ± 3.74	65.79 ± 33.06	92.22 ± 8.00	82.07 ± 14.96	99.29 ± 0.81
	$A_{d_1} \rightleftharpoons A_{d_2}$	91.37 ± 2.84	73.14 ± 20.83	89.99 ± 7.47	95.58 ± 4.57	99.95 ± 0.15
	$B_{d_1} \rightleftharpoons B_{d_2}$	90.56 ± 1.77	70.67 ± 28.74	94.22 ± 2.44	84.41 ± 15.78	97.89 ± 2.29
LBP	$A_{d_1} \rightleftharpoons B_{d_1}$	96.66 ± 1.69	90.08 ± 7.86	96.72 ± 3.24	98.06 ± 2.23	99.60 ± 0.55
	$A_{d_2} \rightleftharpoons B_{d_2}$	97.38 ± 1.57	96.06 ± 3.40	97.43 ± 2.15	94.41 ± 4.77	100.00 ± 0.00
	$A_{d_1} \rightleftharpoons B_{d_2}$	87.79 ± 1.76	78.78 ± 28.61	97.19 ± 2.65	54.11 ± 9.06	100.00 ± 0.00
	$A_{d_2} \rightleftharpoons B_{d_1}$	80.79 ± 0.56	77.42 ± 22.96	86.42 ± 7.03	38.72 ± 33.16	98.27 ± 1.89
	$A_{d_1} \rightleftharpoons A_{d_2}$	88.89 ± 5.51	79.86 ± 26.01	98.23 ± 1.60	63.88 ± 34.98	100.00 ± 0.00
	$B_{d_1} \rightleftharpoons B_{d_2}$	81.94 ± 5.26	77.00 ± 23.32	89.08 ± 7.05	37.87 ± 2.31	98.47 ± 1.71
Granulometry	$A_{d_1} \rightleftharpoons B_{d_1}$	81.06 ± 1.26	60.68 ± 32.15	91.65 ± 5.38	46.23 ± 3.54	99.92 ± 0.18
	$A_{d_2} \rightleftharpoons B_{d_2}$	81.99 ± 1.49	73.26 ± 20.48	92.96 ± 5.00	40.52 ± 11.56	99.30 ± 0.91
	$A_{d_1} \rightleftharpoons B_{d_2}$	70.85 ± 1.20	62.98 ± 21.09	68.98 ± 3.16	38.62 ± 21.33	99.21 ± 0.89
	$A_{d_2} \rightleftharpoons B_{d_1}$	72.62 ± 6.88	59.34 ± 30.91	78.66 ± 13.00	30.45 ± 23.85	99.23 ± 0.74
	$A_{d_1} \rightleftharpoons A_{d_2}$	76.77 ± 8.17	69.08 ± 21.50	72.53 ± 9.36	55.89 ± 33.59	99.35 ± 0.43
	$B_{d_1} \rightleftharpoons B_{d_2}$	78.38 ± 6.80	61.48 ± 24.89	79.80 ± 11.87	52.87 ± 12.35	99.72 ± 0.55

From the results in Table 1 it is clear that the Granulometric features did not perform well, even when both training and test data sets were selected from the same date (e.g., $\mathrm{A}_{d_1} \rightleftharpoons \mathrm{B}_{d_1}$). The spatio-temporal generalization accuracy (e.g., $\mathrm{A}_{d_1} \rightleftharpoons \mathrm{B}_{d_2}$, $\mathrm{A}_{d_2} \rightleftharpoons \mathrm{B}_{d_1}$) obtained with the Granulometric features was only about 71%, which is not sufficient for an automated mapping application. Both the GLCM and LBP methods performed well on the spatial generalization (same date) tasks, with LBP faring slightly better. The GLCM features performed well on both temporal generalization and spatio-temporal generalization tasks, achieving accuracies close to 90% in these tasks.

Overall accuracy does not tell the whole story, though. Examination of the per-class true positive rates (Table 1, right hand side) show that LBP and GLCM each have a weak class (OIS and FS, respectively) whenever the training and test data sets were selected from different dates, as represented by the last four rows of results for each algorithm. The OIS and FS classes turn out to be the smallest two classes in this study, which may have biased the SVM training process, though.

A significant result of this initial robustness study is thus that the GLCM features may be more suited to this problem than an earlier study indicated. Although the LBP method performs slightly better than GLCM when both training and testing patterns were collected on the same date, it fails to impress in the across-date classification experiments.

The main purpose of the study, though, was to quantify the influence of spurious differences on the generalization performance of classifiers using texture features as input, represented directly by the $\mathrm{A}_{d_1} \rightleftharpoons \mathrm{A}_{d_2}$ and $\mathrm{B}_{d_1} \rightleftharpoons \mathrm{B}_{d_2}$ classes. Here, the GLCM features consistently produced the best overall classification accuracies, indicating that these features may be somewhat robust to changes in viewing- and illumination geometry. Even so, the performance of the GLCM features do appear to have deteriorated under these influences. The LBP features appear to have experienced an even more pronounced drop in generalization performance across dates. The results of this study indicate that texture feature algorithms are indeed influenced by viewing- and illumination geometry of satellite images.

Lastly, the effects of seasonal changes on vegetation do not appear to affect generalization performance much. The temporal as well as spatio-temporal generalization performance of all three the methods on the non-built-up (NBU) class remained comparable to the accuracies obtained in the same-date experiments.

5. CONCLUSION

An ideal image feature would be sensitive to real differences between the various settlement classes, while remaining insensitive to spurious differences caused by viewing- and illumination geometry differences. This paper presented the results of a study to investigate whether well-known texture features behave in this manner.

The results of a classification experiment involving two scenes of the same area acquired under different conditions indicate that the GLCM method may yet produce the best features under these conditions. Unfortunately, the generalization performance of the GLCM features were still far from ideal, yielding overall generalization classification accuracies of less than 95%. Future work will focus on attempts to directly improve the performance of these algorithms by explicitly treating the shadowed parts of the images differently.

6. REFERENCES

[1] M. Pesaresi, "Texture Analysis for Urban Pattern Recognition Using Fine-resolution Panchromatic Satellite Imagery," *Geographical and Environmental Modelling*, vol. 4, no. 1, pp. 43–63, 2000.

[2] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, no. 9, pp. 1940–1949, 2003.

[3] S. May and J. Inglada, "Urban area detection and segmentation using OTB," in *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*. IEEE, 2010, vol. 4.

[4] L. Abeigne-Ella, F. van den Bergh, B. J. van Wyk, and M. A. van Wyk, "A Comparison of Texture Feature Algorithms for Urban Settlement Classification," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*. IEEE, 2009, vol. 3.

[5] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 610–621, 1973.

[6] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[7] Y. Chen and E. R. Dougherty, "Gray-scale morphological granulometric texture classification," *Optical Engineering*, vol. 33, pp. 2713–2722, 1994.