Adaptive Threshold-Based Shadow Masking for Across-Date Settlement Classification of Panchromatic QuickBird Images

F. P. S. Luus, F. van den Bergh, and B. T. J. Maharaj

Abstract—Multitemporal land-use analysis is becoming increasingly important for the effective management of earth resources. Despite that, consistent differences in the viewing- and illumination geometry in satellite-borne imagery introduce some issues in the creation of land-use classification maps. The focus of this study is settlement classification with high-resolution panchromatic acquisitions, using texture features to distinguish between settlement classes. The important multitemporal variance component of shadow is effectively removed before feature determination, which allows for minimum-supervision across-date classification. Shadow detection based on local adaptive thresholding is employed and experimentally shown to outperform existing fixed threshold shadow detectors in increasing settlement classification accuracy. Both same and across-date settlement accuracies are significantly improved with shadow masking during feature calculation. A statistical study was performed and found to support the hypothesis that the increased accuracy is due to shadow masking specifically.

Index Terms—Image texture analysis, feature extraction, urban areas, remote sensing.

I. INTRODUCTION

MULTITEMPORAL satellite-borne image acquisition introduces complex variances due to a conflation of differences in viewing angles, illumination characteristics and environmental factors of the captured scenes. Multitemporal land-use analysis has to filter out these artificial differences to obtain an accurate account of actual land-cover changes. In the case of supervised settlement classification with limited ground truth data, such calculated invariance can substantially improve across-date classification accuracy. The varying viewing- and illumination geometry in multitemporal imagery must be accounted for since those differences will normally become embedded in the classification features and compromise the supervised labelling.

The QuickBird satellite acquires scenery at potentially very different azimuth angles due to its maximum 45° off-nadir wide accessible ground swath. Viewing geometry variations

F. P. S. Luus, and B. T. J. Maharaj are with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: luus@ieee.org).

F. van den Bergh is with the Remote Sensing Research Unit, Meraka Institute, Council for Scientic and Industrial Research, Pretoria 0001, South Africa.

aggravates the strong directional differences seen in urban surface features [1]. The sun elevation and solar illumination characteristics during acquisition may also have significant across-date differences, which produce considerable illumination variance of which shadowing is a dominant factor. Changes in dynamic range of the scene intensity is another effect of illumination variance, but shadowing is more adverse since its presentation is locally coupled with surface features and thus more difficult to remove.

Shadow detection is the initial step after which effective shadow removal is achieved through mainly shadow correction or masking. Augmenting data such as digital surface models have been used to perform topographic correction of surface reflectance, and terrain illumination correction with shadow and occlusion detection [2]. Most shadow detection methods depend on rich multispectral imagery and the shadow property of maintained relative color but reduced intensity (retinex theory) [3], and the use of near-infrared information [4].

Tsai presented an algorithm which uses the ratio value of the hue over the intensity to construct the ratio map for shadow detection in color aerial images [5]. The detection accuracy was improved with a successive thresholding scheme utilized by Chung et al. [6]. The shadow property of low intensity can be used to good effect for shadow detection [7], where fixed intensity thresholding is a predominant panchromatic detection method [8]. Bimodal histogram splitting is suitable for large shadow detection [9], but finding a suitable threshold for the smaller shadows of urban structures is non-trivial. This paper focuses on panchromatic-only imagery, since multispectral methods are well studied, and for the research to also be useful for imagery from satellites like WorldView-1.

The authors previously explored fixed thresholding for panchromatic shadow removal [7] and aim in this new study to expand the investigation into the relative performance benefits of local adaptive thresholding for shadow detection. Adaptive thresholding allows for more robust shadow detection, especially for gradient shadows, and its threshold parameters require less across-date modification than with fixed thresholding. The causal relationship between shadow detection accuracy and the accuracy improvement in across-date settlement classification with shadow masking is also statistically analysed in this work.

The original image and its detected shadow mask are used to perform shadow masking, and in this study has the purpose of effectively removing multitemporal shadow variances by removing the shadows. Histogram matching,

gamma correction and linear-correlation correction [10] are popular methods of shadow correction, where shadows are lifted to have similar intensity to that of the surroundings. Corrected shadows suffer from posterization and the effect on classification accuracy is of concern. Alternatively this study will opt for shadow masking, where shadows are ignored during feature calculation.

Pattern classification relies on differential measures to distinguish between different classes, but in the instance of across-date settlement classification the potentially severe shadow variances can cause intra-class feature variation. This leads to a higher incidence of class confusion and overlap, so it is the objective of shadow invariance to remove the significant illumination variance from the feature calculation. Texture features are used as it has been shown to be a good measure of settlement patterns (see [11]).

The methodology is explained in Section II and a description of the study area is given in Section III. The experimental setup and objectives are elaborated on in Section IV and the results and a discussion thereof are provided in Section V, before conclusions are drawn in Section VI.

II. METHODOLOGY

The effects of across-date shadow variances on settlement classification accuracy are investigated by removing shadows before calculating settlement features. Same-date and across-date settlement classification are analysed with the shadows intact and with the shadows removed to ascertain the influence of shadow differences, primarily on classification accuracy. The modified features then capture the settlement distinctions minus the temporal illumination geometry variances.

Gray-level co-occurrence matrix (GLCM) [12] and local binary pattern (LBP) [13] texture features have been shown to perform well in settlement determination (see [11]) and are used in this study. However, texture features such as GLCM and LBP are sensitive to viewing- and illumination geometry differences [11], so these features have the potential to benefit from calculated feature invariance.

The shadow-effect mitigation method involves shadow detection and removal as shown in Figure 1, where detection is the determination of a binary shadow mask indicating the perceived occlusion of sunlight. Fixed intensity thresholding can be applied to produce a rudimentary shadow mask [7] by declaring all pixels with an intensity less than the fixed threshold as shadow, but here the accuracy of the mask has a strong dependence on the specific threshold chosen. A range of experiments is thus performed for fixed thresholding with a threshold of up to 40% of the maximum intensity as determined by the image bit depth.

Local adaptive thresholding is better suited as a panchromatic detector since the actual threshold is determined relative to the mean intensity in a local window. A pixel is declared as shadow in the mask if its intensity is less than the mean pixel intensity in the square window centered at that pixel, minus a given offset intensity value, which gives local adaptive thresholding two main parameters, i.e. window size (number of edge pixels) and the adaptive offset. Shadow intensity
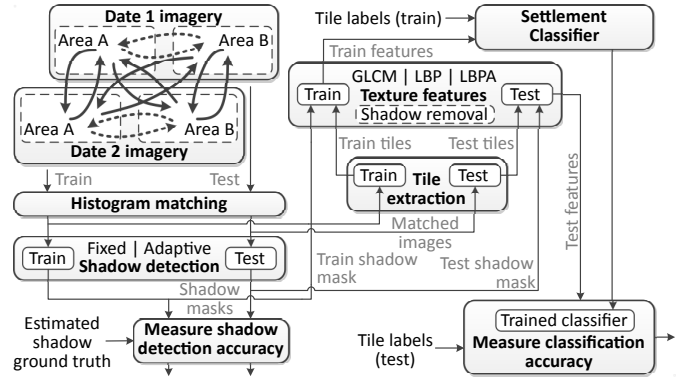

Fig. 1. An outline of the experiment methodology, and the separation of same-date (– –) and across-date (——) experiments (top left).

perturbations in the same scene and across-date dynamic range differences can thus be better accommodated than with a fixed threshold. To simplify analysis the shadow detection experiments are performed in terms of only one threshold, so multi-date images are initially histogram matched to reduce dynamic range variances and then the threshold detector classifies every pixel with an intensity less than the threshold as shadow. Fixed intensity thresholding is used as a benchmark to test local adaptive thresholding against, where the adaptive threshold is tested for the range of 0% to 40% of the maximum intensity subtracted from the local window intensity average.

The binary shadow mask provided by the detector is then directly used during texture feature calculation in order to mask out shadow areas. Every GLCM pixel pair is skipped if one of those pixels is located in shadow, which then effectively removes off-diagonal entries in the co-occurrence matrix that would have resulted with no masking. The LBP features are calculated in a similar manner so that a pattern is not placed when its central pixel forms part of a shadow area. The shadow masking is intensified in a variant named LBPA (LBP-All) where a pattern placement would be skipped if any of its pixels falls in the shadow mask. This differentiation between LBP and LBPA is examined to enable further conclusions to be drawn regarding the efficacy of stricter shadow masking.

III. DATA DESCRIPTION

The settlement classifier is evaluated in a study area with three main settlement types as shown in Figure 2, namely formal settlements (FS), formal settlements with backyard shacks (FSB) and ordered informal settlements (OIS). A fourth non-builtup class (NBU) that includes natural vegetation is added to test classifier separability between builtup and non-builtup areas. Formal settlements are characterised by permanent residential structures that are positioned in a planned manner, while formal settlements with backyard shacks have larger residential structures accompanied by smaller backyard shacks. When permanent and semi-permanent residential structures are arranged in a planned manner an ordered informal settlement is constituted.

A 4.85×9.86 km^2 section of the subtropical highland of Soweto (Gauteng, South Africa) was selected as the study site. Panchromatic QuickBird imagery of the site was captured at

a 0.6×0.6 m^2 pixel resolution, with a nominal 30° off-nadir wide accessible ground swath, on 18 October 2005 (d_1, early summer, rain season) and 30 May 2006 (d_2, early winter). The across-date settlement classification is investigated for these two dates denoted by d_1 and d_2.

The 11-bit panchromatic scenes were across-date histogram matched before representative polygons of each settlement type were selected in an assisted manner to produce multiple co-registered sample pairs over both dates. Image tiles of 120×120 m^2 areas were then automatically extracted from the polygon pairs with tile overlap to maximise polygon coverage.

IV. EXPERIMENTAL SETUP

The purpose of the study is to determine how shadow masking can increase settlement classification accuracy, so experiments are performed to determine the shadow detection accuracy and to measure the settlement classification accuracy with and without shadow masking. The correlation between shadow detection accuracy and settlement classification is also measured to make inferences about the mechanism underlying settlement classification accuracy improvements.

1) Settlement classification accuracy: A distinction is made between same-date and across-date experiments to test the hypothesis that differences in shadow profiles contribute to settlement classification inaccuracy. The tiles from each scene (d_1 and d_2) are separated into areas A and B, which constitute interchangeable training and testing datasets. For two acquisitions of the study area there are 4 same-date and 8 across-date experiments, as shown in Figure 1. The first datasets A_{d_1} and A_{d_2} consist of 125, 202, 142 and 335 co-registered labelled tiles respectively for the FS, FSB, OIS and NBU classes. The remaining datasets B_{d_1} and B_{d_2} contain 91, 263, 154 and 318 tiles for the FS, FSB, OIS and NBU classes, respectively.

Texture features are calculated per labelled tile, using either GLCM or LBP features. The first 13 of Haralick's GLCM features [12] are determined with the GLCM window having the tile dimensions of 200×200 pixels. Using the first 13 GLCM features proved to be more accurate for the study area, compared to using a feature subset of size 6 based on information gain or correlation feature selection, or only the most relevant features (energy, contrast, variance, correlation, entropy and the inverse difference moment). GLCM pairs are used in all cardinal and ordinal directions with respective ℓ_1-norms of 1 and 2. LBP features utilise the 10 basic patterns [13] in an 8-point circle with radii of 1, 4 and 8 pixels to render a total of 30 features. When shadow masking is performed the GLCM pairs or LBP patterns that fall within a shadow area are ignored during feature calculation. Both fixed and local adaptive thresholding shadow detectors are tested and compared as part of the shadow masking process.

A multi-layer perceptron (learning rate of 0.3, momentum rate of 0.2, 500 training epochs, and unipolar sigmoid activation functions) is used as classifier with texture features as input and with four classes FS, FSB, OIS and NBU as separate output units. The number of input units is equal to the number of texture features (13 for GLCM and 30 for LBP) and the number of units in the single hidden layer is the sum of the number of attributes and classes, divided by two. The perceptron is trained with each one of the datasets (A_{d_1}, A_{d_2}, B_{d_1} or B_{d_2}) in turn, and tested on each of the remaining datasets. In this manner each experiment pairs different datasets and is performed for 10 repetitions, where the perceptron network weights are reinitialised with a different random seed each time.

2) Shadow detection accuracy: Shadow detection accuracy is measured with a Jaccard similarity coefficient (also known as the Jaccard index), which is the ratio of the intersected area to the union of the two shadow areas given by the ground truth and detected shadow masks. A representative set of ground truth shadow masks is used, determined for all of the settlement classes over both dates. The shadow detection accuracy is determined for both the fixed and local adaptive thresholding detectors over a range of threshold values.

V. RESULTS AND DISCUSSION

Shadow variances are amongst the more acute of across-date differences, and this is indicated in the comparison shown in Figure 2. Sample areas of the three main settlement types are shown for each of the study dates and the shadows detected with local adaptive thresholding are shown in color. For a particular settlement type the same area is shown for both dates so that the shadow differences may be directly compared. Since d_1 was acquired during the summer season, and d_2 during winter, the shadows of d_2 are longer due to the Sun being in a more northerly position.

1) Settlement classification accuracy: When across-date shadow differences are removed during feature calculation a notable improvement in settlement classification accuracy is observed, as shown in Figure 3. The LBP features have also shown benefit from shadow masking on same-date training and testing dataset pairs, but the across-date accuracy improvements are more significant due to greater feature variance being reduced. The same-date accuracy increases with LBP were only possible when the higher detection accuracy of local adaptive thresholding was used and the across-date improvements were also greater than with fixed thresholding.

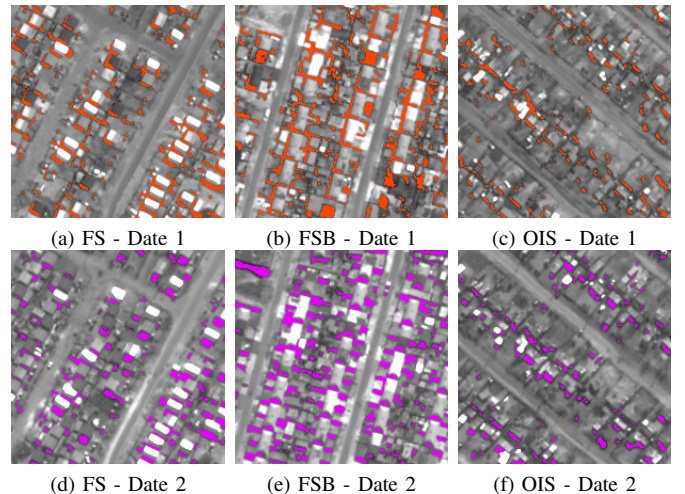| (a) FS - Date 1 | (b) FSB - Date 1 | (c) OIS - Date 1 |
| (d) FS - Date 2 | (e) FSB - Date 2 | (f) OIS - Date 2 |

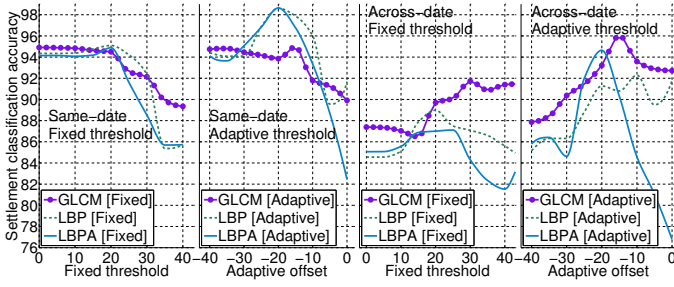Fig. 2. Examples of the across-date shadow differences for three of the settlement classes found in Soweto.

Fig. 3. Settlement classification percentage accuracy with GLCM, LBP and LBPA with fixed and adaptive shadow detection thresholds (25-pixel window).

TABLE I
SHADOW DETECTION AND SETTLEMENT CLASSIFICATION ACCURACIES

Feature	Mask type	Max. accuracy		Jaccard index	
		Same-date	Across-date	Same-date	Across-date
GLCM	No mask	94.91	87.39	0	0
	Fixed	94.91	91.73	0	0.484
	Adaptive	94.97	96.04	0.401	0.401
LBP	No mask	94.35	84.55	0	0
	Fixed	95.10	88.98	0.446	0.446
	Adaptive	98.61	94.37	0.388	0.342
LBPA	No mask	94.15	85.07	0	0
	Fixed	94.88	87.11	0.446	0.530
	Adaptive	98.64	94.64	0.388	0.388

LBP/LBPA features with local adaptive threshold shadow masking posted statistically significant accuracy increases in all of the 4 same-date experiments. Posterization in the shadow areas potentially increments randomized patterns in the LBP histogram which decreases the valid to randomized pattern ratio. Fixed threshold detection may also remove non-shadow areas which may negate the accuracy improvement of shadow masking, while adaptive thresholding can lower the pattern noise floor with more precision. In same-date experiments shadows and the associated posterization affect GLCM similarly without degrading the non-shadow features to the extent seen with LBP and LBPA. However, in across-date experiments the shadow contrast variance negatively affects GLCM features, so shadow masking causes improvement here.

In the across-date experiments with fixed threshold shadow masking the GLCM, LBP and LBPA features achieved 8, 5 and 4 statistically significant accuracy improvements out of the 8 possible experiments, respectively. Here LBP features significantly outperformed LBPA and GLCM features respectively in 4 and 5 of the 8 across-date experiments, despite GLCM features leading to a higher mean accuracy value. The maximum mean settlement classification accuracies and corresponding shadow detection accuracies for the various experiments are given in Table I.

When local adaptive thresholding is used the across-date settlement classification accuracies with GLCM, LBP and LBPA features increase from 87.39% (κ =0.82), 84.55% (κ =0.78) and 85.07% (κ =0.79) to 96.04% (κ =0.94), 94.37% (κ =0.92) and 94.64% (κ =0.92), respectively. Here LBPA features significantly outperform LBP features in 4 of the 8 across-date experiments, but GLCM features achieve a higher settlement classification accuracy than LBPA features with statistical significance in 2 of the 8 experiments.

The settlement classification accuracy for the various local adaptive threshold parameters are shown in Figures 4(a), (b) and (c) for the GLCM, LBP and LBPA features, respectively. The accuracy is relatively stable throughout the range of local window sizes, but it is more sensitive to threshold offset especially for LBPA. Considering the shadow accuracy graph shown in Figure 4(d) it can be asserted that LBPA features are more sensitive to shadow accuracy than GLCM or LBP features. The fact that any pattern point in shadow excludes the placement for LBPA, in contrast to just the center point for LBP, means that the detected shadow mask will have a greater influence on the LBPA calculations than with LBP.

(a) GLCM - Classification accuracy   (b) LBP - Classification accuracy

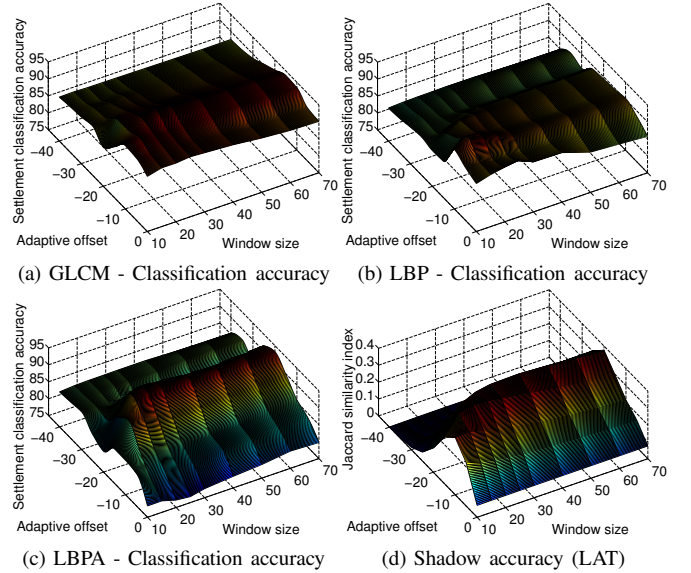(c) LBPA - Classification accuracy   (d) Shadow accuracy (LAT)

Fig. 4. Across-date settlement classification percentage accuracy with local adaptive thresholding shadow detection and a) GLCM, b) LBP and c) LBPA features. d) The shadow detection accuracy as measured from ground truth data for the various adaptive threshold parameters.

2) Shadow detection accuracy: In Table I the Jaccard index measures the shadow accuracy, which is 0 in the case of no mask being applied since the mask area is zero. It is noted that the Jaccard index with fixed thresholding is greater than with local adaptive thresholding, and that the largest index of 0.53 is achieved with fixed thresholding. However, since shadow masking with local adaptive thresholding performs significantly better than fixed thresholding it is clear that the Jaccard index, calculated with the available ground truth shadow masks, is not an accurate predictor of the possible improvements in settlement classification accuracy. This is most likely due to the subjectivity of the human-aided ground truth shadow mask generation, where it is hard to separate shadows from darker surfaces and to delineate the shadow boundaries of partially translucent objects like trees.

3) Statistical analysis: The scatter graphs in Figure 5 display the relationship between settlement classification and shadow detection accuracy for the same-date and across-date experiments, with added linear regressions. With the exception of same-date experiments with GLCM features, there is a positive correlation for all other experiments where the settlement
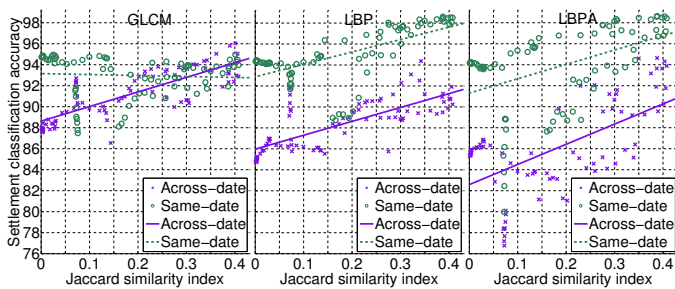
Fig. 5. The relationship between classification accuracy and shadow detection accuracy (local adaptive thresholding) based on the Jaccard index.

TABLE II
CORRELATION COEFFICIENTS AND STATISTICAL SIGNIFICANCE

Feature	Measure	Same-date	Across-date	Δ p-val
GLCM	ρ	-0.062	0.806	**0**
	p-val	0.581	0	
LBP	ρ	0.629	0.759	**0.049**
	p-val	0	0	
LBPA	ρ	0.467	0.602	**0.003**
	p-val	0	0	

classification accuracy improves as the Jaccard index becomes greater. So as the shadow masks become more accurate the shadow masking reduces settlement feature variations to a greater degree.

Pearson's correlation coefficients (denoted by ρ) given in Table II show that the across-date experiments have stronger correlations than same-date experiments, due to stronger across-date shadow profile differences. The correlation for same-date experiments with GLCM features is not reliable as indicated by its p-value of 0.581, which gives the statistical significance of the correlation as calculated using Student's t-distribution. For a null-hypothesis of same and across-date correlations being equal, it can be seen that there are relatively small probabilities Δ p-val of obtaining correlation differences at least as extreme as were observed, given that the null-hypothesis is true. The case of correlated correlation coefficients [14] is used to determine Δ p-val, where the Jaccard index is the shared variable between same and across-date samples. At a significance level of 0.05 all of the features show a definitely greater across-date correlation, which supports the hypothesis that it is the shadow masking specifically that can improve settlement classification accuracy.

VI. CONCLUSION

Multi-date satellite imagery embeds acquisition-related variations, which are difficult to eliminate in panchromatic scenes without augmenting information. A means of removing shadow differences was explored and local adaptive thresholding was employed for shadow detection and was shown to be superior to fixed thresholding. Shadow masking was integrated into the feature determination layer for effective shadow invariance by performing masking during GLCM and LBP feature calculation. GLCM features outperformed LBP features during across-date classification, but LBP features were better for same-date classification with shadow masking using adaptive thresholding.

Minimum-supervision multitemporal classification is thus possible, where a classifier may be trained on a single date and tested on other dates. Experimental results showed significant increases in settlement accuracy, where same-date tests improved from 94.15% to 98.64% and across-date experiments from 87.39% to 96.04%. A statistical correlation study indicated that across-date experiments benefited more from shadow detection accuracy, and support was obtained for the theory that it is the removal of shadows specifically that improved settlement classification accuracy. Future research may consider the use of multiband information and an extension to the study with an expanded multitemporal image set.

REFERENCES

[1] J. A. Voogt and T. R. Oke, "Effects of urban surface geometry on remotely-sensed surface temperature," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 895–920, 1998.
[2] X. Wu, S. Collings, and P. Caccetta, "BRDF and illumination calibration for very high resolution imaging sensors," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS 2010)*, 2010, pp. 3162–3165.
[3] L. Lorenzi, F. Melgani, and G. Mercier, "A complete processing chain for shadow detection and reconstruction in VHR images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 9, pp. 3440–3452, Sept. 2012.
[4] M. Teke, E. Başeski, A. Ö. Ok, B. Yüksel, and Ç. Şenaras, "Multispectral false color shadow detection," in *Photogrammetric Image Analysis*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6952, pp. 109–119.
[5] V. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 6, pp. 1661–1671, June 2006.
[6] K.-L. Chung, Y.-R. Lin, and Y.-H. Huang, "Efficient shadow detection of color aerial images based on successive thresholding scheme," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 2, pp. 671–682, Feb. 2009.
[7] F. P. S. Luus, F. van den Bergh, and B. T. J. Maharaj, "The effects of shadow removal on across-date settlement type classification of QuickBird images," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS 2012)*, July 2012.
[8] Y. Wei, Z. Zhao, and J. Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS 2004)*, vol. 3, 2004, pp. 2008–2010.
[9] P. M. Dare, "Shadow analysis in high-resolution satellite imagery of urban areas," *Photogrammetric engineering and remote sensing*, vol. 71, no. 2, pp. 169–177, 2005.
[10] W. Liu and F. Yamazaki, "Object-based shadow extraction and correction of high-resolution optical satellite images," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. PP, no. 99, pp. 1–7, 2012.
[11] F. van den Bergh, "The effects of viewing- and illumination geometry on settlement type classification of QuickBird images," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS 2011)*, July 2011, pp. 1425–1428.
[12] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 3, no. 6, pp. 610–621, Nov. 1973.
[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, July 2002.
[14] X. Meng, R. Rosenthal, and D. Rubin, "Comparing correlated correlation coefficients," *Psychological Bulletin*, vol. 111, no. 1, pp. 172–175, January 1994.