

SHADOW DETECTION FROM VHR AERIAL IMAGES IN URBAN AREA BY USING 3D CITY MODELS AND A DECISION FUSION APPROACH

K. Zhou^{a,*}, B. Gorte^a

^a Dept. of Geoscience and Remote Sensing, Delft University of Technology, the Netherlands - (k.zhou-1, b.g.h.gorte)@tudelft.nl

Commission VI, WG IV/3

KEY WORDS: VHR aerial images, 3D city model, free training samples, mislabel, QDA, decision fusion, fuzzy membership, entropy

ABSTRACT:

In VHR (very high resolution) aerial images, shadows indicating height information are valuable for validating or detecting changes on an existing 3D city model. In the paper, we propose a novel and full automatic approach for shadow detection from VHR images. Instead of automatic thresholding, the supervised machine learning approach is expected with better performance on shadow detection, but it requires to obtain training samples manually. The shadow image reconstructed from an existing 3D city model can provide free training samples with large variety. However, as the 3D model is often not accuracy, incomplete and outdated, a small portion of training samples are mislabeled. The erosion morphology is provided to remove boundary pixels which have high mislabeling possibility from the reconstructed image. Moreover, the quadratic discriminant analysis (QDA) which is resistant to the mislabeling is chosen. Further, two feature domains, RGB and ratio of the hue over the intensity, are analyzed to have complementary effects on better detecting different objects. Finally, a decision fusion approach is proposed to combine the results wisely from preliminary classifications from two feature domains. The fuzzy membership is a confidence measurement and determines the way of making decision, in the meanwhile the memberships are weighted by an entropy measurements to indicate their certainties. The experimental results on two cities in the Netherlands demonstrate that the proposed approach outperforms the two separate classifiers and two stacked-vector fusion approaches.

1. INTRODUCTION

Very high resolution aerial images reveal very rich and valuable details on the earth surface. The very detailed color information allows distinguishing many essential objects, such as individual buildings, trees, streets, etc. The point clouds can be also acquired from image matching (Hirschmuller, 2008) with high accuracy by using aerial images (Stal et al., 2013). With these advantages and high updating frequency, aerial images become very good resources for change detection, update or validate of the 3D models. Because the quality of point clouds from image matching still suffers from incompleteness and noises, the performance of detection changes on 3D models by using these point clouds can be largely affected. The shadows in the images which indicate valuable height information has been proposed to be very helpful to find height changes on 3D models (Rathje et al., 2005). On the other hand, radiometric distortion in the shadow area would deteriorate performance of classification and give out false alarms on change detection (Adeline et al., 2013). Therefore, shadows in VHR images especially in urban environment have to be detected.

The existing literature presents three main categories of shadow detection (Adeline et al., 2013, Lorenzi et al., 2012): property-based, supervised learning based and model-based approaches. Property-based methods do not need any prior knowledge and often combine with automatic thresholding. They focus on exploring property of spectral information to separate shadow from others. The shadow regions hold some properties: shadows have low luminance because of occlusion of direct light (Tsai, 2006); higher hue value due to the radiance received from shadows decreases from short (blue) to long wavelength (red) due to Rayleigh scattering (Tsai, 2006, Adeline et al., 2013). These properties

can be better explained by using invariant color model consisting of hue, saturation and intensity. Tsai (2006) provides a comparative study on five different invariant color models: *HIS*, *HSV*, *HCV*, *YIQ* and *YCaCr*. The shadows are expected to have high value of ratio of hue over intensity. A thresholding method (Otsu, 1975) is used to separate shadows from non-shadows. Based on this finding, a new ratio is designed (Chung et al., 2009) in order to stretch the gap between shadow and black objects with a successive thresholding. However, the automatic thresholding is often case-dependent. With training samples, supervised learning are expected with better performance on shadow detection on images. The SVM with a polynomial kernel of degree 3 performs well for regular images on RGB bands by using training examples selected by users (Arbel and Hel-Or, 2011). Apart from RGB features, the texture features presented by four space-frequency features for each band from symlet wavelet are used by an SVM for shadow detection in remote sensing images (Lorenzi et al., 2012). However, the supervised approach requires lots of manual work to generate good training examples with large variety. A closed-form solution is provided to reduce the large amount of user inputs to identify shadows from regular images by image matting (Levin et al., 2008). Still, a large amount of manual work is required. Model-based methods take a 3D model to reconstruct shadow for the image with sun position and azimuth angle by ray tracing (Tolt et al., 2011) or z-buffer (Gorte and van der Sande, 2014). However, the 3D information which is often not very accurate and matching with images results in the poor shadow detection in the image. An supervised learning approach (Tolt et al., 2011), a support vector machine (SVM), is applied to improve shadow detection by using reconstructed image to provide free training examples. In order to reduce mislabeled samples, the interior of large shadows or non-shadows in the reconstructed images are chosen for training. However, if the 3D model is outdated or not complete, the big

*Corresponding author



Figure 1. The 3D city models of Amersfoort with 200,000 triangles is created by using a hydrological map and AHN dataset around 2009. The corresponding area of the aerial image in 2010 is shown in the red polygon areas. The different colors in the 3D model show different classes of objects.

shadow or non-shadow area are still mislabeled and can provide a large amount of mislabeling samples.

The performance of shadow detection is obviously strongly related to the features chosen to be separated from non-shadows. Shadow detection based on RGB features is very suitable to find shadows with low radiometric reflection, but dark roofs can be misclassified. On the other hand, shadow detection based on ratio of hue over intensity performs better on dark roofs which receive more radiations from long wavelength, but bluish solar panels with high hue values can be more likely classified as shadows and red objects under the shadows with low hues may be classified as non-shadows. Thus, reliability assessment of different features should be considered in the classification. Stacked-vector approach is a straight-forward way to combine different features for classification. However, a multivariable statistic model, e.g. multivariate Gaussian model, does not have the mechanism to include weight of features according to their reliability (Benediktsson et al., 1990). Many more nonparametric approaches, such as neural networks, decision trees and SVM takes weights of different features into account (Waske and Benediktsson, 2007). Instead of one-step stacked approach, decision fusion proposes to fuse the information deduced from preliminary classification on several individual feature domains (Benediktsson et al., 1990, Fauvel et al., 2006). The approach allows to choose different classifiers which might be more suitable to different feature domains, e.g. an accurate multivariate statistical model may provide better classification than these non-parametric approaches. The reliability can be derived globally for each classifier (Benediktsson et al., 1990) or locally for pixels (Fauvel et al., 2006) from preliminary classifications.

In this paper, we propose a novel and fully automatic approach for shadows detection from the image by using 3D city models. The reconstructed image from 3D city model by ray tracing is treated as a pool of training samples with large variety for a supervised learning. However, the mislabeling effects introduce serious mixtures between shadows and non-shadows. Erosion is applied to remove the unreliable labels on the boundaries, while QDA is tested with better capability than SVM of solving the mixture problem. In order to further addressing the problem, the two widely used feature domains, RGB and ratio, are analyzed and the complementary effects are found. The decision fusion approach of using fuzzy membership function and pixel-wise en-

tropy measurement is proposed to solve the conflicting situations and make most of the complementary effects. This paper is organized as follows. Section 2 describes study area and data. Section 3 illustrates the methodology proposed for shadow detection, followed by experimental results and comparisons in section 4 and conclusions in section 5.

2. STUDY AREA AND DATA

The two study areas are urban areas located in Amersfoort and Assen/Delft, Netherlands. The reasons of choosing these two areas are: 1) The residential buildings and trees contribute to serious shadows in both areas. 2) Many dark roofs and reddish roads in Amersfoort increase difficulties of shadow detection from classification on either individual feature domain described in Section 3.3, while more bluish solar panels existed in Assen/Delft would also increase difficulties in classification in ratio domain. Each aerial images is acquired by UltraCam on 23 April, 2010 or 14 April, 2015 with RGB bands of 3.5cm resolution covering around 0.2km^2 for Amersfoort and Assen/Delft with a size of 11310×17310 pixels or 7500×11500 pixels respectively. The automatic 3D city modeling is intensively investigated in the last decades and summarized in (Haala and Kada, 2010). Oude Elberink (Elberink et al., 2013) creates a national-wide general 3D city model by using topographic maps and point clouds, e.g. Top10NL and AHN2 respectively. Both Top10NL and AHN2 are open sources in the Netherlands. Top10NL is an object oriented topographic dataset at scale 1:10,000, while AHN2, a point cloud, provides height information for these objects. The other topographic and height data can be also used as input. For Amersfoort, a hydrological map and AHN2 point clouds around 2009 are provided to construct the 3D model, while for Assen/Delft, a BGT, large scale topographic map, and AHN2 point clouds from 2010, are used. Specific rules are defined to assign the point clouds to polygons for reducing large amounts of points for model reconstruction. The simplified model include five general classes of objects, but not restrictedly: water, road, terrain, buildings and forest. Normally, forest only includes low vegetations without the trees as they are very difficult to model and can be various in different seasons. More details of model reconstruction can be found in (Elberink et al., 2013). A LoD2 3D model with buildings, terrain and road classes and the corresponding aerial image for partial city area of Amersfoort are shown in

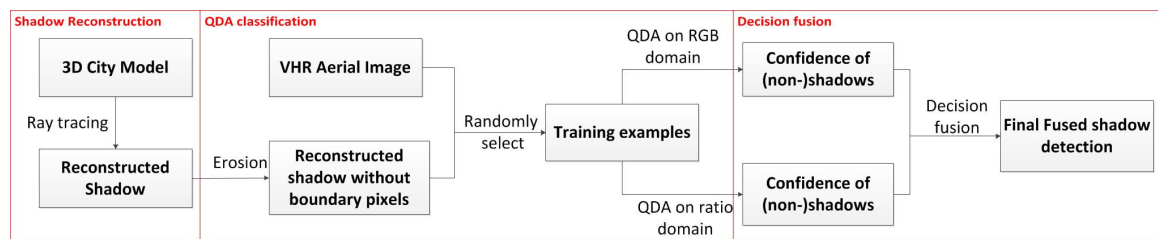


Figure 2. The flow chart of methodology.

Figure 1.

3. METHODOLOGY

The proposed methodology as shown in Figure 2 for shadow detection consist of (1) shadow reconstruction (2) QDA shadow classification (3) decision fusion on two feature domains. In the following sections, the details will be discussed.

3.1. Shadow Reconstruction

Shadow reconstruction for aerial images needs the prior knowledge of sun position and camera parameters. The sun position can be calculated from the time the image was taken, while internal and external orientation of camera can be precisely estimated by bundle adjustment including several ground control points. If these prior information are known, the shadow can be reconstructed from ray tracing. It is very straight forward and can give very accurate results. At first, a ray from each pixel in the image plane is generated from camera to 3D model. Then first intersection between the ray and scene is found, which indicates the point can be by the camera through the pixel. Another shadow ray from the intersection point to light source is generated and if the ray is intersected with scene, the pixel is in shadow. A Kd tree which is a space-partitioning data structure for organizing triangles in a k -dimensional space is designed in order to fast the intersection between ray and 3D model to a reasonable time. Constructing a efficient Kd tree for triangular meshes is also time consuming, so a $O(N \log^2 N)$ approach is adopted and more details can be found in (Wald and Havran, 2006).

However, there are several reasons that the reconstructed shadows may not exactly match with shadows in the image.

- 1) The time of 3D model may not match with the image.
- 2) The trees which can cause significant shadows are often missed or not well modeled in the 3D models.
- 3) The accuracy of the 3D modeling are often affected in the areas with low point density and missing data due to occlusions. The shadows around boundary region are often not reliable due to many spurious triangles.

Although shadow reconstruction is not enough for shadow detection in the images, these shadows are the prior knowledges of where the shadow can be in the image as the objects in the city would not change dramatically. Therefore, large portions of these shadows are labeled correctly and show very large variety. By using them as training examples, supervised classification methods are expected to reveal the characteristics of shadows and improve the shadow detection. However, the mixture problem aggravated by mislabeling should be properly addressed.

3.2. QDA Shadow Classification

In order to reduce the mislabeling effects, a erosion morphological filter is applied in both shadow and non-shadow areas in order to remove the unreliable boundary pixels mentioned in section 3.1. As most of reconstructed shadows are caused by buildings, a disk-shaped structure, which is often used for filtering while preserving building structure, is used to remove the artifacts from spurious triangles in 3D models. However, the mislabeled areas still exist due to the trees and different time-line. The mixture between shadows and nonshadows in the training examples become a serious problem. While SVM is used in many researches (Guo et al., 2011, Tolt et al., 2011, Lorenzi et al., 2012), SVM aims to find an optimal boundary hyperplane between classes and would put lots of effort on the mislabeled samples, resulting in the over fitting problem. However, if the multivariate models of two classes can be estimated properly from relatively large portions of correctly labeled samples, a better classification can be obtained. By randomly choosing training examples from reconstructed shadow map and displaying them in RGB or ratio of hue over intensity domain, both classes show Gaussian-like distribution as shown in Figure 3 even serious mixture of two classes exists. With large portion of correct labeled examples, the estimation of distribution of two classes can be resistant to the mislabeling and mixture problem. A commonly used QDA (Theodoridis and Koutroumbas, 1999) which assumes a Gaussian distribution for each classes with different covariances is chosen for classification.

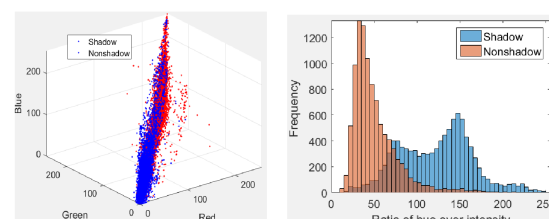


Figure 3. 10,000 training examples are randomly selected for each class and displayed in RGB and Ratio domains in left and right image respectively.

Given a d -dimensional feature space R^d , a feature vector \mathbf{x} is a training examples with a corresponding class $\omega_i, i = 1, 2, \dots, M$. In binary classification problem, $M = 2$. The posterior probabilities are form as $P(\omega_i|\mathbf{x}), i = 1, 2, \dots, M$ and can be decomposed according to Bayes' theorem:

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})} \quad (1)$$

where $p(\mathbf{x}|\omega_i)$ is conditional probability density function(pdf)

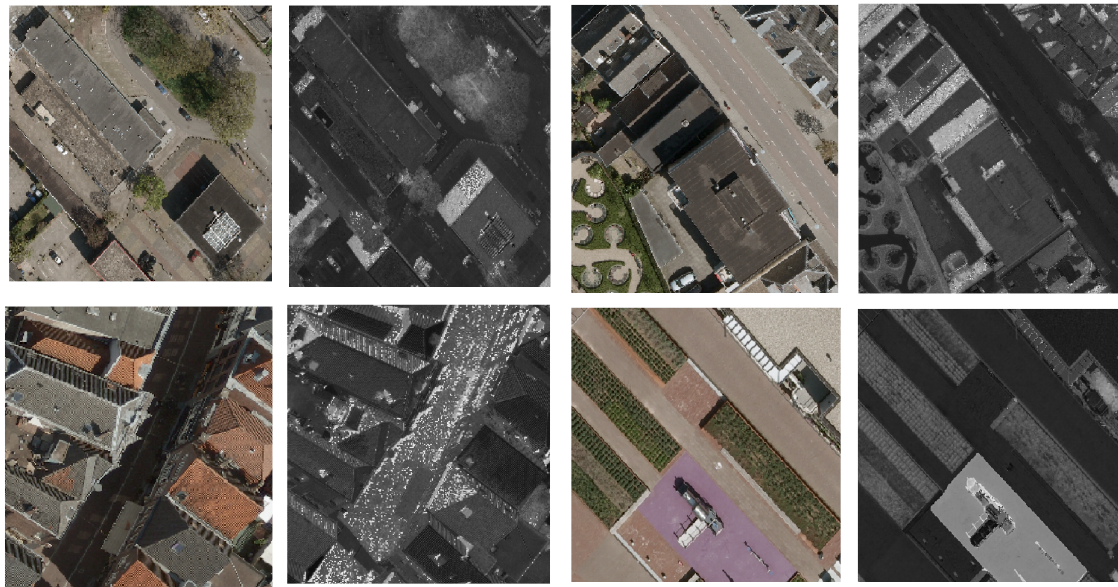


Figure 4. Four cases are displayed in RGB and ratio domain. In ratio image, shadows are more white. The images in upper row shows that roofs are easily detected in ratio domain, while the images in lower row shows that shadow in city canyon and purplish objects can be correctly detected in RGB domain.

$p(\omega_i)$ is prior probability of a class ω_i
 $p(\mathbf{x})$ is the joint distribution $p(\mathbf{x}) = \sum_{i=1}^M p(\mathbf{x}|\omega_i)p(\omega_i)$

The denominator is not dependent on ω_i and can be treated as a constant. If pdfs $p(\mathbf{x}|\omega_i)$ are Gaussian multivariate distributions, $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, the logarithmic function $g_i(\mathbf{x})$ of the posterior probability becomes:

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln p(\omega_i) \quad (2)$$

or

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln p(\omega_i) - \frac{l}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| \quad (3)$$

The classification rule:

$$\hat{\omega}_k = \arg \max_k g_k(\mathbf{x}) \quad (4)$$

The decision boundaries are quadratic equations in \mathbf{x} . QDA with much more flexible covariance assumption can fit the data better than linear discriminant analysis (LDA). The unknown parameters in multivariate Gaussian distributions are estimated by using maximum likelihood.

3.3. Decision Fusion on Two feature domains

Even the QDA can be resistant to mislabeling, accurate estimation of Gaussian distributions is affected apparently. The pixels far away from the center of Gaussians or in the mixture region is less reliable. However, if the pixels are not reliable in one feature domain but are better located in another domain, the mixture problem in either feature domains can be further mitigated.

The property of shadow is studied intensively (Tsai, 2006, Chung et al., 2009, Adeline et al., 2013) and can be summarized into two main properties:

- 1) Shadows have low radiometric reflection.
- 2) Radiation received from shadow area decrease from short (blue-violet) to long (red) wavelength because of Rayleigh scattering.

According to these properties, two feature domains, RGB and the ratio of hue over intensity from invariant color models, are widely used. According to property (1), shadows have the low RGB values or intensity. Moreover, according to the properties (2), shadows are expected with higher hue values than nonshadow for the same object. Therefore, the ratio of hue over intensity is more effective to separate the two classes than only using intensity. Shadows are expected to have large ratios. The *HSI* color space is tested with best performance in Tsai's work and the ratio is defined as $\frac{H+1}{I+1}$ and the scaled to $[0, 255]$, where range of H and I are both in $[0, 1]$ (Tsai, 2006).

The two feature domains are chosen as they have complementary characteristics explained by using several cases in Figure 4. Different objects can be better detected correctly in different feature domains. Dark roofs are much easier to be detected correctly in ratio domain. In the upper row of the figure, dark buildings can be easily misclassified as shadow in RGB domain. On the other hand, as most of roofs under sunlight are expected to receive more radiation from long wavelength, the value ratios of dark roofs becomes lower and can be more far away from shadows. So the confidence of it belongs to non-shadow in ratio domain becomes high. As the material of objects also plays a very important role on pixel value in the image, reddish objects under the shadow can have low hue value and bluish objects in the sun can have high hue value. Even the adverse effects from hue can be reduced by low intensity, still these objects can be easier detected from RGB domain with intensity property alone. In the lower left of the figure, some parts of street with red material in the city canyon receiving less Rayleigh effects and reflecting more red color can be misclassified as non-shadows in ratio domain. Parts

of shadows on the street show gray colors with relatively low ratio values, however, they can be easily identified as non-shadows with low RGB values. In the lower right of figure, the bluish or purplish object under the sun with high hue values may have high ratio values mixed with shadows. However, in RGB domain, it is easier to distinguish it from shadows.

It is obvious that many conflicting situations are introduced by the two classifications and a decision fusion approach is to make a wise choice on these situations. Instead of crispy classification, a fuzzy membership presenting a partial membership to a class (Mather and Tso, 2009) can be used to define the confidence of the pixel belonging to a class. The higher this value, the more likely the pixel belongs to the class (Fauvel et al., 2006). Therefore, the fuzzy membership explains whether the objects can be better detected as a certain class in a classifier. As the QDA assumes the two classes are Gaussian distributed, the Gaussian curve membership function (Kaufmann and Swanson, 1975) is chosen.

$$\mu(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (5)$$

where $\mu(\mathbf{x})$ is a fuzzy membership degree

When a conflict situation happens, a max operator on fuzzy membership degrees are better applied to solve the problem when the reliability of classifications are different. The fused result for each class from classifiers is obtained and then final classification can be determined.

$$\mu_i^f = \max(\mu_i^j(\mathbf{x}), \quad j = 1, 2, \dots, N); \quad (6)$$

where μ_f is the fused membership degree for each class from several classifiers
 $\mu_i^j(\mathbf{x})$ is the membership degree of a pixel \mathbf{x} to class i given classifier j . In 2-class problem with 2 classifiers, $i \in \{1, 2\}$, $j \in \{1, 2\}$

However, it is unreliable to fully rely on the membership degree, when the classes have serious mixing problem. For a 2-class problem with 2 classifications, the membership degree of a pixel can be very high to both classifiers in the mixing area between two classes. The large uncertainty of classification is especially in the mixing area. A α -quadratic entropy is introduced to measure the pixel-wise reliability to each classifier (Fauvel et al., 2006). However, it does not include measurement of mixing extent of the pixel between two classes. A new entropy is provided introduced to measure the mixture extent of pixel in each classifiers:

$$H^j = -\sum_i \mu_i^j(\mathbf{x}) \log_b(\mu_i^j(\mathbf{x})) \quad (7)$$

where H^j is the entropy of a pixel to class j
 $\mu_i^j(\mathbf{x})$ is normalized membership degree among different classes.
 b is the base of the logarithm $b = 2$ for 2-class problem

For 2-class case, the equation becomes:

$$H^j = -\mu^j(\mathbf{x}) \log_2(\mu^j(\mathbf{x})) - (1 - \mu^j(\mathbf{x})) \log_2(1 - \mu^j(\mathbf{x})) \quad (8)$$

If $\mu(\mathbf{x}) = 0.5$, the membership degree of two classes are $[0.5, 0.5]$. The entropy is 1 as shown in Figure 5. Therefore, the classifier is totally uncertain about the pixel. The corresponding membership functions of each classifier are weighted by entropies and then the same max operator is applied:

$$w_j = \frac{\sum_{k=0, k \neq j}^N H^k}{(N-1) \sum_{k=0}^N H^k} \quad (9)$$

$$\mu_i^f = \max(w_j \mu_i^j(\mathbf{x}), \quad j = 1, 2, \dots, N); \quad (10)$$

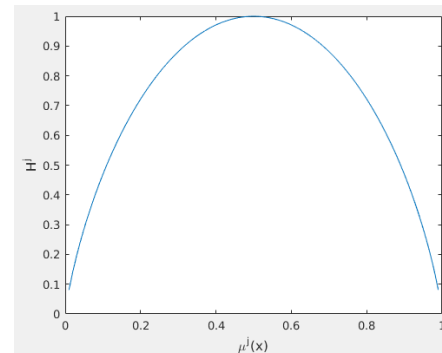


Figure 5. Entropy curve of 2-class case.

4. EXPERIMENT RESULTS AND COMPARISON

The experiments in this paper were applied on two city areas in Amersfoort and Assen/Delft, Netherlands. The datasets are described in Section 2. The experiments are fully automatic without any manual work of selecting training examples. The proposed approach was compared with many other approaches in order to show the effectiveness of our approach.

4.1. Experiment on Amersfoort

The 3D model consists of 200,000 triangles and the size of reconstructed shadow image is 11310×17310 with pixel size of $6 \mu\text{m} \times 6 \mu\text{m}$. The altitude and azimuth of sun position is 44.58° and 140.410° . The camera position is $(154698.489 \text{ m}, 462637.358 \text{ m}, 623.512 \text{ m})$. The external orientation of the camera is $[-0.2801^\circ, 0.0032^\circ, -89.2683^\circ]$, while the internal parameters, focal length and position of principle points, are $(100.5 \text{ mm}, 0 \text{ mm}, 0 \text{ mm})$. The KD tree and parallel computing was adopted to fast the ray tracing. With a HP computer with a 8GB ram and quadcore, the KD tree construction need 15s, while the time for reconstruction was 40 min. The reconstructed image is a black and white image with value in $\{0, 1\}$. The black pixels are shadows while the white pixels are non-shadows.

A erosion filtering with a disk-shaped structure with 10 pixels was applied both to black and white pixels. For each class, 10,000 training examples are randomly selected for QDA analysis on RGB and ratio feature domain. Then fuzzy membership functions were derived from the two QDAs. The decision fusion approach was applied to obtain the final shadows. Figure 6 shows that the decision fusion approach performs well on the 4 cases shown in Figure 4. In the first two cases in the first two rows, the black roofs are not well detected in RGB domain, but well detected in ratio domain. The decision fusion relies more on the decision

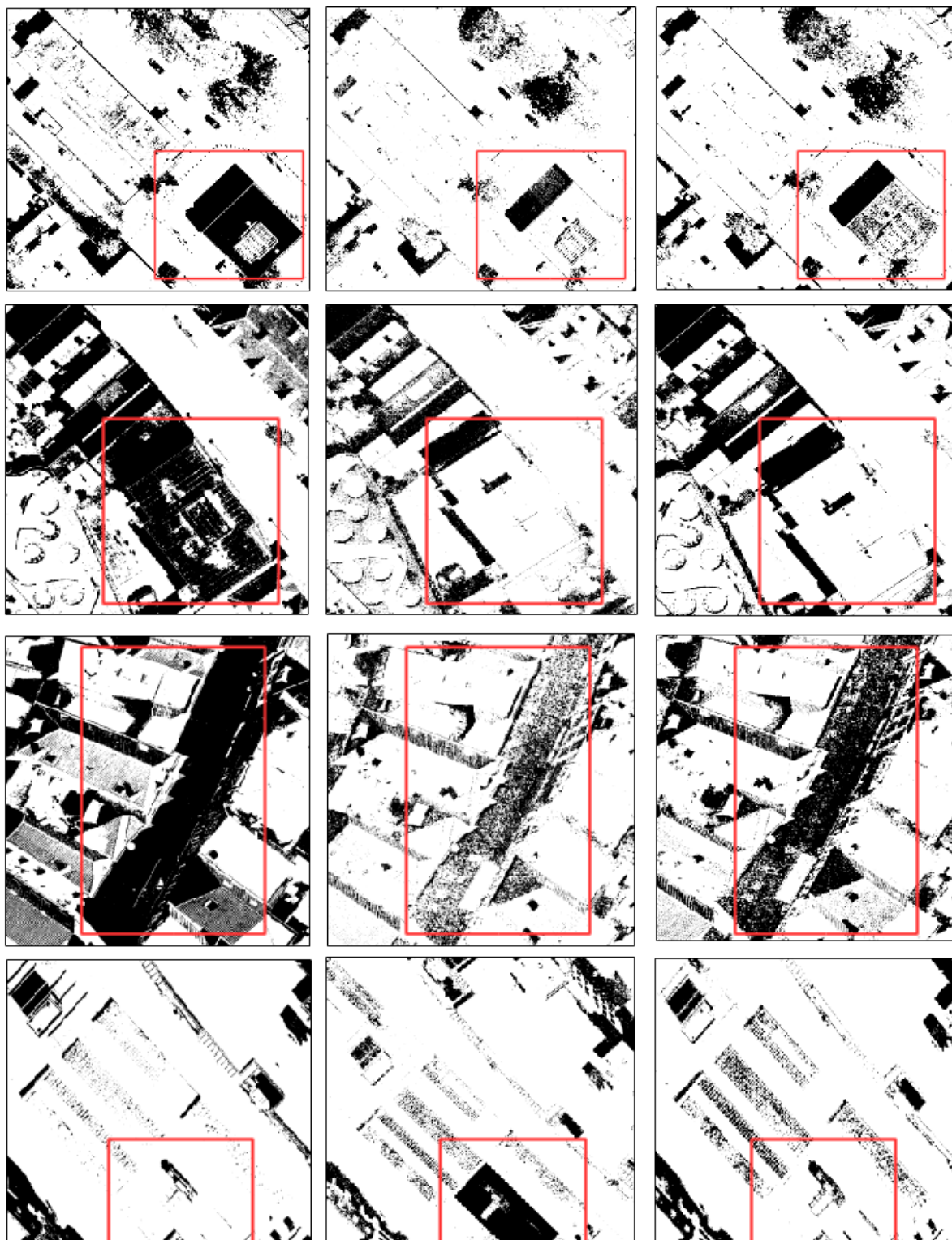


Figure 6. The shadow detection images for the four cases listed in Figure 4 from Amersfoort with different approaches. The left, middle and right column of images are obtained from QDA on RGB domain, QDA on ratio domain and decision fusion respectively.

	TPR %	FPR %	FNR %	Correctness %	KC
QDA+RGB	97.13	18.67	2.87	81.33	0.7668
QDA+Ratio	79.15	2.94	20.85	97.06	0.7795
QDA+(RGB+Ratio)	96.45	14.72	3.55	85.28	0.8122
SVM+(RGB+Ratio)	73.24	7.91	26.76	92.09	0.6860
Proposed	90.10	5.50	9.90	94.50	0.8576

Table 1. The shadow detection results for different classifications on Amersfoort dataset.

from ratio domain to derive the final results. But still small portion of roofs are detected as shadow in the first case. The reason can be that some dark roof are in the serious overlap area and they can get high confidence value to the shadow class. Therefore, some roof pixels are still inclined to choose decision from RGB domain. In the third case, the shadows in the city canyon are well detected in RGB domain, but very poorly detected in the ratio domain. The fusion makes a wise choice to detect more shadows, while still a small portion of shadows are misclassified. The reason can be these pixels has reddish color and are brighter due to reflection from windows in the buildings. The resulting low ratio values would have high confidence to be assigned to nonshadow in ratio domain. In the last case, the purple object is detected as non-shadow in RGB domain, but treated as shadow in ratio domain. The decision fusion chooses to believe the result from RGB domain. Overall, it is obvious that our proposed approach is very effective to solve the conflicting situations.

The accuracy of proposed approach was compared with the two separate classifiers and two stacked-vector fusion approaches quantitatively in Table 1 with four measurements: TPR(true positive rate), FPR(false positive rate), FNR(false negative rate), correctness and KC(Kappa coefficient). TPR describes completeness of the detection, FPR describes commission errors, and FNR describes omission errors. QDA with RGB domain has highest completeness of detection with 97.31%, however, many dark roofs are also misclassified as shadows, so commission error is also highest with 18.67%. Therefore, it has lowest the correctness of shadow detection. QDA with ratio domain has an opposite result that it has lowest misclassification on non-shadows with lowest commission error, however, many shadows are not detected with a completeness of 79.15%. Even with a highest correctness of shadow detection, the KC is low. The reason may be that the city canyon effects are quite strong and many objects with strong red materials are under the shadows in this area. QDA with stacked vector improves the classification, however, SVM with stacked vector become worse and has lowest KC. An SVM with RBF kernel may have overfitting problem due to serious mixture between two classes. The proposed approach have both high completeness and correctness on shadow detection, low rate on misclassification. It means the decision fusion performs well in the conflicting situations. With the highest KC, 0.8576, the proposed approach outperforms the other approaches.

4.2. Experiment on AssenDelft

The 3D model consists of 1,000,000 triangles and the size of reconstructed the shadow image is 7500 * 11500 with pixel size of 9 μ m * 9 μ m. The altitude and azimuth of sun position is 42.68° and 216.310°. The camera position is (1111487.246 m, 500278.593 m, 338.912 m). The external orientation of the camera is [−0.275°, 0.223°, 178.472°], while the internal parameters, focal length and position of principle points, are (101.4 mm, −0.18 mm, 0.27 mm). As field of view of the camera only covers partial of the 3D model, the triangles falling into the camera are selected for KD tree construction and ray tracing. With the same

computer, the KD tree construction need 3s, while the time for reconstruction was 30 min.

After applying the proposed approach, the results were also compared in Table 2. The QDA with RGB domain still detect many non-shadows as shadows with second highest commission error 16.32% because of dark roofs, but QDA with Ratio domain works quite well in shadow detection with high completeness and correctness while has low misclassification rate. The reason can be that red road under shadows are much less. Still, a small amount of bluish solar panels are misclassified. The SVM with the stacked vector perform worst comparing with all the other methods. QDA with the stacked vector has a high completeness on detecting shadows, but still the dark roofs are not properly addressed with 13.57% commission errors. The proposed approach also performs very good with the highest KC, 0.9023.

In conclusion, the proposed approach performs better than the other listed methods on the datasets from different cities with complex environments. It means the proposed method is effective for various environments, while the performance of other methods are more dependent on the environment. The QDA with ratio domain performs quite well on AssenDelft dataset, however, has a low rate of shadow detection on Amersfoort dataset. The QDA with stacked vector works quite stable in both environments, however, is still not as good as proposed approach.

5. CONCLUSION

This paper employs an existing 3D city model for shadow detection in the aerial image. The ray tracing approach with accurate camera settings and sun position can reconstruct a very accurate shadow image from the model. As the reconstructed image can present large portion of image correctly, the image is treated to provide free training examples with large variety for supervised classification. But the mislabeling problem aggravates mixture problem between two classes. Firstly, erosion is applied to the reconstructed image to remove the inaccurate pixels on the boundaries. Secondly, QDA is chosen as it is quite effective to the mixture problem by exploring distributions from large portions of right training samples. Still, the estimation of distributions is affected and may affects the pixels far away from the center of distribution and in the mixture regions. Two widely used feature domains are studied and they have complementary characteristics: dark building in the sun are more easily correctly detected in ratio domain, while many red objects under the shadow and bluish objects under the sun can be better detected in RGB domain. The complementary characteristics leads to the conflicting situations from the two preliminary classification. Finally, the decision fusion approach is proposed to wisely make a choice in these situations. Fuzzy membership is chosen to define the confidence of pixel belonging to each class for each classifier. The max operator on fuzzy membership degree can provide the proper choice when facing the conflicting situation. As two classes have serious mixture, the membership degree of the

	TPR %	FPR %	FNR %	Correctness %	KC
QDA+RGB	98.89	16.32	1.11	83.67	0.8275
QDA+Ratio	92.88	6.02	7.12	93.97	0.8866
QDA+(RGB+Ratio)	98.69	13.57	1.32	86.43	0.8572
SVM+(RGB+Ratio)	86.12	19.23	13.88	80.76	0.7048
Proposed	96.85	7.82	3.15	92.18	0.9023

Table 2. The shadow detection result for different classifications on AssenDelft dataset.

pixels falling into these mixture regions are not reliable. An entropy measuring pixel-wise reliability in each classifier is provided. The fuzzy membership degree is weighted in each classifier according to the entropies. Then the max operator is applied to obtain the fused fuzzy membership degree for each class and a final decision can be made to conflicting situations. By comparing with other separate classifiers and two stacked-vector approaches, the proposed approach makes wise choice on conflicting situations and makes better use of complementary characteristics of two feature domains. With two complex city environments, the proposed approach is proved to be more adaptive to different scenarios.

A variety of further works can be anticipated. First, as a 3D city model includes the prior knowledge of the different classes of objects, the class-wise reliability of each classifier can be derived from training samples from each class. The reliability will help to minimize the effects of poor classifiers. Second, the texture features can be considered as another domain for classification or the Markov random field analysis considering relations adjacent pixels can be applied before the classification or after fusion. Finally, the further research on using the output of shadows for change detection on the 3D model can be promising.

REFERENCES

- Adeline, K., Chen, M., Briottet, X., Pang, S. and Paparoditis, N., 2013. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, pp. 21–38.
- Arbel, E. and Hel-Or, H., 2011. Shadow removal using intensity surfaces and texture anchor points. *IEEE transactions on pattern analysis and machine intelligence*, 33(6), pp. 1202–1216.
- Benediktsson, J. A., Swain, P. H. and Ersoy, O. K., 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. In: *Geoscience and Remote Sensing Symposium*, pp. 489–492.
- Chung, K.-L., Lin, Y.-R. and Huang, Y.-H., 2009. Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE Transactions on Geoscience and Remote sensing*, 47(2), pp. 671–682.
- Elberink, S. O., Stoter, J., Ledoux, H. and Commandeur, T., 2013. Generation and dissemination of a national virtual 3d city and landscape model for the netherlands. *Photogrammetric engineering & remote sensing*, 79(2), pp. 147–158.
- Fauvel, M., Chanussot, J. and Benediktsson, J. A., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10), pp. 2828–2838.
- Gorte, B. and van der Sande, C., 2014. Reducing false alarm rates during change detection by modeling relief, shade and shadow of multi-temporal imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(4), pp. 65.
- Guo, R., Dai, Q. and Hoiem, D., 2011. Single-image shadow detection and removal using paired regions. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2033–2040.
- Haala, N. and Kada, M., 2010. An update on automatic 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), pp. 570–580.
- Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2), pp. 328–341.
- Kaufmann, A. and Swanson, D. L., 1975. *Introduction to the theory of fuzzy subsets*. Academic press, New York. Vol. 1 Number 8,
- Levin, A., Lischinski, D. and Weiss, Y., 2008. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp. 228–242.
- Lorenzi, L., Melgani, F. and Mercier, G., 2012. A complete processing chain for shadow detection and reconstruction in vhr images. *IEEE transactions on geoscience and remote sensing*, 50(9), pp. 3440–3452.
- Mather, P. and Tso, B., 2009. *Classification methods for remotely sensed data*. CRC press, Boca Raton.
- Otsu, N., 1975. A threshold selection method from gray-level histograms. *Automatica*, 11(285–296), pp. 23–27.
- Rathje, E. M., Woo, K., Crawford, M., Neuenschwander, A. et al., 2005. Earthquake damage identification using multi-temporal high-resolution optical satellite imagery. In: *International Geoscience and Remote Sensing Symposium*, Vol. 7, pp. 5045.
- Stal, C., Tack, F., De Maeyer, P., De Wulf, A. and Goossens, R., 2013. Airborne photogrammetry and lidar for dsm extraction and 3d change detection over an urban area—a comparative study. *International Journal of Remote Sensing*, 34(4), pp. 1087–1110.
- Theodoridis, S. and Koutroumbas, K., 1999. *Pattern recognition*, Academic press. New York.
- Tolt, G., Shimoni, M. and Ahlberg, J., 2011. A shadow detection method for remote sensing images using vhr hyperspectral and lidar data. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pp. 4423–4426.
- Tsai, V. J., 2006. A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE transactions on geoscience and remote sensing*, 44(6), pp. 1661–1671.
- Wald, I. and Havran, V., 2006. On building fast kd-trees for ray tracing, and on doing that in $O(n \log n)$. In: *2006 IEEE Symposium on Interactive Ray Tracing*, pp. 61–69.
- Waske, B. and Benediktsson, J. A., 2007. Fusion of support vector machines for classification of multisensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12), pp. 3858–3866.