AUTOMATIC BUILDING CHANGE DETECTION THROUGH LINEAR FEATURE FUSION AND DIFFERENCE OF GAUSSIAN CLASSIFICATION

Thesis
Submitted to
The School of Engineering of the
UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for
The Degree of
Master of Science in Computer Engineering

By
Daniel P. Prince

UNIVERSITY OF DAYTON

Dayton, Ohio
December, 2016
AUTOMATIC BUILDING CHANGE DETECTION THROUGH LINEAR FEATURE FUSION AND DIFFERENCE OF GAUSSIAN CLASSIFICATION

Name: Prince, Daniel P.

APPROVED BY:

Vijayan K. Asari, Ph.D.
Advisor Committee Chairman
Professor, Department of Electrical and Computer Engineering

Theus H. Aspiras, Ph.D.
Committee Member
Research Scientist, Department of Electrical and Computer Engineering

Russell C. Hardie, Ph.D.
Committee Member
Professor, Department of Electrical and Computer Engineering

Robert J. Wilkens, Ph.D., P.E.
Associate Dean for Research and Innovation
Professor, School of Engineering

Eddy M. Rojas, Ph.D., M.A., P.E.
Dean, School of Engineering
ABSTRACT

AUTOMATIC BUILDING CHANGE DETECTION THROUGH LINEAR FEATURE FUSION
AND DIFFERENCE OF GAUSSIAN CLASSIFICATION

Name: Prince, Daniel P.
University of Dayton
Advisor: Dr. Vijayan K. Asari

Many applications in infrastructure planning and maintenance are currently aided by the collection of aerial image data and manual examination by human analysts. The increasing availability and quality of this image data presents an opportunity for computer vision and machine learning techniques to aid in infrastructure planning and maintenance. Due to the immense effort required for human analysts to view and organize the data, there is great demand for computer automation of these tasks.

A strategy for detecting changes in known building regions in multitemporal visible and near-infrared imagery based on a linear combination of independent features and a difference of Gaussian based classification approach is proposed. Initial building candidates are discovered using a linear combination of features representing vegetation intensity, image texture, shadow intensity and distance from known road areas. The resulting building candidates are classified by shape using a unique difference of Gaussians technique and a standard Support Vector Machine classifier. Building regions reported in the reference data set from the prior observation time are revisited using the
same classification approach to minimize the number of false positive detections from the feature fusion strategy.

The effectiveness of the proposed technique is evaluated on five wide area real-world images. Ground truths for the building regions in all five images are manually created and used to measure the accuracy of the building detection and change detection results. Detection statistics and visualized results of the proposed algorithm are presented, and it is observed that the results are promising compared to the manually created ground truth. As a possible continuation of this research, a brief discussion on parameter estimation for building change detection based on image characteristics is included.
Dedicated to my parents, who gave me everything I need to be successful.
ACKNOWLEDGMENTS

I would like to thank Dr. Asari, Sidike Paheding, Almabrok Essa, and the rest of the Vision Lab for their constant support in my research and education. Their encouragement and feedback had a profound impact on my thesis work and professional development as an engineer.

This work is supported by Pacific Gas and Electric (PG&E). I would like to thank PG&E for its support and for providing the image datasets used to develop and validate the algorithms presented in this paper. Especially, I would like to thank Francois Rongere for his invaluable suggestions during the development of this technology.
# TABLE OF CONTENTS

ABSTRACT ................................................................. iii

DEDICATION .............................................................. v

ACKNOWLEDGMENTS .................................................. vi

LIST OF FIGURES ....................................................... viii

LIST OF TABLES ......................................................... xi

I. INTRODUCTION ..................................................... 1

II. LITERATURE SURVEY ............................................... 3

III. BUILDING REGION CANDIDATE CREATION AND CLASSIFICATION .......... 8

   3.1 Feature fusion for building candidate creation ................................. 8
   3.2 Building candidate classification .................................................. 12
   3.3 Reference region reidentification ................................................ 13
   3.4 Building change detection ........................................................... 15

IV. PERFORMANCE EVALUATION ...................................... 19

   4.1 Dataset description ............................................................... 19
   4.2 Parameter selection ................................................................ 20
   4.3 Evaluation methods ................................................................. 20
   4.4 Building detection and change detection results ............................ 25
   4.5 Discussion of results ................................................................ 26

V. CONCLUSION ............................................................ 39

BIBLIOGRAPHY .............................................................. 41
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Block diagram of the overall automatic building detection algorithm.</td>
<td>9</td>
</tr>
<tr>
<td>3.2</td>
<td>An example road intensity image compared to the original RGB image.</td>
<td>11</td>
</tr>
<tr>
<td>3.3</td>
<td>Block diagram for the separation of background regions from true building regions.</td>
<td>12</td>
</tr>
<tr>
<td>3.4</td>
<td>RGB image showing a representative area of buildings.</td>
<td>14</td>
</tr>
<tr>
<td>3.5</td>
<td>Visualized DoG intensities for multiple sigma scales with the manually identified building ground truth regions outlined in white. Note that building regions of different sizes and shapes are characterized by regions of low intensity for the varying sigma values.</td>
<td>14</td>
</tr>
<tr>
<td>3.6</td>
<td>Block diagram for DoG feature extraction and classification of a building candidate region.</td>
<td>15</td>
</tr>
<tr>
<td>3.7</td>
<td>Block diagram for the reidentification of building regions from the reference set.</td>
<td>16</td>
</tr>
<tr>
<td>3.8</td>
<td>Block diagrams for producing final building change detection using Boolean operations on binary images.</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Training and testing images, displayed in increasing order of number of building regions.</td>
<td>21</td>
</tr>
<tr>
<td>4.2</td>
<td>Block diagram for calculating the True Positive (TP) removed region detections.</td>
<td>23</td>
</tr>
<tr>
<td>4.3</td>
<td>Block diagram for calculating the False Positive (FP) removed region detections.</td>
<td>23</td>
</tr>
<tr>
<td>4.4</td>
<td>Block diagram for calculating the False Negative (FN) removed region detections.</td>
<td>24</td>
</tr>
</tbody>
</table>
4.5 The resulting fusion image for Image 3. Pixels with lower intensity, marked in blue, are more likely to represent building regions than pixels with higher intensity, marked in yellow. 27

4.6 Building candidate regions created through the thresholding of the fusion intensity of Image 3 shown in Figure 4.5 as compared to the manually created ground truth. Blue indicates ground truth pixels that have not been detected by the algorithm, green indicates ground truth pixels that have been detected by the algorithm, and red indicates false positive pixels detected by the algorithm that do not indicate true buildings. 28

4.7 Automatically detected building regions versus the ground truth for Image 3. Blue indicates ground truth areas that have not been detected by the algorithm, green indicates ground truth areas that have been detected by the algorithm, and red indicates false positive areas detected by the algorithm that do not indicate true buildings. 29

4.8 Detected building regions after reidentification of reference building regions from the initial observation versus the manually created ground truth for Image 3. The same color coding is used as in Figure 4.7, with additional green areas that indicate building regions from the previous time that have been reidentified as being buildings. 30

4.9 Detected building regions after reidentification of reference building regions versus the reference building regions for Image 3. Blue indicates reference building areas that have not been detected by the algorithm, green indicates reference building areas that have been detected by the algorithm, and red indicates new areas detected by the algorithm that were not included in the reference set. Note that the reference building regions for this image are neither entirely complete nor entirely accurate for the current time. 31

4.10 Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 1. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9. 32

4.11 Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 2. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9. 33
4.12 Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 4. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.  

4.13 Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 5. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.
## LIST OF TABLES

4.1 Cross validation evaluation metrics for building detection accuracy without the reidentification step. .......................................................... 26

4.2 Cross validation evaluation metrics for building detection accuracy including the reidentification step. .................................................. 26

4.3 Change detection evaluation results. .................................................... 36

4.4 Total building region area for each set of regions in square feet. .......... 36
CHAPTER I

INTRODUCTION

Ever-increasing advances in sensor technology provide attractive research opportunities for remotely monitoring building changes using aerial or satellite imagery. Such development brings a great deal of applications in urban planning, cartographic mapping, disaster assessment, infrastructure safety monitoring, and others. Since manual analysis of the huge datasets captured by various sensors requires extensive effort, there is a demand to develop automatic algorithms that can perform analysis tasks with minimum human intervention.

A particular application of aerial imagery that has wide applicability is the detection and recognition of objects. This problem is well addressed by recent developments in feature extraction and machine learning that are designed to characterize and classify sets of data that are difficult to identify using simple image processing techniques. The ability for machine learning classifiers to study immense data sets in order to recognize patterns and later classify new data based on these patterns is well suited to augmenting the efforts of human analysts in their analysis of large data sets. Unlike a human analyst, a classification algorithm does not grow tired as it continues to search for patterns, and it follows a well defined set of behaviors in all circumstances. Using these technologies, analysts are able to monitor data sets that are orders of magnitude larger and more complex than what they would be able to do on their own.
One example of this kind of object recognition that has been studied extensively is the detection of buildings and changes to building areas over time. Some works propose to extract buildings from visible Red-Green-Blue (RGB) images, while others utilizes a variety type of sensors information such as multispectral, Synthetic Aperture Radar (SAR), Light Detection And Ranging (LiDAR), Digital Surface Models (DSM), and so on. In this study, we use two-dimensional RGB and Near InfraRed (NIR) satellite images for building extraction and change detection, plus a shapefile record of known road regions in the areas around the example imagery.

In general, the main contributions of this paper can be summarized as follows:

- A new feature fusion technique for identifying likely building characteristics in a linear combination for building detection is proposed.

- A Difference of Gaussian (DoG) histogram feature is effectively utilized to differentiate true building candidates from false positives based on their shape.

- A reidentification strategy of previously existing building regions through a Support Vector Machine (SVM) classifier.

This thesis is organized in the following manner: Chapter II summarizes the existing literature on the topic of building detection and change detection. Chapter III provides a detailed description of the proposed framework, including adaptive local feature extraction, multi-stage background elimination and building change identification. Chapter IV presents the experimental results with performance validation using statistical evaluation metrics. Finally, conclusion of the paper is drawn in Chapter V.
CHAPTER II

LITERATURE SURVEY

The basic approach of 2-D building change detection is to first identify building candidates and then perform a comparison with existing building reference maps. Tanathong et al. [1] detected building changes using an objected oriented change detection approach, where object level comparison is performed from two different time instances. This method is also referred to Post Detection Comparison (PDC) [2]. However, the performance of this method may be decreased by image color discrepancies. Benedek [2] improved the PDC method by extracting buildings from on the sparsely populated multitemporal images based on Marked Point Process (MPP) [3, 4]. Bouziani et al. [5] presented anther object oriented change detection algorithms which consists of several stages, including object class modeling, segmentation, object properties (i.e., geometric and contextual properties) learning, and change identification. In this method, multispectral image with the panchromatic image fusing is required prior to change detection. Moe and Sein [6] proposed unsupervised building changed detection techniques by utilizing mathematical morphological operations to extract building spectral-structural information from the image. In this method, color histogram was first applied to determine if there is change between images captured in two different epochs, and a Speeded Up Robust Features (SURF) based image registration technique is employed. Followed that buildings was detected by morphological filters and changes is decided
a object matched based change rule. Nevertheless, image perspective variation such as noise artifacts or resolution distortion may affect this algorithm performance during the color histogram computation and the image registration processes. Overall, aforementioned two-dimensional (2-D) change detection methods are limited due to image spatial or spectral resolution quality. Studies by Huang et al. [7, 8] and Bovolo et al. [9] showed that even high-resolution data may be ineffective to conventional change detection techniques due to the effects of image illumination, sensor viewing angles, orientation and the residual false registration between multitemporal images.

On the other hand, similar color characteristics often makes 2-D based changed detection algorithm difficult to distinguish buildings from other man-made constructions, like roads and artificial architects, during the change detection procedure. Incorporating 3-D information during building extraction and change detection has been explored in the literature. Digital Surface Models (DSMs) have been widely recognized as an effective tool for building change detection since it provides volumetric information [10] of objects in a scene. DSM can be generated either by Very High-Resolution (VHR) images or Airborne Laser Scanning (ALS) [11, 12]. Jung [13] compared DSM from multitemporal aerial images to eliminate background objects and then detect changes based on the decision tree based classifier. Rottensteiner et al. [14] proposed a building change detection method using DSM and a normalised difference vegetation index. They employed a Dempster-Shafer fusion model [15] to detect building from a DSM and compared the detection results with an existing building data base to find changes, the resultant changes were specified as confirmed, new, demolished, or changed. Similarly, a method presented by Tian et al. [16] also employed Dempster-Shafer fusion technique for refining the detected change, in contrast, changes was computed by using height information and Kullback-Leibler divergence similarity measure. Chaabouni-Chouayakh and Reinartz [17] introduced a post-processing method to avoid DSM associated artifacts, where shape features
were extracted for describing the detect changes and a Support Vector Machine (SVM) [18] classifier was used for predicting real changes, whereas Tian et al. [19] improved the change detection by applying image denoising and thresholding on the difference image. Qin [20] effectively fused multi-channel indicator and self-organizing maps (SOM) to obtained building changes, and the ambiguous changes was further analyzed using Markov Random Field (MRF). Recently, Qin et al. [21] combined height, texture and spectral information from orthophotos and DSM to perform 3D building change detection. They applied the mean-shift segmentation and machine learning techniques such as decision tree and SVM for classifying building segments, and changes is determined by change indicator and dual thresholding. However, The main disadvantage of above-mentioned 3-D building change detection algorithms is that their performance largely relies on the quality of stereos images and DSMs. Furthermore, image repositories may lack of DSM, stereo or multi-sensor information. This would result in many methods infeasible in real life applications. Therefore, in this research, we simply use RGB and NIR images for building change detection, and there is no any additional spectral or 3-D data is required, which gives our algorithm wider applicability. The proposed method is not only capable of detecting building changes from sparsely populated rural regions but also from very dense urban areas.

Recent techniques have been developed for building detection without the goal of multitemporal change detection that may be used to provide a comparison for the building detection techniques presented in this research. For example, research by Ok et al. [22] detects buildings with arbitrary shapes using their spatial relationships with nearby shadows. Their work also extends the GrabCut partitioning technique, which existed previously as a semi-automated image segmentation technique that required user input to indicate the location of building areas to be segmented [23]. In their work, building regions are detected by first detecting the shadow regions in an image, and searching in the direction opposite of the shadow angle for building regions. GrabCut is then applied to the detected
building regions to segment them and obtain an accurate boundary around each building region. This technique works well when the only elevated objects of interest in a scene are buildings, and these buildings shadows are non-overlapping.

Another recent building detection technique by Manno-Kovacs and Sziranyi [24] involves the detection of buildings based on assumptions of consistent geometric orientation of buildings in a local region. After detecting the relevant edges and corners in a scene, feature points that match the defined mixture of Gaussians orientation model are likely to represent building regions with 90 degree edges and whose neighboring building orientations are consistent. This technique works well for detecting buildings, as long as their orientations are consistent in a local area.

This research extends our previous research into building detection from visible imagery based on subtraction of background region characteristics and the use of shadow direction estimation for elimination of false positive detections [25]. A similar approach for finding building candidate regions based on pixel-based and texture-based considerations is used to create the initial set of buildings. However, in this latest work, a model of continuous values identifying the likelihood of each pixel being a building is creating using various features in a linear combination. The use of entropy as the defining feature for finding building candidates with smoothed textured roofs has been substituted for a gradient magnitude based approach, which provides similar information with clearer edge definition and less computational complexity. Additionally, the shadow direction filtering step has been substituted for a shape-based classification approach based on SVM to give stronger performance in areas where building shadows are not easily identifiable, such as dense urban regions. A Difference of Gaussian (DoG) histogram feature is used to classify the initial building candidate regions. The effectiveness of multi-scale DoG filtering has been used to great effect in the seminal Scale Invariant Feature Transform algorithm [26]. In this work, the DoG
filtering process is instead used as a sophisticated shape-based feature extraction technique that identifies the presence of regions of different sizes.
CHAPTER III

BUILDING REGION CANDIDATE CREATION AND CLASSIFICATION

The block diagram shown in Figure 3.1 outlines the structure of the proposed algorithm. This building detection scheme begins with identifying the pixels in an image that are likely to represent building regions based on their color characteristics through Normalized Difference Vegetation Index (NDVI) and shadow intensity, their texture characteristics using gradient magnitude, and their distance from known road areas. These building candidate areas are further analyzed based on their shape and edge characteristics through a DoG classification process.

3.1 Feature fusion for building candidate creation

The initial process for finding building candidate regions is based on estimating likely building regions through a linear combination of multiple features that identify building characteristics. This process consists of computing an image’s NDVI, gradient magnitude, shadow intensity, and road intensity, resulting in a combined intensity image identifying regions likely to be buildings and those likely to be non-building background regions. The combination of strong building features creates a robust model of likely building areas that are later thresholded based on a desired tolerance of the combined intensity image. The feature fusion technique is also computationally simple in order to allow for fast execution.
All features are normalized using min-max normalization in order to ensure that their values are within the range of [0,1] before performing the following weighted sum:

$$I_F(x, y) = w_1 |I_N(x, y)| + w_2 I_G(x, y) + w_3 I_S(x, y) + w_4 I_D(x, y)$$  \hspace{1cm} (3.1)

where $I_N(x, y)$ is the NDVI of the image, $I_G(x, y)$ is the gradient magnitude of the image, $I_S(x, y)$ is the shadow intensity of the image, and $I_D(x, y)$ is the road intensity of the image.

The NDVI for an image characterizes the difference between vegetation and non-vegetation areas, such as stone, dirt, roads, or other non-vegetation areas. True building areas are very likely to have low NDVI [27]. However, in the dataset used for this work, areas with negative NDVI were found to be dark shadows cast by large trees or other objects, which are also unlikely to be buildings, so the absolute value of the NDVI was used in place of the signed NDVI value. The NDVI for an image can be computed based on the relationship between the images’ red and NIR channels using the following equation:

$$I_N(x, y) = \frac{I_{NIR}(x, y) - I_R(x, y)}{I_{NIR}(x, y) + I_R(x, y)}$$  \hspace{1cm} (3.2)
where \( I_{\text{NIR}}(x, y) \) is the NIR channel of the image and \( I_R(x, y) \) is the red channel of the image.

The gradient magnitude is a commonly used image feature to find the strength of the edges in an image. It can be computed using the following equation:

\[
I_{G}(x, y) = \sqrt{\left( \frac{\partial I_{\text{gray}}(x, y)}{\partial x} \right)^2 + \left( \frac{\partial I_{\text{gray}}(x, y)}{\partial y} \right)^2} \tag{3.3}
\]

where \( I_{\text{gray}}(x, y) \) is the grayscale converted image based on the visible components of the input image. In this research, the Sobel kernel was used for the computation of the gradient magnitude.

The shadow intensity for an image characterizes the confidence that a given region consists of shadows. Since building areas are unlikely to exist in shadows, they are conversely likely to have a low shadow intensity. The equation for computing the shadow intensity is based on the one presented by Shorter et al. [28], but has been modified through the addition of an absolute value calculation so that shadow regions will have high intensity, and building regions will have low intensity. The shadow intensity \( I_S(x, y) \) for an image can be computed as follows:

\[
I_S(x, y) = \left| \frac{4}{\pi} \arctan \frac{\frac{I_R(x, y)}{I_{\text{gray}}(x, y)} - \sqrt{I_R(x, y)^2 + I_G(x, y)^2 + I_B(x, y)^2}}{I_R(x, y) + \sqrt{I_R(x, y)^2 + I_G(x, y)^2 + I_B(x, y)^2}} \right| \tag{3.4}
\]

where \( I_G(x, y) \) and \( I_B(x, y) \) are the green and blue channels of the image respectively.

Road areas have been found to be a source of false positive detections due to their similar texture and shape with some building areas. To eliminate false positives on road areas, existing road data is mapped onto the input imagery from vector shapefiles. Since these road areas are not guaranteed to be completely updated or accurate with respect to the input imagery, the road data is filtered in order to express the increasing confidence that areas near the road data are unlikely to be buildings. The process for determining the road intensity image \( I_D(x, y) \) is given in the following equation,
which uses Gaussian filtering for distributing the mapped road data into the image coordinates:

\[
I_D(x,y) = [I_{BR}(x,y) \oplus M(x,y)] * \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}
\]  

(3.5)

where \( \oplus \) is the morphological dilation operator, \( M(x,y) \) is a disk structuring element of radius \( r \), \(*\) is the convolution operator, \( \sigma \) is the standard deviation of the Gaussian function, and \( I_{BR}(x,y) \) is a binary image constructed from known road data mapped to the image. The parameters \( r \) and \( \sigma \) must be adjusted for a given set of data in order to reflect the expected width of roads and the confidence of the road data in these images respectively. An example road intensity image is shown in Figure 3.2.

After the building candidate regions have been created as per the feature fusion intensity relationship given in Eq. (3.1) and subsequent thresholding, it is possible that true building regions may be connected to background regions. Therefore, we separate the two classes of regions so that they may be classified accurately in later steps. A scheme for segmenting the large regions that are likely to include background regions is described in Figure 3.3. The threshold for the gradient magnitude
Figure 3.3: Block diagram for the separation of background regions from true building regions.

in this operation should be set so that it segments all building edges in the given imagery with a greater strength than the process performed as part of the feature fusion process. The goal of this operation is to create a set of building candidate regions where no single building candidate region includes both true buildings and background areas.

3.2 Building candidate classification

While the feature fusion process is able to identify the majority of building regions in an image, it also contains many false positive non-building areas which have texture and color characteristics that are similar to true buildings. To reduce the false positives in this detection, a SVM classifier is used to classify each building candidate region as being a true building or background through a number of descriptive features.

For the SVM classifier to distinguish between the true building and non-building regions, it must be presented with features that represent each class of region. At this stage in the algorithm, a multi-scale DoG feature is used to identify the two classes of image region. A 120-bin histogram of the intensity in each building region is calculated to signify the distribution of the intensity at each
DoG scale. After the histograms for each scale have been calculated, they are concatenated in order to form the final feature vector for classification using SVM. Since 6 scales are used for the dataset in this research, the resulting feature vector will be 720 elements long. By recognizing the distribution of thin and thick regions in a building candidate, the geometric signature of buildings can be extracted and used to classify candidates using the SVM classifier. The process for calculating the DoG features is described in the following equation:

\[
\Gamma_{\sigma, k_i}(x, y) = I_{\text{gray}}(x, y) \ast \left[ \frac{1}{2\pi\sigma^2_i} e^{-\frac{x^2 + y^2}{2\sigma^2_i}} - \frac{1}{2\pi\sigma^2_i} e^{-\frac{x^2 + y^2}{2\sigma^2_i}} \right] 
\]

for \( i = 1, 2, 3... \)

where \( I_{\text{gray}}(x, y) \) indicates the grayscale image, \( \sigma \) indicates the variance of the Gaussian functions, and \( k \) is the Gaussian scale factor. The variance values should be selected based on the size of building regions and the resolution of the imagery.

As an example of this feature extraction process, Figure 3.4 shows a representative RGB area of interest image, and Figure 3.5 shows the intensity of the DoG filtering process at different scales of \( \sigma \). Figure 3.6 depicts the algorithm for feature extraction and classification of a single building candidate region.

### 3.3 Reference region reidentification

Since it is critical that the algorithm minimize false negative detections as much as possible, all building regions from the initial observation time that are not identified in the current imagery are revisited to ensure that they are not in fact building areas. Specifically, reference building regions that have not been identified through the algorithm are classified using the previous DoG feature extraction and classification process. If these regions are classified as non building areas, then they are marked as having been removed from the reference region. If these regions are classified as building areas, they are reidentified and added to the current detected building regions. This
Figure 3.4: RGB image showing a representative area of buildings.

Figure 3.5: Visualized DoG intensities for multiple sigma scales with the manually identified building ground truth regions outlined in white. Note that building regions of different sizes and shapes are characterized by regions of low intensity for the varying sigma values.
technique allows for buildings that are difficult to detect with the current feature fusion algorithm, especially buildings with roofs that exhibit high gradient magnitude textures, to be considered in the change detection more accurately. Figure 3.7 depicts a block diagram describing this approach.

### 3.4 Building change detection

After the final building candidate regions have been produced, the detections from the current imagery are compared against the reference building regions from the initial observations. Areas that exist in only the initial observation and not the current imagery are marked as having been removed, areas that exist in only the current imagery and not in the initial observation are marked as newly added areas, and areas that exist in both the initial observation and the current imagery are marked as unchanged. Color-coded graphics are produced in order to visualize the changes in a
region and to check the results of the algorithm. These graphics are shown for the final results in all images in Section 4.4.

The calculations for determining the changes between building regions are based on simple pixel-wise comparisons. These changes can be calculated through a series of Boolean operations on the binary images representing the final detected building regions from the building detection algorithm and the reference building regions. Unchanged regions are calculated as the logical AND of the detected building regions and the reference building regions, removed regions are calculated as the logical AND of the logical NOT of detected building regions and the reference building regions, and the added regions are calculated as the logical AND of the detected building regions and the logical NOT of the reference building regions. Figure 3.8 shows a block diagram for the final change detection algorithm.

This simple approach of detecting changes with respect to the binary reference regions allows the algorithm to work with a wide range of formats for the building regions in the initial observations. In fact, in the dataset used for the development of this algorithm, no imagery was provided for the initial observation, only a geospatial vector shapefile indicating polygons representing the
building regions. Hence, for detecting changes in building regions from this dataset, an approach that operates completely independently of the imagery in the initial observation is required.
Figure 3.8: Block diagrams for producing final building change detection using Boolean operations on binary images.
CHAPTER IV

PERFORMANCE EVALUATION

4.1 Dataset description

A diverse set of images were used for the evaluation of the proposed algorithm. This dataset includes 5 RGB and NIR images at a resolution of 25 megapixels with a Ground Sample Distance (GSD) of 1 foot per pixel. These images each represent a total area of 25 million square feet, although they all have a varying amount of imagery excluded from them in the form of whitespace that indicates areas where building regions are not of interest. The areas shown in these images include rural, suburban, and urban areas of California. Figure 4.1 shows thumbnails of the images used to train the classifier, which represent a diverse range of areas including urban, suburban, and industrial areas. The known road areas were provided in the form of vector polyline shapefiles which were read to find the relevant vertices for a given image, then mapped from geocoordinates to pixel coordinates.

A potential source of ambiguity is the difference between the building regions indicated by the initial observation that are used to report changes against and the regions manually marked as being buildings in the imagery under analysis. In this paper, the former regions are referred to as the “reference” regions, and the latter as the “ground truth” regions. Note that the imagery associated with the reference regions are not made available to the algorithm presented here, only
the geocoordinates of the building regions from that reference dataset. Additionally, in every image in the evaluation dataset, the reference building regions are both incomplete and incorrect to varying degrees with respect to the ground truth regions.

4.2 Parameter selection

The following weights were used for the relationship in Eq. 3.1 and were determined empirically to represent the most appropriate relationship between the features for the imagery being tested: $w_1 = 0.40, w_2 = 0.25, w_3 = 0.10, w_4 = 0.25$. A threshold of 0.11 was selected for creating the initial binary building candidate regions from the feature fusion image.

In the road mapping procedure described in Eq. 3.5, a radius $r$ of 14 feet is selected for $M(x, y)$ (according to the specified Ground Sample Distance for the imagery being processed) in order to create the assumptions of the resulting region, and $\sigma$ is selected to be 15 to determine the spread of probability of the road regions.

The Gaussian variance values $\sigma_i$ in Eq. 3.6 are selected to be 2, 4, 8, 16, 32, and 64 to create the multi-scale features. All values of $k_i$ are set to 2 so that the larger Gaussian function have twice the variance of the next smaller function. As discussed in Section 3.2, a 120-bin histogram will be calculated for the DoG features at each scale, and all 6 will be concatenated in order to create a 720 element feature vector for classification by SVM.

4.3 Evaluation methods

The evaluation procedure in this research consists of two different sets of metrics: one set where the building regions detected by the algorithm are compared to the building regions in the ground truth, and one set where the changes detected by the algorithm with respect to the reference regions are compared to the changes between the reference regions and the ground truth. The former will
Figure 4.1: Training and testing images, displayed in increasing order of number of building regions.
be referred to as the building detection results, and the latter as the change detection results. An algorithm with perfect building detection results would have perfect change detection results and vice versa. Both metrics are presented here in order to provide a more complete description of the algorithm’s performance.

The building detection results are calculated by comparing the automatically detected building regions generated by the algorithm with the manually created ground truth images. The ground truth images were created by drawing binary regions on all observed building regions in a graphics editing application.

The change detection results are calculated by comparing the changes between the detected regions with respect to the reference regions with the changes between the ground truth regions with respect to the reference regions. This comparison process requires additional calculations to find the relationship between each set of changes. For example, in order to find the true positive removed regions, or the areas that have been correctly indicated as being removed by the algorithm from the reference regions as confirmed by the ground truth, the detected removed regions are combined with the true removed regions. This process is described in Figures 4.2 through 4.4. The same general process described in these block diagrams is used to compute the true positive, false positive, and false negative added regions by starting with the added region calculation described in Figure 3.8.

After determining the pixel-wise true positive, false positive, and false negative detections for both the building detection and change detection results, a completeness and correctness metric can be calculated for each in order to provide a compact description of performance. The following equations describe the metrics used to create the pixel-wise completeness and correctness scores for this method:
Figure 4.2: Block diagram for calculating the True Positive (TP) removed region detections.

Figure 4.3: Block diagram for calculating the False Positive (FP) removed region detections.
Figure 4.4: Block diagram for calculating the False Negative (FN) removed region detections.

Completeness = \( \frac{TP}{TP + FN} \) \hspace{1cm} (4.1)

Correctness = \( \frac{TP}{TP + FP} \) \hspace{1cm} (4.2)

where a True Positive (TP) is defined as a pixel that indicates a detection of a building region or detection of a correct change, a False Negative (FN) is defined as a pixel that indicates a missed detection of a building region or a missed detection of a change, and False Positive (FP) is defined as an incorrect identification of a background region as being a building region or an incorrect detection of a change that did not in fact occur.
4.4 Building detection and change detection results

Tables 4.1 and 4.2 show a cross validation evaluation of building detection performance on the set of training images selected without the reidentification step, and with the reidentification step respectively. In each round of these tests, the detection results for a single image is selected for evaluation, and the remaining images are used for creating the training set for the SVM model. By repeating the experiment five times (in order to use each of the five images for testing), an overall impression of the algorithm’s performance across a range of images is determined. In each pair of columns, a different method is evaluated. The columns labeled “NDVI” indicate that all coefficients in Eq. (3.1) are set to zero, besides the coefficient corresponding to NDVI, which was set to one, effectively ensuring that only NDVI is used to generate the initial building candidates. Likewise, the columns labeled “Gradient” indicate that all coefficients are set to zero besides the one corresponding to the gradient magnitude. The columns labeled “Fusion” indicate that the coefficients described in Section 4.3 were used. Note that the reference regions from the initial observation time in each image used for reidentification have a varying level of completion and correctness, but no set of reference regions for any image is fully accurate.

Table 4.3 shows the change detection performance of the algorithm with reidentification. The “Added regions” section indicates the completeness and correctness of the reported changes where building regions were added in each of the images, and the “Removed regions” section indicates the completeness and correctness of the reported changes where building regions were subtracted in each of the images. A perfect building detection that entirely matched the ground truth would result in values of 100% in all fields in this table.

Table 4.4 shows the total building region area in terms of square feet for the ground truth regions, reference regions, and detected regions across all five images. The column labeled “Detection”, the total building region area is shown for the detected building regions without reidentification, and
Table 4.1: Cross validation evaluation metrics for building detection accuracy without the reidentification step.

<table>
<thead>
<tr>
<th></th>
<th>Gradient</th>
<th>NDVI</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completeness</td>
<td>Correctness</td>
<td>Completeness</td>
</tr>
<tr>
<td>Image 1</td>
<td>44.07%</td>
<td>59.00%</td>
<td>44.81%</td>
</tr>
<tr>
<td>Image 2</td>
<td>48.45%</td>
<td>75.14%</td>
<td>40.38%</td>
</tr>
<tr>
<td>Image 3</td>
<td>50.91%</td>
<td>63.24%</td>
<td>45.22%</td>
</tr>
<tr>
<td>Image 4</td>
<td>38.47%</td>
<td>59.29%</td>
<td>30.04%</td>
</tr>
<tr>
<td>Image 5</td>
<td>36.06%</td>
<td>52.36%</td>
<td>43.59%</td>
</tr>
</tbody>
</table>

Table 4.2: Cross validation evaluation metrics for building detection accuracy including the reidentification step.

<table>
<thead>
<tr>
<th></th>
<th>Gradient</th>
<th>NDVI</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completeness</td>
<td>Correctness</td>
<td>Completeness</td>
</tr>
<tr>
<td>Image 1</td>
<td>77.99%</td>
<td>70.39%</td>
<td>86.72%</td>
</tr>
<tr>
<td>Image 2</td>
<td>59.33%</td>
<td>79.81%</td>
<td>52.98%</td>
</tr>
<tr>
<td>Image 3</td>
<td>80.39%</td>
<td>74.42%</td>
<td>86.35%</td>
</tr>
<tr>
<td>Image 4</td>
<td>87.15%</td>
<td>67.54%</td>
<td>87.49%</td>
</tr>
<tr>
<td>Image 5</td>
<td>62.73%</td>
<td>60.97%</td>
<td>79.19%</td>
</tr>
</tbody>
</table>

The column labeled “Detection Reid.” shows the total building region area for the detected building regions with reidentification.

4.5 Discussion of results

The result of the feature fusion process shown in Figure 4.5 and the thresholded version shown in Figure 4.6 indicates that the fusion of the selected features with the chosen weights correctly separates vegetation, roads, shadows, and regions with rough textures from other likely building regions. Note that some building regions do appear undersegmented in this result, particularly in the lower right corner of the image, due to having a comparatively small region with low gradient magnitude intensity.
Figure 4.5: The resulting fusion image for Image 3. Pixels with lower intensity, marked in blue, are more likely to represent building regions than pixels with higher intensity, marked in yellow.
Figure 4.6: Building candidate regions created through the thresholding of the fusion intensity of Image 3 shown in Figure 4.5 as compared to the manually created ground truth. Blue indicates ground truth pixels that have not been detected by the algorithm, green indicates ground truth pixels that have been detected by the algorithm, and red indicates false positive pixels detected by the algorithm that do not indicate true buildings.
Figure 4.7: Automatically detected building regions versus the ground truth for Image 3. Blue indicates ground truth areas that have not been detected by the algorithm, green indicates ground truth areas that have been detected by the algorithm, and red indicates false positive areas detected by the algorithm that do not indicate true buildings.
Figure 4.8: Detected building regions after reidentification of reference building regions from the initial observation versus the manually created ground truth for Image 3. The same color coding is used as in Figure 4.7, with additional green areas that indicate building regions from the previous time that have been reidentified as being buildings.
Figure 4.9: Detected building regions after reidentification of reference building regions versus the reference building regions for Image 3. Blue indicates reference building areas that have not been detected by the algorithm, green indicates reference building areas that have been detected by the algorithm, and red indicates new areas detected by the algorithm that were not included in the reference set. Note that the reference building regions for this image are neither entirely complete nor entirely accurate for the current time.
Figure 4.10: Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 1. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.
Figure 4.11: Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 2. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.
Figure 4.12: Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 4. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.
Figure 4.13: Comparison between final detection results with reidentification with respect to the ground truth and reference regions for image 5. The color coding for the ground truth comparison images is the same as in Figure 4.8, and the color coding for the reference comparison images is the same as in Figure 4.9.
Table 4.3: Change detection evaluation results.

<table>
<thead>
<tr>
<th></th>
<th>Added regions</th>
<th>Removed regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completeness</td>
<td>Correctness</td>
</tr>
<tr>
<td>Image 1</td>
<td>45.96%</td>
<td>66.00%</td>
</tr>
<tr>
<td>Image 2</td>
<td>58.61%</td>
<td>52.98%</td>
</tr>
<tr>
<td>Image 3</td>
<td>53.60%</td>
<td>58.97%</td>
</tr>
<tr>
<td>Image 4</td>
<td>48.16%</td>
<td>60.45%</td>
</tr>
<tr>
<td>Image 5</td>
<td>47.05%</td>
<td>48.00%</td>
</tr>
</tbody>
</table>

Table 4.4: Total building region area for each set of regions in square feet.

<table>
<thead>
<tr>
<th></th>
<th>Ground Truth</th>
<th>Reference</th>
<th>Detection</th>
<th>Detection Reid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image 1</td>
<td>1,303,812</td>
<td>1,254,234</td>
<td>1,168,977</td>
<td>1,564,721</td>
</tr>
<tr>
<td>Image 2</td>
<td>2,809,254</td>
<td>1,132,687</td>
<td>2,423,131</td>
<td>2,663,315</td>
</tr>
<tr>
<td>Image 3</td>
<td>3,706,332</td>
<td>3,198,304</td>
<td>2,653,549</td>
<td>3,761,094</td>
</tr>
<tr>
<td>Image 4</td>
<td>3,820,579</td>
<td>3,478,476</td>
<td>3,368,791</td>
<td>4,886,000</td>
</tr>
<tr>
<td>Image 5</td>
<td>2,643,841</td>
<td>2,538,850</td>
<td>2,008,920</td>
<td>3,154,699</td>
</tr>
</tbody>
</table>

The thresholded result in Figure 4.6 is further improved in Figure 4.7 through the use of the SVM classifier with the DoG features generated for every building candidate region. The majority of non-building regions have been eliminated from the candidate regions in this step due to their shape characteristics. However, some true building regions have also been removed in this classification step due to their DoG features being uncharacteristic of most buildings.

Figure 4.8 shows the increase in building detection performance through the use of reidentification using the reference building regions. Note that many small building regions that were missing or undersegmented before reidentification have been identified due to their DoG features. Figure 4.9 indicates that despite some reference regions marked in blue do not indicate building regions in the imagery, the reidentification algorithm is able to identify non-building regions using their DoG features to prevent the addition of false positives in the final detection results.
The evaluation metrics shown in Tables 4.1 and 4.2 illustrate the improved completeness and correctness scores offered by the proposed feature fusion strategy over gradient magnitude or NDVI based approaches. While many of the NDVI results score high in correctness, their completeness scores are generally lower than those in the feature fusion approach, indicating that the NDVI approach results in fewer false positive detections, but more false negative detections.

Table 4.3 shows that the algorithm was able to detect a majority of both added and removed regions in the dataset. In general, the higher scores of the removed region metrics versus the added region metrics indicate that the algorithm is more capable of showing when building regions have been removed than when they have been added. This difference in metrics is unsurprising, as the task of identifying removed regions is simpler than identifying added regions. The process for finding removed regions is aided by the reidentification algorithm and the reference region data, which in general contains an accurate segmentation of the reference building regions. In contrast, a correct identification of an added region by definition requires searching an area of the image that has no reference region information available and accurately deciding what portions of this area contain building regions.

However, Table 4.3 shows low correctness in removed regions for the evaluations from Images 4 and 5. This is due to the relatively low amount of true removed regions in each, as well as the density of the true buildings in these images. These two factors combine together to create challenging conditions for change detection of removed regions, as there are relatively few regions that could be correctly identified as negative change, and relatively many regions that could be incorrectly identified as negative change.

Table 4.4 provides an objective comparison on the size of the regions marked in each image for the ground truth, reference, and detection results. It also shows the relative density of the building regions in each image, with Image 4 having the highest area of true building regions and Image
1 having the least. An examination of the difference between the detection results with and without reidentification is useful in understanding how much area was added to each detection result through the process of reidentification. The result of Image 4 in particular increased quite a lot from reidentification, but much of it was due to a false positive detection in the upper right corner of the image where a large region of a sports complex was imprecisely identified as a building region.

The reported change detections shown in the final result visualizations in Figures 4.10 through 4.13 indicate a meaningful detection of many building regions that have been changed or marked inaccurately in the reference regions. In particular, the detected regions in the bottom half of Image 2 indicate a significantly more complete set of building regions than the reference regions. Additionally, in this image, a baseball diamond and a dirt patch under construction were included in the set of reference building regions, which have been correctly determined by the algorithm to not be true building regions. Figure 4.9 indicates a particularly useful result in the bottom left corner of the image, where a large commercial building has been marked as an added building region in this imagery, as well as marking a large dirt area that may have been of concern in the initial observation as not being a true building region in the current imagery. Across many of the images, the algorithm has marked many of the regions around the perimeter of the visible imagery as being added buildings, which may have been ignored in the creation of the reference building region data.
CHAPTER V

CONCLUSION

A new technique for the automatic building change detection is presented. Through a linear combination of various features that are known to represent a building’s pixel, texture, and location characteristics, a wide range of buildings can be detected, even in challenging imagery. By combining a number of distinctive features, effective detections can be performed across a diverse range of data. Since many real world applications for building change detections will have a reference set of buildings to compare the current detections against, the current detections can be made more complete through the use of reidentification algorithms as developed in this research. This technique also does not rely on preexisting knowledge of shadow direction in an image, as many competing techniques do, allowing it to be used on a wider range of data sets.

The accuracy of the proposed methodology may be improved by incorporating more advanced features that are may add to the confidence of the linear feature fusion in identifying regions that are likely to be buildings. In particular, a robust process for the automatic identification of road areas would increase the accuracy in general, since the road intensity image that is created relies on complete and updated road data imported from Geographic Information System (GIS) software that may not exist in all areas where building detection is performed.

The proposed algorithm also relies on the tuning of the coefficients in order to perform the best detections. Simple adaptive techniques such as Otsu’s thresholding method were investigated during
the development of the proposed algorithm, but their results were inadequate for the complexity of
the imagery included in the dataset. This parameter selection process may be automated in future
work by extending the algorithm with an adaptive data analysis process.

Another area of future work should include a more sophisticated segmentation process for se-
lecting a more precise region in the final decision of which pixels belong in a building region. Simple
segmentation techniques like the active contour model were investigated for this algorithm, but it
was found to be challenging to select parameters that resulted generally in improved segmentations
across all possible building regions.
BIBLIOGRAPHY


