Building extraction in satellite images using active contours and colour features

Gregoris Liasis & Stavros Stavrou

To cite this article: Gregoris Liasis & Stavros Stavrou (2016) Building extraction in satellite images using active contours and colour features, International Journal of Remote Sensing, 37:5, 1127-1153, DOI: 10.1080/01431161.2016.1148283

To link to this article: http://dx.doi.org/10.1080/01431161.2016.1148283

Published online: 18 Feb 2016.

Submit your article to this journal

Article views: 167

View related articles

View Crossmark data

Citing articles: 1 View citing articles
ABSTRACT

Obtaining the segmentation of building footprints from satellite images is a complex process since building areas and their surroundings are presented with various color intensity values and complex features. Active contour region-based segmentation methods can be used to establish the corresponding boundary of building structures. Typically, these methods divide the image into regions that exhibit a certain similarity and homogeneity. However, using the traditional active contour algorithms for building structures detection, in several cases where spectral heterogeneity exists, over-detection or under-detection are usually noticed. In this work, the Red, Green and Blue (RGB) representation and the properties of the Hue, Saturation and Value (HSV) color space have been analyzed and used to optimize the extraction of buildings from satellite images in an active contour segmentation framework. Initially, the satellite image was processed by applying a clustering technique using color features to eliminate vegetation areas and shadows that may adversely affect the performance of the algorithm. Subsequently, the HSV representation of the image was used and a new active contour model was developed and applied for building extraction, utilizing descriptors derived from the value and saturation images. A new energy term is encoded for biasing the contours to achieve better segmentation results. An effective procedure has been designed and incorporated in the proposed model for the active contour initialization. This process enhances the performance of the model, leading to lower computational cost and higher building detection accuracy. Additionally, statistical measures are used for designing optimum morphological filters to eliminate any misleading information that may still exist. Qualitative and quantitative measures are used for evaluating the performance of the proposed method.

1. Introduction

Built-up areas on satellite images display diverse features, different colors and intensities. Furthermore, buildings may be obscured by environmental objects and may occlude each other. For a successful segmentation and building extraction process,
this complexity has to be handled properly. Building detection from monocular satellite images can serve in land use analysis, creation and update of maps or Geographic Information System (GIS) databases and development of urban monitoring applications. The manual processing of such images is tedious and time-consuming, thus very expensive. This leads to the need for designing effective methods for the automated detection of buildings using remote sensing or satellite images.

Designing and implementing methods for detecting building structures is an active field of research and a large number of studies have been reported in literature. The existing methods can be classified into two categories with respect to the type of the images that are used. The first category includes the detection of buildings using monocular remote sensing images and the second category covers the detection of buildings using an additional channel of data such as height information. Detailed reviews can be found in Chen et al. (2013), Haala and Kada (2010), Baltsavias (2004), Ünsalan and Boyer (2005). Since this work deals with the detection of buildings through single monocular optical remote sensing images, the discussion of the previous studies will be focused on this domain.

Initial attempts in the monocular context mostly used edge-based methods and relied on the extraction of features such as lines and corners (Huertas and Nevatia 1988; Irvin and McKeown 1989). In these studies, the cast shadows were also utilised for the detection of the corresponding buildings. Shadow information was used for estimating the location, shape and height of buildings by Irvin and McKeown (1989). Liow and Pavlidis (1990) used an edge detector and features such as shadows for extracting building boundaries in aerial images. Furthermore, local edge information and global shape information was utilized to eliminate the segmentation errors or to enhance the resultant contours. Shufelt and McKeown (1993) combined several separated systems relied on shadow information for the detection of a complete set of building boundaries. McGlone and Shufelt (1994) have used image orientation information and shadow evidence to verify a building hypothesis based on the projective geometry and vanishing point calculations. Shufelt (1999) performed a detailed evaluation and a comparison study of the edge-oriented methods. His work revealed that with edge oriented methods it is difficult to handle the entire building complexity because these methods utilize information from only a single band and are mostly based on a simple building shape hypothesis such as a parallelogram structure. Thus, new methods are required to achieve better results.

The launch of remote sensors with the ability to acquire very high resolution (VHR) multispectral images, inspired many researchers to use information from the additional bands in their proposed models. Classification methods were utilized in a number of research works. Lee, Shan, and Bethel (2003) have developed a supervised classification method, where Hough transformation was used in an automated building detection process using imagery acquired from the earth observation satellite system IK-ONOS. Geometric image features were extracted and a Support Vector Machine (SVM) classification framework was used for detecting artificial objects in Inglada's (2007) model. Texture features were utilized in an SVM framework by Koc-San and Turker (2014). Ghaffarian and Ghaffarian (2014a) collect training areas using shadow evidence and design an automated parallelepiped supervised classification method for detecting building objects. Graph-based theory was also used in implementing methods for the
detection of buildings. The Scale Invariant Feature Transform (SIFT) was used in a graph
cut model by Sirmacek and Unsalan (2009). A two level graph partitioning framework
was developed by Ok (2013), for increasing the performance of his previous proposed
model (2013), which was based on fuzzy logic and Grab Cut.

Markov Random Fields (MRFs) were thoroughly investigated in building detection
models from satellite images. Krishnamachari and Chellappa (1996) used MRFs to group
line segments and in the later years Katartzis and Sahli (2008) extended this model by
creating a stochastic framework for detecting rooftops of the presented buildings.
Independent component analysis (ICA) was utilized by Ghaффarian and Ghaффarian
(2014b), for developing a model, named Purposive Fast-ICA (PFICA) to detect buildings
from monocular high-resolution images acquired from Google Earth. In this study, a
colour-based fusion technique was proposed to approximate shadow, vegetation, bare
soil, road and building patches. This information was then used to initialize the Fast-ICA
algorithm for detecting the presented buildings in a given image.

A number of studies reported in the literature used a different approach by analysing
and incorporating multiple cue data to segment satellite images. Tsai (2006) analysed
and evaluated the most common colour spaces for shadow detection in aerial images
using a threshold technique applied to the colour spaced transformed images. Sirmacek
and Unsalan (2008) used invariant colour features and shadow information for building
detection. The red, green, blue (RGB) colour space was used and shadows were detected
by applying a threshold on the blue colour invariant image. In addition, buildings with
red rooftops were detected by applying a similar method on the red colour invariant
image. The evaluation and subsequently the selection of the most appropriate features
capable to characterize all or most of the buildings presented in the analysed scene, is
important for the performance of all the above methods.

Building extraction is a demanding, however a fundamentally important process in
many telecommunication or remote sensing applications such as natural disaster mon-
itoring, urban change detection, urban scene reconstruction, cartography update, urban
inventory and wireless planning (Hu, You, and Neumann 2003; Karantzalos and Paragios
2010; Chen et al. 2012). Image segmentation becomes very difficult in the presence of
complex features and spectral heterogeneity that often occurs in satellite images due to
factors, such as spatial variations in illumination, the use of various surface materials and
imperfections in the overall acquisition process. The problem becomes more compli-
cated if the regions of interest in the analysed scene are presented with a large scale of
different spectral characteristics, which is common in satellite images. Curve evolution
schemes such as the active contour without edges (Chan and Vese 2001), has the ability
to work well with natural images, however, in several cases multiple phases are needed
to achieve good performance (Vese and Chan 2002). The more complexity and variety of
intensity scales exist in the regions of interest, the more phases are needed to be
incorporated. This leads to the iteration of multiple partial differential equations (PDE)
simultaneously for approximating the segmentation of a given image. This time-con-
suming process limits the effectiveness of the method (Chan and Zhu 2003; Cremers,
Sochen, and Schnörr 2006; Karantzalos and Paragios 2010).

In the recent years, the idea of evolving an active contour for segmenting images in
the presence of spectral heterogeneity using local instead of global image statistics
has become popular (Li et al. 2008, 2011, Zhang, Song, and Zhang 2010; Zhang et al.
The use of local image properties leads to methods that are more efficient. This is because global region-based image segmentation methods typically rely on a specific region descriptor such as intensity mean values in each region to be segmented and is difficult to be estimated for images that include multiple regions of interest with various mean colour intensity values. A popular method is the local binary fitting model (LBF) by Li et al. (2008) and the extended version of the LBF, which incorporates a so-called local intensity clustering (LIC) property (2011). Another approach is the model, active contours driven by local image fitting energy (LIF) proposed by Zhang, Song, and Zhang (2010), where local features are incorporated in an active contour model to address the spectral heterogeneity drawback. Zhang et al. (2010) also proposed the extended versions of LIF, active contours with selective local or global segmentation and the variational approach to simultaneous image segmentation and bias correction (2015). These models integrate local and global features for image segmentation using active contours. Wang et al. (2015) have followed a similar approach in which local statistical features such as local intensities and global similarity measurements such as Bhattacharyya coefficient are utilized for segmenting images in the presence of spectral heterogeneity.

The above-mentioned active contour models detect all regions that exhibit a certain similarity and homogeneity. Therefore, by tuning their parameters, different regions of interest with various degrees of similarity and homogeneity can be detected from the image. The models have the ability to detect and extract the boundaries of all objects in the image, however it is difficult to selectively detect special features such as building structures (Zhang et al. 2010, Wang et al. 2015; Ahmadi et al. 2010; Elbakary and Iftekharuddin 2014). Furthermore, they seem to produce good results in those cases where the spectral heterogeneity varies slowly (Dai, Ding, and Yang 2015). For these reasons, it is necessary to optimize active contour segmentation models for building boundaries detection purposes in the presence of spectral heterogeneity and feature complexity.

The work presented in this paper involves with the evaluation and optimization of the active contour models for building structures detection in the presence of spectral heterogeneity and feature complexity. RGB colour space and Hue, Saturation, Value (HSV) colour features are thoroughly investigated and utilized for the development of an Optimized Active Contour (OAC) level set segmentation framework for the detection of buildings from monocular Google Earth satellite images. The Google Earth application is selected mainly because of free access to a large volume of images. The objective of the model was to handle properly the complexity of the scenes in satellite images and obtain an optimized segmentation of building footprints automatically. A well-defined initialization process able to enhance the detection of buildings accuracy and drive the evolving curves faster to the boundary of the buildings that exist in the analysed scene, has been designed. The shadows and vegetation regions are eliminated by implementing a colour clustering method leading to a more simplified scene. Subsequently, the HSV colour space components are utilized and an optimized active contour segmentation model is implemented and applied for the detection of buildings. The proposed model includes a new energy term based on the descriptors derived from the saturation and value representation of the image, for biasing, the active contours to selectively
detect building structures. Finally, statistical measures are used for designing optimum morphological filters to eliminate any misleading information that may still exist on the segmentation mask.

2. Active contour models for satellite image segmentation

The proposed building detection framework depends on optimizing the active contour without edges or Chan and Vese (CV) model (2001). Therefore, the traditional implementation of the model is briefly introduced in the following section. More details can be found in the corresponding referenced paper.

2.1 Chan-Vese active contour model

Active contours are piecewise polynomial curves that match a deformable model to an image by means of energy minimization (Kass, Witkin, and Terzopoulos 1988). In active contour models, a parametric or non-parametric representation of an initial curve is evolved using explicit and implicit forces. By energy minimization, these models are able to detect objects or obtain the image segmentation. The level set method for computing and analysing the evolution of a contour using partial differential equations is a very popular approach (Osher and Sethian 1988). The contour is implicitly represented, via a two-dimensional continuous function \( \phi(x, y, t) : \Omega \rightarrow \mathbb{R} \), defined in the image domain where \( \Omega \) denotes an open bounded subset of \( \mathbb{R} \). The function \( \phi(x, y, t) \), where \( (x, y) \) is the coordinates in the image plane and \( t \) is the time, is called the level set function. At any given time, the level set function define an edge contour and a segmentation of the image. The edge contour is taken to be the zero level set where \( \phi(x, y, t) = 0 \) and the segmentation is given by the two regions \( \{ \phi > 0 \} \) and \( \{ \phi < 0 \} \). CV model is formulated on level set and is a popular method for segmenting images containing objects with and/or without weak boundaries. This model uses a geometric representation of the contour. The method does not use an edge function based on gradient information to stop the evolving curve. The CV is an energy based minimization segmentation model that separates the grey-scale single band representation of an image into regions based on the mean intensities inside and outside the evolving curve. The optimal segmentation is found by minimizing the Mumford-Shah energy functional (Mumford and Shah 1989), for partitioning a given image \( I \) with a variable closed curve \( C \) using the following equation:

\[
E(c_1, c_2, C) = \lambda_1 \int_{\text{inside}(c)} |l(x, y) - c_1|^2 dx dy + \lambda_2 \int_{\text{outside}(c)} |l(x, y) - c_2|^2 dx dy + \mu L|C| + \nu A|C|
\]

where \( \lambda_1, \lambda_2, \mu, \nu > 0 \) are fixed parameters responsible to weight the different terms in the energy while \( c_1, c_2 \) are the intensity mean values of \( l \) inside and outside \( C \) respectively. In the CV model two regularization terms \( L|C| \) and \( A|C| \) are incorporated for controlling the smoothness of the detected boundaries. \( L \) and \( A \) represent the length and area of the variable closed curve \( C \) respectively.
The CV model has the ability to split and merge for detecting objects with various mean intensities and delineate objects with or without weak boundaries. This attribute makes it a good choice for handling the complexity of the scenes in satellite images and extract the presented buildings.

In several remote sensing applications, among the various methods already discussed, curve evolution schemes like active contours, deformable models and level set were used to segment satellite images. These methods revealed promising results due to their ability to cope with topological changes (Samson et al. 2001; Besbes, Belhadj, and Boujemaa 2006). Niu (2006) used a geometric active contour formulation, which was based on level set methodology in a semi-automatic framework, where the initial curves were manually defined in order to detect certain objects like buildings. In this formulation, the energy minimization function is based on image edges or gradients, a process that can limit the success of the method especially when the scenes are incorporating objects with weak boundaries. Cao, Yang, and Zhou (2005) used an energy function, which was based on a modified Mumford-Shah segmentation model for artificial objects detection from aerial images. They used a coarse-to-fine strategy and a fractal error metric at the evolutionary stage of the algorithm in order to detect urban or semi urban areas in aerial images. Karantzalos and Argialas (2009) also proposed a region-based level set segmentation method for artificial object detection in aerial or satellite images. A variational geometric level set function based on active contours without edge formulation was used and artificial objects such as roads and buildings were detected. These methods in some cases are vulnerable to misleading information derived from the presence of a large scale of intensities, shadows or occlusions, which is a common scenario in satellite images. Several studies have been reported to address the above issues by incorporating prior building shape information to label the segmented regions as buildings (Karantzalos and Paragios 2009). These schemes depend on the efficiency of the pre-defined building shape templates. A different approach can be seen in Peng, Zhang, and Liu (2005) study, where an improved active contour model is proposed by modifying the traditional snake to be attracted from the radiometric and geometric building features by customizing the snake energy function. In a similar approach, Ahmadi et al. (2010) design an innovative model based on active contours for extraction of building boundaries from high-resolution aerial images where the radiometric properties of the building classes were predefined.

The CV model has been used extensively in literature for the detection of buildings. Despite the good results, it is associated with several limitations such as extensive initialization, under-detection and over-detection. The model depends on initial curve placement and number. If the initial curves do not cover effectively, all building structures that may exist in the analysed scene, some of them may be missed. Nevertheless, if the initial curves over-cover the whole scene including objects or regions that do not consist building structures, then this may lead to erroneous or over-segmentations. The initial curve placement and number in the CV model also influences the computation cost. Very few initial curves lead to higher computation cost. Furthermore, spectral heterogeneity also may lead to misleading segmentation of building structures. A typical urban area in satellite images includes a variety of classes of buildings presented with different colour intensity values that makes it difficult for the traditional CV model to detect all buildings correctly. Several experiments have been performed in this work using the traditional CV model, with different initialization
Figure 1. Traditional CV segmentation method: (a), (b), (c), (d), (e) original RGB image, (i) initialization scheme using a series of regular circles all over the image, (iv) initialization scheme using a rectangle, (ii), (v) active contour building boundaries detection, (iii), (vi) segmentation mask.
schemes and sample results of heterogeneous and complex scenes are presented in Figure 1. All the above conclusions about the performance of the CV model are also in line with the findings of Liasis and Stavrou (2013), Ahmadi et al. (2010), Karantzalos and Paragios (2010) and Niu (2006) work.

The approach that this work follows to address the above limitations is based on designing an effective automated procedure for active contour initialization. Subsequently, the available information from colour features is utilized for enhancing

Figure 1. (Continued)
3. The development of an optimized active contour model to automatically detect buildings

3.1 Active contour initialization

The estimation of the optimal position of the initial curves can prove to be a difficult task without prior knowledge of the scene and the objects to be detected. In the urban building boundary extraction process using active contours that was proposed by Ahmadi et al. (2010), the initial curves were generated automatically as a series of regular circles all over the image. The number and size of the initial curves were experimentally estimated. Karantzalos and Paragios (2009, 2010) used an arbitrary elliptical curve to initialize their proposed curve evolution scheme and a data term based on prior knowledge of the analysed scene was incorporated to drive the evolving curve to the building boundaries. In the above publications, it was noticed that the initialization schemes influence the models speed of execution and accuracy.

A k-means clustering algorithm which aims to segment $n$ observations into $k$ partitions, in which, each observation is assigned to the nearest partition (with respect to partition centre) can be incorporated and pre-segment the image (Hartigan and Wong 1979). This information can be used to develop the initial curves of the level set segmentation procedure for addressing the initialization issue. Using k-means algorithm, along with morphological operations erosion and dilation for the refinement of k-means clustering result, the initial curves can be defined within or near the buildings that may exist in the image and the active contour model may lead to a better building segmentation result. The form and the parameters of a small shape or template called structuring element (SE), to be used in erosion and dilation, are important for the success of the process. Since the aim was to detect buildings, which in most of the cases are presented squares, rectangles or a combination of the above, a structuring element in the form of a rectangle is naturally more suitable. In this work, the length and width of the structuring element are estimated using statistical measures based on the objects that have been detected using k-means algorithm. After a number of experiments, it was found that the square root of the average length and width of the detected objects are optimum values for the structuring element size. Figure 9 presents the results of these experiments.

The k-means algorithm calculates data features and tries to find the natural clustering among them (Lloyd 1982). The input data points $s_i, i = 1 \ldots n$ where $n$ is the total number of points, are classified into multiple classes based on their inherent distance from each other. As has been noticed in literature this method is sensitive to initialization (Peña, Lozano, and Larrañaga 1999) thus the method so-called k-means++ proposed by Arthur and Vassilvitskii (2007) has been incorporated for choosing initial cluster centroid positions or seeds. The k-means++ algorithm uses an heuristic to find centroid seeds for k-means clustering. According to Arthur and Vassilvitskii, k-means++ improves the running time of Lloyd’s algorithm, and the quality of the final solution. All the points
are clustered around centroids or mean points $m_i (m_i \forall = 1 \ldots k)$ that are obtained by minimizing the following objective function:

$$\sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - m_i)^2$$

(2)

where $k$ is the number of clusters.

The algorithm steps of the proposed pre-segmentation method for designing the initial curves are as follows:

Step 1: Use the appropriate single band (saturation or value) representation of a given image and compute the intensity histogram.

Step 2: The parameters such as the number of iterations and the number of the different clusters are defined.

Number of iterations = 10, Number of clusters = 2. The number of iterations was calculated experimentally and kept static for all the under evaluation images. Two classes are defined in order to represent as foreground objects all the buildings and as background objects all the rest.

Step 3: Initialize the $k$ centroids using the $k$-means ++ algorithm.

Step 4: Repeat the following steps until the cluster labels of the image become static:

4.1 Cluster the points based on the distance of their intensities from the centroid intensities

$$c' = \text{argmin}_j |x' - m_j|^2$$

(3)

4.2 Compute the new centroid for each cluster

$$m_j = \left( \sum_{i=1}^{k} 1(c_i = j)x' \right) / \left( \sum_{i=1}^{k} 1(c_i = j) \right)$$

(4)

where $j$ and $i$ iterates over all the centroids and all the intensities respectively, $m_i$ are the centroid intensities and $k$ is the number of clusters.

Step 5: Apply morphological operations on foreground objects using a structuring element in the form of a rectangle.

5.1 Compute the number of connected components or objects $(n)$, their length $X_i$ for $i = 1 \ldots n$ and width $Y_i$ for $i = 1 \ldots n$.

5.2 Use the length and width of the detected objects for the calculation of the structuring element (SE) length and width denoted as $X_{SE}$ and $Y_{SE}$ respectively, based on the following functions:

$$X_{SE} = \sqrt{\text{average}(X_i)}, \quad Y_{SE} = \sqrt{\text{average}(Y_i)}$$

(5)

5.3 Apply erosion and then dilation using the same SE.

Step 6: Apply boundary detection for developing the final contours to be used as the initial curves for the active contour segmentation model.

Figure 2 presents sample results of initial curves implementation using the proposed method.
3.2 RGB and HSV colour space features

In most cases, building structures in satellite images are presented with various colour intensity values and different classes of buildings are created in the grey scale representation of the image. Thus, some buildings may be missed or detected incorrectly by the active contour CV formulation segmentation model even encoded with a well-defined initialization process. During this work, it has been observed that in those cases where a large scale of colour intensity values exist, the buildings presented with mean intensity values lower than the majority of the existing buildings, are not usually accurately detected by the CV model. This was also observed by Ahmadi et al. (2010) in their study where it was noticed that the number of building and background classes which are formed by the variety of colour intensity values is an important factor in achieving good building detection results using the CV model. Obtained segmentation results using the traditional CV method, where buildings presented with lower mean intensity values than the majority of the existing buildings are not detected, are shown in Figure 1.

This limitation can be overcome using features from the different colour spaces such as the RGB and HSV. Concrete, asphalt and white surfaces are low saturation areas while colour surfaces are high saturation areas (Arévalo, González, and Ambrosio 2008). After the initial experiments performed in this work, it has been observed that the saturation image, captures well building structures presented with colours such as red, brown and blue. The value image seems to capture well the buildings with high colour intensity values in all three RGB bands. The above observations are utilized in this work for simplifying the capture scenes and the regions of interest in satellite images become more homogeneous. Figure 3 presents, the value and the saturation representation of Google Earth images. As can be seen buildings presented with red, brown or very similar colour roofs that are common in urban areas are highlighted in saturation images.
However, it should be also noticed that vegetation regions and shadows are also highlighted in the saturation image and this may affect the accuracy of the building feature detection. Thus, the applied method processes the satellite image by applying a
colour clustering technique to eliminate vegetation regions and shadows that are typically classified as areas with low illumination in several scenes and can adversely affect the building extraction process. To eliminate these regions, indices and information obtained from the visible bands are used. The relationships between the red, green and blue bands in these areas are utilized and the pixel colour intensity values of shadows and vegetation regions are identified and subsequently eliminated using the \textit{k-means} clustering algorithm as described in \textbf{Section 3.1} and thresholding. Vegetated regions are classified based on their relatively high intensity value in the green band while shadows are classified based on their low intensity values in all three bands. After the iterative procedure using Equations (3) and (4) applied on the saturation component of the image, the detected regions are labelled as vegetation or shadow areas and are eliminated if the following criterion functions are satisfied, respectively:

\[ c^R < c^G, c^B < c^G \text{ and } c^G < T_V \]  \hspace{1cm} (6)

\[ c^R \approx c^G \approx c^B \text{ and } c^R, c^G, c^B < T_S \]  \hspace{1cm} (7)

where \( c^R, c^G, c^B \) are the intensity values for the RGB bands and \( T_V, T_S \) are threshold values defined experimentally. \textbf{Figure 3} shows sample results of the saturation images after the application of the clustering procedure based on \textit{k-means} and thresholding for eliminating the shadows and vegetation areas.

\textbf{3.3 Optimized active contour segmentation model}

The piecewise constant model of CV as was defined in Equation (1) is extended in this work for partitioning a given image \((I)\) with a variable closed curve \((C)\) for the detection of buildings. The level set formulation \( C = \{(x, y)|\Phi(x, y) = 0\} \) is used to solve the energy minimization problem following the principles as presented by Equation (1). A new energy term is incorporated utilizing the intensity descriptors of the value \((I)\) and saturation \((S)\) representation of a given image after eliminating shadow and vegetation regions using the proposed scheme as presented in \textbf{Section 3.2}. The proposed active contour model is described by the following equation:

\[
E_{OAC}(c_1, c_2, d_1, d_2, \Phi) =

\lambda_1 \int_{\Omega} (I(x, y) - c_1)^2 H(\Phi) \, dx \, dy + \lambda_2 \int_{\Omega} (I(x, y) - c_2)^2 (1 - H(\Phi)) \, dx \, dy

\]  

\[
+ \lambda_1 \int_{\Omega} (S(x, y) - d_1)^2 H(\Phi) \, dx \, dy

\]  

\[
+ \lambda_2 \int_{\Omega} (S(x, y) - d_2)^2 (1 - H(\Phi)) \, dx \, dy

\]  

\[
+ \mu \int_{\Omega} \delta(\Phi(x, y))|\nabla(\Phi(x, y))| \, dx \, dy

\]  

where \( \lambda_1, \lambda_2, \mu \) are the weights for the different terms and \( \delta(\Phi(x, y)) \) is the Dirac delta function.
where $c_1, c_2$ are the intensity mean values of the value image inside and outside $C$, respectively and $d_1, d_2$ are the intensity mean values of saturation image inside and outside $C$, respectively. $H(\Phi)$ and $\delta(\Phi)$ are the Heaviside and Dirac functions. These functions are estimated using the following equations:

\[
H_\varepsilon(z) = \frac{1}{2} \left( 1 + \frac{1}{\pi} \arctan\left( \frac{z}{\varepsilon} \right) \right), \quad \delta_\varepsilon(z) = \frac{1}{\pi \varepsilon^2 + z^2}, \quad z \in \mathbb{R}
\]  

Keeping $\Phi(x)$ fixed and minimizing the CV energy with respect to the constants $c_1, c_2, d_1, d_2$ then these constants are estimated as follows:

\[
c_1(\Phi) = \frac{\int_{\Omega} I(x,y)H(\Phi(t,x,y))dx\,dy}{\int_{\Omega} H(\Phi(t,x,y))dx\,dy}, \quad c_2(\Phi) = \frac{\int_{\Omega} I(x,y)(1 - H(\Phi(t,x,y)))dx\,dy}{\int_{\Omega} (1 - H(\Phi(t,x,y)))dx\,dy}
\]

\[
d_1(\Phi) = \frac{\int_{\Omega} S(x,y)H(\Phi(t,x,y))dx\,dy}{\int_{\Omega} H(\Phi(t,x,y))dx\,dy}, \quad d_2(\Phi) = \frac{\int_{\Omega} S(x,y)(1 - H(\Phi(t,x,y)))dx\,dy}{\int_{\Omega} (1 - H(\Phi(t,x,y)))dx\,dy}
\]

Subsequently, keeping $c_1, c_2, d_1, d_2$ fixed and minimizing the CV energy with respect to $\Phi(x)$ where the Euler–Lagrange equation is used to represent $\Phi(x)$, the following level set formulation model is derived:

\[
\frac{\partial \Phi(x,y,t)}{\partial t} = \delta_\varepsilon(\Phi) \left[ -\lambda_1 \left( (1 - c_1)^2 + (S - d_1)^2 \right) + \lambda_2 \left( (1 - c_2)^2 + (S - d_2)^2 \right) \right] + \mu \text{div} \left( \frac{\nabla \Phi}{|\nabla \Phi|} \right) - \nu
\]

where div is the divergence operator.

The algorithm steps of the active contour segmentation method are as follows:

Step 1: The initial contour $C$ is implemented and placed on the given image by initializing the level set function using the following equations:

\[
\Phi(x,y) = C, \quad \Phi(x,y) = 1_{\text{inside} C}, \quad \Phi(x,y) = -1_{\text{outside} C}
\]

Step 2: Set the parameters for the level set formulation and energy minimization:

Number of iterations $n = 30 - 60$, $\mu = 1$, Time step $dt = 0.1$

\[
\nu = 10^{-3}255^2, \lambda_1 = \lambda_2 = 1, \varepsilon = 1
\]

The above settings are defined based on the best practice as reported in literature for natural images (Chan and Vese 2001; Vese and Chan 2002; Chan and Zhu 2003; Karantzalos and Argialas 2009; Peng, Zhang, and Liu 2005; Liapis and Stavrou 2013; Ahmadi et al. 2010) and kept the same for all the under evaluation images.
Step 4: Evolve the level set function $\Phi$ according to Equation (11). At each time step, the level set function is reinitialized to be the signed distance function to its zero level set curves (Sussman, Smereka, and Osher 1993) and the constants $c_1, c_2, d_1, d_2$ are updated according to equations as presented in 10.

Step 5: Extract the zero level set from $\Phi_t(x, y)$ if the level set evolution terminates.

Step 6: Spatial morphological operations using a structuring element are applied on the segmentation mask. The process aims to remove small areas that are not part of a building, as well to create better-structured buildings. The same approach as was discussed in Section 3.1 for designing the morphological filter and specify its parameters is followed.

The suggested optimized active contour segmentation algorithm for building structures detection is presented in Figure 4 and all the images derived from the different processing steps are shown in Figure 5.

![Segmentation algorithm flowchart.](image-url)
4. Experimental results

In this work, a large number of experiments have been performed in order to assess the performance of the algorithm. In addition, the CV traditional model has been implemented and evaluated against the proposed method.

4.1 Image datasets

The algorithm is evaluated on 96 randomly chosen satellite Google Earth and/or Google maps images (2013–2015) from various countries and area types. The spatial resolution of the selected images varies from 1 to 10 m. All the tested images are 8-bit images with pixel values range from 0 to 255 and contain three bands (RGB). They were selected to represent diverse building attributes such as size, shape and colour in challenging environmental conditions from different countries. The images that were used contained 5063 buildings in total. Three different area types, high-, semi- and low-density urban regions, are defined by considering the land use along with the density of the existing buildings. For the classification of the different area types, contextual information such as the density, building size and distance between buildings were utilized to quantify urban morphological patterns (Baltsavias 2004, Chen et al. 2013).

4.2 Accuracy assessment strategy

Quantitative object or pixel level measures based on the work of Shufelt (1999), Ok, Senaras, and Yuksel (2013) and Ghaffarian and Ghaffarian (2014b) are used for the performance evaluation. For defining the quantitative measures, the buildings were delineated manually by human operators in order to create reference or ground truth (GT) images. If both the manual and the evaluation method detected a pixel as building,
then the pixel is denoted as a true positive (TP). If both the manual and the under evaluation method detected a pixel as a non-building, then the pixel is denoted as a true negative (TN). If the under evaluation method incorrectly presents a pixel as building while the ground truth image presents it as a non-building, then the pixel is denoted as a false positive (FP). Finally, if the under evaluation method incorrectly detects a pixel as a non-building while the ground truth image presents it as a building, then the pixel is denoted as a false negative (FN). The following well-known pixel based statistical metrics, precision \( (p) \), recall \( (r) \) and weighted harmonic mean score \( F \) of precision and recall are computed to evaluate the performance of the proposed method:

\[
p = \frac{TP}{TP + FP} \times 100, \quad r = \frac{TP}{TP + FN} \times 100
\]

\[
F = \left(1 + \frac{B^2}{2 \times p \times r}ight) \times 100
\]

where \(|\cdot|\) denotes the number of pixels or objects assigned to each distinct category and \(B\) is a non-negative real constant parameter to trade-off the precision and recall.

The precision metric gives the percentage of the true detected pixels against the true positive pixels along with the misclassified pixels as buildings. The recall metric, where the denominator stands for the total number of true building pixels, is the building detection accuracy and indicates the percentage of building pixels correctly detected by the evaluation algorithm. The weighted harmonic mean \( (F) \) of precision and recall values is an important measure since it combines these two metrics together into a single one and can be though as a quality metric. Metrics, precision and recall are equally important in a detection of building structures process, thus specifying the \(B\) parameter equal to one is the most appropriate. The same metrics are applied to the object-based assessment, where a building object is identified as TP if at least some percentage of its pixels is correctly detected. Subsequently, if a detected object corresponds to less than some percentage of an object in the ground truth image, then is denoted as FN. Finally, a resulted object is identified as FP if it does not coincide with any of the building objects in the ground truth image.

### 4.3 Results and discussion

The proposed model has been implemented by Matlab R2013b on a computer with Intel Core i5 2.6 GHz CPU, 4 GB RAM and Windows 7 64 bit operating system. The traditional CV model has been also implemented for comparing its performance against the proposed model. Table 1 presents the object-based performance of initial curves coverage. Figures 6 and 7, Tables 2 and 3 illustrate the visual and quantitative object- and pixel-based building structures detection evaluation results. The time needed by the

<table>
<thead>
<tr>
<th>Urban area type</th>
<th>Total number of buildings</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>( F ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High density</td>
<td>2898</td>
<td>95.31</td>
<td>98.39</td>
<td>96.83</td>
</tr>
<tr>
<td>Semi density</td>
<td>1728</td>
<td>93.22</td>
<td>94.83</td>
<td>94.02</td>
</tr>
<tr>
<td>Low density</td>
<td>437</td>
<td>91.43</td>
<td>94.12</td>
<td>92.75</td>
</tr>
<tr>
<td>Overall</td>
<td>5063</td>
<td>93.86</td>
<td>96.40</td>
<td>95.11</td>
</tr>
</tbody>
</table>
traditional active contour and the proposed model to detect all the corresponding buildings per image were recorded. The average time in seconds using all the tested images is presented in Table 4.

In several cases, as can be seen in Figure 1, where multiple buildings with different colour intensity values exist in satellite images, darker than the average building roofs, are not usually detected using the traditional level set CV segmentation method applied in the grey-scale representation of the image. In this work, it was observed that a saturation image is able to highlight building features that are presented with at least a high-intensity colour value in one of the RGB bands. This was expected since the pure colour surfaces have high saturation representation (Tsai 2006, Arévalo, González, and Ambrosio 2008). However, in the case of vegetation regions and/or shadows, these were also highlighted in several scenes on the saturation representation. By utilizing the intensity properties of the saturation representation of the image after eliminating the vegetation areas and shadows in the proposed active contour segmentation model that also encompass a well-defined initialization process, several buildings that were missed by the traditional CV active contour model were detected and extracted in most of the examined cases. In this work, along with the general initialization $k$-means ++ scheme that has been utilized, additional steps such as threshold and morphological filters have been incorporated for refining further the results of the $k$-means clustering method. The threshold values $T_V$ and $T_S$ of the two criterion functions as presented in Equations (6) and (7) have been estimated using a training process. Samples of shadow and
Figure 7. Building detection in satellite images. (a), (c), (e), (g), (i), (k), (m), (o), (q), (s), (w) Original images, (b), (d), (f), (h), (j), (l), (n), (p), (r), (t), (v) building segmentation masks using the propose method.
vegetation regions were used to define the range of the colour intensity values. It has been observed that these values vary for different capture areas. However, the estimation of the most appropriate values is not a very difficult task since in Google Earth images, large areas are presented with similar radiometric characteristics. Thus, the range of the threshold values can be estimated only once using a small number of representative images for each area. The median value of the observed range for each area is calculated and used as threshold value in the corresponding area. As can be seen in Figure 3, the proposed clustering method for eliminating shadows and vegetation regions works well in various scenes. Furthermore, buildings with green or variations of green roofs are not affected when the elimination of vegetation and shadow regions process is applied (Figure 3, Image a). This can be explained since the radiometric characteristics of green artificial objects as presented in satellite images are different from the radiometric characteristics of vegetation regions. Nevertheless, some buildings are not detected when they have very similar radiometric characteristics with shadows. These types of building structures are expected to be eliminated as shadows. However, this drawback is limited since only those buildings that entirely eliminated as shadows are finally missed. The visual interpolation of all figures suggests that this type of building roofs is very rare in the evaluation scenes. As can be seen in Figure 6 Image (k), building structures with similar radiometric characteristics with shadows are successfully detected if a small region of the corresponding building roofs still exists in the cluster image. The proposed active contour model achieves to detect such building structures utilising more iterations in the curve propagation process ($n \approx 60$), since initial curves are created for these buildings by the proposed initialization scheme.

The segmentation results of the presented method demonstrate that the building detection algorithm works well to satellite images of urban areas. Results presented in

| Table 2. Object-based building structures detection evaluation using 60% overlap threshold. |
|---------------------------------|----------|----------|----------|----------|----------|----------|
| Urban area type                  | Total number of buildings | CV      | OAC     | CV      | OAC     | CV      | OAC     |
| High density                     | 2898     | 78.72    | 95.06   | 71.60   | 90.37   | 74.99   | 92.66   |
| Semi density                     | 1728     | 68.60    | 90.39   | 69.27   | 89.81   | 68.93   | 90.10   |
| Low density                      | 437      | 67.56    | 88.24   | 69.11   | 85.81   | 68.33   | 87.01   |
| Overall                          | 5063     | 74.03    | 92.83   | 70.59   | 89.79   | 72.27   | 91.29   |

| Table 3. Pixel-based building structures detection evaluation. |
|---------------------------------|----------|----------|----------|----------|
| Dense urban area type           | Total number of images | Precision (%) | Recall (%) | F (%) |
| High density                    | 32       | 92.94    | 88.12    | 90.47   |
| Semi density                    | 32       | 88.05    | 86.02    | 87.02   |
| Low density                     | 32       | 82.33    | 84.42    | 83.36   |
| Average                         | 32       | 87.77    | 86.19    | 86.95   |

| Table 4. Computation cost evaluation. |
|---------------------------------|----------|
| Method                          | Computation cost average (s) |
| CV                              | 17.42    |
| OAC                             | 12.96    |
all figures suggest that the proposed methodology can detect almost all the buildings and their basic structure and shape, even when these buildings have irregular shapes, interact with environmental objects and occlude each other. In addition, the proposed method discards most of the non-building areas with similar building characteristics and any misleading artefacts. Applying optimized morphological post-processing functions ensure that the building shapes are retained.

The algorithm’s overall per object segmentation results, reach a building detection accuracy or recall of 89.79%, precision 92.83% and a radio $F$ equal to 91.29% as seen in Table 2. The values in Table 2 correspond to 60% overlapping threshold. This value has been selected following the best practice used in other studies in literature (Ghaffarian and Ghaffarian 2014b, Ok, Senaras, and Yuksel 2013). The algorithm performs well in all the predefined density urban area types. The pixel-based evaluation results as can be seen in Table 3 reveals also good and balance results, which are in line with the object-based results. The recall metric estimated at 86.19%, the precision at 87.77% and the $F$ score at 86.95%. All the results reveal that the performance of the algorithm decreases as building density decreases. This has been expected since active contour models based on CV formulation detect regions of interest that exhibit a certain similarity and homogeneity. In high dense urban areas, the building structures are presented with greater homogeneity and similarity. In addition, the results revealed that the performance of the proposed method is superior when compared against the traditional CV model. The computation cost is an important factor in the practical application of the method and has been proved that for the proposed method it is significantly lower than the CV model (Table 4). After the application of the proposed initialization scheme, less iterations compared to the traditional CV model are needed to detect the building structures since as can be seen in Figure 2, the initial contours are near the boundaries of the building structures. The proposed initialization scheme also enforces the active contours to avoid several misleading regions during the propagation process, leading to better segmentations. The coverage of the buildings as presented in the analysed scenes using the proposed initialization method has been thoroughly assessed. The aim of the initialization development curves process was to create curves of any form and structure within or near the building objects. Thus, in this assessment if any part of the initial curve coincides with a building in the ground truth image, it is considered as a TP. An initial curve is denoted as a FP if does not coincide with any of the building objects as presented in the ground truth image. A FN object is denoted a building object presented in the ground truth image whereas an initial curve has not been developed. As can be seen from Table 1, the overall recall accuracy has been calculated at 96.40%, the precision at 93.86% and the $F$ score at 95.11%.

Finally, it should also be noticed that if the image do not adhere to quality standards the resulting segmentation might suffer, as is the case with most of the building segmentation algorithms (Ghaffarian and Ghaffarian 2014b, Ok, Senaras, and Yuksel 2013; Karantzalos and Paragios 2010; Karantzalos and Argialas 2009; Peng, Zhang, and Liu 2005, Liasis and Stavrou 2013; Ahmadi et al. 2010). However, the algorithm has been shown to work well even in such cases, as is shown in Figure 7 and the presented results for the Images (e) and (h).

A limitation of the proposed method that should be noticed is that in high-density urban environments, some non-building objects such as bridges and parking places or
even parts of the roads are classified as buildings. This behaviour is expected and is noticed in other schemes proposed in literature (Sirmacek and Unsalan 2009; Ok 2013, Ghaffarian and Ghaffarian 2014b, Karantzalos and Paragios 2010), because of the similarity between these objects and the buildings. Finally, the results of this work revealed that in some cases where two or more buildings are extremely close to each other are classified as a single building. This influences the object based results as reported in Table 2 for the TPs buildings where two or more merged buildings based on the 60% coverage threshold are classified as two or more TPs even though one of them might has less than 60% TP pixels. This has been observed mainly in high-density areas and very few buildings less than 1.5% of all the existing buildings in all the tested images were affected. This percentage can be thought as the highest margin of error of the object-based results.

In this work extensive analysis of those parameters that might affect the building detection results was performed. One parameter, which is important for the object-based measures, is the object-overlapping threshold of 60%. Figure 8 presents this relationship and as can be seen, even for high overlap threshold such as 80% and more the proposed method achieves satisfactory detection performance where the accuracy of precision, recall and $F$ score estimated near 90%, 85% and 87%, respectively.

Another important factor for the overall performance of the proposed method is the automated estimation of the structuring element size used in erosion and dilation morphological operations. It is expected that if the structuring element is too large or too small the results will be affected, leading to under-segmentation or over-segmentation respectively. Figure 9 presents how the pixel based performance metrics are affected when the size of the structuring element is increased or decreased by a
The detection of buildings from monocular satellite images using an optimized active contour model is proposed. Active contour models are highly dependent on the spectral homogeneity of the objects in an image and may often lead to erroneous segmentation of satellite images where spectral heterogeneity and complexity exist. The proposed method works efficiently and has the ability to detect buildings with arbitrary shapes, sizes and even with significantly different coloured rooftops. It addresses limitations of the traditional active contour without edges model for buildings detection and optimizes the overall segmentation procedure. A colour clustering method has been developed and used to remove the vegetation areas and shadows resulting in less complex scenes. The image is decomposed in HSV colour components and the intensity properties of the saturation and value representations are used to develop the optimized active contour model. An additional energy term has been encoded for biasing the active contours to delineate several buildings that were represented with lower intensity values than the average of the buildings in the grey-scale representation of the image and often missed by the traditional active contour models. The proposed automated
active contour initialization process drives the contours to a better segmentation of the image faster avoiding misleading regions. Post-processing operations for retaining the building’s structure are also applied. The estimation of the parameters used in the morphological erosion and dilation operations proved to be appropriate for retaining the shape of the detected buildings and remove any remaining small misleading artefacts. Extensive experiments were performed on a large number of images and the produced segmentation results for heterogeneous and complex satellite images verify the efficiency of the suggested method. The proposed method achieves better accuracy when compared to traditional active contour without edges segmentation model. Furthermore, the method is time efficient, there is no dependency on any additional data and user interaction is not needed.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Gregoris Liasis  http://orcid.org/0000-0002-7734-8385

References


学霸图书馆

www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具