Detecting the number of buildings in a single high-resolution SAR image

Yongfeng Cao1*, Caixia Su1 and Guangbin Yang2

1School of Mathematics and Computer Science, Guizhou Normal University, 550001 Guiyang, China
2School of Geography and Environmental Science, Guizhou Normal University, 550001 Guiyang, China

*Corresponding author, e-mail address: yongfengcao.cyf@gmail.com

Abstract
This paper proposes an effective scheme for detecting the number of buildings in a scene from a single high-resolution Synthetic Aperture Radar (SAR) image. The layover and double bounce echoes in SAR images are detected first as building elements, which are then split, merged, or discarded to make each patch correspond to one building. A model describing the statistical relationship between the number of buildings and the features of the detected building elements is constructed. Based on this model, large building patches are split into the proper number of small patches. This scheme is tested on 3-m and 6-m resolution TerraSAR-X images that cover two sites in different provinces of China, and its advantages and limitations are discussed.

Keywords: Number of buildings; high-resolution SAR image; building detection.

Introduction
High-resolution Synthetic Aperture Radar (SAR) sensors, which deliver images with metre or sub-metre resolution, have made it possible to extract and analyse detailed urban information [Soergel et al., 2006; Michaelsen et al., 2010]. In the metre-level geometric resolution SAR data collected by modern space-borne sensors such as TerraSAR-X and Cosmo-SkyMed, the geometric extent of individual objects such as bridges, buildings, and roads becomes visible.

This paper addresses detecting the number of buildings in a single high-resolution SAR image. This information has promising applications in many fields, such as urban and rural planning, natural-resource protection, disease spread and population distribution, and environment evaluation. It is also now possible to analyse the dynamic development of the number of buildings in a specific region with high spatial and temporal resolution SAR images. These space-time patterns can be used to understand the effects of landscape change and to inform new policies (such as the reformulation of policies concerning home development, water quality, and biodiversity sustainability) to guide future growth.

Although building footprint detection and 3-D reconstruction using high-resolution SAR
images has recently become an active research topic, few studies have focussed on detecting the number of buildings from SAR images.

Related studies
This section introduces some examples of SAR image-based building information extraction that are relevant to the work of extracting the number of buildings that is described in this paper.
Some studies distinguish town/city areas from others or further classify town/city structures into several types. None of these studies can neglect the features of buildings as the main elements of a town/city. Gouinaud and Tupin [1996] successfully detected urban areas from ERS-1 SAR images (with a spatial resolution of 30 m) based on a statistical method called ffmax-filter, which considers the global backscattering law of urban areas as a result of the composite effects of three types of surfaces (smooth surfaces, rough surfaces, and conductor surfaces). Additionally, based on ERS-1 SAR images (here including multi-temporal images), Dell’Acqua and Gamba [2003] refined urban areas into three subclasses (city centre, residential areas, and suburban areas) by combining textural features and a neuro-fuzzy classifier. From Radarsat-1 and ALOS PALSAR images (all with a spatial resolution of 6.25 m), Stasolla and Gamba [2008] distinguished built-up areas from the background by using local spatial autocorrelation features and the morphological method. The main assumption in this work is that built-up areas can be considered as agglomerates of hot spots, which are the bright responses created by buildings owing to multiple reflections and double-bounce reflectors.
Some studies focus on detecting the areas that specifically correspond to buildings. Tison et al. [2004] chose the Fisher distribution as the statistical model for high-resolution SAR images of urban areas and used this model in a Markovian classification method to distinguish six different classes: ground, dark vegetation, bright vegetation, dark roof, medium roof, and bright roof/strong reflector. In this work, an X-band interferometric airborne SAR image (with a resolution of 0.37 m in range and 0.18 m in azimuth) was used. To detect building areas, He and Hellwich [2009] used Bayesian network modelling of the interactions among surface evidence, building evidence, and orientation angles. In their study, fully polarimetric airborne SAR data (with a resolution of 1.499 m in range and 0.748 m in azimuth) were used. Wegner et al. [2011] detected building areas using a combination of orthophoto (with a spatial resolution of 0.31 m) and high-resolution InSAR data (with a resolution of 0.38 m in range and 0.18 m in azimuth) with a conditional random field (CRF) framework. Reliable SAR features (building corner lines, and the InSAR phase distribution along the corner lines) were used.
Recently, many researchers have attempted to reconstruct three-dimensional (3-D) information about buildings. Sportouche and Tupin [2010, 2011] reconstructed buildings by coupling high-resolution optical images (Quickbird images with a spatial resolution of 0.7 m) and SAR data (TerraSAR-X images with a spatial resolution of 1.1 m) using a three-step scheme (building detection, height estimation, and qualification). The building footprints in an SAR image are generated around the pairs of associated double bounce and shadow that fulfil directional and spatial constraints. Building height is estimated by optimising a statistical criterion that is based on the partitioning of the SAR image
into different regions (layover, single roof, and shadow) depending on the supposed height and a radiometric criterion. However, this method can treat only large, isolated, and rectangular buildings. Xu and Jin [2007] reconstructed 3-D buildings from multi-aspect SAR images (airborne Pi-SAR polarimetric images with metre resolution). In this work, building edges were first detected by a POL-CFAR detector. Then, a Hough transform and numerous post-processing steps were performed to derive parallelogram-like building images for all aspects. Finally, the maximum likelihood estimations of building objects were derived from these building images. This work considered only simple buildings with rectangular shapes. Thiele et al. [2007] reconstructed small and large buildings from multi-aspect InSAR images (with a spatial resolution of 0.3 m) from two orthogonal flight directions. For small-building detection (i.e., buildings with a maximum extent [length × width × height] of 8 m × 8 m × 4 m), only the frequently observed lines of bright double-bounce scattering are used. For large-building detection (with a minimum extension of 20 m × 20 m × 4 m), all the building information (layovers, corner lines, roof signatures, and shadows) is exploited. This work also deals only with rectangular buildings. Simonetto et al. [2005] automatically extracted 3-D buildings from stereoscopic high-resolution images (with decimetre spatial resolution) recorded by the SAR airborne RAMSES sensor. The structural information from strong echoes generally produced from dihedral corners between ground and buildings was used for building footprint detection. This work restricted itself to treating only large rectangular buildings. Poulain et al. [2011] created and updated building objects in a cartographic database using high-resolution SAR (TerraSAR-X images with 1 m resolution) and optical images (with resolutions from 0.6 m to 2.5 m). The features used in this method were mostly from optical images, and the only feature from an SAR image was the contrast, which is defined as the ratio of the means of the layover and shadow regions. Ferro et al. [2013] detected and reconstructed building radar footprints from single very high-resolution SAR images (TerraSAR-X images taken in the Spotlight mode with 1 m resolution). Layover, double-bounce, and shadow effects were all used to detect primitives and generate a radar footprint hypothesis. The concepts of semantic meaning and membership grade were introduced for each primitive and footprint hypothesis and used to calculate a score for selecting the most reliable hypothesis.

From the analysis of previous studies, several useful conclusions can be derived. For town/city/built-up area detection, SAR images with coarse, medium, and fine spatial resolution are all useful. For the detection of building areas, at least metre-level-resolution SAR images should be used, and most studies rely on height measurements, shadow analysis, and bright-pixel detection. These approaches for detecting building areas may be used as the preprocessing step for detecting the number of buildings. Most approaches to the reconstruction of a 3-D building share some common properties. First, they use at least metre-level-resolution SAR images. Second, they fuse information from multiple images. Third, they address only simple, large, and rectangular buildings. Fourth, most methods are restricted to treating isolated buildings or areas with sparsely distributed buildings. For effectively detecting the number of buildings, the requirements of the first two points are excessive, and the limitations of the last two points will cause poor building-detection performance for most urban areas, which are densely packed with buildings.
Detecting the number of buildings

The main contribution of this paper is to propose a scheme for detecting the number of buildings from a single high-resolution SAR image. This work differs from the studies mentioned in the previous section in several ways. First, it focuses on the number of buildings, not the position, footprint, size, or height of these buildings. Second, instead of multi-source image data (such as multi-aspect SAR, PolSAR, InSAR, or a combination of optical and SAR images), a single high-resolution SAR image is used. It will be demonstrated that satisfactory information accuracy can be achieved from a single high-resolution SAR image. Clearly, using multi-source images will improve accuracy, but this approach may be more expensive than using a single image, which is the most easily available data format. Third, this scheme uses 3-metre-level spatial resolution data, which are available from most commercial systems, such as TerraSAR-X, Cosmo-SkyMed, and Radarsat-2. It can be observed from the real SAR images in Figure 1 that a coarser resolution causes multiple buildings to merge into one bright block, whereas finer resolution causes more small bright spots and lines from one building; thus, resolution in the range 3 to 6 metres is a useful compromise.

The scheme for detecting the number of buildings consists of the following four steps (see Fig. 2):

- Step 1: Building detection. This step detects building pixels and attempts to merge pixels that belong to a single building and to split pixels that belong to different buildings;
- Step 2: Modelling the number of buildings. This step models the relationship between the features of a detected building region and the real number of buildings in it;
- Step 3: Splitting large bright patches. This step cuts each large bright patch detected in Step 1 into small patches of the number estimated by the model described in Step 2;
- Step 4: Calculating the number of buildings. A method for calculating the number of buildings in an arbitrary region is given. A method to obtain another type of information,
Building detection

Because the proposed approach must be generic, it is necessary to find features that are common to most types of buildings and insensitive to imaging parameters such as the imaging aspect and incidence angle. According to the standard SAR-imaging mechanism, the building area in an SAR image can be divided into several subparts: layover, double bounce, roof, and shadow. Based on our data, the most reliable structures that can be extracted from an SAR image are the strong echoes (layover and double bounce). The information in these strong echoes has been widely used for urban analysis. Layovers are always adjacent to double bounces; thus, we combine them as a feature for building detection. Shadows can also appear often and are of great importance. However, their presence depends on the noise-equivalent radar cross-section of the sensor and the backscattering level of the background [Simonetto et al., 2005]. The interference between adjacent buildings, especially in dense urban scenarios, may cause some shadows to be missing. Roofs do not appear very different from the ground, and sometimes the appearance of two building roofs can be completely different because of their different geometric structures, materials, and roughness [Simonetto et al., 2005; Thiele et al., 2007].

Bright building pixels that correspond to the layover and double bounce can be obtained using a classification method such as a threshold. However, a high threshold will cause several isolated regions to be detected from one building, and a low threshold will connect freestanding buildings. It is difficult to determine a threshold that is optimal for all building regions because of the nonlinear relationship between the double-bounce
response, the building orientation angle, and complex surroundings [Ferro et al., 2011].

In this section, a method that is mainly based on morphological operators is proposed for detecting building regions in high-resolution SAR intensity images. The goal of this method is that one detected region corresponds to one building. Our method consists of the following three sequential steps:

- Step 1: Thresholding with hysteresis: this process requires two thresholds, high and low. We begin by applying a high threshold, \( T_{\text{high}} \). This threshold identifies the building points for which we can be confident that most are genuine. (Other objects, such as cars, street lampposts, and trees may also cause high bright points. Some of these false positives will be eliminated in Step 3.) Starting from these points, the building area can be expanded through the image. While expanding, we apply the lower threshold \( T_{\text{low}} \), thereby allowing us to find the faint portions of buildings as long as we have a starting point. This process is similar to part of the work of the famous Canny edge detector [Canny, 1986]. We accomplish this process using a morphological reconstruction operator [Gonzalez and Woods, 2007] with the high-threshold result as the mark image and the low-threshold result as the mask image;

- Step 2: Closing and opening: closing is designed to further connect previously detected building points that belong to a single building and to remove small background holes. Opening is designed to disconnect buildings that are misconnected by the previous steps and to make the shape of these detected bright patches more regular;

- Step 3: Area thresholding: the above processes prepare a label image in which the bright patches are labelled as buildings. Now it is necessary to eliminate some false positives. It was determined through observing our high-resolution SAR image scenes that most bright patches caused by cars, street lamps, isolated trees, and other man-made objects such as statues have a smaller area than buildings. To remove these unreasonably small patches, we compare the area of the patches to a threshold \( T_{\text{area}} \), which is an empirical limit that corresponds to the minimum size of a building.

**Modelling the number of buildings**

Using a previously developed building-detection method that is based on thresholding and mathematical morphology, we can, to a certain extent, reduce false connections (i.e., bright patches that are connected but belong to different buildings) and false disconnections (i.e., a bright patch that denotes a single building but is divided into several isolated patches). However, the false-connection phenomenon is still common in many regions in which buildings are located very close to one another or where the buildings are very tall. Therefore, it is inaccurate to directly use the number of bright patches in a region as the number of buildings. By analysing the bright-patch samples, we find an obvious correlation between the features and the real number of buildings in the bright patches. In this section, we use this helpful correlation to obtain more-accurate information regarding the number of buildings in a bright patch by constructing a linear regression model of the dependent (outcome) variable \( y \) (the number of buildings in a bright patch) and the predictor variables \( x_i, i = 1, \ldots, n \) (features of the bright patch). Using the constructed regression model, we can calculate \( y \) for every bright patch as long as we have obtained the \( x_i \)'s of the patch.
Description of the model
A linear regression function is chosen to model the correlation between the features and the real number of buildings in bright patches. According to regression analysis theory [Freedman, 2005], our model, which includes k predictor variables, is as follows:

\[ y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_ix_i + \ldots + a_kx_k + e \]  

where \( y \) is the dependent variable (in our case, \( y \) is defined as the number of buildings), \( x_i \) are the predictor variables (in our case, they are the features of the bright patch), \( a_0, a_1, \ldots, a_i, \ldots, a_k \) are model parameters, and \( e \) is a random error, which is assumed to obey a normal distribution conditioned on dependent variable. Sufficient sample data (at least \( k+1 \), but for the sake of reliability, 5–10 times the number of predictor variables would be better), \( \{y, x_1, \ldots, x_i, \ldots, x_k\}_j \), \( j=1,\ldots,n \) are needed to estimate the parameters of the regression model. We obtain these sample data by selecting bright patches from the results of the building detection step, computing the region features, and counting the freestanding buildings in these regions by referring to Google Earth maps. Using these sample data, model parameters can be estimated by the ordinary least squares method. It should be noted that the output of the regression model is decimal, whereas the number of buildings in a patch is an integer. Thus, rounding is usually required when we use this information.

Features and modelling process
Some accessible features that correlate with the number of buildings in a bright patch should be chosen for modelling. According to our data, certain shape-features of a bright patch meet the conditions. These features are as follows:
- **Area** – the number of pixels in the bright patch;
- **MinorAxisLength** – the length (in pixels) of the minor axis of the ellipse with the same normalised second central moments as the patch;
- **MajorAxisLength** – the length (in pixels) of the major axis of the ellipse with the same normalised second central moments as the patch;
- **Solidity** – the scalar that specifies the proportion of the pixels in the convex hull that are also in the patch;
- **Eccentricity** – the eccentricity of the ellipse with the same second-moments as the patch. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1, where 0 corresponds to a circle region and 1 to a line segment;
- **Compactness** – the ratio of the square of the perimeter to the area of the patch.

If we directly use all of these features for modelling, there will be a multicollinearity problem [Schroeder, 1990] because some features may be correlated with each other. In the presence of serious multicollinearity, regression estimates are unstable, and the possibility of overfitting the data increases. The best regression models are those in which each of the predictor variables correlates highly with the dependent (outcome) variable but minimally with the other predictor variables. Because multicollinearity diagnostics are easily obtained, the results of regressions with strong multicollinearity problems should never be reported. Factor analysis or principal component analysis can be used to produce new independent
variables from these correlated variables, but these new variables have no actual meaning. Here, we use the stepwise algorithm [Draper and Smith, 1981] provided by the mainstream statistical software SPSS (Statistical Product and Service Solutions) to construct our regression model. At each step, the predictor variable that is not in the equation and has the smallest probability of $F$ is entered if that probability is sufficiently small ($<0.05$). Variables that are already in the regression equation are removed if their probability of $F$ becomes sufficiently large ($>0.1$). The method terminates when no more variables are eligible for inclusion or removal. VIF (variance inflation factors) [Wooldridge, 2000] are used to examine multicollinearity (if any of the VIF are greater than 5, to be conservative). The goodness of fit of the constructed model is checked using the Adjusted R Square (ARS) [Harel, 2009]. The statistical significance [Mood et al., 1974] is assessed by performing an $F$-test of the overall fit, which is followed by $T$-tests of the individual parameters. This implementation uses the functions that are available in SPSS.

**Splitting large bright patches**

![Figure 3 - (a) A large bright patch; (b) distance image of (a); (c) marker image obtained by ultimate erosion (markers are denoted by white pixels); (d) marker image after selecting building-kernels; (e) final segmentation result.](image)

In this section, we propose a hybrid of morphological techniques to split each large bright patch into small patches of the number estimated using the model in the previous Section, “Modelling the number of buildings”, and attempt to make each small patch correspond to a freestanding building. This procedure makes it simple for the end user to extract the number of buildings in an arbitrary region (the detailed reason for this will be explained in the next Section). Additionally, this procedure is helpful for estimating the position of the actual building footprints, although such estimation is not the topic of our paper. We choose morphological operators instead of spatial clustering methods (such as K-Means and ISODATA) because of their advantage in the shape preservation of buildings. The ordinary watershed segmentation method [Beucher and Meye, 1993] should not be used here because it cannot exactly control the number of patches in the final segmentation.
result. Accordingly, we use a marker-controlled watershed transform. Several techniques, such as ultimate erosion, the distance transform, and the basin dynamic are used to seek the best markers. For the sake of convenience for description, we suppose that the input binary image we use contains only one large bright patch. This specific process consists of the following three steps:

-Step 1: Obtaining the building-kernel candidates. This step finds markers for the watershed transform using ultimate erosion. Because each marker, if used for the watershed transform, will obtain a small patch (which corresponds to a building) in the final segmentation result, we here call it a building-kernel candidate. Morphological erosion shrinks the foreground object (bright patches) by removing pixels on the object boundaries. The number of pixels removed from the object in an image depends on the size and shape of the structure element. We use a 3×3 unit matrix as the structure element of erosion. For a given large bright patch, erosion operations are performed repeatedly until no foreground pixels remain. During this process, some disconnected regions will appear and every region will finally disappear. All the regions that will disappear in the next erosion loop are called the ultimate erosion result. In our case, the ultimate erosion result (see Fig. 3c) is simply the building-kernel candidates.

-Step 2: Selecting building kernels. We first obtain the estimated number of buildings, N, in the processed large bright patch by using the model constructed in the Section “Modelling the number of buildings”. Then, we select, based on the basin dynamic criterion, N building kernels from the candidates obtained by the above step. If N is greater than the number of candidates, all of the candidates are selected as building kernels.

![Figure 4 - Basin dynamic.](image)

To calculate the basin dynamic for all candidates, a watershed transform should first be performed in a distance image, with all the candidates as markers. The distance image is obtained by a morphological distance transform on the input binary image. The distance transform calculates the distance between each pixel, which is set to 1, and the nearest 0 pixel for the binary image. Let Dis denote the distance image. We finally use a new distance
image nDis=max(Dis)-Dis (see Fig. 3b). The watershed transform on a distance image can be explained as a flooding process as follows: imagine the distance image as a terrain that consists of mountains (local maximum areas) and basins (local minimum areas). For the case with no markers, all basins connect with the same groundwater at their bottom points. For the case with markers, only the markers connect with the groundwater. The process, which is such that the groundwater level continuously ascends from the zero level, thereby flooding all the basins and then all the mountains, is called the flooding process. If we build a dam at each site where water from different basins/markers meets during the flooding process, then we obtain a partition of this terrain (a segmentation of the distance image) according to the final dams.

The basin dynamic was first proposed by Grimaud [1992] as a measure for the basin’s saliency and has achieved good results for choosing markers in watershed segmentation [Dougherry, 1992; Najman and Schmitt, 1994]. Suppose M is a basin in the terrain (that corresponds to the distance image nDis). Then, the basin dynamic of M can be defined as follows [Najman and Schmitt, 1994]:

\[
\min(\max_{s\in[0,1]}(f(\gamma(s)) - f(\gamma(0))) | \gamma : [0,1] \to \mathbb{R}^2, f(\gamma(1)) < f(\gamma(0)), \gamma(0) \in M) \tag{2}
\]

where \(\gamma\) is a path that links two points. In practice, we can define the basin dynamic of M as the minimum altitude difference of all paths that link the bottom point of basin M to the closest basin’s bottom point with a lower altitude, as illustrated in Figure 4. The lowest basin has no dynamic; thus, we give it the maximum value. During the water-flooding process on a terrain, the basin dynamic can be easily calculated as follows: let \(A()\) denote the altitude of some basin’s bottom point. When the water from different basins meets somewhere at water level \(h\), let \(\min A()\) denote the minimum \(A()\) of all these meeting basins. Then, the basin dynamic of the basin M that satisfies \(A(M) > \min A()\) can be calculated to be \(h - A(M)\), and the basin M affiliates itself with the basin that has \(\min A()\).

The specific process is as follows: let all candidates in Step 1 be markers, and perform the watershed transform on the distance image nDis. During the flooding process, all the basin dynamics of the candidates can be obtained using the method described above (i.e., let \(A()\) denote the altitude of some marker’s bottom point. When the water from different markers meets somewhere at water level \(h\), let \(\min A()\) denote the minimum \(A()\) of all these meeting markers. Then, the basin dynamic of the marker M that satisfies \(A(M) > \min A()\) can be calculated to be \(h - A(M)\), and the marker M affiliates itself with the marker that has \(\min A()\). We suppose that real building kernels always have a larger dynamic than the false positives and thus choose the first \(N\) candidates after sorting all candidates by their dynamics (see Fig. 3d). If \(N\) is greater than the number of candidates, all the candidates are selected.

It can be observed that the dynamic measure considers not only the absolute altitude of a basin but also the relative altitude of the basin compared with other lower basins around it. Its merit can be observed clearly from Figure 4, in which basin-4 and basin-5 are simply terrain undulations in a large basin, b45. Suppose we must split the whole terrain into three parts using the watershed transform. Based on the altitude measurements, basin-5, basin-4, and basin-1 should be selected as markers, and these markers will cause over-segmentation.
of b45. Based on the basin dynamics, basin-5, basin-1, and basin-6 should be selected as markers, and these markers will produce a more reasonable segmentation result.

-Step 3: Building-kernel-controlled watershed segmentation. The final segmentation of a large bright patch is achieved by performing a marker-controlled watershed transform on the distance image nDis, with the building-kernels chosen in Step 2 as markers. (Note: a marker that corresponds to the background should be added when performing the marker-controlled watershed segmentation). The large bright patch is divided into several small ones using the borders of the watershed segmentation (see Fig. 3e).

**Determining the number of buildings**

Let \( O1 \) denote the output map of the Section “Building detection”, let \( MD() \) denote the model derived in the Section “Modelling the number of buildings”, let \( O2 \) denote the output map of the Section “Splitting large bright patches”, and let \( X=\{x_{1i},x_{2i},...,x_{ki}\}, i=1,...,Q \) denote the feature data set of all \( Q \) bright patches in \( O1 \). Then, there are two methods for determining the number of buildings in a specified region. The first method counts the bright patches in the defined region from \( O2 \). The second method, which is based on \( O1,MD() \) and \( X \), estimates \( y \), the number of buildings in the region, by the following model:

\[
y = \text{round}\left( \sum_{i=1}^{m} y_i \right)
\]

\[
y_i = MD(i) = a_0 + a_1 x_{1i} + ... + a_j x_{ji} + ... + a_k x_{ki}
\]

where \( m \) denotes the number of bright patches in the region from \( O1 \), \( y_i \) is the calculated number of buildings in the \( i \)th bright patch using the regression model \( MD(i) \), and \( x_{ji} \) is the \( j \)th feature of the \( i \)th bright patch. It can be observed that the first method requires fewer data and is simpler to use for end users than the second method. That benefit is the main contribution of the procedure for splitting large bright patches. Another benefit of the procedure is that when a large bright patch caused by several buildings is located on the border of a specified region, the number of buildings of the patch that are within this region can be easily counted after splitting. If not otherwise specified, the first method is used in this paper.

![Figure 5](image)

**Figure 5** - (a) An SAR image scene; (b) building detection result before splitting large bright patches; (c) building detection result after splitting large bright patches. The number of buildings in the specified region (with dashed-line outline) can be derived by counting the bright patches in it. The density of the number of buildings for the pixel (denoted by a star) can be determined by counting the bright patches in a circle region centred on this pixel.
Next, it will be demonstrated how to determine the density of the number of buildings for every site (pixel). We calculate the pixel-level density using the concept of local window processing: build a circular or rectangular local window (that corresponds to a one-square-kilometre area) centred on each pixel; then, determine the number of bright patches in the window from $O_2$ and assign this number to the central pixel. The output map of the density of the number of buildings is very useful for finding interesting urban structures (which can be clearly observed in Fig. 10), which may be of potential value in many fields, such as commerce and environmental protection.

**Experimental results and analysis**

**Data set description**

The effectiveness of the proposed scheme has been tested on a TerraSAR-X image of the city of Wuhan, Hubei Province, China. The image was acquired in the HH polarisation and strip modes, thereby resulting in a geometrical resolution of approximately 3 m×3 m (azimuth×ground range). The data product is the EEC level with 16 bit in DataDepth and a spacing unit of 1.25 m in columns and rows. The incidence angle varies from 33.9° to 36.6°. The original scene has been cut to a subset of 9000×2600 pixels. The cut includes built-up areas, lakes, forests, and farms. The radar brightness, $\beta^0_{\text{dB}}$, is derived from the image pixel values or digital numbers (DN) applying the calibration factor $k_s$ as $\beta^0_{\text{dB}} = 10 \cdot \log_{10}(k_s \cdot |\text{DN}|^2)$. The $\beta^0_{\text{dB}}$ value is independent of the processor and product type of the data. If suitable geocoded incidence angel mask (GIM) data are available, $\beta^0_{\text{dB}}$ can be further calibrated to produce a $\sigma^0_{\text{dB}}$ that is independent of the relative orientation of the illuminated resolution cell, the sensor, and the distance in range between them. Although the calibration step is not strictly necessary, it allows the definition of the algorithm parameters to be used with SAR images of different data sets and data products acquired by the same or different sensors.

**Selection of parameters and training samples**

All the user-defined parameters of our scheme are in the building-detection step. The parameters of thresholding with hysteresis, $T_{\text{high}}$ and $T_{\text{low}}$, must be selected by analysing the amplitudes of sample pixels that belong to building regions in the SAR image. In our experiments, reasonable DN values for $T_{\text{high}}$ were on the order of 600 to 800 (6.1 to 8.6 dB for the $\beta^0_{\text{dB}}$ value); reasonable DN values for $T_{\text{low}}$ were on the order of 350 to 450 (1.5 to 3.6 dB for the $\beta^0_{\text{dB}}$ value). The area threshold, $T_{\text{area}}$, should be selected by analysing the areas of bright patches caused by small non-building structures such as cars, lamps, and isolated trees. In our experiments, 20 to 40 pixels was determined to be reasonable. The size of the structures for closing should be smaller than the size of most holes caused by connecting adjacent bright structures from different buildings and of most intervals between buildings. A 5 pixel×5 pixel structure for closing is suitable for our experiments. The aim of opening is to remove faint connection structures between two or more large bright patches and make every bright patch regular. The size of the structure for opening is determined by the definition of the “faint connection-structure”. In our experiments, a 3 pixel×3 pixel structure for opening was determined to be reasonable.

When selecting samples for modelling the number of buildings, two main points should be noted. First, based on the accepted principle of regression analysis, the number of samples
should be at least 5~10 times the number of independent variables. Second, try to choose samples that include buildings of general sizes and types and try to make the number-of-buildings values of these samples uniformly cover the range of possible values.

**Building-detection result**

Because it is difficult to see the details of buildings from a large-scene image, we show only the building detection result for a small area (see Fig. 6) of our entire test scene. As observed in Figure 6, the low-threshold result includes almost all bright building pixels and some false positives, whereas the high-threshold result consists only of building pixels with very high values. Through morphological reconstruction (with the high-threshold result as the marker image and the low-threshold result as the mask image), almost all false positives are eliminated. The closing operation connects isolated small bright patches into large patches, and the opening operation makes patches have more-regular shapes and splits some patches by removing their thin structures (some of which are caused by the closing operation). Finally, the area-thresholding step eliminates all of the small patches and makes the number of the remaining bright patches approximate the real number of buildings. Although we try to make every bright patch in the result correspond to one building, some bright patches from different buildings will be connected as a large building (such as the two large bright patches on the right part of Fig. 6h), which is why we need further processing steps.

![Figure 6](image)

**Modelling the number of buildings**

Six features (Area, Solidity, MajorAxisLength, MinorAxisLength, Eccentricity, and Compactness) are selected as possible predictors. Forty-eight bright patches are selected to obtain the sample data set \( X = \{ y, x_1, \ldots, x_i, \ldots, x_k \} \) for estimation of our model.
Here, \( n=48 \), \( y \) denotes the real number of buildings in a specified bright patch, which is obtained by counting the freestanding buildings in the bright-patch region by referring to Google Earth. Based on \( X \), a regression model is constructed using the stepwise regression algorithm provided by the SPSS software. At each step of the stepwise algorithm, the predictor variable that is not in the equation that has the smallest probability of \( F \) is entered if that probability is less than 0.05. Variables that are already in the regression equation are removed if their probability of \( F \) becomes greater than 0.1. Information regarding the constructed model is presented in Table 1 and Figure 7. Three shape features, Area, MinorAxisLength, and Compactness, are chosen as the final predictors. Their VIF are all less than 5, which is a very conservative threshold for diagnosing multicollinearity. The final regression model is \( y=a_0+a_1x_1+a_2x_2+a_3x_3 \), where \( x_1, x_2, \) and \( x_3 \) denote the three shape features and \( y \) denotes the number of buildings in the specified bright patch from which the above features are calculated. The coefficients of the model are calculated from data set \( X \) as \( a_0=0.427624, a_1=0.000475, a_2=0.006759, \) and \( a_3=0.019766 \). For this regression model, the value of the ARS is 0.818, which is approximately 1; this result implies that the fitting degree of the regression model is rather good. The \( F \) statistic is 71.47016, which is far greater than \( F_{0.05}(3,44) = 2.82 \) at the given significance level \( \alpha = 0.05 \). Thus, we can conclude that the regression equation is significant and the overall linear relationship is valid. The \( T \) statistics for all coefficients are greater than \( t_{0.05}(44) = 2.01 \) at the given significance level \( \alpha = 0.05 \); thus, we can infer that the relationship between each dependent variable and independent variable is significant.

Figure 7 - Fitting curves and residuals of (a) Area vs. number of buildings, (b) MinorAxisLength vs. number of buildings, and (c) Compactness vs. number of buildings.
Table 1 - Information regarding the regression model constructed by the stepwise algorithm.

| Model          | Co.      | VIF | t     | Sig. | F      | Sig.   | ARS |
|----------------|----------|-----|-------|------|--------|--------|-----|
| Constant       | 0.4276424 |     | 1.851 | .071 | 71.47  | 6E-17  | 0.818 |
| Area           | 0.0004754 | 3.315| 4.148 | 1.5E-4 |        |        |     |
| Compactness    | 0.0067597 | 1.722| 3.022 | .004 |        |        |     |
| MinorAxisLength| 0.0197668 | 3.721| 2.467 | .018 |        |        |     |

Qualitative analysis suggests that an obvious correlation between the selected features and the number of buildings in a bright patch exists: if a bright patch is the composition of more than one building’s radar footprints, then it is more likely for the patch to have a larger Area and MinorAxisLength than a patch that consists of only one building-radar footprint. The building-detection step of our proposed scheme makes the shape of these one-building bright patches regular (with small Compactness) by merging adjacent bright pixels and removing small background holes and thin foreground structures. A higher Compactness is more likely caused by unwanted structures, such as background holes and irregular foreground structures, produced by false connections between bright patches that belong to different buildings. These unwanted structures are larger than the element structure used in the closing and opening operations and thus cannot be removed in the building-detection step of our scheme. They can be clearly observed from the bright patches in Figure 8.

A better method for testing for errors in the regression model is to assess the model against a set of data that was not used to create the model. The data in Table 2 and Table 3 are not used to create the model. Our regression model performs well on them.

Figure 8 - (a) Optical images; (b) SAR images; (c) large bright patches; (d) results of the splitting-large-bright-patches procedure of our scheme; (e) results of the K-means method; (f) results of the ISODATA method.
Table 2 - Estimated number of buildings in the six bright patches of Figure 8.

|                  | patch1 | patch2 | patch3 | patch4 | patch5 | patch6 |
|------------------|--------|--------|--------|--------|--------|--------|
| Real number      | 12     | 4      | 3      | 3      | 7      | 11     |
| ISODATA          | 6      | 5      | 2      | 2      | 5      | 5      |
| Our Method       | 12     | 4      | 4      | 3      | 9      | 13     |

**Splitting large bright patches**

For each large bright patch in the building detection result, first, the number of buildings \( y \) is calculated using the model derived in the previous section. Then, if \( \text{round}(y) \geq 2 \), the patch is split into \( \text{round}(y) \) small patches. Figure 8 shows the segmentation results of our method, the K-means method, and the ISODATA method [Ball and Hall, 1967] for six large bright patches. The classic K-means method, which randomly selects \( K \) points as the initial centroids, is used. The parameter \( K \) is set to \( \text{round}(y) \). The ISODATA method can estimate the optimal \( K \) dynamically but needs many user-defined parameters. In our case, we set the searching number of classes from Min=2 to Max=20, the minimum pixels in the class to 50, the maximum intra-class standard variance to 5, the minimum inter-class distance to 20, the maximum merge pairs per iteration to 2, and the maximum iteration time to 10. It can be observed from Figure 8 that K-means and ISODATA poorly preserve the spatial shape of the building regions. This deficiency is because they all prefer spherical clusters and clusters are expected to be of similar size. Our method performs well at preserving shape. From Table 2, it can be observed that our model can estimate the number of buildings in a large bright patch reasonably well.

**Performance evaluation**

To assess the stability and accuracy of the proposed method, we choose eight small regions from our large test scene. In each region, there are many bright patches. The average relative error (ARE) of the number of buildings is used as a performance indicator. The ARE is defined as

\[
ARE = 1 / N \times \sum_{i=1}^{N} \frac{|n^i_{d} - n^i_{t}|}{n^i_{t}} \quad [4]
\]

where \( n^i_{t} \) is the real number of buildings in the ith region, \( n^i_{d} \) is the number of buildings in the ith region estimated by the model of the number of buildings, and \( N \) is the total number of regions for the test. The real number of buildings in a region is derived by counting the freestanding buildings in the region by referring to Google Earth maps. For each test region, there is an ARE (given in Tab. 3), and for all the test regions, the average ARE equals 0.13.
Table 3 - Number of buildings in the eight test regions.

| Region  | Real number | Estimated number | ARE  |
|---------|-------------|------------------|------|
| Region 1| 9           | 9                | 0    |
| Region 2| 9           | 7                | 0.22 |
| Region 3| 13          | 14               | 0.08 |
| Region 4| 9           | 11               | 0.22 |
| Region 5| 8           | 11               | 0.38 |
| Region 6| 5           | 5                | 0    |
| Region 7| 6           | 6                | 0    |
| Region 8| 7           | 6                | 0.14 |
| Average |             |                  | 0.13 |

It should be noted that the ARE value may be somewhat misleading for users who are less interested in the number information than other information, such as the detection rate and positions of buildings. Let us consider the following extreme example: a scene contains 100 buildings. The proposed scheme does not detect any of the buildings and detects 100 false positives. In this case, we calculate ARE=0, which is very good in the sense of ARE, but this is very bad in the sense of building detection. Thus, we next use some conventional measurements to demonstrate the performance of our scheme. Another test site (Fig. 9a) is chosen. We manually draw building footprints in the test site (Fig. 9c) by referring to the SAR image, the output of the bright-patch splitting procedure, and Google Earth maps. It should be noted that these building footprints are still approximate because we lack exact scene information and imaging-geometry analysis. If a bright patch partially overlaps any building footprint or if the imaging geometry clearly indicates that this patch is from buildings, we identify it as a true positive; otherwise, we identify it as a false positive. If there is a building footprint but no bright patch that overlaps at least partially with it, then there is a missed positive. The percentage of false positives (PFP) is defined as the ratio of false positives to total positives identified. The percentage of missed positives (PMP) is defined as the ratio of missed positives to real buildings. In this site, there are 93 buildings total. Our scheme detects 100 bright patches and thus produces ARE=0.08. Fourteen buildings are missed (PMP=0.15) by our scheme. It can be observed from Figure 9(a) that most of these missed buildings are single-floor buildings with sloping roofs and many nearby trees. These features make few structures that produce the strong echoes (layover and double bounce, which are the key features indicating a building) in the SAR image. There are 8 false positives among the 100 detected bright patches. From Figure 9, it can be observed that approximately half of these false positives are likely caused by trees, whereas the others are caused by man-made structures.

Table 4 - Performance of our scheme on the test site shown in Figure 9.

| Real Num. of Buildings | False Positives | Missed Positives | Estimated Num. | PFP  | PMP  | ARE  |
|------------------------|-----------------|------------------|----------------|------|------|------|
| 93                     | 8               | 14               | 100            | 0.08 | 0.15 | 0.08 |
Figure 9 - (a) Optical image of a test site; (b) SAR image of the test site. (c) Building footprints. The footprints in green are not detected by our scheme. (d) The output of our bright-patch-splitting procedure. Blue patches are false positives.

**Density of the number of buildings**

A pixel-level density map of the number of buildings (Fig. 10b) is calculated using an 801-pixel × 801-pixel local window (which corresponds to a region of one square kilometre) on the result of the splitting-large-bright-patches procedure. Each pixel value represents the number of buildings per square kilometre. Figure 10(c) shows an $N$-level density map, which is obtained by classifying pixels into $N$ different classes according to their density value (here, we set $N=8$). The density map and eight-level density map exhibit some interesting spatial structures in Wuhan City. This information may have potential value for commerce and for environmental evaluation and protection.

**Additional check using a different scene**

The relationship between the features extracted from the building patches and the real number of buildings in the corresponding area may depend strongly on the specific geographical zone. It may even be very different for different provinces in China. Therefore, we added an additional check using a different test scene from Sichuan Province of China. A TerraSAR-X data with HH polarisation, 16 bit in DataDepth, geometrical resolution of approximately 6 m×6 m (azimuth×ground range), and incidence angle that varies from 26.1° to 27.8° is used. By analysing the amplitude of sample pixels that belong to building regions in the SAR image, the parameters of thresholding–with-hysteresis, $T_{\text{high}}$ and $T_{\text{low}}$, are set to 600 and 350. The image data of this scene are re-sampled to achieve the same spacing unit in columns and rows.
as the image data of the Hubei province scene. The area threshold $T_{area}$ is set to 40. The size of the structures for closing and opening is set to 5 pixels×5 pixels and 3 pixels×3 pixels. Forty-two bright patches are selected as samples for modelling the relationship between the features extracted from the building patches and the real number of buildings in the corresponding area. The information from the regression model constructed by the stepwise algorithm is presented in Table 5 and Figure 11. It can be observed that in this instance, only area and compactness are selected as the predictor variables. The MinorAxisLength feature is discarded by the stepwise algorithm because of its high correlation with area. Compared with the relationship for the test scene of Hubei Province, this new relationship does not change markedly, which may be explained in part by the widely criticised phenomenon that cities in China often look similar to each other.

A small area of 870×500 pixels is used for performance evaluation (Fig. 12). In this test site, there are a total of 151 buildings. Our scheme detects 175 bright patches and thus has ARE=0.16. There are 18 buildings that are missed (PMP=0.12) by our scheme. It can be observed from Figure 12 a, c that three of these missed buildings are single-floor buildings with sloping roofs and many trees around them, and six of them have large orientation angles (they are arranged almost parallel to the range direction). There are 26 false positives among the 151 detected bright patches; thus, PFP=0.17. From Figure 12, it can be observed that most of the false positives are caused by trees and small man-made structures.

Figure 10 - (a) TerraSAR-X image of part of Wuhan City, China. (b) The pixel-level density map of the number of buildings. Each pixel value represents the number of buildings in a local window (which corresponds to a region of one square kilometre) centred on this pixel. (c) Eight-level density map obtained by classifying the pixels of image (b) into eight different classes according to their density values.
Figure 11 - Fitting curves and residuals of (a) area vs. number of buildings and (b) compactness vs. number of buildings.

Figure 12 - (a) Optical image of the test site; (b) SAR image of the test site; (c) building footprints. (the footprints in green are missed by our scheme); (d) the output of our bright-patch-splitting procedure (blue patches are false positives).
Table 5 - Information regarding the regression model constructed by the stepwise algorithm.

| Model   | Co.       | VIF | t    | Sig. | F         | Sig. | ARS |
|---------|-----------|-----|------|------|-----------|------|-----|
| Constant| 0.4863514 | 2.214 | 0.032 | 144.58 | 9E-19    | 0.875 |
| Area    | 0.0005258 | 2.298 | 6.378 | 1.5E-7 |          |      |     |
| Compactness | 0.0209807 | 2.298 | 5.608 | 1.8E-6 |          |      |     |

Conclusions

In this paper, a complete scheme for detecting the number of buildings in a single high-resolution SAR image has been proposed. It has been demonstrated that building number information of satisfactory accuracy can be obtained from the residential and commercial scenes in our test sites in different provinces of China. This scheme is effective and easy to implement.

We found that high-resolution SAR data cannot provide precise detection of the number of the buildings in the case of false connections (when bright patches that belong to different buildings are connected) and false disconnections (when multiple bright patches that belong to the same building are isolated). Thus, a correction based on visual analysis is required (in our work, the correction is based on morphological operators and a regression function that models the relationship between the visual features of a detected building region and the real number of buildings in it).

Some limitations of this scheme should be noted to enable valid usage: this scheme is not suitable for scenes (such as industrial scenes) with very large and flat buildings. In our scheme, there is an underlying assumption that the bright structures of one building converge. That assumption will not apply to a large, flat building because its bright structures tend to spread over a large space. Another underlying assumption of our scheme is that buildings are separated from each other. Thus, our scheme may not work well in many parts of the world, such as in Europe, where buildings are usually not separated from each other. However, it may work well in many developing countries, especially China, in which the buildings of most cities are usually similar and separated from each other. This scheme may miss some single-floor buildings with sloping roofs and many trees around them and some buildings that are arranged almost parallel to the range direction because these buildings seldom cause strong double-bounce structures in SAR image. Trees, cars, bushes, and man-made objects, such as walls, will cause some false positives if the bright patches they create are larger in area than the user-defined threshold for eliminating false positives.

The potential for using multi-temporal high-resolution SAR images to analyse the space-time pattern of the number of buildings in a region will be studied in the future. SAR image filtering, shadow features, and context features should be properly used to further improve the performance of this scheme. Furthermore, the scheme should be tested on several different SAR data sets after performing complete radiometric calibration to test its applicability and generality.

Acknowledgements

We acknowledge the feedback we received from the referees. Many of their suggestions have been implemented, thereby improving the quality of the paper. This work was supported by the National Natural Science Foundation of China (grant No. 40901207, No. 41161065 and
No.41361091) and the Joint Funds of Department of Science and Technology of Guizhou Province and Guizhou Normal University (LKS[2013]28).

References
Ball G., Hall D. (1967) - A clustering technique for summarizing multivariate data. Behavioral science, 12: 153-155. doi: http://dx.doi.org/10.1002/bs.3830120210.
Beucher S., Meyer F. (1993) - The morphological approach to segmentation: the watershed transformation. In: Mathematical morphology in image processing, Dougherty E.R., pp. 433-481.
Canny J. (1986) - A computational approach to edge detection. IEEE Transaction on Pattern Analysis and Machine Intelligence, 8: 679-698. doi: http://dx.doi.org/10.1109/TPAMI.1986.4767851.
Dell’Acqua F., Gamba P. (2003) - Texture-Based Characterization of Urban Environments on Satellite SAR Images. IEEE Transactions on Geoscience and Remote Sensing, 41: 153-159. doi: http://dx.doi.org/10.1109/TGRS.2002.807754.
Dougherry E.R. (1992) - An Introduction To Morphological Image Processing. SPIE Press, Bellingham, Wa.
Draper N., Smith H. (1981) - Applied Regression Analysis. 2nd Edition, New York: John Wiley & Sons Incorporation.
Ferro A., Brunner D., Bruzzone L., Lemoine G. (2011) - On the relationship between double bounce and orientation of buildings in VHR SAR images. IEEE Transactions on Geoscience and Remote Sensing Letter, 8: 612-616. doi: http://dx.doi.org/10.1109/TGRS.2010.2097580.
Ferro A., Brunner D., Bruzzone L. (2013) - Automatic detection and reconstruction of building radar footprints from single VHR SAR images. IEEE Transactions On Geoscience and Remote Sensing, 51: 935-952. doi: http://dx.doi.org/10.1109/TGRS.2012.2205156.
Freedman D.A. (2005) - Statistical Models: Theory and Practice. Cambridge University Press, pp. 41-70.
Gonzalez R.C., Woods R.E. (2007) - Digital Image Processing. Third Edition, Prentice Hall, pp. 656-664.
Gouinaud C., Tupin F. (1996) - Potential and use of radar images for characterization and detection of urban areas. IEEE Transactions on Geoscience and Remote Sensing Symposium, Lincoln, NE, 474-476. doi: http://dx.doi.org/10.1109/IGARSS.1996.516377.
Grimaud M. (1992) - A New Measure of Contrast: Dynamics. In: Proceedings of Image Algebra and Morphological Processing III, San Diego, SPIE 1769: 292-305. doi: http://dx.doi.org/10.1117/12.60650.
Harel O. (2009) - The estimation of R2 and adjusted R2 in incomplete data sets using multiple imputation. Journal of Applied Statistics, 36: 1109-1118. doi: http://dx.doi.org/10.1080/02664760802553000.
He W., Hellwich O. (2009) - Bayesian Building Extraction from high resolution polarimetric SAR data. IEEE Transactions on Geoscience and Remote Sensing Symposium, Cape Town, pp. 402-405. doi:http://dx.doi.org/10.1109/IGARSS.2009.5417398.
Michaelsen E., Stilla U., Soergel U., Doktorski, L. (2010) - Extraction of Building Polygons from SAR Images: Grouping and Decision-Level in the GESTALT system. Pattern Recognition
Letter, 31 (10): 1071-1076. doi: http://dx.doi.org/10.1016/j.patrec.2009.10.004.
Mood A.M., Graybill F.A., Boes D.C. (1974) - Introduction to the Theory of Statistics. Third Edition, McGraw-Hill, pp. 246-249.
Najman L., Schmitt M. (1994) - Watershed of a Continuous Function, Signal Processing. Special issue on mathematical morphology, 38: 99-112. doi:http://dx.doi.org/10.1016/0165-1684(94)90059-0.
Poulain V., Inglada J., Spigai M., Tourneret J., Marthon P. (2011) - High-Resolution Optical and SAR Image Fusion for Building Database Updating. IEEE Transactions On Geoscience and Remote Sensing, 49: 2900-2910. doi: http://dx.doi.org/10.1109/TGRS.2011.2113351.
Schroeder M.A. (1990) - Diagnosing and dealing with multicollinearity. Western Journal of Nursing Research, 12: 175-187. doi: http://dx.doi.org/10.1177/019394599001200204.
Simonetto E., Oriot H., Garello R. (2005) - Rectangular Building Extraction From Stereoscopic Airborne Radar Images. IEEE Transactions On Geoscience and Remote Sensing, 43: 2386-2395. doi: http://dx.doi.org/10.1109/TGRS.2005.853570.
Soergel U., Thoennessen U., Brenner A., Stilla U. (2006) - High-resolution SAR data: new opportunities and challenges for the analysis of urban areas. IEEE Proceedings of Radar Sonar and Navigation, pp. 294-300. doi: http://dx.doi.org/10.1049/ip-rsn:20045088.
Sportouche H., Tupin F. (2010) - A processing chain for simple 3D reconstruction of buildings in urban scenes from high resolution optical and SAR images. 8th European Conference on Synthetic Aperture Radar, Aachen, Germany.
Sportouche H., Tupin F., Denise L. (2011) - Extraction and Three-Dimensional Reconstruction of Isolated Buildings in Urban Scenes From High-Resolution Optical and SAR Spaceborne Images. IEEE Transactions on Geoscience and Remote Sensing, 49: 3932-3946. doi: http://dx.doi.org/10.1109/TGRS.2011.2132727.
Stasolla M., Gamba P.(2008) - Spatial Indexes for the Extraction of Formal and Informal Human Settlements From High-Resolution SAR Images. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing, 1: 98-106. doi: http://dx.doi.org/10.1109/JSTARS.2008.921099.
Thiele A., Thoennessen U., Cadario E., Schulz K., Soergel U. (2007) - Building Recognition From Multi-Aspect High-Resolution InSAR Data in Urban Areas. IEEE Transactions On Geoscience and Remote Sensing, 45: 3583-3593. doi: http://dx.doi.org/10.1109/TGRS.2007.898440.
Tison C., Nicolas J., Tupin F., Maitre H. (2004) - A new statistical model for markovian classification of urban areas in high-resolution SAR images. IEEE Transactions On Geoscience and Remote Sensing, 42: 2046-2057. doi: http://dx.doi.org/10.1109/TGRS.2004.834630.
Wegner J.D., Hansch R., Thiele A., Soergel U. (2011) - Building detection from one orthophoto and high-resolution InSAR data using conditional random fields. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4: 83-91. doi: http://dx.doi.org/10.1109/JSTARS.2010.2053521
Wooldridge J.M. (2000) - Introductory Econometrics: A Modern Approach. South Western Publishing.
Xu F., Jin Y. (2007) - Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images. IEEE Transactions On Geoscience and Remote Sensing, 45: 2336-2353. doi: http://dx.doi.org/10.1109/TGRS.2007.896614.

© 2014 by the authors; licensee Italian Society of Remote Sensing (AIT). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).