



HAL
open science

Detecting blind building façades from highly overlapping wide angle aerial imagery

Jean-Pascal Burochin, Bruno Vallet, Mathieu Brédif, Clément Mallet, Thomas Brosset, Nicolas Paparoditis

► **To cite this version:**

Jean-Pascal Burochin, Bruno Vallet, Mathieu Brédif, Clément Mallet, Thomas Brosset, et al.. Detecting blind building façades from highly overlapping wide angle aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 96, pp.193 - 209. 10.1016/j.isprsjprs.2014.07.011 . hal-01559518

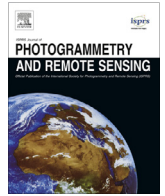
HAL Id: hal-01559518

<https://hal.science/hal-01559518v1>

Submitted on 10 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Detecting blind building façades from highly overlapping wide angle aerial imagery



Jean-Pascal Burochin^{a,*}, Bruno Vallet^a, Mathieu Brédif^b, Clément Mallet^a, Thomas Brosset^b, Nicolas Paparoditis^a

^a Université Paris-Est, IGN/SRIG, MATIS, 73 avenue de Paris, 94160 Saint-Mandé, France

^b QuelleVille?, 5 avenue de la République, 75011 Paris, France

ARTICLE INFO

Article history:

Received 31 January 2014

Received in revised form 4 July 2014

Accepted 14 July 2014

Available online 13 August 2014

Keywords:

Vertical imagery

Aerial

Urban

Building façade

Opening detection

Feature

Classification

Learning

ABSTRACT

This paper deals with the identification of blind building façades, *i.e.* façades which have no openings, in wide angle aerial images with a decimeter pixel size, acquired by nadir looking cameras. This blindness characterization is in general crucial for real estate estimation and has, at least in France, a particular importance on the evaluation of legal permission of constructing on a parcel due to local urban planning schemes.

We assume that we have at our disposal an aerial survey with a relatively high stereo overlap along-track and across-track and a 3D city model of LoD 1, that can have been generated with the input images. The 3D model is textured with the aerial imagery by taking into account the 3D occlusions and by selecting for each façade the best available resolution texture seeing the whole façade. We then parse all 3D façades textures by looking for evidence of openings (windows or doors). This evidence is characterized by a comprehensive set of basic radiometric and geometrical features. The blindness prognostic is then elaborated through an (SVM) supervised classification.

Despite the relatively low resolution of the images, we reach a classification accuracy of around 85% on decimeter resolution imagery with 60 × 40% stereo overlap.

On the one hand, we show that the results are very sensitive to the texturing resampling process and to vegetation presence on façade textures. On the other hand, the most relevant features for our classification framework are related to texture uniformity and horizontal aspect and to the maximal contrast of the opening detections.

We conclude that standard aerial imagery used to build 3D city models can also be exploited to some extent and at no additional cost for facade blindness characterisation.

© 2014 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Urban areas have been extensively analyzed over the last twenty years using remote sensing data, with an emphasis on building detection (Rottensteiner et al., 2014).

Due to local urban planning schemes,¹ the existence of openings (windows or doors) on a façade determines whether a building can be erected in front of this façade or at a given distance. This information can therefore have an important impact on the evaluation of

legal permission of constructing on an area and, more generally, on the estimation of the potential of an area for urban development. Such studies can be implemented into urban simulation tools such as (Brasebin et al., 2011).

This paper describes a method to automatically identify *blind façades*, *i.e.* façades which have no openings, using standard aerial imagery (without oblique configuration). Compared with terrestrial acquisition systems, this solution decreases façade occlusions due to foreground objects. The challenge of this problem is that the rectified texture of the façades can be stretched by the perspective transformation: this is mainly due to the opening angle of the camera lens and the location of the façade in aerial images (*cf.* Section 2). Fig. 1 illustrates this issue: the rectified texture of the openings are dark spots.

* Corresponding author.

E-mail addresses: jean-pascal.burochin@ign.fr (J.-P. Burochin), bruno.vallet@ign.fr (B. Vallet), mathieu.bredif@ign.fr (M. Brédif), clement.mallet@ign.fr (C. Mallet), thomas.brosset@gmail.com (T. Brosset), nicolas.paparoditis@ign.fr (N. Paparoditis).

¹ Plan local d'urbanisme (PLU) in French.

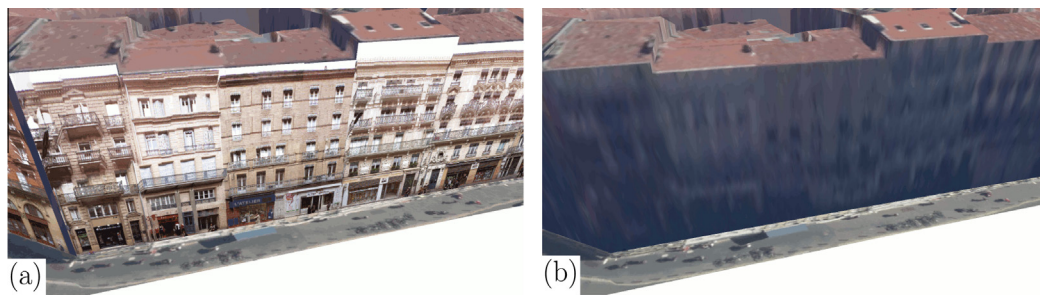


Fig. 1. Impact of the image resolution in rectified images. (a) Image of façades acquired by the Mobile Mapping System *Stéréopolis* (Paparoditis et al., 2012), 1 cm GSD. (b) The same façades acquired by standard aerial imagery, 10 cm GSD: pixels are stretched over 1 m in the vertical direction on several vertical planes.

The *blind façade* detection is a task required by the e-PLU project (<http://www.e-plu.fr>), which aims at transforming urban planning documents into a 3D digital urban simulation platform (Brasebin et al., 2011). Morphological and urban criteria are both taken into account to estimate the legal permission of constructing on a parcel. Such a system is at high stakes because of the complexity of the urban planning documents. It can also contribute to sustainable development and energy management (Bozonnet et al., 2013).

An overview of façade analysis approaches from the literature is proposed in Section 1.1. Then the general analysis strategy for *blind façades* classification is presented in Section 1.2.

1.1. Related work on façade analysis

Image-based façade analysis is a key issue which has been increasingly important in the two last decades. Many applications are concerned such as building reconstruction, semantic labeling or realistic content generation.

Here, two common approaches are discussed about terrestrial and aerial imagery: scene description by local feature detection (Section 1.1.1), or by more general features (Section 1.1.2).

1.1.1. Local feature extraction

Mostly, target objects must have structural properties which accurately match the *a priori* model in the data so as to be correctly detected.

Local features are generally extracted in images acquired by optical terrestrial systems because of their high spatial resolution (Ground Sample Distance, GSD, of about 2 centimeters). A common approach is to match the objects of the image scene with *a priori* models. A widespread scheme, namely the *bottom-up* method, consists in starting with very local information, e.g. pixel-level, so as to independently analyze different elements in the whole image (e.g. using texture filter banks or semantic features like corners or edges of specific elements), and then progressively group them into more sophisticated configurations.

Starting with a set of independent, possibly overlapping window hypotheses, Korč and Förstner (2008a) found an optimal configuration of non-overlapping windows using a graph-based binary integer programming solution. They also evaluated the relevance of graphical models for such an issue (Korč and Förstner, 2008b). Similarly, Reznik and Mayer (2008) started with an independent window corner detection process, then inferred individual window hypothesis localization from this corner set by employing an Implicit Shape Model and finally grouped them into rows and columns. Haugeard and Philipp-Foliguet (2012) detected image contours and regions in order to feed adjacency graphs and recognize windows using a *Support Vector Machines* classifier – SVM, (Cortes and Vapnik, 1995). Farabet et al. (2012) parsed urban scenes from a set of multi-scale feature vectors, which provides a classification of each pixel into object categories such as building, tree, grass or

sky. They showed that both low-level and mid-level features are relevant for efficient discrimination. A study of texture filter banks commonly employed in pixel-based urban object classification frameworks was proposed by Drauschke and Mayer (2010).

Using vertical aerial images, Meixner and Leberl (2011) were able to extract the façade items (e.g., floors and windows). Their approach was based on window redundancy and gradient accumulation profiles.

Another way to use local features is to directly start with the global object scene model and then progressively refine it by analyzing more and more low-level information (*top-down* approach). Müller et al. (2007), for instance, segmented a façade image according to a hierarchical subdivision pipeline in order to obtain a textured 3D model including semantic structures. Burochin et al. (2010) recursively split façade images with a into elementary models. Generally speaking, a widespread modeling tool consists in *grammars* (Ripperda and Brenner, 2009; Teboul et al., 2013), which is often adopted to introduce *a priori* information about specific spatial arrangement of windows. However, this is not sufficient for our specific purpose.

1.1.2. General feature extraction

As mentioned above, openings and façades may not be always accurately described using local features. Therefore, the latter ones have limited contribution for subsequent detection and classification.

General features mainly consist in a more flexible model which summarizes the statistical layout properties. They are used, for instance, in the classification of image scenes into open–close, indoor–outdoor, or natural–urban. Instead of specific elements detection, the point is to extract general information about the geometrical or the radiometric properties often at different scales and orientations. Gradient orientation histograms (*HOG*) are commonly used to obtain such descriptions (Dalal and Triggs, 2005). Contrary to local feature extraction schemes, the use of general features does not depend on an accurate element localization but is based on strong discriminating information statistically contained in the image. Thus, this additional kind of measurement theoretically supplies high stability. Furthermore, it does not require high image resolution (feasible with 10 cm GSD). In terrestrial images, Kumar and Hebert (2003), for instance, detected man-made structures with multi-scale gradient orientation features based on *intra-scale* and *interscale* measurements. Zhang and Košecká (2007) also used *HOG* to match façade images. Han and Liu (2013) categorized natural scene spatial layout by a Spatial Envelope which consisted of a set of output magnitude of a bank of multiscale-oriented filters.

In oblique aerial images, Nyaruhuma et al. (2012) generated façade clues by a measure of contour orthogonality and texture matching. The result was used to update the cadastral building outline. Similarly, Xiao et al. (2012) generated façade descriptors

based on a direction map in order to detect and reconstruct buildings. The direction map was generated from the detection of structure edges of the façade within 10–16 cm images.

1.2. Proposed strategy

The work presented in this paper is dedicated to very high resolution rectified images. Here, the resolution in the vertical direction is far from the native 10 cm GSD due to perspective transformation (up to 1 m). These images are noisy and contain shadows (Fig. 1). Hence, façades cannot be represented with sophisticated models: statistical radiometric and geometrical analysis with general features is preferred in conjunction with the detection of dark rectangular areas. As discussed in Section 1.1, façade analysis require feature computation at various scales, that is why our approach is composed of both *top-down* and *bottom-up* extracted information.

An outlook of the full work-flow is presented by Fig. 2. First, each façade polygon of the 3D model was textured by selecting the best available resolution texture of the aerial imagery seeing the whole façade. Then, statistical information was extracted, namely general features. Additionally, openings were detected by a stochastic process based on their contrast with the wall background. Finally, a supervised classifier was trained from all these general and local features in order to recognize *blind façades*. This learning step was required in order to capture the heterogeneity of urban structures and façade occlusions.

The main contributions of this work are threefold:

1. Façade-based feature extraction in standard airborne imagery. In this context, general radiometric and geometrical information of the images was computed: radiometry uniformity, *structuredness*, and *repetitiveness* of the texture. Local information based on contrasted openings was also inserted.
2. Analysis of the façade features. All these general and local features were combined into a *blind façade* descriptor. For efficient discrimination, the most relevant ones were selected.
3. Deadlock identification of the problem. The results demonstrates the sensitivity to the texturing resampling process and to vegetation presence on façade textures.

In the next section, the aerial image input and pre-process steps – rectification in façade planes and creation of the reference data set – are described. The general feature used – radiometry uniformity, *structuredness*, and *repetitiveness* of the texture – are described in Section 3. Besides, Section 4 introduces an energetic formulation to find the best configuration of opening detections: the optimization process consists in a combination of a *reversible jump Markov Chain Monte Carlo sampling (rjMCMC)* and of *simulated annealing*. The classification model used is introduced in Section 5: *Support Vector Machines – SVM*, (Cortes and Vapnik, 1995) – adopted because of its high generalization performance (Abu-Mostafa et al., 2012).

2. Data

The e-PLU project collaborates with the PLAINE COMMUNE conurbation, in the North-East of Paris, France. Several typical building blocks were selected within this area for deep analysis.

First, the use of standard imagery is motivated and the acquisition process is explained in Section 2.1. Typical building blocks in the study area are then described in Section 2.2. The resampling process into the façade plane is presented in Section 2.3. Finally, these resampled façade images were visually interpreted to build a reference data set as described and discussed in Section 2.4.

2.1. Relevance of standard airborne imagery

The acquisition of urban images by aerial platforms with vertically directed cameras represents a widespread solution to cover large areas with reduced economical costs.

Compared with terrestrial acquisition systems, this solution decreases façade occlusions due to foreground objects (vegetation, vehicles, urban furniture, etc.). Furthermore all façades are of interest, including those inside building blocks, which are often not easily accessible by terrestrial means. A comparison between oblique aerial imagery and vertical imagery is provided by Meixner and Leberl (2010b). They showed that vertical imagery with high overlap is sufficient to façade analysis.

Images were acquired by an opening angle of 60° and a ground sample distance of 10 cm. The overlap is about 60% in the flight direction and about 40% across this direction. Fig. 3 shows two extreme situations. When the acquisition angle is about 30°, the image texture is properly acquired with a distortion factor of 2 in the vertical direction. The worst situation occurs with a sharp grazing angle of about 6° with a distortion factor of 10: a 10 × 10 meter façade is represented by a texture of 100 × 10 pixels. Under this limit angle, the façade can be analyzed in the neighboring image (with an angle of 30°).

2.2. Typical building block selection

An expert selected 70 typical building blocks which refer to several urban constructions corresponding to different administrative area types (based on urban densities and activities, building shapes and sizes, etc.).

These building blocks are characteristic of a large-scale urban variety. Indeed several architectural periods were selected, old districts as well as modern amenities and infrastructures under construction. Different activities are represented: individual and community residential houses, industrial areas, car parks, gardens, schools, etc. The road network is diversified: boulevards, streets, dead ends, squares, etc.

Areas that are prone to change in the forthcoming years were also selected. Projects, such as tramway line construction, new malls, consist in challenging insertion issues inside existing constructions. Opening location knowledge on such changeable

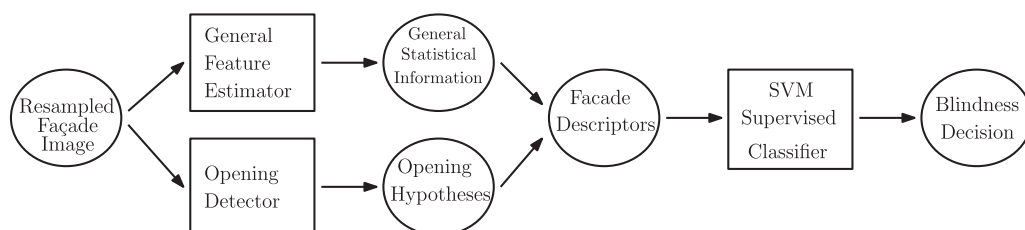


Fig. 2. Proposed methodology for blind façade detection.

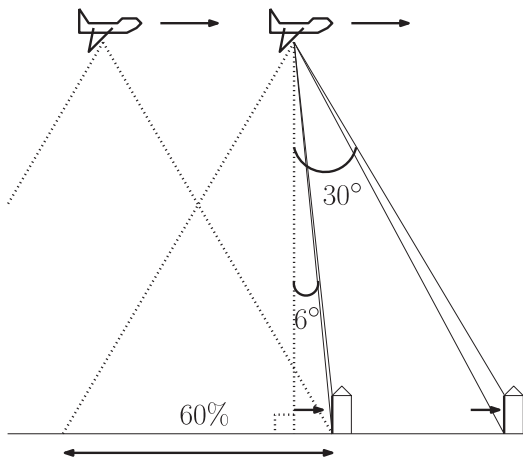


Fig. 3. Acquisition protocol: each façade is acquired with a minimum angle of 6° and a maximum angle of 30° .

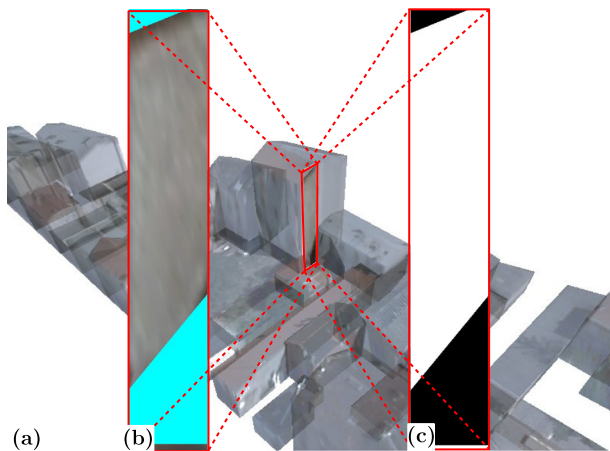


Fig. 4. Projection of the best viewing image in the façade plane: (a) Bounding rectangle in the façade plane; (b) rectification of the texture in the bounding rectangle; and (c) façade rectification mask.

constructions is an important issue to show how the project e-PLU can help urban planners and architects to estimate the impact of urban regulation.

2.3. Façade image rectification

The selected representative building blocks contained 735 façades. For each façade polygon of the 3D Model,² a bounding rectangle in the façade plane was defined, aligned with the vertical direction (Fig. 4a). The best viewing aerial image – i.e. the image in which the façade was acquired with the largest angle – was rectified in this rectangle (Fig. 4b). For non rectangular façades, masks were created so as to indicate which pixels actually belong to the façade (Fig. 4c).

Because of the imperfections of the 3D model – (Durupt and Taillandier, 2006), precision: 1 m – the resulting rectified façade images may be negatively impacted:

Border alignment

Building edges in the images might not fit exactly with the model edges. In this case, several building edges may appear within the

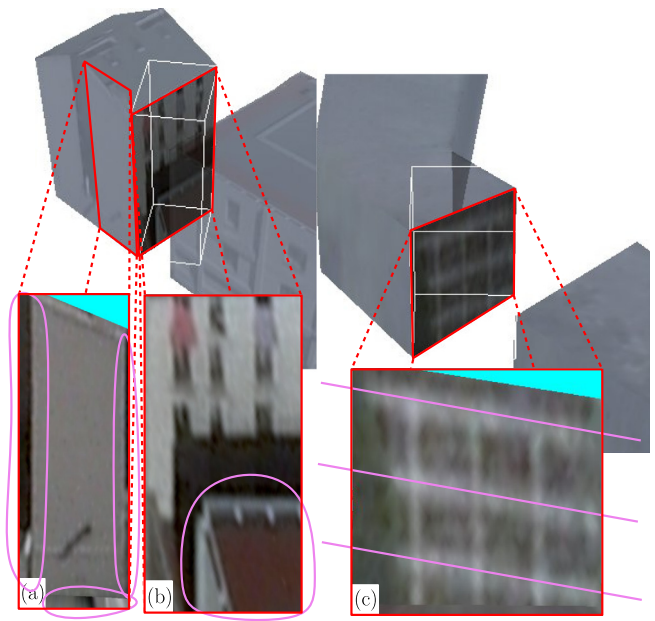


Fig. 5. Image rectification problems (circled in red lines): (a) Border alignment; (b) geometrical occlusion: a close building projected inside the rectified façade texture; and (c) contour axis alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

rectangle, that might be confused with opening edges (Fig. 5a). Conversely, meaningful edges within the façade might fall outside the rectangle.

Contour axis alignment

If the façade in the model is not perfectly parallel to the real façade, horizontal/vertical contours may not be axis-aligned in the image (Fig. 5c).

Both problems will hinder the opening detection process and the horizontal/vertical gradient accumulation. Note that close buildings or trees might be projected inside the rectified façade texture, which also disrupts the analysis (Fig. 5b).

2.4. Reference data set

A reference data set is required in order to evaluate the analysis results (see Section 5). Hence, an expert (a town planning consultant) visually located openings (doors and windows) as rectangular boxes inside the resampled images.³ 862 openings were then identified in 271 façades and 182 façades were considered *blind* (not containing any opening).

The operator had to cope with two kinds of problems. On the one hand, several images were acquired with a very small angle (6° in the worst situations), which implies a great distortion of the façade textures. On the other hand, there also are radiometric problems related to shadows and under-exposition. Therefore, all façade textures which cannot be interpreted by a human eye were considered *non classifiable*. This referred to 282 separate façades.

Moreover, as mentioned above, even if images were acquired from an aerial platform, above main street objects, façade textures are often occluded by vegetation and close building projection. Consequently, the operator also located those occlusion areas as

³ An editor was especially implemented so as to carry out this task. It is available as a sample of the GILVIEWER free open-source library (<https://code.google.com/p/gilviewer>).

² BATI-3D®, <http://professionnels.ign.fr/bati-3d>.

rough rectangular inputs. Thus, the analysis process was oriented to the façade area only.

Although these occlusion masks were manually built in this study, automatic alternatives exist to large scale purposes. Standard *Normalized Difference Vegetation Index*, *NDVI* (Kriegler et al., 1969) could be used to generate the vegetation mask (after a correct threshold fitting). The visibility problem could be solved by computer graphics approaches such as *z-buffering* based on the building 3D model (Vallet and Houzay, 2011).

The analysis of these data is presented in the three next sections: feature extraction in Sections 3 and 4, and *blind façade* classification in Section 5.

3. General Blindness Indicators

As detailed in Section 1.1, several general features were used in the literature to analyze façade images. Some of them were used in this study. The *façade texture uniformity* was measured with radiometry statistics (Section 3.1), the *structuredness* with gradient orientation histogram (Section 3.2), and the *repetitiveness* by gradient accumulation profiles (Section 3.3).

Only the façade area outside the occlusion mask was analyzed (see Section 2.4). Furthermore, the image was cropped. Indeed a margin was added at each side of the façade rectangle in order to cope with the border projection issue (see Section 2.3). The width of this margin was empirically set to 20 cm. It is represented by a line of the same color as the mask (cyan) on the illustrations of this paper.

3.1. Uniformity indicator

If openings are assumed to be contrasted with wall background, the radiometry distribution is more likely uniform with *blind façade* textures than with the textures of façade with openings. Fig. 6 shows radiometric histograms of a *blind façade* and a façade with openings.

The *uniformity indicator* is defined by the intensity standard deviation of the resampled image *I*:

$$\sigma_I^2 = \sum_{p \in M(I)} (I(p) - \bar{I})^2, \tag{1}$$

where $p \in M(I)$ is a pixel inside the mask $M(I)$ of the façade area. \bar{I} is the intensity mean defined by:

$$\bar{I} = \frac{1}{\#M(I)} \sum_{p \in M(I)} I(p). \tag{2}$$

3.2. Structuredness Indicators

A *blind façade* is likely to have a poor *structuredness* look. *Structuredness* refers to the man-made structure detection studied by Kumar and Hebert (2003). They noted that the gradients seem to be low in smooth areas whereas the distribution of its magnitude is likely to be approximately high and uniform in textured ones. Furthermore, in an area that contains few straight edges embedded in a smooth background, the gradient orientation

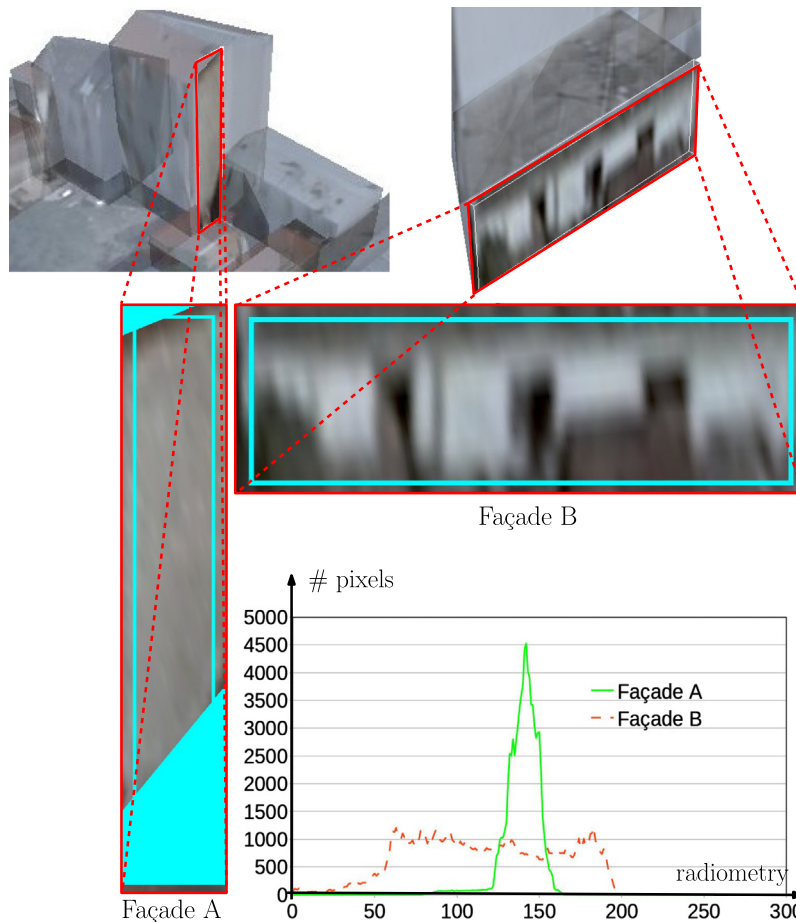


Fig. 6. Radiometry histograms of façade image A (*blind*) and façade image B (with openings).

histogram contains significant peaks. Hence, the average *spikeness* of the gradient orientation histogram was computed, under the assumption that the main contrasts of façade textures are grouped in opening areas. Let $\Upsilon(i)$ be the magnitude of the orientation histogram at the bin i . These measurements consisted of *heaved central-shift moments*. The p th ordered moment v_p was *heaved* because only positive differences were accumulated (values greater than the mean):

$$v_p = \frac{\sum_i (\Upsilon(i) - v_0)^{p+1} \mathbb{1}_{\Upsilon(i) > v_0}}{\sum_i (\Upsilon(i) - v_0) \mathbb{1}_{\Upsilon(i) > v_0}}, \quad (3)$$

where v_0 is the mean of the orientation distribution and $\mathbb{1}_x$ is the unit step function such that $\mathbb{1}_x = 1$ for $x > 0$, and 0, otherwise. In this work, the histogram was not smoothed as proposed in (Kumar and Hebert, 2003), but it was weighted by gradient module in order to reduce noise effect as they recommended. The three first moments were used (v_0 , v_1 and v_2).

For façade with openings, orientation histograms exhibit two main peaks that correspond to horizontal and vertical directions. Consequently, features have been derived in order to evaluate the significance of such peaks (Kumar and Hebert, 2003). This information was commonly used in the literature: Benedek et al. (2013) modeled vertical and horizontal orientations peaks by the correlation of the histogram with a bimodal Gaussian mixture. Here, the histogram was split into three parts: horizontal orientation, vertical orientation and other directions. Magnitude differences between these parts were computed. Let $\Upsilon_{[a,b]}$ be the magnitudes of the histogram between the orientation a and the orientation b . The indicator $\Delta_{\text{horizontal}}$ of horizontal orientations is defined by the difference between the horizontal orientation mean, $\overline{\Upsilon_{[0^\circ, \phi]}}$, and the mean of other orientations, $\overline{\Upsilon_{[\phi, 90^\circ - \phi]}}$:

$$\Delta_{\text{horizontal}} = \frac{1}{v_0} (\overline{\Upsilon_{[0^\circ, \phi]}} - \overline{\Upsilon_{[\phi, 90^\circ - \phi]}}), \quad (4)$$

where the parameter ϕ is the orientation tolerance. It was empirically set to 10° (sufficient value to isolate the peaks). This indicator was normalized by the moment v_0 . In the same way, an indicator Δ_{vertical} of vertical orientations was computed by the difference between the vertical orientation mean, $\overline{\Upsilon_{[90^\circ - \phi, 90^\circ]}}$, and the mean of other orientations:

$$\Delta_{\text{vertical}} = \frac{1}{v_0} (\overline{\Upsilon_{[90^\circ - \phi, 90^\circ]}} - \overline{\Upsilon_{[\phi, 90^\circ - \phi]}}). \quad (5)$$

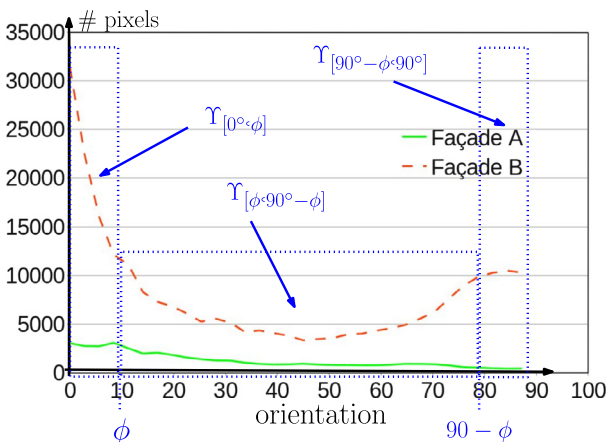


Fig. 7. Weighted gradient orientation histogram of façade image A (*blind*) and façade image B (with openings). All orientation values are fitted between 0° (horizontal) and 90° (vertical).

Fig. 7 shows the orientation histograms of a *blind façade* and a façade with openings. A sixth *structuredness* indicators was finally computed with the standard deviation of the orientation histogram:

$$\sigma_{\Upsilon}^2 = \sum_i (\Upsilon(i) - v_0)^2. \quad (6)$$

Low values indicate low contour orientation in the texture and high values occurs for structured textures (*i.e.* façade with openings).

3.3. Repetitiveness Indicators

An indicator of structure *repetitiveness* was also experimented similarly to (Meixner and Leberl, 2010a). Here openings are assumed to appear vertically and horizontally aligned in the façade texture. Gradient accumulation profiles of *blind façades* are more uniform than those of façade with openings. Therefore, the profiles $V(\nabla_y)$ and $H(\nabla_y)$ were respectively computed from the y -gradient accumulations in rows and columns, and, in the same way, $V(\nabla_x)$ and $H(\nabla_x)$ from x -gradients (Fig. 8).

The periodicity of these profiles is analyzed with *Fast Fourier Transform*, *FFT*, as introduced by the window width estimation (Liu and Galagowicz, 2010). If a periodicity exists in the structure alignment, then the *FFT* contains one significant peak. This was measured by computing the entropy \mathcal{E} of the spectrum, which provided us with four *repetitiveness* indicators. In this work, only the entropy of the profile $\mathcal{E}(V(\nabla_x))$ was analyzed.

3.4. Discussion

The computing time of the extraction of the set of these nine general features was about a fraction of a second for an image of 100,000 pixels with a single-core processor of 2.83 GHz.

Here gradient-based features were mainly used. In further work, an advanced signal processing method would be useful in order to better analyze the histograms and in particular detect specific cases such as vegetation occlusion, shadows, dirty spots, etc.

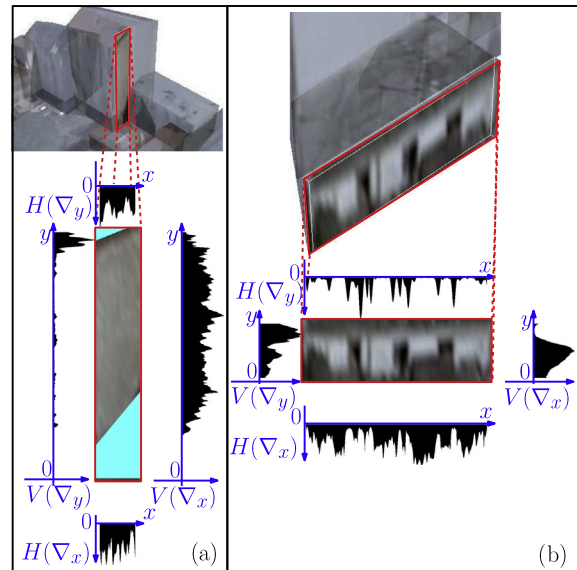


Fig. 8. (a) Gradient accumulation profiles of façade image A (*blind*). (b) Gradient accumulation profiles of façade image B (with openings). Top, left, bottom and right profiles: $V(\nabla_y)$, $H(\nabla_y)$, $V(\nabla_x)$ and $H(\nabla_x)$ respectively.

With a more general purpose, the noise influence could be reduced by using a pre-segmentation result as input of the whole feature extraction. The features could be inferred from the analysis of a set of segments, such as the one generated by *ultimate opening* introduced by (Hernández and Marcotegui, 2008; Marcotegui and Hernández, 2011) or by the *elementary models* of Burochin et al. (2010).

The two façade textures shown here are ideal cases. Other textures in the data are not so clear and easy to analyze on the basis of such general features. Thus, additional features are introduced in Section 4 based on the façade opening detection.

4. Opening detection

In addition to the general features described above, some indicators were computed so as to directly detect façade opening areas.

The aerial image resolution does not allow the use of specific models to represent them. Indeed, they often consist of mere dark spots. Thus, the opening detection was formulated as the generation of a set of dark rectangles on the image. A rectangular shape was used because it represents a suitable trade-off between simplicity and reality-fitting. This geometry also provides computation simplifications in order to independently work in horizontal and vertical directions.

Section 4.1 introduces an energetic formulation to evaluate the relevance of such an opening detection configuration. Section 4.2 describes the optimization process to find the best configuration which consists in a combination of a *reversible jump Markov Chain Monte Carlo sampling (rjMCMC)* (Hastings, 1970; Green, 1995) and of *simulated annealing* (such as described by Salamon et al. (2002)). This *top-down* approach can solve the problem of the low resolution by including a regularization component. Section 4.3 presents detection results.

As described in Section 3, the façade area is analyzed only outside the vegetation mask and outside the margins. In the illustrations of this paper, the margins are represented by same color than the area outside the façade (cyan).

4.1. Energetic formulation

Let the configuration R be composed of non-overlapping rectangles which are darker than the wall pane. This problem was formulated as the minimization of a configuration energy $E(R)$ of any configuration R :

$$E(R) = E_{\text{data}}(R) + \theta E_{\text{reg}}(R). \quad (7)$$

The configuration energy E is the linear combination of a *data term* (Section 4.1.1), which depends on image radiometry inside each rectangle, and a *regularization term* (Section 4.1.2), which is related to the rectangle layout relevance *a priori* independent of image data. The θ coefficient balances the two terms.

4.1.1. Data term

The data energy term is related to the relevance of each opening detection. A contrast energy E_{rect} (Eq. (9) and a constant term $E_{\text{rect}}^{\text{min}}$ are associated for each one:

$$E_{\text{data}}(R) = \sum_{r \in R} (E_{\text{rect}}^{\text{min}} - E_{\text{rect}}(r)). \quad (8)$$

The constant term $E_{\text{rect}}^{\text{min}}$ prevents the process from generating noisy dark rectangles, in a same way as the *Minimum Description Length (MDL)*, (Rissanen, 1983). This threshold is empirically fixed at a value greater than the façade wall radiometry noise (see Section 4.3.2 in order to get further information).

Let $E_{\text{rect}}(r)$ be the cost of rectangle r depending on contrast E_{edge} of its four edges.

The process analyzes the four edges independently of each other. This was essential in the problem as only the rectangles that really had four contrasted edges were required. The contrast product was used according to Eq. (9) so as to give the same weight to each edge and to set the cost to zero if one of the edges has a too low contrast.

$$E_{\text{rect}}(r) = \prod_{e \in \text{edge}(r)} E_{\text{edge}}(e). \quad (9)$$

A contrast measurement close to the one defined by Benedek et al. (2013) was used in order to robustly characterize rectangles by edge orthogonality. It was adapted to the problem of this work where horizontal and vertical directions were known. The x -gradient mean \bar{G}_x was used to compute vertical contrasts according to Eq. (10) (respectively y -gradient mean \bar{G}_y to compute horizontal ones).

$$E_{\text{edge}}(e) = \bar{G}_x(e) \rho_v(\bar{O}(e)). \quad (10)$$

The vertical contrast was weighted by a factor $\rho_v(\bar{O})$ which linearly decreases with the gradient orientation mean \bar{O} along the edge according to Eq. (11) (Fig. 9a).

$$\rho_v : \begin{cases} [0, \frac{\pi}{2}] & \rightarrow [0, 1] \\ \bar{O} & \mapsto \max(0, 1 - \frac{4}{\pi} \bar{O}) \end{cases} \quad (11)$$

Note that vertical gradient orientation ($O = \pi/2$) does not affect the contrast and that an edge with an orientation superior to $\pi/4$ is set to zero.

In the same way, the horizontal contrast was computed with a weight $\rho_h(\bar{O})$ expressed by Eq. (12) (Fig. 9b).

$$\rho_h : \begin{cases} [0, \frac{\pi}{2}] & \rightarrow [0, 1] \\ x & \mapsto \max(0, \frac{4}{\pi} x - 1) \end{cases} \quad (12)$$

Conversely to Benedek et al. (2013), whose analysis concerned aerial images of building roofs, the contrast measurement was not normalized because of local noise inside wall textures. Then façades was assumed to be acquired with correct illumination conditions: too dark façades were visually classified as *non classifiable* (cf. Section 2.4).

4.1.2. Regularization term

All opening detection were independently generated in the image. The regularization term prevents them from overlapping. This consists in the sum of all intersections according to Eq. (13).

$$E_{\text{reg}}(R) = \sum_{r \in R} \sum_{s \in R, s \neq r} r \cap s \quad (13)$$

The θ coefficient was fixed to a large value in Eq. (7) in order to remove any opening detection overlapping.

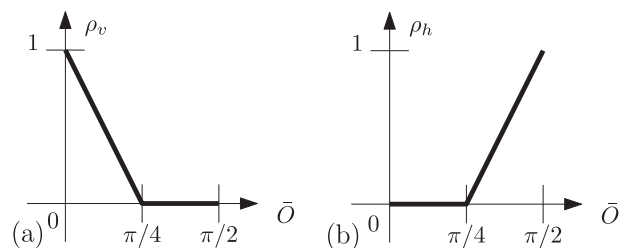


Fig. 9. Weight of contrast in function of the gradient orientation mean. (a) Vertical direction and (b) horizontal direction.

4.2. Configuration optimization

4.2.1. Related work

This subsection describes the generation process of the best non-overlapping opening detection configuration. It could not be performed by iteratively studying all configurations without using a very long computing time. Indeed, if L is the image length (width or height) and N is the maximal opening detection number, then the complexity of such an algorithm is exponential: $O(L^{4N})$, which is not possible in the current application.

The configuration energy E does not have any interesting properties such as convexity but there are optimization solutions for this kind of problem. First, deterministic methods could be used, such as greedy approaches: recursively detecting the darkest rectangle and removing its area from the search space until there are no more rectangles with a contrast energy greater than the threshold $E_{\text{rect}}^{\text{min}}$. The result of such an algorithm is not globally optimal because there is a risk of under-segmentation (one big dark rectangle located on several small less contrasted openings) but complexity is polynomial ($O(NL^4)$). There are also heuristics, inspired by natural processes, such as evolutionary algorithms (cf. Jong, 2006) that may converge more quickly toward interesting solutions.

The solution of the reversible jump Markov Chain Monte Carlo (rjMCMC) sampling scheme (Hastings, 1970; Green, 1995) was chosen combined with simulated annealing (Salamon et al., 2002) because of its rigorous stochastic scheme. This consisted of a detection method of objects modeled by geometrical parameters in a large dimension search space. The reversible jumps were required because the number of openings is unknown. The complexity was independent from the image area and the opening detection number (about a few million of configuration energy computations). The use of such a process has been developed in the literature during the last decade (Alegre and Dellaert, 2004; Ortner et al., 2007; Ripperda and Brenner, 2009; Tournaire et al., 2010).

The generic and efficient implementation of Brédif and Tournaire (2012)⁴ was used: see (Casella and Robert, 2004; Descombes, 2011) for additional information on the theoretical background.

This scheme consists in a marked point process which is a Markov Chain Monte Carlo sampling (MCMC) of unordered object sets. The process iteratively generates, moves and removes several points in the image. Each point corresponds to an opening detection offset (top left corner (x, y)), to which two marks are associated (a width w and a height h). Each state of the Markov Chain contains parameters of the whole configuration (points and marks). Transitions between states are the proposition kernels of the process (called reversible jumps (K_i)). At each iteration, the probability of accepting or rejecting the new configuration R' from the configuration R is determined by the acceptance Green ratio $\Gamma(R, R')$.

These successive rectangle configurations is sampled according to a Gibbs target probability distribution π_T at temperature T (See Section 4.2.2) and with a reference probability distribution π_{ref} :

$$\pi_T(R) = \frac{1}{Z} \pi_{\text{ref}}(R) e^{-E(R)/T}, \quad (14)$$

where Z is an unknown normalization constant.

Algorithm 1 summarizes this method at a relative constant temperature T and from a configuration R .

Algorithm 1. Pseudocode of the Metropolis–Hastings–Green sampling (Green, 1995).

repeat

Select a proposition kernel K_i according to probability $\pi_{kk'}$;
 Generate the sequence $u_{kk'}$ according to probability $\varphi_{kk'}(u)$;
 Build the candidate configuration $(R', v_{k'k}) = T_{kk'}(R, u_{kk'})$;
 Accept or reject R' according to probability $\Gamma(R, R')$;

until stationary regime

A suitable decrease of temperature T leads to an optimal configuration $R^* = \arg \max_R E(R)$ as described in Section 4.2.2. The reference probability distribution π_{ref} is explained in Section 4.2.3. Kernels and their respective Green ratio $\Gamma(R, R')$ computation are detailed in Appendix A.

4.2.2. Simulating annealing

In order to minimize the configuration energy E , the simulated annealing principle (Salamon et al., 2002) is the progressive penalty of the acceptance rate Γ toward a higher state of energy. This process is obtained by decreasing the temperature parameter T . At system initial state, T is the Infinite (a large initial value T_0 in practice). The sampling distribution π_T is then totally independent of the configuration energy E and follows the reference distribution π_{ref} . At the end of the process, T is nearly zero and the distribution π_T converges toward a Dirac distribution centered on global minima of the configuration energy E . This scheme provides an exploration of the largest possible part of search space so as to detect global minima.

Salamon et al. (2002) pointed out that the process convergence toward a global optimal is assured by a logarithmic temperature decrease and a high enough initial temperature (in order to let the process explore the whole search space and not to exclusively focus in a local optimum at the beginning). Yet, this theoretical scheme is too slow. In this work, a geometrical temperature was used to decrease $T_n = a^n T_0$, with the real a is related to the decrease speed and n is the iteration of the process. This choice was a good compromise between theory and practice and ensure us to reach the minimum close to the optimal one.

4.2.3. Reference distribution

The probability distribution function π_T (Eq. (14)), which samples opening detection configurations, depends on a reference distribution function π_{ref} . This reference distribution must sample the whole configuration search space and not to be zero on global minima of the configuration energy E .

In the application of this work, the opening detection number k was sampled by a Poisson process $P(k) = \lambda^k e^{-\lambda} / k!$, which only depends on a parameter λ , related to an a priori opening number mean ($\lambda = 5$ in the implementation: there are few openings). In addition, each opening detection $r = R[i] \in R, 1 \leq i \leq k$ was uniformly and independently sampled according to Eq. (15).

$$\pi_{\text{ref}}(R) = P(k) \prod_{i=1}^k P(r = R[i]). \quad (15)$$

Finally, both parameter pairs (x, w) and $(y, h) : P(r) = P(x, w)P(y, h)$ were uniformly and independently sampled so as to generate each opening detection r .

4.3. Results

The quality of the results depends on the value of contrast threshold $E_{\text{rect}}^{\text{min}}$ (Section 4.1.1) and on the evaluation method. Both

⁴ <http://logiciels.ign.fr/?-librjmcmmc.19->

choices are respectively discussed in Sections 4.3.1 and 4.3.2. The performance of the detection is showed in Section 4.3.3.

4.3.1. Evaluation method

There are different object detection evaluation methods. The simplest one consists in comparing an opening detection mask with a reference one. Bel-Hadj-Ali (2001) analyzed evaluation methods in the situation of precise geometrical data. He especially defined distances between detections and reference objects based on their intersection area.

Those methods are dedicated to very precise geometrical evaluation. In the application of this work, image textures were so stretched that even the reference data set was very inaccurate (about a third of the data set was considered *non classifiable*). Therefore, the evaluation was based on opening centers of the detections: an opening detection is a true positive if its center lay inside a reference opening (Fig. 10).

4.3.2. Contrast threshold

The influence of the contrast threshold, E_{rect}^{min} (Section 4.1.1, Eq. (8)) was analyzed based on the evolution of the precision and the recall. These two measures are well described in (Fawcett, 2006), and refer to the ratio between the true positive number – respectively the reference opening number – and the opening detection number. Precision linearly decreases with the recall until one threshold, then drastically falls, as shown in Fig. 11.

The contrast threshold was empirically set to the value 0.1 in order to obtain a suitable tradeoff between a relatively large recall and a sufficient precision.

4.3.3. Performance

The recall and the precision were about 64% and 42% respectively. This performance was relatively low mainly due to border problems (consequence of the rectification process, Section 2.3): it is discussed in Section 4.3.4.

In order to relax the constraint of the temperature decrease by reducing energy gaps between the bad detection location and the good one, the square root $\sqrt{E_{rect}}$ of rectangle energy was used in the contrast energy E_{rect} (Eq. (9)). It was a good compromise to fasten the detection process. The computing time was about twenty minutes per image with a single-core processor of 2.83 GHz, which

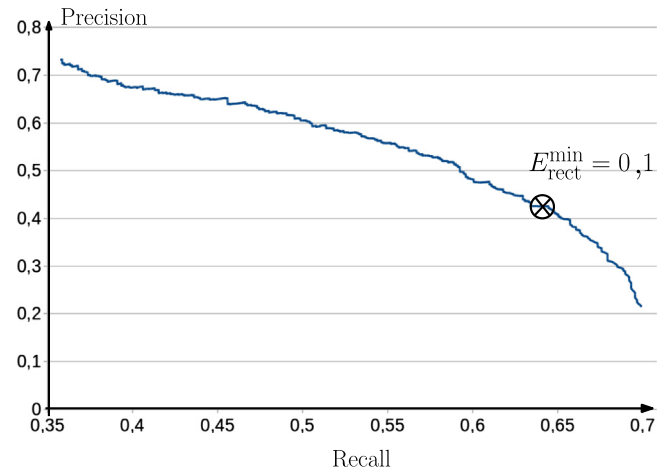


Fig. 11. Detection evaluation results as a function of contrast threshold E_{rect}^{min} .

is time-consuming with regard to the extraction time of the general features. The implementation was not especially optimized but the Section 4.3.5 propose future work.

4.3.4. Quality assessment

The recall and the precision are low but these results are biased by border problems which refer to the two thirds of the erroneously segmented images. On the one hand, the margin added to reduce the rectification problem is often not large enough to compensate for the texture overhang of outside elements (see Fig. 12a). On the other hand, reference openings have edges located inside the margin (see Fig. 12b). Therefore, these openings are not accessible by the sampling which only detects openings with enough gradient at each edge (Eq. (9)).

These two problems are linked with each other and the choice of the margin width seems not to significantly impact the quality of the result.

Remaining problems are due to the contrast. First, several façade textures contain dark spots which are not openings (because of the façade structure or piece of dirt or occlusion). When these spots are rectangular shapes, they may cause false alarms

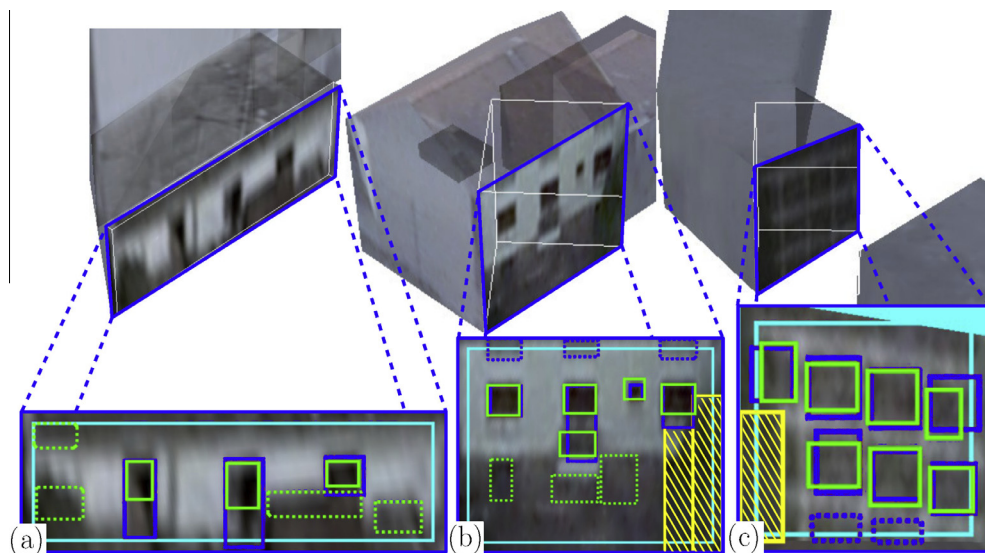


Fig. 10. Detection result. Reference openings are shown by blue rectangles (false negatives are shown in dashed lines). Continuous green rectangles are true positives (they contain a center of a reference opening). Dashed green rectangles are false positives. The occlusions (vegetation) are shown by hatched yellow rectangles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

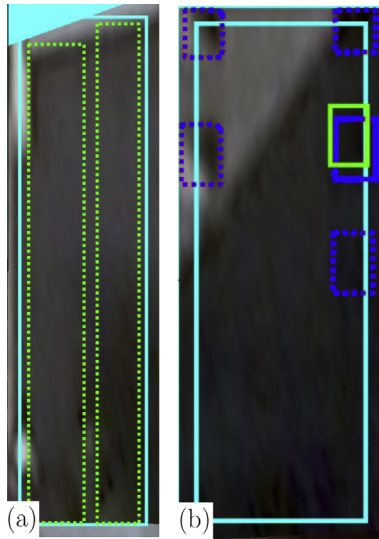


Fig. 12. Border problems: the reference data set consists of the continuous blue rectangles (false negatives are shown in dashed lines), the continuous green rectangle is a true positive whereas the dashed green ones are false positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 13a). Besides, several reference openings are very low-contrasted with the wall background (Fig. 13b). Hence their detection by the gradient-based approach was not ensured. Finally, there are few images with bright openings on a dark wall background (Fig. 13c). Contrary to the misleading contrast and the low-contrasted opening problems, this situation is relatively uncommon (note that even if the opening texture is brighter than the wall background, there are often perspective effects which result in dark shadows).

In order to significantly increase the quality of the detection, perspectives are proposed:

- To increase the recall, the contrast energy E_{rect} (Eq. (9)) could be adapted to weight the product of the contrast energies by the number of edges inside the façade area (instead of simply setting 0 at edges inside the margins). Border masks (more precise than the margin) also could be visually or automatically generated on a pre-processing step.
- This precision rate could be increased by the use of model fitting techniques or coarse classifier so as to reject non relevant detections once the configuration is built. Note that the model used in the generation step has to be as simple as possible in order to sample the space without spending a lot of time on sophisticated computation.

4.3.5. Computing time

The computing time of the detection process is twenty minutes per image (with a single-core processor of 2.83 GHz), which is highly superior to the general features extraction step. The implementation was not optimized but several hints on how to do it for larger scale purposes are developed.

First, initial temperature of the simulated annealing was deliberately set at a value higher than required. Indeed, because of a variety of exposure conditions, the process was configured to sample the whole radiometry dynamic without normalizing the contrast energy E_{rect} . For instance, opening detections in the image in Fig. 10a were validated sooner than the ones in the image in Fig. 10c. In the first case, the detection was nearly entirely obtained in the first iteration phase (in 1 min). Thanks to the high initial temperature, detection locations are relevant (global optimal configuration). No more opening detection were generated for some time. In the other case, the process waited until it reached a low enough temperature to keep the opening detections at high contrast location (maybe after fifteen minutes without detecting anything). In order to shorten the execution time of the detection technique, the decisive improvement would be to automatically set the initial temperature with regard to the statistical content of the whole texture as explained in (Salamon et al., 2002).

Furthermore, the efficient implementation (Verdié and Lafarge, 2014) is particularly well suited to the issue because each façade

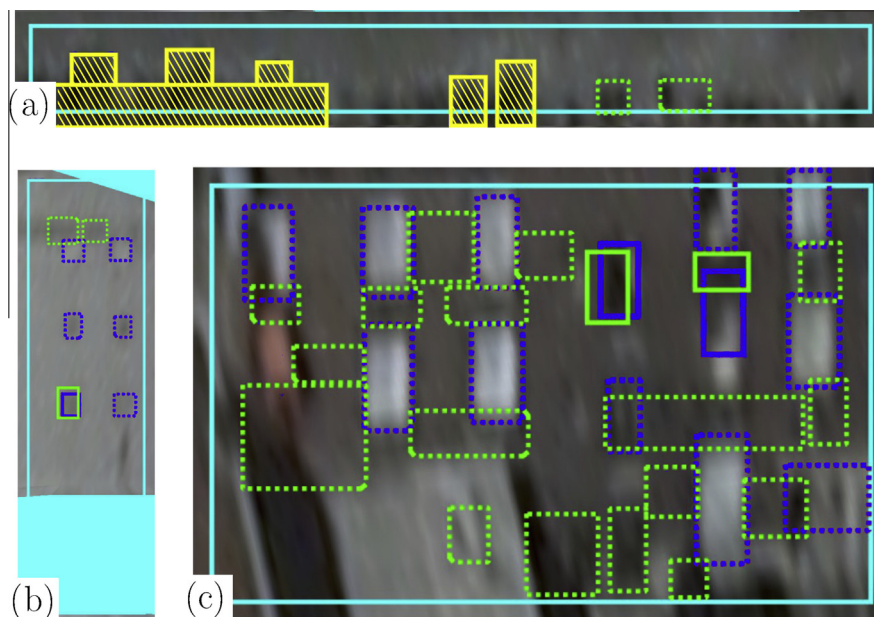


Fig. 13. Contrast challenge (the occlusions are shown by hatched yellow rectangles, other legend elements are explained in Fig. 12): (a) Misleading contrast; (b) low-contrasted openings; and (c) reverse-contrasted openings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

texture is independent from each other. In addition, a rough classification of the façades using one of the method mentioned in Section 1.1 could be used to narrow the detection process. Other sampling approaches might also improve the results such as the *Multiple Birth and Death* scheme – *MBD*, (Descombes, 2011; Benedek et al., 2013) since we deal with simple patterns.

The data term E_{data}^* of the optimal configuration and the contrast energy E_{rect}^* of the most contrasted rectangle were used as local indicators. Despite the limited relevance of their location, they were proved to provide meaningful *blind façade* indicators to this study, as described in the following section.

5. Blind façade classification

Blind façades are those without an opening. This section introduces a classification approach to identify them. First, the feature set (Section 5.1) is represented by a space \mathcal{X} related to all the façade features which have been described above. The choice of a supervised learning process is motivated in Section 5.2 and the partition of the data set is explained. The evaluation method is also introduced in this section. Section 5.3 introduces the classification model: *Support Vector Machines – SVM*, (Cortes and Vapnik, 1995) – and the kernel and the parameter selection are described. An optimization is added to this classification framework (Pal and Foody, 2010) by selecting features that are relevant and removing the other ones, as explained in Section 5.4. The trained classifier is then tested on the test set as demonstrate in Section 5.5: about 85% of correct classification. Finally, this result is discussed in Section 5.6 and perspectives are proposed.

5.1. Data representation

A representation of façade images is proposed by using features previously introduced: number of opening detections and their contrast as described in Section 4 and general features explained in Section 3. Each image is related to a vector x in a feature space \mathcal{X}_{11} according to Eq. (16):

$$x = \begin{pmatrix} \bar{I} \text{ (radiometry mean, Equation 2)} \\ \sigma_I \text{ (uniformity, Equation 1)} \\ v_0 \text{ structuredness (Section 3.2)} \\ v_1 \text{ structuredness ((Equation 3)} \\ v_2 \text{ structuredness (Equation 3)} \\ \Delta_{vertical} \text{ (vertical aspect, Equation 5)} \\ \Delta_{horizontal} \text{ (horizontal aspect, Equation 4)} \\ \sigma_{\mathcal{T}} \text{ structuredness (Equation 6)} \\ \Xi(V(\nabla_x)) \text{ (repetitiveness, (Section 3.3)} \\ E_{rect}^* \text{ (opening contrast max, Equation 9)} \\ E_{data}^* \text{ (opening contrast sum, Equation 8)} \end{pmatrix} \in \mathcal{X}_{11} \quad (16)$$

5.2. Supervised learning

The high variety of the façade and their representation in the images require a supervised approach in order to better capture this high heterogeneity. Indeed there exists human rules which cannot be automatically inferred from the images without additional information. Sometimes *blind façades* look like façades with openings because of local noise (vegetation, dirt, ornament, shadows, etc.): see Fig. 18 for example. Therefore the classifier was fed by a reference data set visually built by an expert (see Section 2.4).

As training set, the 80% of the *classifiable façade* set were selected, which corresponds to 363 façade images. The process

was tested with the other *classifiable façades* (90 images). This proportion is a good compromise between generalization capacity and reliability of the error estimate, as explained in (Abu-Mostafa et al., 2012). Both façade types are equally represented in each set.

In this paper, the proportion of correctly classified *blind façades* was practically as critical as the proportion of correctly classified façades with openings. Thus, rates such as recall, precision or *F*-measures could not be used without adding a bias in the evaluation. Only the classification accuracy (overall accuracy) was used with the cross validation, that is to say, the ratio between the number of correctly classified images (*blind* or with openings) and the total number of images (Fawcett, 2006).

5.3. Model parameter setting

A standard solution is to adopt *Support Vector Machines – SVM* (Cortes and Vapnik, 1995). This classifier projects the feature space into a higher-dimensional space where an hyperplane was computed so as to separate the two classes. Its main assets is to maximize the margin between these two classes, which theoretically reduces the generalization error. A review of such a tool is proposed by Mountrakis et al. (2011).

The regularized-classification implementation (C-SVC) of Chang and Lin (2011) was chosen and both polynomial and radial basis kernel function were tested. These two kernels are considered in (Hsu et al., 2010) to be the most appropriate ones when the feature space geometry is unknown.

The kernel parameters and the regularization parameter were selected by a 10-fold cross-validation in the training set. This step simply sampled the values in a finite set in order to select an approximate solution. The cross validation error of both kernels where approximately the same. The radial basis kernel function was selected because it has only one parameter.

5.4. Feature selection in forward search

In order to reduce the generalization error (Pal and Foody, 2010), a feature selection step was implemented so as to retrieve the best feature space \mathcal{X}^* . A forward search was used by iteratively adding the features in the SVM, starting by the most relevant one.

At the first step, all the feature singleton spaces \mathcal{X}_1 were tested. Fig. 14 shows the cross validation error of each of them: the *uniformity* feature obtained the best result.

Next, each feature couple space $\mathcal{X}_2 = \mathcal{X}_1^* \times \mathcal{X}'_1$ was tested based on the best feature singleton space \mathcal{X}_1^* of the previous step and another one $\mathcal{X}'_1 \neq \mathcal{X}_1^*$.

The best feature space \mathcal{X}_i^* was recursively extended in order to test each $\mathcal{X}_{i+1} = \mathcal{X}_i^* \times \mathcal{X}'_i$ in this way until testing the full feature space \mathcal{X}_{11} . Finally, Fig. 15 shows the evolution of the cross validation error at each step of the feature selection.

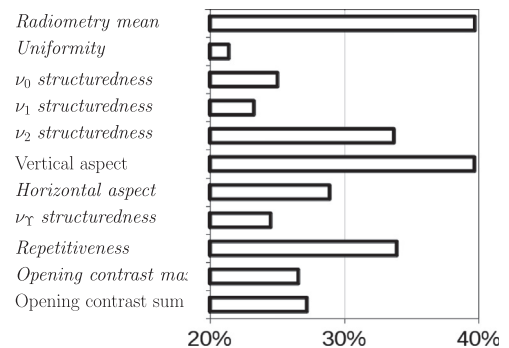


Fig. 14. Cross-validation error of each feature singleton (\mathcal{X}_1).

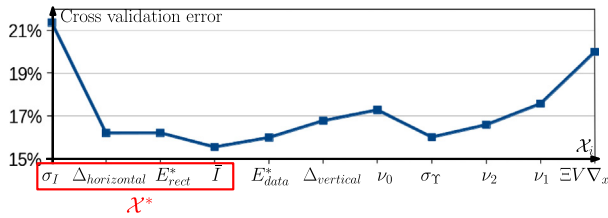


Fig. 15. Feature selection in forward search based on the cross validation error.

For the training set, the best feature space was obtained at the fourth step with both radiometric features, horizontal aspect ($\Delta_{horizontal}$, Eq. (4)) and maximal contrast (E_{rect}^*). With this best feature set, cross validation error was decreased at less than 16%.

$$x = \begin{pmatrix} \text{uniformity} \\ \Delta_{horizontal} \\ E_{rect}^* \\ \text{radiometry mean} \end{pmatrix} \in \mathcal{X}^* \quad (17)$$

No other feature could be added without increasing the cross validation error.

5.5. Results

The images of the test set were classified with the best feature set (in about one second with a single processor of 2.83 GHz). The accuracy was about 86% and the algorithm classified better the façades with openings than *blind façades* as shown in the contingency table (Table 1).

74% of the *blind façades* was correctly classified and 90% of the façade classified as *blind* were actually *blind* (Figs. 4b and 16). Note that it may happens that a façade is classified as *blind* even if several openings were detected (see Fig. 16).

Generally, when the façade has openings, then these latter are detected but are not correctly delineated (opening detection precision of 42%). Correct classification results of this kind of façades are presented in Fig. 17. Three façades with openings were classified as *blind* because their texture was as uniform as the basic *blind façades* and the contrast of their openings was less than the majority of the contrast of the openings in the train sample (Figs. 18 and 21). Regarding *blind façades*, the process failed because of border problems (Figs. 5a and 19) and vegetation (Fig. 20). On Fig. 19, it does not really detect any opening but the features related to the radiometric features and the *horizontal aspect* were biased by the projection of close buildings at the border.

5.6. Discussion

5.6.1. Generalization performance

The generalization performance of the classifier can be estimated by the two error measures described above:

- The 10-fold cross validation error in the training set: about 16%.
- The validation accuracy in the test set: about 14%.

Note that the proportion of support vectors is about 40%, which predicts a theoretical generalization performance lower than these

Table 1 Contingency table of the classification of the test set: number of façades.

	Blind	Not blind
classified as <i>blind</i>	28	3
classified as <i>not blind</i>	10	49

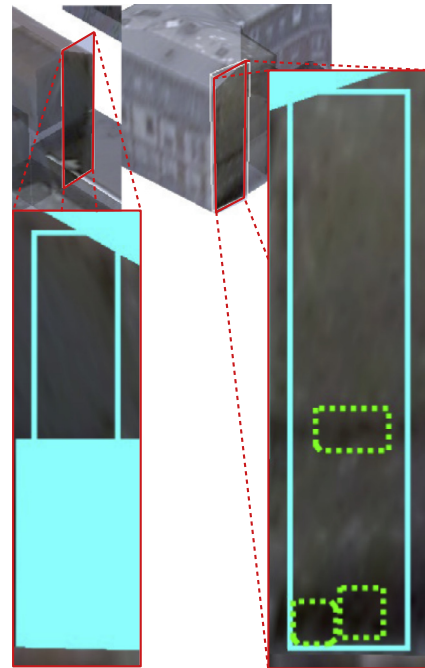


Fig. 16. Correct classification of *blind façades*. The dashed green rectangles are opening detections (false positives).

two empirical estimates (Cortes and Vapnik, 1995). Misclassification are mainly due to pre-processing problems: vegetation and border problems but an improvement could be added to the classification framework itself, as discussed in Section 5.6.2.

5.6.2. Feature selection consistency

The generalization error was decreased by the feature selection. Features which have been set apart are discussed in this paragraph. First, *Repetitiveness* was a poor discriminative feature by itself in this situation (Fig. 14) because the data contained few repetitive structures. But, when the situation occurs, this measure must be adapted in order to focus on façade which really do have regular opening repetitions such as the one shown in Fig. 21.

The process genuinely detected all the openings but the classifier took in consideration only the highest contrast. If this highest contrast is lower than the contrast of other openings and if it has the same value than the contrast of occlusions (shadows, vegetation, dirt or little structures such as Fig. 16), then the classifier failed without further analysis. An interesting solution would be to carry on the whole opening set if the façade have several ones. Indeed, the *repetitiveness measure* is currently based on the entropy of the *FFT* spectrum, which is maybe too progressive among the façade features. The distinction between very repetitive openings and façades with few openings is too soft. A harder *repetitiveness measure* would probably be accepted by the feature selection process and would have prevent the classifier from failing (by a mode extraction technique, for example, such as Gaussian regression).

Apart from this feature, the accumulated contrast of the opening detections was not selected: this probably means that several *blind façades* have an accumulated contrast noise as high as the contrast of most openings. Regarding structuredness and gradient orientation, these features might have redundancy with the horizontal aspect feature.

Naturally, once the resampling border problems are fixed and the vegetation detected, different features than these ones will certainly be selected.

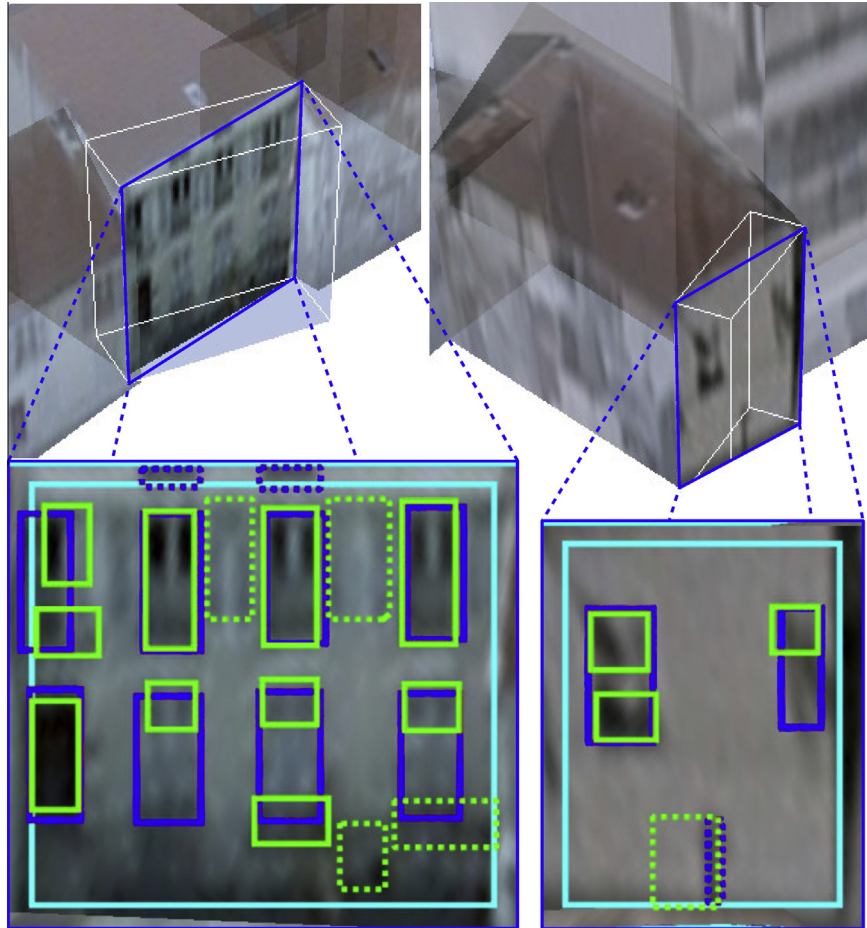


Fig. 17. Correct classification of façades with openings. The reference opening location consists of the continuous blue rectangles (opening false negatives are shown in dashed lines), the continuous green rectangles are opening true positives whereas the dashed green ones are false positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

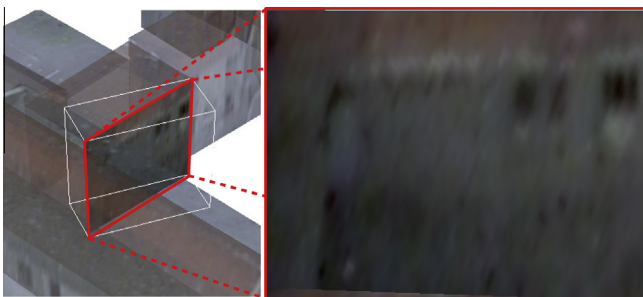


Fig. 18. Classification error due to low contrast. A façade with a uniform background and one opening with very low contrast will not lead to a successful detection.

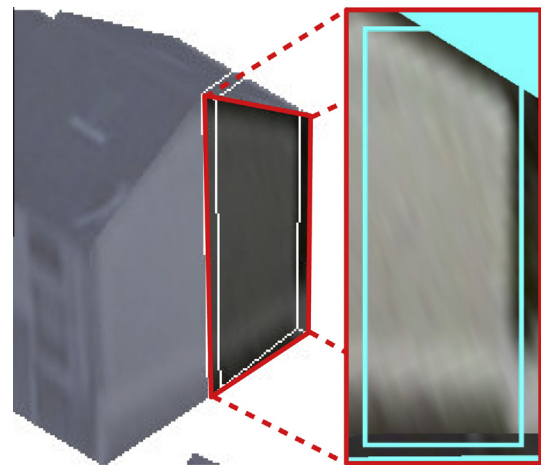


Fig. 19. Classification error due to geometrical occlusion. The projection of close buildings may hinder the classification of the blind façade.

5.6.3. Non classifiable façades and probability estimate

Finally, the *non classifiable façade image set* (cf. Section 2.4) was not analyzed in details because of its very bad classification capacity: about 40% with SVM and three-classes (*blind*, *not blind*, *non classifiable*) and about 30% for the two-classes problem (*classifiable vs non classifiable*). Additionally, we can note that:

- None of the features has sufficient discrimination power.
- Neither the direction of the camera, nor the direction of the theoretical sunlight with regard to the façade plane seemed to have an impact on the nature of this third class.

For further work, a probability distribution on the classification result could be generated from the *non classifiable façade set* by merging the *classifiable score* with a *blindness score* provided by the *bagging* framework or the distance to the hyperplane described in (Wu et al., 2004). Thus, when the *blindness probability* is less than a fixed probability, the process would ask an operator to

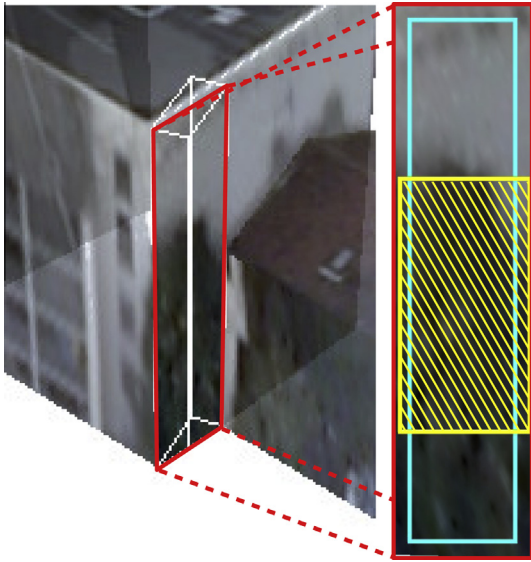


Fig. 20. Classification error due to vegetation occlusion. The occlusion mask (hatched yellow rectangle) is not large enough to cover the entire vegetation area on the *blind façade*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

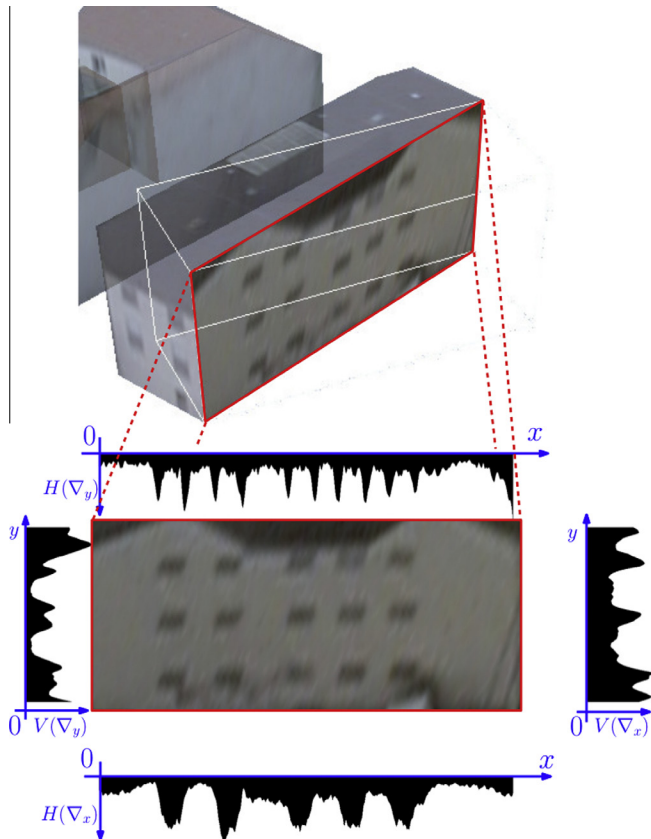


Fig. 21. Classification error: façade with a uniform background and repetitive openings with low contrast. Gradient accumulation computations are explained in Section 3.3. The process detects all the openings but fails in the classification: the *repetitiveness* measure must then be adapted so as to prevent this situation.

classify the border line façade. In this sense, Li et al. (2012) proposed an active learning approach based on an interesting SVM extension: *HintSVM*. This latter locates a query boundary among the unlabeled instances according to a measure of uncertainty and a measure of representativeness.

6. Conclusion

A supervised approach for detecting *blind façades* (i.e. without opening – window or door), was proposed in this paper using wide angle yet vertical aerial images with a decimeter pixel size acquired by standard nadir looking cameras. Such discrimination is highly at stake for urban planners so as to estimate the potential impact of urban regulations and setting up public policies.

Two complementary façade feature sets were designed in this very specific context, namely general and local features. The first one corresponded to general radiometric and geometrical information about the façade texture, tailored to deal with the lack of contour sharpness. The second one was extracted by a stochastic opening detection based on the local contrast in the images. This study demonstrated that the combination of these features is highly relevant in order to provide meaningful supervised classification results. Indeed, the classification accuracy was about 85%, using such challenging data containing a large number of stretched façade textures. These results were very sensitive to the texturing resampling process and to vegetation presence on façade textures.

The execution time of the full work-flow was about twenty minutes per image, mainly due to the opening detection step. In particular, it can be reduced by further tuning the initial temperature of the simulated annealing step with regard to the statistical content of the whole texture as explained in (Salamon et al., 2002).

Furthermore, other relevant façade features could be added such as those proposed by Yang et al. (2012) based on more sophisticated statistical information about the color space, the geometry and the texture. Finally, various camera configurations and positions (oblique VS vertical) can be tested so as to find the most relevant configurations for such façade analysis.

Acknowledgment

We wish to thank VECTUEL® for providing the aerial photos used in this study.

Appendix A. Reversible jump implementation

In order to detect openings, the rectangle configuration energy E was minimized (Eq. (7)). A *reversible jump Markov Chain Monte Carlo* (rjMCMC) sampling scheme (Hastings, 1970; Green, 1995) was combined with a *simulated annealing* (Salamon et al., 2002). This appendix provides the definition of *reversible jumps* and *Green ratio* in Section A.1 and their computation in Section A.2.

A.1. Definitions

Casella and Robert (2004) formalized the following reversible jump structure:

A probability $\pi_{kk'}$ to the kernel selection from state (k, R) toward state (k', R') . Let $\pi_{k'k}$ be the reciprocal probability ($\pi_{k'k} = 1 - \pi_{kk'}$).

A variable sequence $u_{kk'}$. If the starting and ending configuration spaces do not have the same dimensions, Green (1995) proposed to add additional random variables inside the lower-dimension space. Thus, a sequence $u_{kk'}$ of such variables was defined which completed other parameters of the k rectangles of configuration R . Respectively, an additional random variable sequence $v_{kk'}$ of the k' rectangles of configuration R' was defined. In addition, the probability density functions $\varphi_{kk'}(u)$ and $\varphi_{k'k}(v)$ respectively were defined, related to these sequences.

A bijection $T_{kk'}$ between the current state and the candidate state such that $T_{kk'}(R, u_{kk'}) = (R', v_{k'k})$. Let $T_{k'k} = T_{kk'}^{-1}$ be the inverse bijection.

Then the kernel theoretically satisfies the reversibility condition if the configuration R' from configuration R is selected according to acceptance rate $\Gamma(R, R')$ expressed by Eq. (A.1).

$$\Gamma(R, R') = \min \left(1, \frac{\pi_{\text{ref}}(R')}{\pi_{\text{ref}}(R)} \frac{\pi_{k'k} \varphi_{k'k}(v_{k'k})}{\pi_{kk'} \varphi_{kk'}(u_{kk'})} \left| \frac{\delta T_{kk'}(R, u_{kk'})}{\delta(R, u_{kk'})} \right| e^{\frac{E(R) - E(R')}{T}} \right) \quad (\text{A.1})$$

A.2. Computations

The number of opening detections is unknown. Transition kernels of type *birth* and *death* iteratively modified it by respectively adding and removing rectangles (Appendix A.2.1). This is theoretically sufficient to explore the entire search space (Geyer and Møller, 1994). This succession of proposals yields an ergodic Markov Chain which.

Other local modification kernels were also used to speed up the convergence: a kernel type in order to stretch and contract rectangles (Appendix A.2.2) and another one in order to split and merge them (Appendix A.2.3). These two kinds of kernels allow for local variations to a rectangle in one iteration (instead of two: a death, then a birth).

Let $r = (x, y, w, h)$ be the rectangle with top left corner (x, y) , the width w and the height h .

Linear transformations $\hat{r} \rightarrow \hat{r}'$ are expressed related to each kernel, which locally disrupt one object \hat{r} to obtain a candidate \hat{r}' . Their ratio $\Gamma(R, R')$ also is computed so as to accept or reject this transformation according to Eq. (A.1).

A.2.1. Birth and death kernels

The *birth* kernel generates a rectangle candidate \hat{r}' in the k -rectangle configuration R . Then a sequence of four additional variables $u_{kk'}[i] \in [0, 1]$, $1 \leq i \leq 4$ is required to generate this candidate from an additional object $\hat{r} = (u_{kk'}[1], u_{kk'}[2], u_{kk'}[3], u_{kk'}[4])$.

Let W be the image width and H be its height. A new $k' = (k + 1)$ -rectangle configuration candidate R' is then generated, containing the new rectangle \hat{r}' , according to Eq. (A.2). Note that additional variables are sampled such that $u_{kk'}[1] + u_{kk'}[3] < 1$ and $u_{kk'}[2] + u_{kk'}[4] < 1$ in order to constraint the candidate \hat{r}' in the image area.

$$\hat{r}' = (u_{kk'}[1]W, u_{kk'}[2]H, u_{kk'}[3]W, u_{kk'}[4]H) \quad (\text{A.2})$$

Eq. (A.3) then expresses the acceptance rate $\Gamma_{\text{birth}}(R, R')$ of a configuration candidate R' proposed by this kernel.

$$\Gamma_{\text{birth}}(R, R') = \min \left(1, \frac{\lambda}{k+1} \frac{\pi_{k'k}}{\pi_{kk'}} e^{\frac{E(R) - E(R')}{T}} \right) \quad (\text{A.3})$$

The reciprocal operation is obtained by the *death* kernel. Its acceptance rate can be deduced in an analogous way.

A.2.2. Edge translation kernels

One way to move a rectangle or to resize it is to translate one of its edges. Four kernels are used to make such shifts. This paragraph describes the computation of the right edge; translation of other edges can be deduced in an analogous way.

The problem is to horizontally translate the right edge of a candidate \hat{r} (Fig. A.22).

An additional variable $u \in [0, W]$ in the rectangle $\hat{r} = (x, y, w, h, u)$ is required to obtain the new width $w' = u$ of the candidate $\hat{r}' = (x, y, u, h, w)$. Note that the reciprocal operation is the same as the direct one.

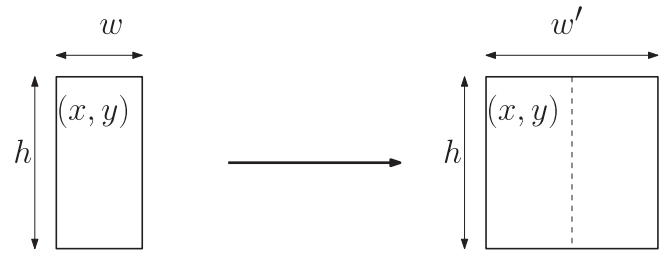


Fig. A.22. Right edge translation kernel.

Eq. (A.4) expresses the acceptance rate $\Gamma_{\text{trans}}(R, R')$ of a configuration candidate R' proposed by this kernel.

$$\Gamma_{\text{trans}}(R, R') = \min \left(1, e^{\frac{E(R) - E(R')}{T}} \right) \quad (\text{A.4})$$

with

$$\begin{cases} \frac{\pi_{\text{ref}}(R')}{\pi_{\text{ref}}(R)} = 1 \\ \frac{\pi_{k'k}}{\pi_{kk'}} = 1 \\ \frac{\varphi_{k'k}(v_{k'k})}{\varphi_{kk'}(u_{kk'})} = 1 \\ \left| \frac{\delta T_{kk'}(R, u_{kk'})}{\delta(R, u_{kk'})} \right| = 1 \end{cases}$$

A.2.3. Split and merge kernels

When the contrast energy $E_{\text{rect}}(r_0)$ of a rectangle r_0 is lower than the contrast energy sum such that $(\sum_n E_{\text{rect}}^{\text{min}} - E_{\text{rect}}(r_n)) < E_{\text{rect}}^{\text{min}} - E_{\text{rect}}(r_0)$ of smaller rectangles $(r_n)_n$ inside r_0 , the optimal process is to split the latter. Reciprocally, when an opening is over-segmented by several rectangles, the optimal process is to merge them in order to cover the whole opening area if its contrast energy is higher. Thus, two split–merge kernel pairs are introduced so as to favor this optimization shortcut.

For the split kernel, two fixed corners are kept: each one determines the location of one of the two new rectangles (Fig. A.23).

Here the first version of the split kernel is described with fixed top-left and bottom-right corners. The demonstration for the other kernel is similar.

From a k -rectangle configuration R , a rectangle \hat{r} is split into a rectangle pair \hat{r}' . A $(k' = k + 1)$ -rectangle configuration R' is obtained with a dimension bigger than the first one: there were four more parameters. Then a 4-additional variables sequence $u_{kk'} = (\alpha_1, \alpha_2, \beta_1, \beta_2) \in [0, w]^2 \times [0, h]^2$ is required in the candidate $\hat{r} = (x, y, w, h, \alpha_1, \alpha_2, \beta_1, \beta_2)$. Here the latter is split into a top rectangle r_1 and a bottom one r_2 according to Eq. (A.5).

$$r_1 = \begin{pmatrix} x \\ y \\ \alpha_1 \\ \beta_1 \end{pmatrix} r_2 = \begin{pmatrix} x + w - \alpha_2 \\ y + h - \beta_2 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \quad (\text{A.5})$$

Note that the additional variables are constrained according to Eq. (A.6) in order to ban overlapping candidates.

$$\alpha_1 + \alpha_2 < w \text{ or } \beta_1 + \beta_2 < h \quad (\text{A.6})$$

Thus, Eq. (A.7) expresses the acceptance rate $\Gamma_{\text{split}}(R, R')$ of a configuration candidate R' proposed by this kernel.

$$\Gamma_{\text{split}}(R, R') = \min \left(1, 12wh \frac{\lambda}{k+1} \frac{\pi_{k'k}}{\pi_{kk'}} e^{\frac{E(R) - E(R')}{T}} \right) \quad (\text{A.7})$$

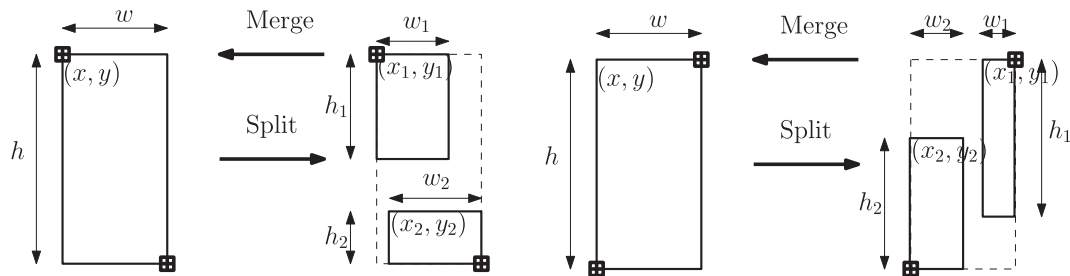


Fig. A.23. Split and merge kernels.

In an analogous way, the merging operation from two rectangles $r_1 = (x_1, y_1, w_1, h_1)$ and $r_2 = (x_2, y_2, w_2, h_2)$ can be realized by Eq. (A.8).

$$\tilde{r}' = (x_1, y_1, x_2 - x_1 + w_2, y_2 - y_1 + h_2, w_1, w_2, h_1, h_2) \quad (\text{A.8})$$

References

- Abu-Mostafa, Y.S., Magdon-Ismael, M., Lin, H.-T., 2012. Learning From Data. AMLBook.com, USA.
- Alegre, F., Dellaert, F., 2004. A probabilistic approach to the semantic interpretation of building facades. In: Proc. International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres. CIPA, Lisbon, Portugal, October 25–27.
- Bel-Hadj-Ali, A., 2001. Positional and shape quality of areal entities in geographic databases: quality information aggregation versus measures classification. In: ECSQARU Workshop on Spatio-Temporal Reasoning and Geographic Information Systems. Toulouse, France, September 19–21.
- Benedek, C., Descombes, X., Zerubia, J., 2013. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34 (1), 33–50.
- Bozonnet, E., Musy, M., Calmet, I., Rodriguez, F., 2013. Modeling method to assess urban fluxes and heat island mitigation measures from street to city scale. Int. J. Low-Carbon Technol.
- Brasebin, M., Perret, J., Haëck, C., 2011. Towards a 3D geographic information system for the exploration of urban rules: application to the French local urban planning schemes. In: Urban Data Management Symposium. UDMS, Delft, The Netherlands, September 28–30, pp. 37–50.
- Brédif, M., Tournaire, O., 2012. Librjmc: An open-source generic c++ library for stochastic optimization. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. XXXIX (Part 3), 259–264.
- Burochin, J.-P., Vallet, B., Tournaire, O., Paparoditis, N., 2010. A formulation for unsupervised hierarchical segmentation of facade images with periodic models. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. XXXVIII (Part 3A), 227–232.
- Casella, G., Robert, C.P., 2004. Monte Carlo Statistical Methods, second ed. Springer Verlag, New York, USA.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (27), 1–27.
- Cortes, C., Vapnik, V.N., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, San Diego, CA, USA, pp. 20–26.
- Descombes, X., 2011. Stochastic Geometry for Image Analysis. Wiley-ISTE, Hoboken, USA and London, UK.
- Drauschke, M., Mayer, H., 2010. Evaluation of texture energies for classification of facade images. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. XXXVIII (Part 3A), 257–262.
- Durupt, M., Taillandier, F., 2006. Automatic building reconstruction from a digital elevation model and cadastral data: an operational approach. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. XXXVI (Part 3), 142–147.
- Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2012. Scene parsing with multiscale feature learning, purity trees, and optimal covers. In: Proc. International Conference on Machine Learning. IMLS, Edinburgh, Scotland, June 26–July 1, pp. 575–582.
- Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.
- Geyer, C.J., Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes. Scandinavian J. Stat. 21 (4), 359–373.
- Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.
- Han, Y., Liu, G., 2013. Biologically inspired task oriented gist model for scene classification. Comput. Vision Image Understand. 117 (1), 76–95.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57 (1), 97–109.
- Haugeard, J.-E., Philipp-Foliguet, S., 2012. Kernel on graphs based on dictionary of paths for image retrieval. In: International Conference on Pattern Recognition. IAPR, Istanbul, Turkey, August 23–26.
- Hernández, J., Marcotegui, B., 2008. Segmentation of facade images using ultimate opening. In: International Society of Stereology. ISS, Paris, France, February 3.
- Hsu, C.-w., Chang, C.-C., Lin, C.-J., 2010. A Practical Guide to Support Vector Classification. <<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>> (accessed 30.01.2014).
- Jong, K., 2006. Evolutionary Computation: A Unified Approach. MIT Press, Cambridge, UK.
- Korč, F., Förstner, W., 2008a. Finding optimal non-overlapping subset of extracted image objects. In: Proc. International Workshop on Combinatorial Image Analysis, Buffalo, NY, USA, April 7–9.
- Korč, F., Förstner, W., 2008b. Interpreting terrestrial images of urban scenes using discriminative random fields. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. 37 (part B3a), 291–296.
- Kriegler, F.J., Malila, W.A., Nalepka, R.F., Richardson, W., 1969. Preprocessing transformations and their effects on multispectral recognition. In: Proc. Remote Sensing of Environment, October 13–16, pp. 97–131.
- Kumar, S., Hebert, M., 2003. Man-made structure detection in natural images using a causal multiscale random field. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Madison, Wisconsin, USA, pp. 16–22.
- Li, C.-L., Ferng, C.-S., Lin, H.-T., 2012. Active learning with hinted support vector machine. JMLR W&CP 25, 221–235.
- Liu, C., Gagalowicz, A., 2010. Image-based modeling of Hausmannian facades. Int. J. Virtual Real. 9, 13–18.
- Marcotegui, B., Hernández, J., 2011. Shape ultimate attribute opening. Image Vision Comput. 29 (8), 533–545.
- Meixner, P., Leberl, F., 2010a. Describing buildings by 3-dimensional details found in aerial photography. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. XXXVIII (Part 3A), 151–156.
- Meixner, P., Leberl, F., 2010b. Vertical or oblique aerial photography for semantic building interpretation. Photogrammetrie, Fernerkundung und Geoinformation 19, 247–256.
- Meixner, P., Leberl, F., 2011. 3-Dimensional building details from aerial photography for internet maps. Remote Sens. 3, 721–751.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. 66 (3), 247–259.
- Müller, P., Zeng, G., Wonka, P., Van Gool, L., 2007. Image-based procedural modeling of facades. ACM Trans. Graph. 26 (3), 85.
- Nyaruhuma, A.P., Gerke, M., Vosselman, G., Mtalio, E.G., 2012. Verification of 2D building outlines using oblique airborne images. ISPRS J. Photogramm. Remote Sens. 71, 62–75.
- Ortner, M., Descombes, X., Zerubia, J., 2007. Building outline extraction from digital elevation models using marked point processes. Int. J. Comput. Vision 72 (2), 107–132.
- Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Trans. Geosci. Remote Sens. 48 (5), 2297–2307.
- Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: a multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. Revue Française de Photogrammétrie et de Télédétection 200, 69–79.
- Reznik, S., Mayer, H., 2008. Implicit shape models, self-diagnosis, and model selection for 3D facade interpretation. Photogramm. – Fernerkundung – Geoinf. 4, 187–196.
- Ripperda, N., Brenner, C., 2009. Application of a Formal Grammar to Facade Reconstruction in Semiautomatic and Automatic Environments. In: Proc. International Conference on Geographic Information Science. AGILE Conference on GIScience, Hannover, Germany, June 2–5.
- Rissanen, J., 1983. A universal prior for integers and estimation by minimum description length. Ann. Stat. 11 (2), 416–431.
- Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3d building reconstruction. ISPRS J. Photogramm. Remote Sens. 93, 256–271.
- Salamon, P., Sibani, P., Frost, R., 2002. Facts, Conjectures, and Improvements for Simulated Annealing. Society for Industrial and Applied Mathematics, No. 7, Philadelphia, PA, United States.

- Teboul, O., Kokkinos, I., Koutsourakis, P., Paragios, N., 2013. Parsing facades with shape grammars and reinforcement learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (7), 1744–1756.
- Tournaire, O., Brédif, M., Boldo, D., Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. *ISPRS J. Photogramm. Remote Sens.* 65 (4), 317–327.
- Vallet, B., Houzay, E., 2011. Fast and accurate visibility computation in urban scenes. *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.* 38 (Part 3/W22), 77–82.
- Verdié, Y., Lafarge, F., 2014. Detecting parametric objects in large scenes by Monte Carlo sampling. *Int. J. Comput. Vision* 106 (1), 57–75.
- Wu, T.-f., Lin, C.-J., Weng, R.C., 2004. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* 5, 975–1005.
- Xiao, J., Gerke, M., Vosselman, G., 2012. Building extraction from oblique airborne imagery based on robust faade detection. *ISPRS J. Photogramm. Remote Sens.* 68, 56–68.
- Yang, M.Y., Förstner, W., Chai, D., 2012. Feature evaluation for building facade images: an empirical study. *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.* XXXIX (Part 3), 513–518.
- Zhang, W., Košecká, J., 2007. Hierarchical building recognition. *Image Vision Comput.* 27 (3), 704–716.