AN UNSUPERVISED HIERARCHICAL SEGMENTATION OF A FAÇADE BUILDING IMAGE IN ELEMENTARY $2D$ - MODELS

Jean-Pascal Burochin, Olivier Tournaire and Nicolas Paparoditis

Université Paris-Est, Institut Géographique National, Laboratoire MATIS
73 Avenue de Paris, 94165 Saint-Mandé Cedex, France
{firstname.lastname}@ign.fr

Commission III/3, III/4

KEY WORDS: street level imagery, façade reconstruction, unsupervised hierarchical segmentation, gradient accumulation, recursive split, model matching

ABSTRACT:

We introduce a new unsupervised segmentation method adapted to describe façade shapes from a single calibrated street level image. The image is first rectified thanks to its vanishing points to facilitate the extraction of façade main structures which are characterized by a horizontal and vertical gradient accumulation which enhances the detection of repetitive structures. Our aim is to build a hierarchy of rectangular regions bounded by the local maxima of the gradient accumulation. The algorithm recursively splits horizontally or vertically the image into two parts by maximizing the total length of *regular edges* until the radiometric content of the region hypothesis corresponds to a given model (planar and generalized cylinders). A *regular edge* is a segment of a main gradient direction that effectively matches to a contour of the image. This segmentation could be an interesting tool for façade modelling and is in particular well suited for façade texture compression.

1 INTRODUCTION

1.1 Context

Façade analysis (detection, understanding and reconstruction) from street level imagery is currently a very active research domain in the photogrammetric computer vision field. Indeed, it has many applications. Façade models can for instance be used to increase the level of details of $3D$ city models generated from aerial or satellite imagery. They are also useful for a compact coding of façade image textures for streaming or for an embedded system. The characterization of stable regions in façades is also necessary for a robust indexation and image retrieval.

1.2 Related work

Existing façade extraction frameworks are frequently specialized for a certain type of architectural style or a given texture appearance. In a procedural way, operators often step in a pre-process to split correctly the image in suitable regions. Studied images indeed are assumed to be framed in such a way that they exactly contain relevant information data such as windows on a clean wall background.

Most building façade analysis techniques try to extract specific shapes/objects from the façade: windows frame, etc. Most of them are data driven, *i.e.* image features are first extracted and then some models are matched with them to build object hypotheses. Some other model-driven techniques try to find more complex objects which are patterns or layouts of simple objects (*e.g.* alignments in $1D$ or in $2D$). Higher level techniques try to generate directly a hierarchy of complex objects composed of patterns of simple objects usually with grammar-based approaches. Those methods generally devote their strategy to a special architectural style.

1.2.1 Single pattern detection Strategies to extract shape hypotheses abound in recent works. (Čech and Šára, 2007), for instance, propose a segmentation based on a maximum *a posteriori* labeling. They associate each image pixel with values linked with some configuration rules. They extract a set of non-overlapping windowpanes hypotheses, assumed to form axis-parallel rectangles of relatively low variability in appearance. This restriction does not take into account lighting variations. With a supervised classification-based approach, (Ali et al., 2007) extracted windows width an *adaboost* algorithm. In the same fashion, (Wenzel and Förstner, 2008) minimize user interaction with a clustering procedure based on appearance similarity.

Assuming the regularity of the façade, (Lee and Nevatia, 2004) use a gradient profile projection to locate window edges coordinates. They first locate valley between two extrema blocks of each gradient accumulation profile and they roughly frame some floors and windows columns. Edges are then adjusted on local data information. Their results are relevant for façades whose background does not contain any contours such as railings, balconies or cornices.

1.2.2 $1D$ or $2D$ grid structures detection (Korah and Rasmussen, 2007, Reznik and Mayer, 2007) use linear primitives to generate rectangle hypotheses for windows. A *Markov Random Field* (*MRF*) is then used to constrain the hypotheses on a $2D$ regular grid. (Korah and Rasmussen, 2007) generate their rectangular hypotheses in a similar way as (Han and Zhu, 2005): they project on image $3D$ planar rectangles. (Reznik and Mayer, 2007) learn windows outline from training data and use as hypotheses for window corners characteristic points.

1.2.3 Façade grammars A façade grammar describes the spatial composition rules of complex objects (*e.g.* grid structure) and/or simple objects to construct a façade. Approaches based on grammars succeed in describing only façades corresponding to the grammar. Nevertheless, to obtain a detailed description a specific grammar is required per type of architecture (*e.g.* Hausmanian in the case of Parisian architecture). The drawback is that many grammars are necessary to describe the variety of building architectures in a general framework.

For instance, to detect windows on simple buildings, (Han and Zhu, 2005) integrates rules to produce patterns in image space. In particular, this approach integrates a bottom-up detection of rectangles coupled with a top-down prediction hypotheses taken from the grammar rules. A Bayesian framework validates the process. (Alegre and Dellaert, 2004) look for rectangular regions with homogeneous aspect by computing radiometry variance. (Müller et al., 2007) extract an *irreducible region* to summarize the façade by periodicity in vertical and horizontal directions. Their results are significant with façades that effectively contain regular window grid pattern or suitable perspective effects. (Ripperda, 2008) fixes her grammar rules according to prior knowledge: she beforehand computes distribution of façade elements from a set of façade images.

These approaches either use a too restrictive model dedicated to simple façade layout, or are too specialized for a particular kind of architecture. They thus would hardly deal with usual Parisian façades such as Hausmanian buildings or other complex architectures with balconies or decoration elements.

Our process works exclusively on a single calibrated street-level image. Although we could have, we voluntarily did not introduce additional information such as $3D$ imagery (point clouds, etc.) because for some applications such as indexation, image retrieval and localization, we could just have a single photo acquired by a mobile phone.

2 OUR MODEL BASED SEGMENTATION STRATEGY

Most of the aforementioned approaches provide good results for relatively simple single building. Only a few of them have addressed very complex façade networks such as the ones encountered in European cities where the architectural diversity and complexity is large. Our work is upstream from most of these approaches: we do not try to extract semantic information but we just propose a façade segmentation framework that could be helpful for most of these approaches. This framework must firstly separate a façade from its background and neighboring façades, and then, identify intra-façade regions of specific elementary texture models. These regions must be robust to change in scale or point of view.

Our strategy requires horizontal and vertical image contour alignments. We thus first need to rectify images in the façade plane: vertical and horizontal directions in the real world respectively become vertical and horizontal in the image. To do so, we extract vanishing points which provide an orthogonal basis in object space useful to resample the image as required.

Regarding segmentation, the core of our approach relies on a recursive split process and a model based analysis of each subdivided regions. Indeed we do not intend to directly match a model to the whole façade, but we build a tree of rectangular regions by recursively confronting data with some basic models. If a region does not match with any of them, it is split again, and the two sub-regions are analyzed as illustrated by the decision tree on figure 1. Our models are based on simple radiometric criteria: planes and generalized cylinders. Such objects are representative of frequent façade elements like window panes, wall background or cornices.

We start each process with the whole image region. We test if its texture matches our planar model. If it does, then the process stops: we have recognized a planar and radiometrically coherent region in the image. Otherwise, we test if it matches our generalized cylinder model. In the same manner, the process stops on the cylinder model. Otherwise the region is not considered as homogeneous (in the sense of our models) and it is split in two sub-regions. The process recursively analyzes these two sub-regions exactly as the same way as the large region. Thus, we build a segmentation tree whose leaves are planar or generalized cylinder models. The following sections explain each step of this algorithm.

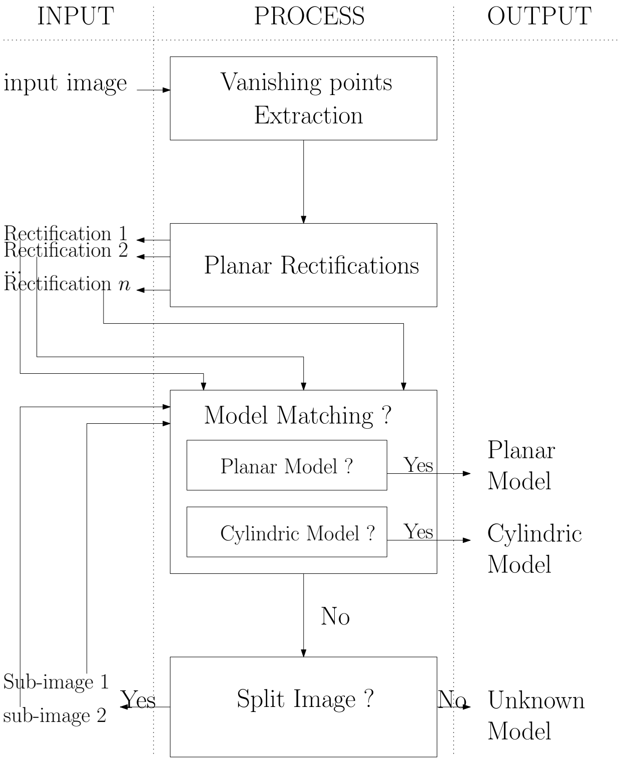

Figure 1: Our algorithm recursively confronts data with models. If region does not match with any proposed model, we split it.

3 RECTIFICATION PROCESS

3.1 Extracting Vanishing Points

Our rectification process relies on vanishing point lines detected by (Kalantari et al., 2008). They project relevant image segments on the Gaussian sphere: each image segment is associated with a point on the sphere. Their algorithm relies on the fact that each circle of such a $3D$-point distribution gathers points associated with the same vanishing point in the image. Then they estimate the best set of circles that contains the highest number of points. The more representative circles are assumed to provide main façade directions: the vertical direction and several horizontal ones. Figure 2 upper-right shows some detected edges that support main vanishing points: segments associated with the same direction are drawn in the same color.

3.2 Multi-planar Rectification Process

We rectify our image in each plane defined by a couple of one of the horizontal vanishing points and the vertical one. We then project the image onto the plane. Figure 2 bottom right shows a rectification result. Figure 2 bottom left shows rectified edges on the façade plane.

Calibration intrinsic parameters are supposed to be known. Rectified image is resampled in grey levels, but such a restriction already provides some interesting perspectives.

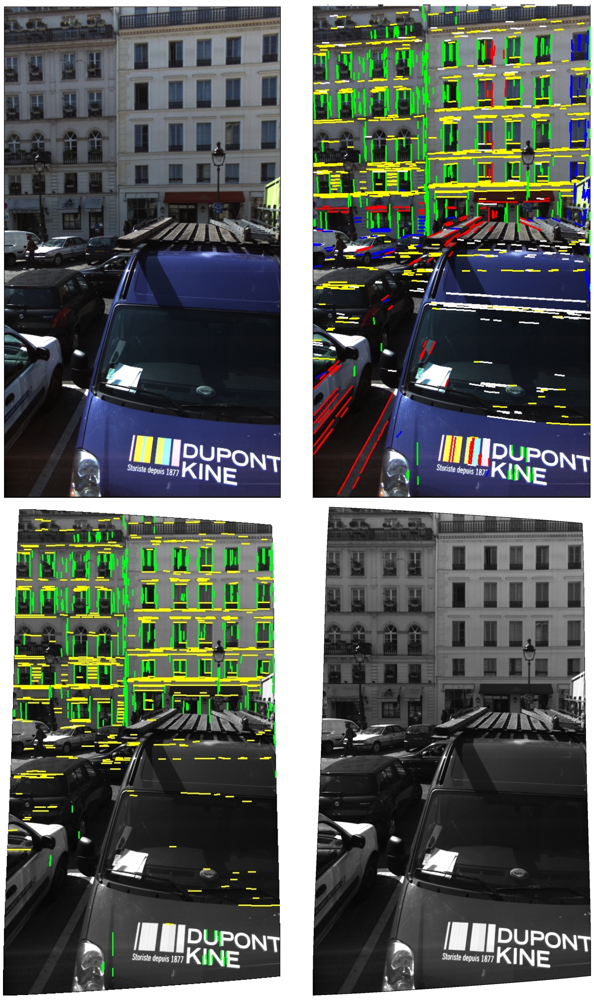

Figure 2: The rectification process. upper left: original image only; upper right: original image with segments that support main vanishing points (green and blue ones are those for the main vertical direction, yellow and white ones for the main horizontal direction and red ones for an aberrant vanishing point); bottom left: rectified image with rectified segments that support the two selected directions; bottom right: rectified image only

4 MODEL MATCHING

Given a region of a rectified image, we try to match two geometric models with data in increasing complexity order: the planar model $\mathcal{M}_1$, then the generalized cylinders one $\mathcal{M}_2$. This decision tree indeed provides a good compromise between quality and compression rate.

To match an image region with a model, we simply count local radiometric differences as follows. Let I_k be the sub-image at region R_k of a façade image I. Sub-image I_k is described by model $\mathcal{M}$ when the deviation $N_{\mathcal{M}}(I_k)$ is small enough and if this model is the simplest one. Deviation $N_{\mathcal{M}}(I_k)$ is defined by the number of pixels whose radiometry differs too much from the model. Radiometric medians provide some significative robustness: the influence of parasite structures such as tree branches or lighting posts, is significantly reduced. Figure 3 illustrates models we use.

4.1 Planar Model

A planar model is an image with an uniform radiometry. Let $\mathcal{M}_1$ be the planar model of a sub-image I_k. It is defined by equation 1.

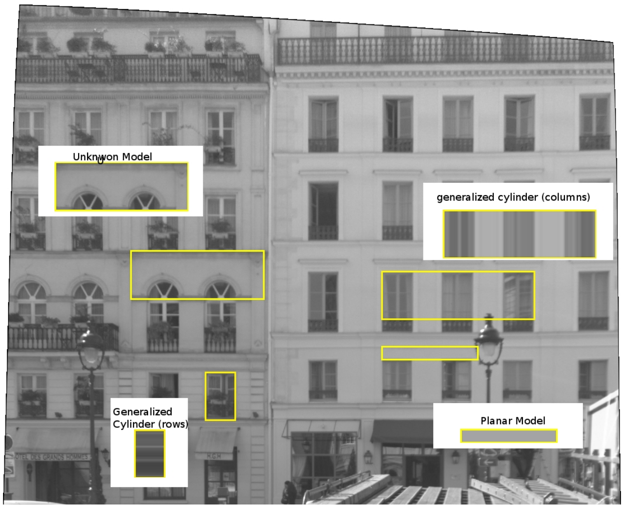

Figure 3: Description of our radiometric $2D$-models

An instance of a planar model is depicted on the lower-right corner of figure 3.

$$\mathcal{M}_1 : \forall p \in R_k, I_k(p) = median(I_k) + \epsilon(p) \qquad (1)$$

where $\epsilon(p)$ is the difference between the image I_k and the model $\mathcal{M}_1$ at the pixel p. If this difference is smaller than an arbitrary threshold, it is tolerated. It refers to the acquisition noise or some texture defects. Otherwise, the deviation $N_{\mathcal{M}}(I_k)$ is incremented.

4.2 Generalized Cylinder Model

A generalized cylinder model is designed either in columns ($\mathcal{M}_2^c$) or in rows ($\mathcal{M}_2^l$). The model in columns is composed of medians of columns and the cylinder model in rows is composed of medians of rows. They is defined by equation 2. Functions $median_x$ and $median_y$ respectively return the median of the column at x abscissa and the row at y ordinate. Figure 3 shows an instance of each generalized cylinder model.

$$\begin{aligned} &\forall(x,y) \in R_k, \\ &\mathcal{M}_2^c : I_k(x,y) = && median_x(I_k(x,y)) + \epsilon(x,y) \\ &\mathcal{M}_2^l : I_k(x,y) = && median_y(I_k(x,y)) + \epsilon(x,y) \end{aligned} \qquad (2)$$

where $\epsilon(x,y)$ is the difference between the image I_k and the model $\mathcal{M}_2$ at the pixel (x,y). In the same manner as planar model, the deviation $N_{\mathcal{M}}(I_k)$ is incremented when this difference is greater than an arbitrary threshold.

5 SPLIT PROCESS BY ENERGY MAXIMIZATION

Given a region of a rectified image that does not match with any model, we try to split it by measuring the internal gradient distribution energy.

5.1 Generating splitting hypotheses

We select split hypotheses with a technique close to (Lee and Nevatia, 2004). We accumulate x-gradient absolute values by column and y-gradient absolute values by row, where x-gradient and y-gradients are related respectively to vertical and horizontal edges. We use convolution with a discrete first order derivative operator. Local extrema of these accumulations are our split hypotheses. This reinforces low but repetitive contrasts. Figure 4 illustrates this process.

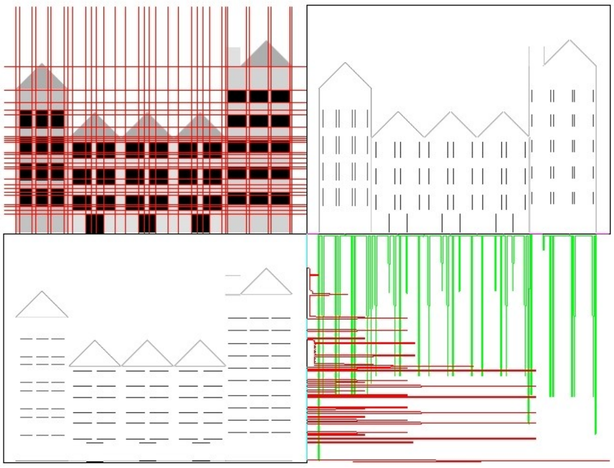

Figure 4: Upper right and bottom left images respectively are x-gradient and y-gradient. Bottom right image presents accumulation profiles: green profile for x-gradient and red one for y-gradient. extrema of this profiles are our split hypotheses: red lines in upper left image.

Such a rough set of hypotheses supplies initial interesting segmentation. (Lee and Nevatia, 2004) base their window detection on similar rough segmentation. They almost use the same procedure except that they do not accumulate gradients in the same orientation: they respectively treat x-gradient and y-gradient horizontally and vertically. Thus they locate valley between two extrema blocks of each gradient accumulation profile and they frame some floors and windows columns. Their results were relevant in façades composed of a fair windows grid-pattern distribution on a clean background.

Main buildings structure are detected. Each repetitive objects are present in vertical or horizontal alignment as common edges generate local extrema in accumulation profiles. Local gradient extremum neighborhood is set *a priori*. In our case, this neighborhood is set to 30 centimeters. However this last grid-pattern usually is not enough by itself to summarize façade texture: repetitive elements of our images are not necessarily evenly distributed. Thus our split strategy relies on breaks between two façades or inside one façade.

5.2 Choosing the best splitting hypotheses

The best splitting hypothesis maximizes its pixels number of *regular edges* in each of the two sub-region. A *regular edge* is a segment of a main gradient direction that effectively matches to a contour of the image. The weight W_H of the split hypothesis H that provides the two regions R_1 and R_2 is given by $W_H = f(R_1) + f(R_2)$, where the function f returns the pixels number of *regular edges* in a region. We select the hypothesis $H^* = \arg\max_H W_H$.

If we try for instance to split image at x_0 location, we reaccumulate y-gradients in left region and in right region separately. Local extrema are detected in each of those y-profiles (cf figure 5).

Previous vertical split hypotheses and those new horizontal split hypotheses constitute two new grid patterns for local split hypotheses. Each edge of these grid patterns is either *regular* or *fictive*. *Regular edges* are located on *significant gradient* where a *significant gradient* keeps its orientation uniform. A *fictive edge* does not match with any significant gradient. Such a distinction is illustrated in figure 6.

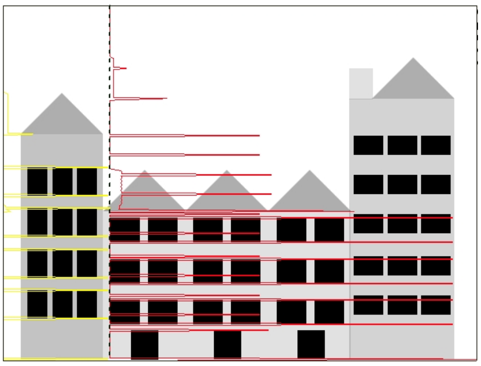

Figure 5: y-gradient profiles are separately accumulated in left region (yellow profile) and in right region (red profile).

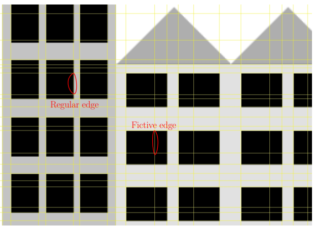

Figure 6: *Regular edges* are located on *significant gradient* where a *significant gradient* keeps its orientation uniform. A *fictive edge* doesn't match with any significant gradient.

The weight of each split hypothesis is the sum of regular edge lengths. Figure 7 illustrates best split selection.

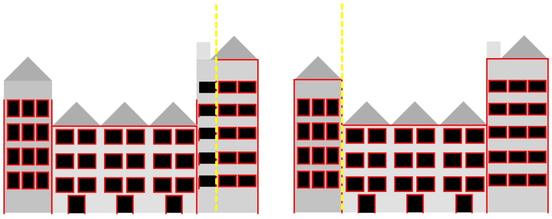

Figure 7: Regular edges are drawn in red. Split hypotheses are drawn in yellow. Right image presents the best split hypothesis whose weight is 8400 regular edge pixels. Left image presents a bad split hypothesis: only 7700 regular edge pixels.

If the given region does not contain any gradient extremum, the process stops. Figure 3 shows a region that do not fit with any model and that is not split.

6 RESULTS

We illustrate our segmentation on a typical instance of our issue: two building façades in the background. We have set maximum model deviation at 15% of each region area. On our images, the depth in the hierarchy of the segmentation tree is represented by the thickness of split lines. First the process detects vertical structure discontinuities (figure 8). The two façades are separated. Then on each of these two new sub-images, background is separated from the foreground (figure 9). At this step we have obtained four images: two façade images and two images of foreground cars. Then the process recursively keeps analyzing these images as shown in figure 10. Figure 11 shows the global segmentation: a tree of about 2000 elementary models.

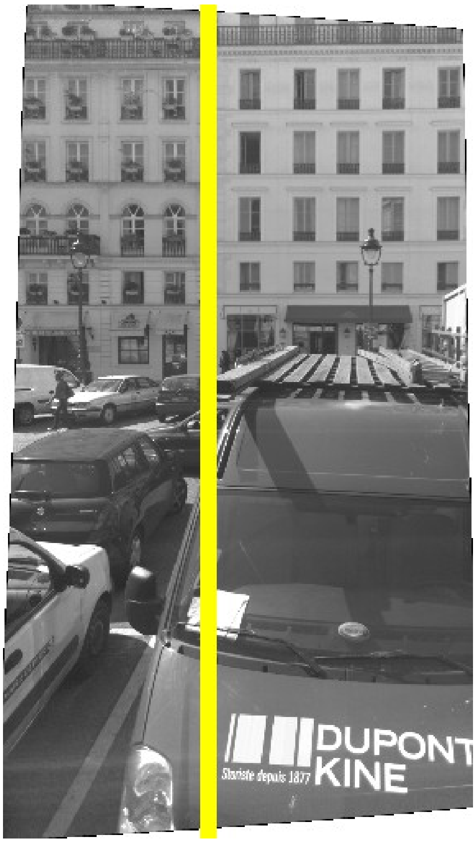

Figure 8: The two façades are separated because of the significative break between their radiometric structure.

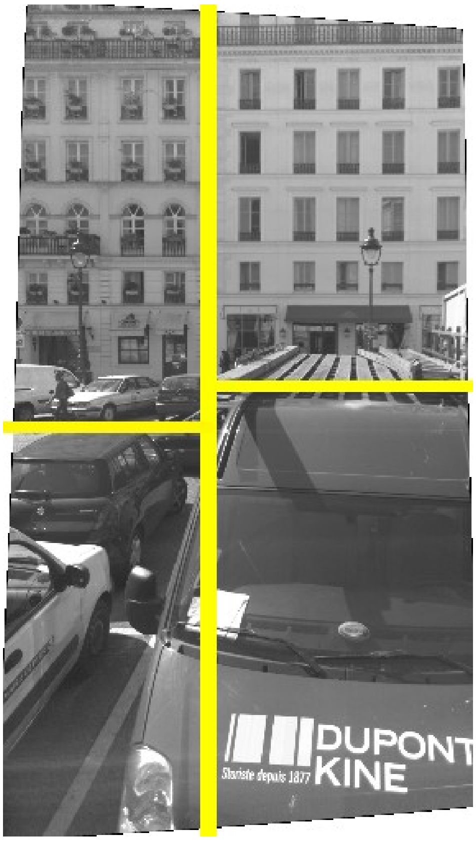

Figure 9: Background is separated from the foreground on each of the two façade images.

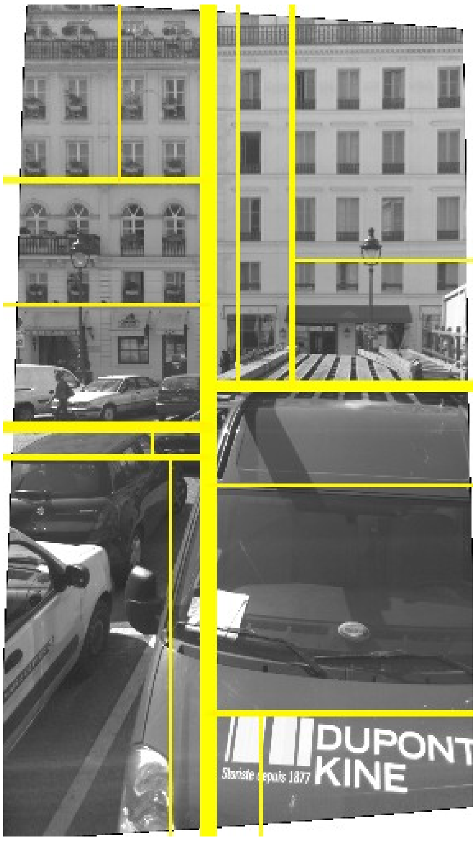

Figure 10: The process recursively segments each of the four subimages. It splits the two façades and the foreground cars.

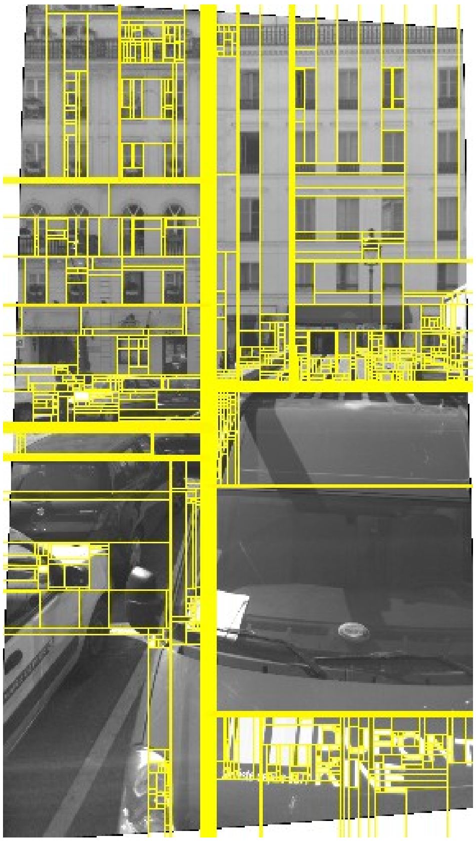

Figure 11: The segmentation result is a tree of 2000 elementary models.

The strength of this process is its ability to localize accurate global structure breaks: it separates façades and foreground. On the one hand, split results at the foreground are not really interesting because the related region is not in the rectified plane: they are based on chaotic gradient distribution. In such a case, the process stops or it oversegments. This phenomenon typically occurs on the cars of figure 11. On the other hand, splits inside façade texture provides some significative information. On figure 10, the left façade is first split between the second and the third floor, whereas the first windows column is extracted from the right façade. This different strategy certainly must be explained by the fact that the process is exclusively based on edges alignment. An other criterion like contour uniformity may direct the split decision toward a more significant separation: favoring floor separation rather than window columns.

Figure 3 shows the region models, the leaves of the segmentation tree. One can see that the synthetic image reconstructed from the $2D$-models is very close to the initial image although the representation is very compact. This shows that our modelling is particularly well adapted for image compression.

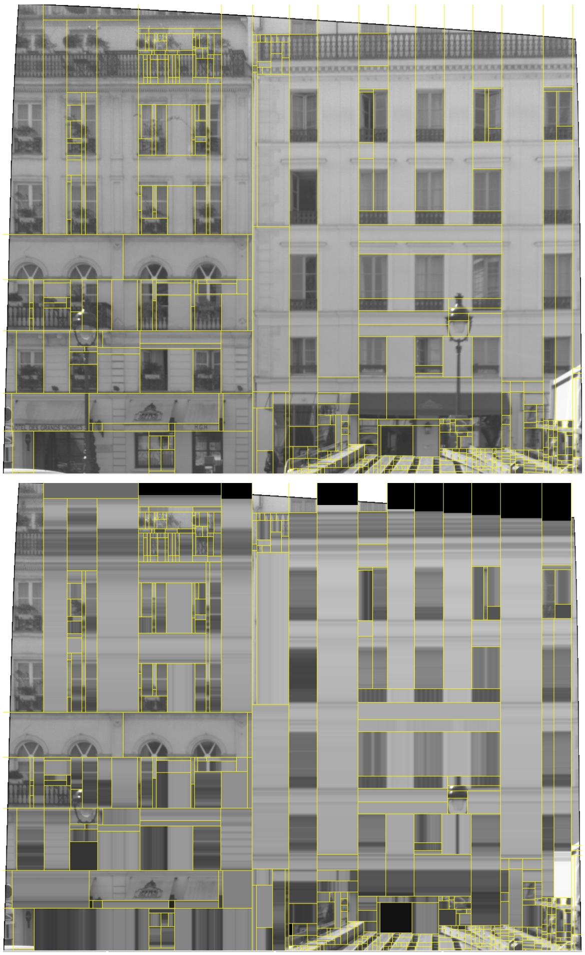

Figure 12: Upper: Rectified façade image. Bottom: Synthetic image reconstructed from 1000 elementary $2D$-models.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a new unsupervised model-based segmentation approach that provides interesting result. It is able to separate a façade from its surroundings but also to organize façade itself in a hierarchy. Still these are first results, thus there are many improvements that could be made. The dictionary of models is currently being extended to periodic textures to manage for instance balconies, building floors or brick texture. Some other objects or specializations of objects could be added such as symmetry computation of (Van Gool et al., 2007). A merger process at each step of the process could also be useful to correct oversegmentations. Besides we could add color information to directly detect difference between two façades ore between two floors in certain cases. We could also use a point cloud to compute an ortho image: displacements due to perspective effects would be avoided.

Such an unsupervised segmentation will provide of course relevant clues to classify the façade architectural style or to detect objects backward or in front of it. It is also intended to give geometrical information that represents relevant indexation features *e.g.* windows gab length or floor delineation.

REFERENCES

Alegre, F. and Dellaert, F., 2004. A Probabilistic Approach to the Semantic Interpretation of Building Facades. Technical report, Georgia Institute of Technology.

Ali, H., Seifert, C., Jindal, N., Paletta, L. and Paar, G., 2007. Window Detection in Facades. In: Proc. of the 14th International Conference on Image Analysis and Processing, pp. 837–842.

Han, F. and Zhu, S.-C., 2005. Bottom-up/top-down image parsing by attribute graph grammar. In: International Conference on Computer Vision, IEEE Computer Society, pp. 1778–1785.

Kalantari, M., Jung, F., Paparoditis, N. and Guédon, J.-P., 2008. Robust and Automatic Vanishing Points Detection with their Uncertainties from a Single Uncalibrated Image, by Planes Extraction on the Unit Sphere. In: IAPRS, Vol. 37 (Part 3A), Beijing, China.

Korah, T. and Rasmussen, C., 2007. 2D Lattice Extraction from Structured Environments. In: International Conference on Image Analysis and Recognition, Vol. 2, pp. 61–64.

Lee, S. C. and Nevatia, R., 2004. Extraction and Integration of Window in a 3D Building Model from Ground View images. In: Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 113–120.

Müller, P., Zeng, G., Wonka, P. and Van Gool, L., 2007. Image-based Procedural Modeling of Facades. In: Proceedings of ACM SIGGRAPH 2007 / ACM Transactions on Graphics, Vol. 26number 3, p. 85.

Reznik, S. and Mayer, H., 2007. Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation. In: Photogrammetric Image Analysis, p. 173.

Ripperda, N., 2008. Determination of Facade Attributes for Facade Reconstruction. In: Proc. of the 21st Congress of the International Society for Photogrammetry and Remote Sensing.

Čech, J. and Šára, R., 2007. Windowpane Detection based on Maximum Aposteriori Probability Labeling. Technical report, K13133 FEE Czech Technical University, Prague, Czech Republic.

Van Gool, L. J., Zeng, G., Van den Borre, F. and Müller, P., 2007. Towards Mass-Produced Building Models. In: Photogrammetric Image Analysis, p. 209.

Wenzel, S. and Förstner, W., 2008. Semi-supervised Incremental Learning of Hierarchical Appearance Models. In: Proc. of the 21st Congress of the International Society for Photogrammetry and Remote Sensing.