AUTOMATIC MODEL-BASED BUILDING DETECTION FROM SINGLE PANCHROMATIC HIGH RESOLUTION IMAGES

Konstantinos Karantzalos[†,§] and Nikos Paragios[†]

[†] Laboratoire de Mathematiques Appliquees aux Systemes (MAS), Ecole Centrale de Paris, Chatenay-Malabry, France
{konstantinos.karantzalos, nikos.paragios}@ecp.fr
[§] Remote Sensing Laboratory, National Technical University of Athens, Athens, Greece

Commission III/4

KEY WORDS: Computer Vision, Pattern Recognition, Variational Methods, Homography, Prior-based segmentation, Competing Priors, Labelling

ABSTRACT:

Model-free image segmentation approaches for automatic building detection, usually fail to detect accurately building boundaries due to shadows, occlusions and other low level misleading information. In this paper, a novel recognition-driven variational framework is introduced for automatic and accurate multiple building extraction from aerial and satellite images. We aim to solve the problem of inaccurate data-driven segmentation. To this end, multiple shape priors are considered. Segmentation is then addressed through the use of a data-driven approach constrained from the prior models. The proposed framework extend previous approaches towards the integration of shape priors into the level set segmentation. In particular, it allows multiple competing priors and estimates buildings pose and number from the observed single image. Therefore, it can address multiple building extraction from single panchromatic images a highly demanding task of fundamental importance in various geoscience and remote sensing applications. Very promising results demonstrate the potentials of our approach.

1 INTRODUCTION

Human visual perception involves a set of processes for distinguishing top-down attention from the stimulus-driven bottom-up [Itti and Koch, 2001]. During our entrance in a crowded classroom in order to localize and recognize someone, we will be looking around, scanning everyone's face without paying much attention to the interior design and room's furniture. However, entering the same classroom with the intention of finding an available desk, we will be looking at pretty much the same scene, and yet our perception will be biased for the arrangement of the furniture, mostly ignoring other people around [Walther and Fei-Fei, 2007]. Several problems/applications exist in computer vision that relate perception with specific-object recognition tasks as well as image segmentation. Variational methods have gained significant attention towards the integration of prior knowledge into the image segmentation processes. Level set algorithms, when extended and formulated towards such a recognition-driven way, became robust to shadows, noise, background clutter or partial occlusions of desired for extraction object [Paragios et al., 2005].

In remote sensing and photogrammetry, among various methods, processing schemes and systems, which have been reported in the literature, conventional variational curve propagation techniques (snakes, active contours, deformable models and more recently level sets) have revealed promising results [Mayer, 1999, Peng et al., 2005, Cao et al., 2005, Karantzalos and Argialas, 2006]. Model-free level sets have been employed to account for the general task of segmenting satellite images [Samson et al., 2001, Ball and Bruce, 2005, Besbes et al., 2006], for the detection of roads in a semi-automatic framework [Keaton and Brokish, 2002, Niu, 2006] and for the automatic detection of buildings and other man-made objects [Cao et al., 2005, Karantzalos and Argialas, 2006]. These methods were purely image-based and therefore were vulnerable to misleading low-level information, due to shadows or occlusions, which is a common scenario observed in remote sensing data.

In this paper, we aim to solve the problem of inaccurate data-driven segmentation caused by misleading low level information due to shadows or occlusions. Looking forward to overcoming above limitations, we propose a novel prior-based variational framework, which can account for automatic building extraction from a single image. An elegant and powerful mathematical formulation, to align prior buildings shapes with the evolving contour shapes, is introduced. Such a term aims to minimize a multi-reference shape-similarity measure that admits a wide range of transformations, beyond similarity and shapes sampling. The objective function involves both the selection of the most appropriate prior model as well as the transformation which relates the model to the image. We propose a dynamic and evolving selection of priors towards accounting for this variation by the use of a labeling function, which controls priors shape effect to specific image regions [Chan and Zhu, 2003, Cremers et al., 2006].

Once -and for every optimization iteration- the level set based segmentation yields to a possible building segment, a prior -from the database- which fits best to that region is applied. The labeling function evolves in time and incrementally determines multiple instances, from the shape prior set, according to the number of the detected objects. Here, the term shape prior refers to building templates, like those shown in Figure 1. Last but not least, neither point correspondence nor direct methods [Irani and Anandan, 1999] were used and thus color or texture compatibility between the prior and the segmented image was needless. Parametrization-free shape descriptions possess a significant advantage over landmark-based and template matching techniques, which represent shapes by collections of points or features.

Our framework fundamentally extends previous work for automatic building detection in single panchromatic images. Performed experimental qualitative and quantitative evaluation demonstrated proposed algorithm's efficiency. The successful recognition-driven results along with the reliable estimation of the transformation parameters suggest that the proposed method forms a highly promising tool for various geoscience segmentation and regis-

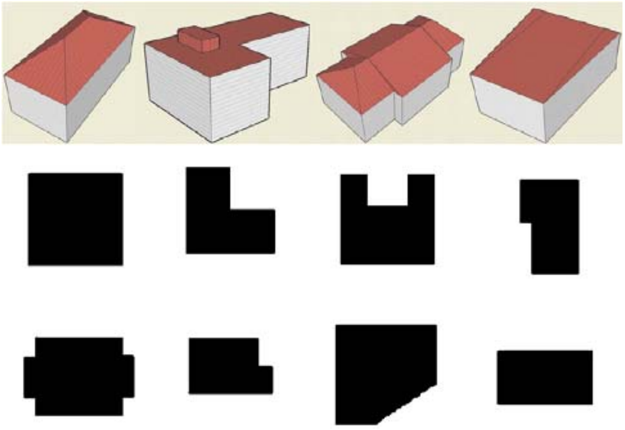


Figure 1: First row: Four different 3d building models with different footprint and roof type. Second and third row: A database of 8 prior binary 2d templates. Each one competes the other one to fit best in a building segment

tration applications. We organize the rest of the paper in the following way. In Section 2, we briefly describe the proposed projective-invariant prior-based formulation for building detection. The generalized variational framework for the integration of multiple competing shape priors for multiple building extraction is detailed in Section 3. Experimental results and the performed qualitative and quantitative evaluation are presented in Section 4. Finally, conclusions and perspectives for future work are in Section 5.

2 PROJECTIVE-INVARIANT SHAPE PRIOR FORMULATION

Given an image $\mathcal{I}(\mathbf{x})$ at domain (bounded) $\Omega \in \mathcal{R}^2$, an interface C and a level set representation $\phi : \Omega \rightarrow \mathcal{R}^+$ one can form a data-driven cost functional $E_{seg}(\phi)$ towards image segmentation. The basic idea in model-based approaches is to extend this data-driven cost functional by adding another energy E_{prior} which favors certain contour formations [Rousson and Paragios, 2008]:

$$E_{total}(\phi) = E_{seg}(\phi) + \mu E_{prior}(\phi) \;\; \mu > 0 \qquad (1)$$

The proposed shape constraints E_{prior} affect the embedding surface ϕ globally (i.e. on the entire domain). In the simplest case (no pose variations between the evolving interface and the prior model), such a prior term can take the following form using the approximations of DIRAC and HEAVISIDE distributions:

$$E_{prior} = \int_{\Omega} \left(H_\epsilon(\phi(\mathbf{x})) - H_\epsilon(\tilde{\phi}(\mathbf{x}))\right)^2 \qquad (2)$$

where $\tilde{\phi}$ is the level set function embedding a given training shape (or the mean of a set of training shapes). Positive and negative values of $\tilde{\phi}$ correspond to object and background regions in $\tilde{\Omega}$, respectively. The prior term is a weighted sum of the non-overlapping positive and negative regions of $\tilde{\phi}$ and ϕ. At each time step, ϕ is modified in image regions where there is inconsistency between the object and background areas indicated by $H_\epsilon(\phi)$ and $H_\epsilon(\tilde{\phi})$. The change in ϕ is weighted by δ_ϵ.

With the above formulation the pose and location of the object of interest is assumed to be identical to the ones of the reference shape. In a realistic segmentation problem and in particular for automatic building detection in aerial and satellite neither the pose nor the location of objects are know. Statistical models of shape variation with respect to the reference frame are a simple approach to deal with this problem [Riklin-Raviv et al., 2007]. However these methods perform well if and only if the underlying assumption for the model is supported from the data. In the case of buildings, that are being observed in remote sensing imagery, the implicit assumption of statistical modeling using a simple Gaussian is rather unrealistic and a real need exists to cope with important variation of the priors.

To this end, the shape-term was extended to incorporate all possible projective transformations between the prior shape and the shape of interest. This was addressed by applying an adequate 2D transformation $\mathcal{T} : R^2 \rightarrow R^2$ to the prior shape $\tilde{\phi}$. The recovery of the transformation parameters, given the prior contour and the curve generated by the zero-crossing of the estimated level-set function, is described subsequently. In order to minimize the energy functional, one has to apply a gradient descent process that calls for the evaluation of ϕ simultaneously with the recovery of the transformation $\mathcal{T}$ for the prior shape $\tilde{\phi}$.

2.1 Planar Projective Homography

To generalize the admissible geometric relation between two corresponding shapes we employ the concept of planar projective homography. The equivalence of geometric projectivity and algebraic homography is supported by a set of theorems presented in [Springer, 1964]. The relation between corresponding views of points on a plane (world plane) in a 3D space can be modeled by a planar homography induced by the plane. Planar projective homography (projectivity) is a mapping $M : \mathcal{P}^2 \rightarrow \mathcal{P}^2$ such that points p_i are collinear if and only if $M(p_i)$ are collinear (projectivity preserves lines) [Springer, 1964], [Hartley and Zisserman, 2003].

Here, similarly to the formulations of [Riklin-Raviv et al., 2007] the homograph is calculated directly in its explicit form:

$$\mathcal{T} = r + \frac{1}{d} t n^T \qquad (3)$$

where $\mathcal{T}$ forms the homography matrix determined by the translation t and rotation r between the two views and by the structure parameters n, d of the world plane. An explicit expression for the induced homography can be derived as follows: Let $\mathbf{y}$ and $\mathbf{y}'$ be the corresponding homogeneous coordinates of two views of a world point in two camera frames ($\mathbf{y} = (x, y, 1)$ and $\mathbf{y}' = (x', y', 1)$), then the transformation from $\mathbf{y}$ to $\mathbf{y}'$ can be expressed as:

$$\mathbf{y}' = \mathcal{T}\mathbf{y}, \;\; \text{where } \mathcal{T} = \begin{bmatrix} h_{11} \, h_{12} \, h_{13} \\ h_{21} \, h_{22} \, h_{23} \\ h_{31} \, h_{32} \, h_{33} \end{bmatrix}$$

The eight unknowns of $\mathcal{T}$ (the ratios of its nine entries from Equation 6) can be recovered by solving at least four pairs of equations of the form:

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \; y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}, \qquad (4)$$

Note that only the ratio t/d can be recovered from $\mathcal{T}$. Classic approaches recover $\mathcal{T}$ by solving an over-determined set of equations like the one above. The translation and rotation (r, t) between the image planes, and the scene structure (n, d), can be recovered by decomposition of the known homography matrix [Faugeras et al., 2001], [Hartley and Zisserman, 2003].

In particular, $(\mathbf{i})$ the translation in the image is described by the vector $t = (t_x, t_y, t_z)$, $(\mathbf{ii})$ the rotation matrix $r \in R^3$ follows the

Weisstein form:

$$r = \begin{bmatrix} c_\beta c_\gamma & c_\beta s_\gamma & -s_\beta \\ s_\alpha s_\beta c_\gamma & -c_\alpha s_\gamma s_\alpha s_\beta s_\gamma & c_\alpha c_\gamma s_\alpha c_\beta \\ c_\alpha s_\beta c_\gamma & -s_\alpha s_\gamma c_\alpha s_\beta s_\gamma & -s_\alpha s_\gamma c_\alpha c_\beta \end{bmatrix}$$

where where s_α is shorthand for sin(α) and c_α for cos(α) and (**iii**) since generally the world plane is not perpendicular to the optical axis of the first camera parameter $n \neq (0,0,1)$, the unit vector n is obtained by: first rotating the vector $(0,0,1)$ by an angle ξ around the y-axis and then by an angle ψ around the x-axis. Hence, n=(-sinξ, sinψcosξ, cosψ cosξ).

The prior shape is matched to the shape being segmented as part of its detection procedure. In order to minimize the energy functional (Eq.1) one has to simultaneously evolve the level set function ϕ and recover the transformation $\mathcal{T}(\mathbf{x})$. At each time step one re-evaluates the homography matrix entries h, based on the estimated transformation parameters. The coordinate transformation $\mathcal{T}$ is applied to the representation $\tilde{\phi}$ of the prior shape. Thus, the transformed representation $\tilde{\phi}\,(\mathcal{T}(\mathbf{x}))$ is substituted for $\tilde{\phi}$ in Eq.2.

The corresponding prior-based energy E_{prior} (Equation 4) now takes the form:

$$E_{prior}(\phi, \mathcal{T}) = \int_\Omega \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\,(\mathcal{T}(\mathbf{x})))\right)^2 \, d\mathbf{x} \tag{5}$$

The transformation parameters $\mathcal{T}(\alpha,\beta,\gamma,t_x,t_y,t_z,\xi,\psi,d)$ are determined via the gradient descent equations obtained by minimizing the energy functional with respect to each of them. The general gradient descent equation for each of the transformation parameters (denoted here by u) is of the form:

$$\frac{u}{t} = 2\mu \int_\Omega \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\,(\mathcal{T}(\mathbf{x})))\right) \frac{\vartheta \mathcal{T}(u)}{\vartheta u} \, d\mathbf{x} \tag{6}$$

However, such a prior formulation can not account for multiple building detection.

3 MULTIPLE PRIORS IN COMPETITION EXTRACTING MULTIPLE OBJECTS

In order to retain the favorable level set property for multiple object segmentation the prior energy of Equation 8 is extended with a labeling (decision) function $L : \Omega \to \{-1, +1\}$, which indicates the regions of the image where the given prior ϕ is to be enforced. The role of the labeling function is to evolve dynamically in order to select these regions in a recognition-driven way during optimization.

Let us now consider the general case of a larger number of building shape priors (like all those of Figure 1) and possibly some further independent unknown objects (which should therefore be segmented based on their intensity only). To this end, we employed a vector-valued labeling function

$$\mathbf{L} : \Omega \to R^k, \quad \mathbf{L}(\mathbf{x}) = (L_1(\mathbf{x}), ..., L_k(\mathbf{x})) \tag{7}$$

towards multi-region segmentation. The $m = 2^k$ vertices of the polytope $[-1,+1]^k$ yield to m different regions $L_j \in \{+1, -1\}$. The indicator function for each of these regions is denoted by $x_i = 1, ..., m$. Each indicator function x_i has the form [Chan and Zhu, 2003], [Cremers et al., 2006]:

$$x_i(\mathbf{L}) = \frac{1}{4^k} \prod_{j=1}^{k} (L_j - w_j)^2, \text{ with } w_j \in \{+1, -1\} \tag{8}$$

With the above k-dimensional labeling formulation, able for the dynamic labeling of up to $m = 2^k$ regions, the following cost functional can account for a recognition-driven segmentation, based on multiple competing shape priors:

$$E_{total} = E_{seg}(\phi, r_{obj}, r_{bg}) + \mu E_{prior}(\phi, \mathcal{T}, \mathbf{L}) \tag{9}$$

where:

$$E_{prior}(\phi, \mathcal{T}, \mathbf{L}) = \sum_{i=1}^{m-1} \int \left(\frac{H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_i\,(\mathcal{T}_i(\mathbf{x})))}{\sigma_i} \right)^2 x_i(\mathbf{L}) d\mathbf{x} + \int \lambda^2 x_m(\mathbf{L}) d\mathbf{x} + \rho \sum_{i=1}^{m} \int |\nabla L| d\mathbf{x} \tag{10}$$

where the terms associated with the two objects are normalized with respect to the variance of the respective template: $\sigma_i^2 = \int \phi_i^2 d\mathbf{x} - \int \phi_i d\mathbf{x}^2$. Contrary to [Vese and Chan, 2002] and [Cremers et al., 2006] the labeling function's dimensionality k is not a priory fixed and is calculated during optimization. Let a positive scalar q denote the number of resulting, from the image-driven functional, segments. Then k is calculated based on the following equation:

$$k = \lceil \frac{\log(1+q)}{\log 2} \rceil \tag{11}$$

In this way, during optimization the number of selected regions $m = 2^k$ depends on the number of the possible building segments according to ϕ and thus the k-dimensional labeling function $\mathbf{L}$ obtains incrementally multiple instances.

3.1 Energy Minimization

The multiple shape prior based segmentation process is generated by minimizing the functional of Equation 9. Minimization is performed by alternating the update of the region intensity descriptors r_{bg} and r_{bg} using a gradient descent evolution with respect to the level set function ϕ, the labeling functions $\mathbf{L}$ and the associated pose parameters $\mathcal{T}_i(\alpha_i,\beta_i,\gamma_i,(t_x)_i,(t_y)_i,(t_z)_i,\xi_i,\psi_i,d_i)$ for every selected prior $\tilde{\phi}_i$:

3.1.1 Evolution of the k-dimensional labeling function For fixed level set function ϕ and transformation parameters, the gradient descent with respect to the labeling functions L_i corresponds to an evolution of the form:

$$\frac{\vartheta E_{total}}{\vartheta L_j} = -\mu \sum_{i=1}^{m-1} \frac{(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_i\,(\mathcal{T}_i(\mathbf{x}))))^2}{\sigma_i^2} \frac{\vartheta x_i}{\vartheta L_j} - \mu\, \lambda^2 \frac{\vartheta x_m}{\vartheta L_j} - \mu\, \gamma\, div \frac{\nabla L_j}{\|\nabla L_j\|}, \tag{12}$$

where the derivatives of the indicator functions x_i are calculated from (14). The first two terms in Equation 12 guide the labeling $\mathbf{L}$ to indicate the transformed priors $\tilde{\phi}_i$ which are most similar to the given function ϕ (i.e. each labeled segment or the background). The last term imposes spatial regularity in the labeling L_j and enforces the selected regions to be compact by preventing flippings with the neighboring locations.

3.1.2 Multiscale Prior Registration For a fixed level set ϕ and labeling function $\mathbf{L}$, the optimization of the projective transformation parameters $\mathcal{T}(\alpha_i,\beta_i,\gamma_i,(t_x)_i,(t_y)_i,\ (t_z)_i,\xi_i,\psi_i,d_i)$ of each selected prior $\tilde{\phi}_i$ was derived from the gradient descent similar to Equation 6. In order, though, to handle both global and local shape deformations a multiscale optimization was introduced.

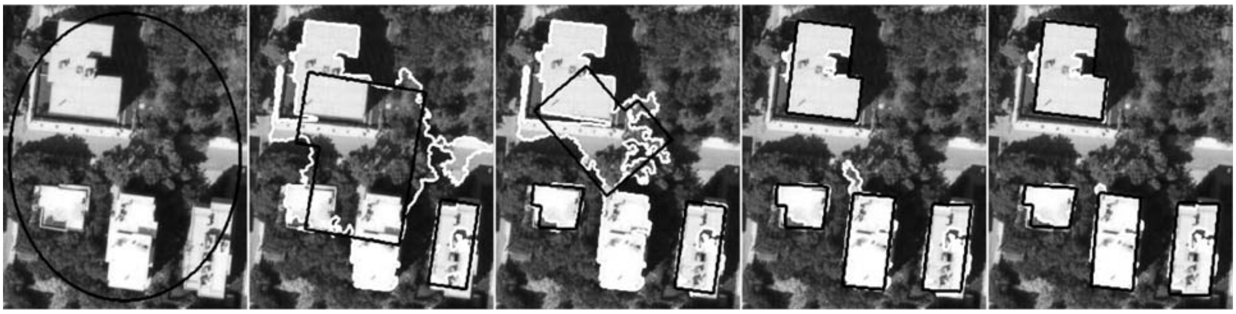


Figure 2: Curve evolution based on the proposed recognition-driven framework. The different steps until algorithms convergence, are shown (in black). The algorithm did manage to extract all fourth buildings and resulting contours (in black) describe accurately buildings boundaries. The evolution of the data-driven term is, also, shown in white.

The multiscale approach is implement via a fixed point iteration on both the level set function ϕ and shape priors $\tilde{\phi}_i$ with a downsampling strategy. Instead of the standard down-sampling factor of 0.5 on each level, it is proposed, here, to use an arbitrary factor $f \in (0, 1)$, which allows smoother transitions from one scale to the next. The full pyramid of images is used ϕ^l, $l = 0, 1, ...$, starting with the smallest possible images ϕ^0 and $\tilde{\phi}_i^0$ at the coarsest grid. Thus, the general gradient descent equation for each of the transformation parameters (denoted by u_i) is of the form:

$$\frac{u_i^l}{t} = 2\mu\, x_i^l(\mathbf{L}^l) \int_\Omega \left(\frac{H_\epsilon(\phi^l) - H_\epsilon(\tilde{\phi}_i^l\left(\mathcal{T}_i^l(\mathbf{x})\right))}{\sigma_i^2} \right) \frac{\vartheta \mathcal{T}_i^l(u_i^l)}{\vartheta u_i^l} \qquad (13)$$

Above equation is analogous to Equation 6 (for the single scale approach), except that (i) the indicator function $x_i(\mathbf{L})$ constrains the integrals to the domain of interest associated with shape $\tilde{\phi}_i$, i.e. to the area where $x_i > 0$ and (ii) moreover, is calculated via fixed point iterations l.

4 EXPERIMENTAL RESULTS [1]

In Figure 2, the curve evolution obtained by the proposed, here, variational framework is presented for automatic building extraction from a high resolution satellite image. Different steps until algorithm's convergence are shown. Starting with an arbitrary elliptical curve (first image from the left) and after a couple of iterations (second image) the data-driven term (shown in white color) yielded to two main segments. The concurrent optimization of the labeling function and the recovery of the appropriate shape priors transformation parameters (α, β, γ, t_x, t_y, t_z, ξ, ψ and d) resulted to the boundaries shown with a black color. Among the competing priors of Figure 1 the fourth from the third row was chosen in order to recover the smaller segment in the bottom right. The competing procedure resulted, also, into the fourth from the second row prior in order to recover the bigger segment in the middle. The later does not corresponds to a semantic image object. Obviously, this state (Figure 2b) was not the global optimum and the algorithm continued until convergence (Figure 2e). All four building were extracted and their detected boundaries are shown in red. Two shape priors from Figure 1 (the same as above) were finally chosen for the recovery of the four detected buildings.

In addition, in the top row of Figure 3, the result of the same prior-based contour evolution (in black) is shown superimposed on the input satellite image. The recognition-driven labelling process did detect, in an unsupervised manner, image building regions and simultaneously the selected priors did permit the reconstruction of the familiar objects. The corresponding 3D plots of the two labeling functions are shown in the middle two rows of the figure. The k-dimensional labeling function allowed automatically multiple instances depending on the number of the detected segments from the data-driven term. For example after a couple of iterations (second column), only one labeling function was able to handle the two detected segments. In algorithms convergence the segmentation result obtained with $k = 2$ labeling functions. Each function controlled which image region has been associated with which label configuration. Thus, by construction, the energy minimization leads to a partition of the image plane into areas of influence associated with each shape model. The two parallelepiped buildings in the bottom right of the image were associated with the second labeling function and the two others with the first one. Such an evolution of the labeling regions (areas of influence) was driven by a competition between the different shape priors. The joint multiscale optimization of the transformation parameters allowed to keep track of the correct pose of each object. Due to such a formulation each location (area of influence) could only be associated with one shape prior and therefore, the algorithm is forced to decide which prior favors most image data.

A visual comparison between the binary output of the purely intensity-based segmentation (Figure 4b) and and the one of the proposed, here, prior-based process (Figure 4c) demonstrates the superior results that were obtained. The resulting output from the proposed here framework did manage to highly match the ground truth (Figure 4e). The algorithm influenced by the labeling function, was robust and managed to surpass the irrelevant non-semantic segments. Above observations are supported by the quantitative evaluation, which indicated that: (i) the purely intensity-based segmentation scored really low with the overall detection quality at about only 70% (Table 1, figure3a) and (ii) the proposed, here, recognition-driven process successfully managed to extract accurately all image buildings with a completeness of about 93%, a correctness of 95% and an overall detection quality of about 88% (Table 1, figure3b). These quantitative results can be compared with the lower rates reported by other automatic algorithms [Doucette et al., 2005, Mayer et al., 2006] but not directly since different data were used and apart from buildings the detection was focused on other man-made objects, too. However, the developed algorithms efficiency should be emphasized.

Moreover, the developed algorithm was applied for the detection of buildings to an aerial (appx. 0.7m ground resolution) test image, which covers a wider area, appears complex and where mul-

[1] http://www.mas.ecp.fr/vision/Personnel/karank/Demos/2D

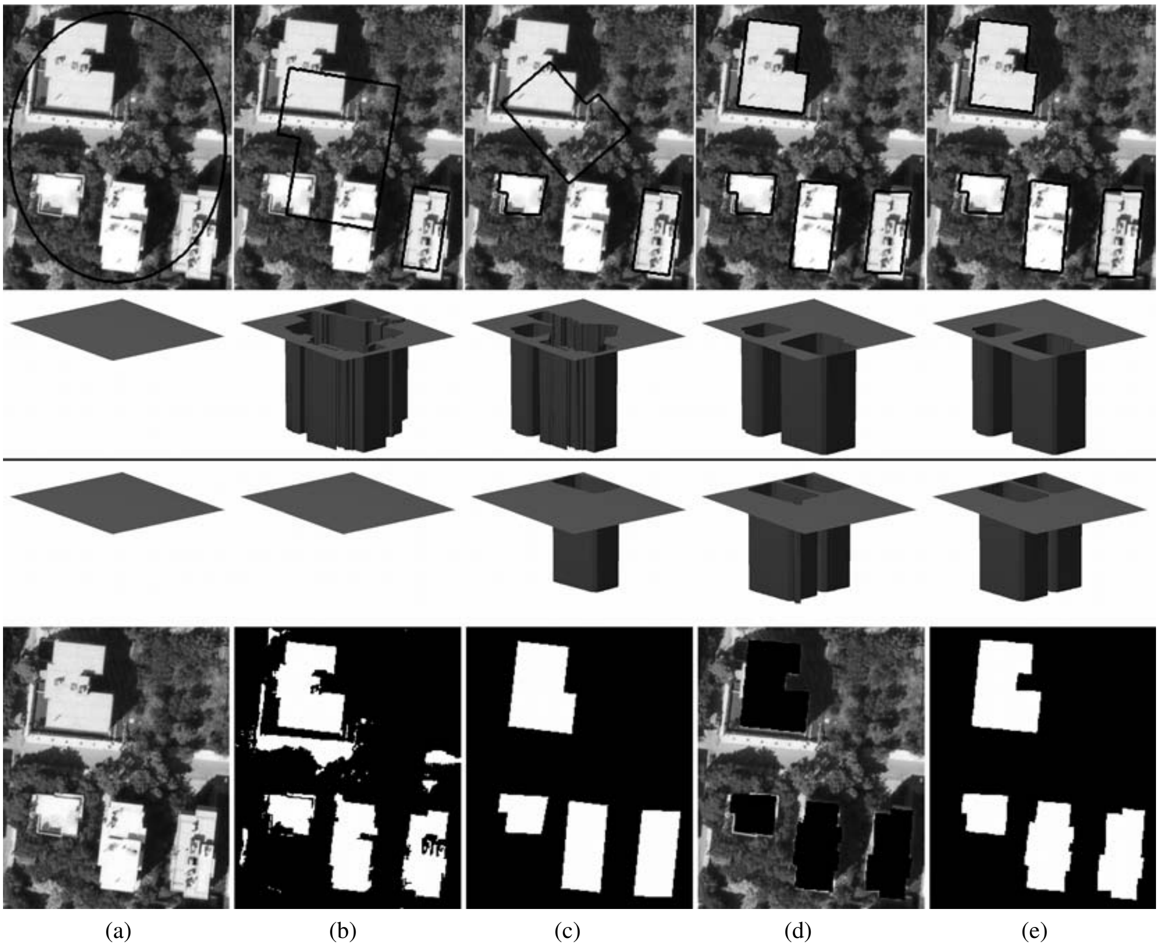


Figure 3: Qualitative evaluation after the application of the proposed recognition-driven segmentation to a high resolution aerial image. First row: The algorithm did manage to extract all fourth buildings and resulting contours (in black) describe accurately buildings boundaries. Second and third row: 3D plots from the evolution of the dynamic labeling. The k-dimensional function allowed automatically multiple instances depending on the number of the detected segments from the data-driven term. After a couple of iterations only one labeling function was needed to handle the two detected segments, while in algorithms convergence the result is obtained with a $k = 2$ labeling functions. Fourth row: Initial image (a), the binary output of the pure image-driven functional (b), algorithm's binary output (c), the ground truth superimposed in black color (d) and the binary ground truth (e).

tiple objects of different classes, shadows, occlusions, different texture patterns and some terrain height variability exists. In Figure 4 the final detected building regions are shown. All buildings, except one, were fully or partly detected. Most of them have been recognized as different identities (are labelled and numbered uniquely) apart from the three-building segment in the top right of the image which i) was poorly detected and ii) appears as one segment in the ground truth data, as well. The correctness of the detection was high at appx. 93% with a completeness at 88% (Table 1, figure 4b). The overall quality of algorithms performance was at 82%, while the detection based on only to the data-term was lower than 76% (Table 1, figure 4a).

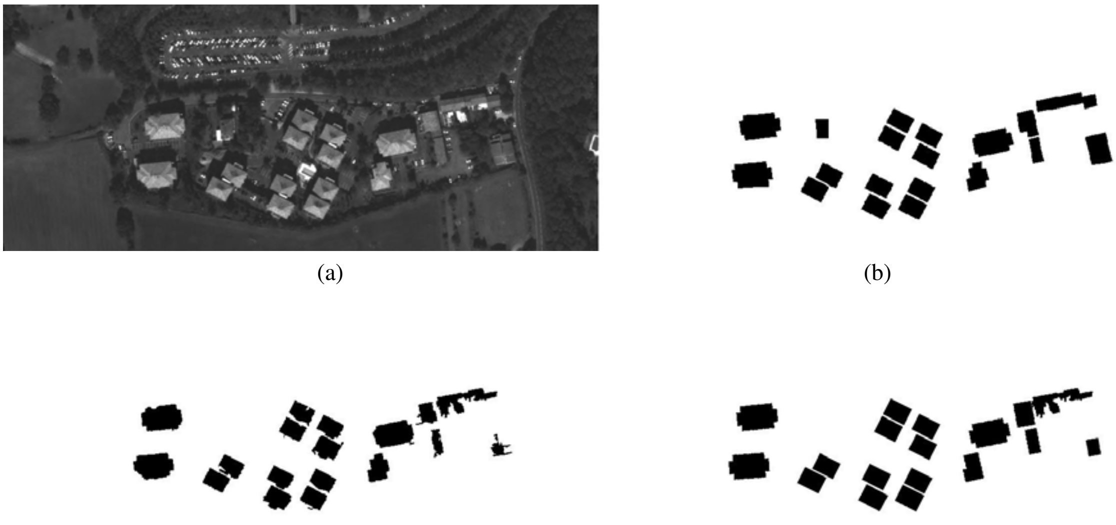
5 CONCLUSIONS AND FUTURE WORK

We have introduced a novel recognition-driven variational framework which accounts for automatic and accurate multiple building extraction from aerial and satellite images. We argued that the proposed framework fundamentally extends previous approaches towards the integration of shape priors into the level set segmentation and in particular (i) by allowing multiple competing priors contrary to [Riklin-Raviv et al., 2007] and (ii) without the need of having a priori knowledge for the pose of objects in image's plane, contrary to [Cremers et al., 2006]. The proposed cost functional is simultaneously optimized with respect to (i) the data-driven term based on the level set function ϕ controlling the segmentation, (ii) the vector-valued labeling function which indicates regions of influence where the competing shape priors should be enforced and (iii) a set of parameters associated with the projective transformation of each prior. The evolution of the labeling function is driven by the competing shape priors and each selected image region is ascribed to the best fitted one. The functional is, also, consistent with the philosophy of level sets as it allows multiple independent object detection.

The successful segmentation results, the reliable estimation of the transformation parameters and the adequate performance of the dynamic labeling encourage future research. A comprehensive solution for general 3D objects would require to extend both the

	Quantitative Measures		
Detection case	Completeness	Correctness	Quality
Figure 3b	0.868	0.790	0.705
Figure 3c	0.926	0.946	0.879
Figure 4c	0.813	0.918	0.758
Figure 4d	0.877	0.927	0.820

Table 1: Quantitative Evaluation



(a) (b)

(c) (d)

Figure 4: Results from the application of the proposed recognition-driven segmentation to a high resolution aerial image. First row: Initial image (a) and the binary ground truth (b). Second row: the binary output of a pure data-driven segmentation (c) and the binary output after the application of the proposed algorithm (d).

transformation model beyond planar projective homography and the labeling function beyond k-dimensional 2D instances. Similarly, for the extension to 4D objects and the reconstruction of buildings in time from several temporal-different data, statistical shape priors (which additionally allow deformation modes associated with each model) are conceivable based on a training sets.

ACKNOWLEDGEMENTS

This work has been partially supported from the Conseil General de Hauts-de-Seine and the Region Ile-de-France under the TERRA NUMERICA grant of the Pole de competitivite CapDigital.

REFERENCES

Ball, J. E. and Bruce, L. M., 2005. Level set segmentation of remotely sensed hyperspectral images. In: IEEE International Geoscience and Remote Sensing Symposium, Seoul, Korea, pp. 5638–5642.

Besbes, O., Belhadj, Z. and Boujemaa, N., 2006. Adaptive satellite images segmentation by level set multiregion competition. Technical Report 5855, INRIA.

Cao, G., Yang, X. and Mao, Z., 2005. A two-stage level set evolution scheme for man-made objects detection in aerial images. In: IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, pp. 474–479.

Chan, T. and Zhu, W., 2003. Level set based shape prior segmentation. Technical Report 03-66, Computational Applied Mathematics, UCLA, Los Angeles.

Cremers, D., Sochen, N. and Schnörr, C., 2006. A multiphase dynamic labeling model for variational recognition-driven image segmentation. International Journal of Computer Vision 66(1), pp. 67–81.

Doucette, P., Agouris, P. and Stefanidis, A., 2005. Automation and Digital Photogrammetric Workstations, Manual of Photogrammetry (5th edition; C. McGlone, ed.). ASPRS, pp. 949-981.

Faugeras, O., Luong, Q.-T. and Papadopoulo, T., 2001. The Geometry of Multiple Images. MIT Press.

Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press, 2nd edition.

Irani, M. and Anandan, P., 1999. All about direct methods. Vision Algorithms: Theory and Practice, Springer-Verlag.

Itti, L. and Koch, C., 2001. Computational modelling of visual attention. Nature Reviews, Neuroscience 2, pp. 194–203.

Karantzalos, K. and Argialas, D., 2006. An image segmentation level set method for building detection. In: International Symposium of Remote Sensing, Busan, Korea, pp. 610–614.

Keaton, T. and Brokish, J., 2002. A level set method for the extraction of roads from multispectral imagery. In: Workshop on Applied Imagery Pattern Recognition, pp. 141–150.

Mayer, H., 1999. Automatic object extraction from aerial imagery-a survey focusing on buildings. Computer Vision and Image Understanding 74(2), pp. 138–149.

Mayer, H., Hinz, S., Bacher, U. and Baltsavias, E., 2006. A test of automatic road extraction approaches. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 36, pp. 209–214.

Niu, X., 2006. A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model. ISPRS Journal of Photogrammetry and Remote Sensing 61, pp. 170–186.

Paragios, N., Chen, Y. and Faugeras, O., 2005. Handbook of Mathematical Models of Computer Vision. Springer.

Peng, J., Zhang, D. and Liu, Y., 2005. An improved snake model for building detection from urban aerial images. Pattern Recognition Letters 26, pp. 587–595.

Riklin-Raviv, T., Kiryati, N. and Sochen, N., 2007. Prior-based segmentation and shape registration in the presence of perspective distortion. International Journal of Computer Vision 72, pp. 309–328.

Rousson, M. and Paragios, N., 2008. Prior knowledge, level set representations & visual grouping. International Journal of Computer Vision 76(3), pp. 231–243.

Samson, C., Blanc-Feraud, L., Aubert, G. and Zerubia, J., 2001. Two variational models for multispectral image classification. In: Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, (Lecture Notes In Computer Science), pp. 344–358.

Springer, C., 1964. Geometry and Analysis of Projective Spaces. Freeman.

Vese, L. and Chan, T., 2002. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. International Journal of Computer Vision 50, pp. 271–293.

Walther, D. B. and Fei-Fei, L., 2007. Task-set switching with natural scenes: Measuring the cost of deploying topdown attention. Journal of Vision 7, pp. 1–12.