Recognition-Driven 2D Competing Priors Towards Automatic And Accurate Building Detection

Konstantinos Karantzalos, *Member, IEEE,* and Nikos Paragios, *Senior Member, IEEE*

***Abstract*—In this paper, a novel recognition-driven variational framework is introduced, towards multiple building extraction from aerial and satellite images. To this end, competing shape priors are considered and building extraction is addressed through an image segmentation approach that involves the use of a data-driven term constrained from the prior models. The proposed framework extend previous approaches towards the integration of multiple shape priors into the level set segmentation. In particular, it estimates the number of buildings as well as their pose from the observed data. Therefore, it can address multiple building extraction from a single optical image, a highly demanding task of fundamental importance in various geoscience and remote sensing applications. Furthermore, it can be easily extended to deal with other remote sensing data through a simple modification of the image term. Very promising experimental results and the performed qualitative and quantitative evaluation demonstrate the potential of our approach.**

***Index Terms*—variational methods, recognition, segmentation, level sets, extraction, registration, object detection**

I. Introduction

RESEARCH towards the automatic extraction of buildings and other man-made objects from aerial and satellite imagery has gain significant attention over the last decade [1], [2], [3], [4], [5], [6]. Among various methods, processing schemes and systems of the literature, curve propagation techniques (snakes, active contours, deformable models and more recently level sets) have revealed promising results [7], [8], [9], [10]. Their main strength is the ability to cope with topological changes. On top of that, they offer natural means of integrating boundary as well as regional information. Conventional level sets have been employed to account for the general task of segmenting satellite images [11], [12], [13], for the detection of roads (in a semi-automatic framework) [14], [15] and for the automatic detection of buildings and other man-made objects [9], [10]. These methods were purely image-based and therefore vulnerable to misleading low-level information, like shadows or occlusions, which is a common scenario in remote sensing data.

In order to overcome this limitation the idea of combining image-based costs with prior constraints, related with the geometry of the objects of interest, was considered in the field of computer vision. It was motivated by the reported observations that human visual perception involves a set of processes for distinguishing top-down attention from the stimulus-driven bottom-up one [16]. Several problems/applications in computer vision relate perception with specific-object recognition tasks and 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, can become robust to shadows, noise, background clutter or partial occlusions [17].

Among the prior-based approaches, the statistical ones encode the object pose through rigid or similarity transformations and aim to account only for the local variations using a small set of non-rigid deformations and a comprehensive training set [18], [19], [20], [21], [22]. Shapes distribution probability is been determined and then the similarity between the evolving object boundary and its projection to the learned distribution upon subtraction of the pose is measured. None of the above methods can account for projective transformations between the prior shapes and the desired for extraction shapes [23].

Furthermore, both the similarity and the affine model do not provide reasonable approximation for the transformation between the prior shape and the shape to segment contrary to [24], where a framework based on homography projective transformations was employed. Such a prior formulation, though, did not account for multiple priors and multiple desired objects, like the approaches proposed in [25], [26] where a labeling function allows the use of multiple priors for segmentation. The main limitation of [26] is that *a priori* knowledge for the pose of objects in image's plane is assumed. Another challenge to be addressed in the context of satellite imaging is that one expects an important number of buildings both in terms of quantity as well shape variation being present in the image. Therefore existing approaches, which were designed to segment a single known object in a given image ([18], [19], [20], [21], [23], [24]) are not suitable.

In this paper, we aim to solve the problem of automatic building extraction/segmentation from satellite images. In particular we would like to overcome limitations of existing inaccurate data-driven segmentation caused by the misleading low level information due to shadows or occlusions. Therefore, we propose a novel prior-based variational framework, which can account for automatic building extraction from a single image. Accurate building boundaries extraction from a single image is a highly demanding task [2] of major importance for supporting several government activities and various GIS applications like map generation and update [27], [28], [2], [29]. The high resolution images are usually available, nowadays, in a single panchromatic channel as in the IKONOS and the QUICKBIRD cases.

K. Karantzalos and N. Paragios are with the MAS Laboratory, Ecole Centrale de Paris, Grande Voie des Vignes, 92 295, Chatenay-Malabry, FRANCE. e-mail: {konstantinos.karantzalos, nikos.paragios}@ecp.fr (see http://vision.mas.ecp.fr/index.html).

Manuscript received June 28, 2008.

Furthermore, the introduced, here, prior-based recognition process is able to efficiently account for automatic multiple building extraction, no matter if their number and shape is familiar or not. To this end, an elegant and powerful mathematical formulation that constrains the propagation of the contour through a partial alignment with a database of prior buildings 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 [25], [26].

The labeling function evolves in time and incrementally determines multiple instances according to the number of the detected objects and the selected shape priors. 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 [30] were used and thus color or texture compatibility between the prior. Parametrization-free shape descriptions possess a significant advantage over landmark-based and template matching techniques, which represent shapes by collections of points or features.

The main contributions of our paper are:

- We propose a variational framework for the integration of multiple competing shape priors that is pose/affine invariant through an explicit estimation of the transformation (opposite to [26]). The proposed functional can explicitly account for the planar projective transformation of any shape prior based on its multiscale optimization process. The segmentation process is carried out concurrently both with the dynamic labeling -contrary to [24]- and the registration of the competing prior shapes overcoming shadows or occlusions. Our labeling function evolves in time allowing multiple instances depending on the number of detected objects.
- Our framework fundamentally extends previous work for automatic building detection in single panchromatic images. We offer an extensive experimental qualitative and quantitative evaluation demonstrating the efficiency of the proposed approach. We show results involving multiple buildings detection on panchromatic high resolution aerial and satellite images. Such results along with the reliable estimation of the transformation suggest that the proposed forms a promising tool for various remote sensing segmentation, registration and object detection applications.

The remainder of the paper is organized in the following way. In Section II, we briefly describe a conventional region-based level set energy functional and demonstrate its limitations. The projective-invariant prior-based formulation for building detection is presented in Section III. The generalized variational framework for the integration of multiple competing shape priors for building extraction is detailed in Section IV. The performed qualitative and quantitative evaluation and an overall discussion on the algorithm's performance are presented in Section V. Finally, conclusions and perspectives for future work are in Section VI.

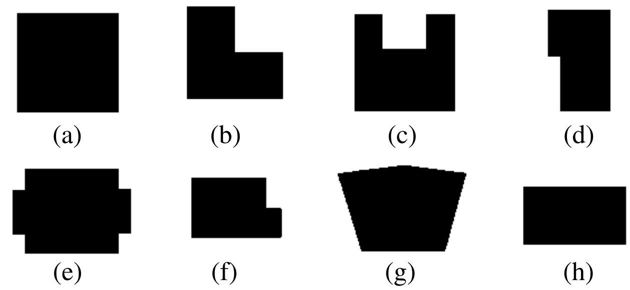


Fig. 1. A database of 8 prior binary templates. Each one competes the other one to fit best in a building segment

II. Data-driven Level Set Based Segmentation

Level set methods [31], [32] have became a popular framework for image segmentation [17]. Given an image $\mathcal{I}(\mathbf{x})$ at domain (bounded) $\Omega \in \mathcal{R}^2$ and an interface C, one can define the level set representation $\phi : \Omega \rightarrow \mathcal{R}^+$ as a Lipschitz function:

$$\phi(\mathbf{x}; C) = \begin{cases} 0 & , \mathbf{x} \in C \\ +\mathcal{D}(\mathbf{x}, C) > 0 & , \mathbf{x} \in [\Omega_C - \Omega] \\ -\mathcal{D}(\mathbf{x}, C) < 0 & , \mathbf{x} \in [\Omega - \Omega_C] \end{cases} \tag{1}$$

where $\mathcal{D}(\mathbf{x}, C)$ is the Euclidean distance between the pixel $\mathbf{x}$ and the interface. The interface C is represented as a zero level set of the ϕ function and for any flow being defined on the interface, an implicit one can be determined [32] that evolves the embedding function ϕ in such a way that the zero-level of ϕ corresponds always to the interface. Such a level set formulation can be considered to define an optimization framework. To this end, one can consider the approximations of Dirac and Heaviside distributions:

$$\delta_\epsilon(\phi) = \begin{cases} 0, & |\phi| > \epsilon \\ \frac{1}{2\epsilon}\left(1 + \cos\left(\frac{\pi\phi}{\epsilon}\right)\right), & |\phi| < \epsilon \end{cases}$$
$$H_\epsilon(\phi) = \begin{cases} 1, & \phi > \epsilon \\ 0, & \phi < -\epsilon \\ \frac{1}{2}\left(1 + \frac{\phi}{\epsilon} + \frac{1}{\pi}\sin\left(\frac{\pi\phi}{\epsilon}\right)\right), & |\phi| < \epsilon \end{cases} \tag{2}$$

These functions can be used to define: (i) a smoothness term penalizing the length of the contour, (ii) energy terms encoding global regional information of the object as well as of the background. These region-based energetic modules aim at separating the object from the background forming in a way adaptive balloon forces:

$$(i) \underbrace{\int_\Omega \delta_\epsilon(\phi)|\nabla\phi| d\mathbf{x}}_{contour\ length}$$

$$(ii) \quad \underbrace{\int_\Omega H_\epsilon(\phi) r_{obj}\left(\mathcal{I}(\mathbf{x})\right) d\mathbf{x},}_{object\ regional\ term}$$

$$(iii) \quad \underbrace{\int_{\Omega} (1 - H_{\epsilon}(\phi)) r_{bg} \left(\mathcal{I}(\mathbf{x})\right) d\mathbf{x}}_{background\ regional\ term}$$

where $r_{obj} : \mathcal{R}^{+} \rightarrow [0,1]$ and $r_{bg} : \mathcal{R}^{+} \rightarrow [0,1]$ are *object* and *background* positive monotonically decreasing data-driven functions. One can imagine the integration of these components to define an image partition that is optimal with respect to some grouping criterion. The first term (i) is a smoothness component and the others two (ii and iii) form a grouping component that accounts for some regional properties (modulo the definition of r_{obj} and r_{bg}) of the area defined by the evolving interface. Such descriptors -which, also, make the approach relatively independent to the initial conditions- measure the quality of matching between the observed image and the expected regional properties of the structure of interest and the background.

Choosing the appropriate region descriptors (r_{obj} and r_{bg}) depends heavily on the nature of the images to be considered. One can model the scene in regions with desired objects and in the background and then, assume that these regions are characterized by Gaussian densities [33]. When such an assumption seems unrealistic one can consider a more flexible parametric density function, like a gaussian mixture [34] or non-parametric densities [35] in order to describe the visual properties of the object and the background. Furthermore, in cases were color information or other remote sensing data like radar or hyperspectral imagery is available, these region descriptors can be accordingly formulated. One should note that (i) the aim of our paper is not to address the image component of the method, and (ii) such image components are defined in a modular and can easily adapted to the image content. Therefore, we will assume a rather simple data component just for demonstration purposes.

Employing the above energy modules a region-based segmentation functional can be considered to account for multiple building detection from aerial and satellite imagery:

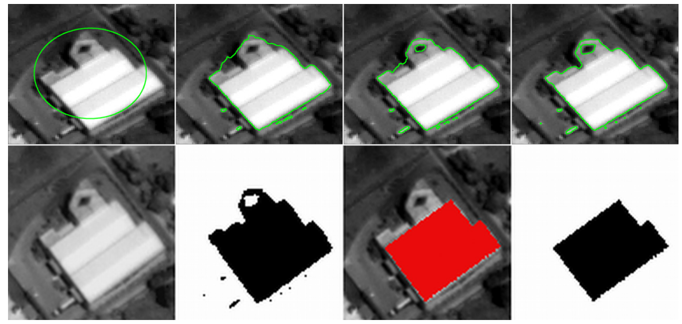
$$\begin{aligned} E_{seg}(\phi, r_{obj}, r_{bg}) \;=\; & \frac{1}{2}\int_{\Omega} \{H_{\epsilon}(\phi)\; r_{obj}\left(\mathcal{I}(\mathbf{x})\right) \;+ \\ & [1 - H_{\epsilon}(\phi)]\; r_{bg}\left(\mathcal{I}(\mathbf{x})\right) \;+\; \nu\delta_{\epsilon}(\phi)|\nabla\phi(\mathbf{x})|\}\; d\mathbf{x} \end{aligned} \quad (3)$$

In all our experiments, in a similar manner with [33], [10], the following region descriptors were employed:

$$r_{obj}\left(\mathcal{I}(\mathbf{x})\right) = \frac{(\mu_{obj} - \mathcal{I}(\mathbf{x}))^2}{\sigma_{obj}^2}, \quad r_{bg}\left(\mathcal{I}(\mathbf{x})\right) = \frac{(\mu_{bg} - \mathcal{I}(\mathbf{x}))^2}{\sigma_{bg}^2}$$

where μ_{obj} is the mean and σ_{obj} the covariance matrix of the object appearance (similar definition for the background). Using such a formulation the scene was modelled as a collection of smooth surfaces and a background, based on observations made at every iteration on the panchromatic image.

Segmentation results from such a data-driven functional (using a gradient descent optimization procedure) are presented in Figure 2. It is clear that for cases where the region-homogeneity criterion is violated then the purely image-based scheme will fail to converge to the desired solution. This will be the case when the term $\left((\mu_{obj} - \mathcal{I}(\mathbf{x}))/\sigma_{obj}\right)^2$ can not sufficiently model the intra-region distributions, and respectively the r_{bg}. The condition of non-homogeneous regions arises frequently in satellite images either due to shadows or other unfavorable lighting conditions (i.e background clutter or partial occlusion of the objects of interest). The algorithm managed to accurately detect buildings boundaries in its bottom right and left part, but failed in its top, where the intensity was more smooth (due to the similar reflectance of the front parking area and building's roof). The intensity information was clearly insufficient to define the object of interest. Above observations correlate well with the performed quantitative evaluation (Table 1: figure 2), which indicates algorithms poor performance scoring low in all measures (i.e the overall detection's quality was below 80%). Such quantitative measures have become standard for man-made object extraction validation [29].



(a) (b) (c) (d)

Fig. 2. Curve evolution based, only, on the data-driven term of Equation 3. First row: Starting with an arbitrary elliptical curve (first image), the algorithm converges (last image) to the detected boundaries shown in green. Second row: Initial image (a), algorithm's binary output (b), the ground truth superimposed in red color (c) and the binary ground truth (d).

To cope with such degraded low-level information, we were motivated to incorporate global shape prior constrains into the level set scheme.

III. Projective-Invariant Shape Prior Formulation

The basic idea lies in the extension of the data-driven cost functional by adding another energy E_{prior} which favors certain contour formations:

$$E_{total} = E_{seg}(\phi, r_{obj}, r_{bg}) + \mu E_{prior}(\phi) \quad \mu > 0 \quad (4)$$

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

$$E_{prior} = \int_{\Omega} \Big(H_{\epsilon}(\phi(\mathbf{x})) - H_{\epsilon}(\tilde{\phi}(\mathbf{x})) \Big)^2 d\mathbf{x} \quad (5)$$

where $\tilde{\phi}$ is the level set function embedding a given training shape (or the mean of a set of training shapes). The representation of the prior shape within the above energy functional is a 3D function $\tilde{\phi} : \tilde{\Omega} \rightarrow R$ that embeds the contour $\tilde{C}$ of the known shape:

$$\tilde{C} = \{\mathbf{x} \in \tilde{\Omega} \mid \tilde{\phi}(\mathbf{x} = 0)\}$$

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. For consistency, the segmenting level set function ϕ is, also, projected to the space of distance functions during the optimization.

With the above formulation the pose and location of the object of interest are assumed to be identical to the ones of the reference shape. In the context of automatic building detection from aerial and satellite imagery, 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 [24]. 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 natural 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 $\tilde{\phi}$.

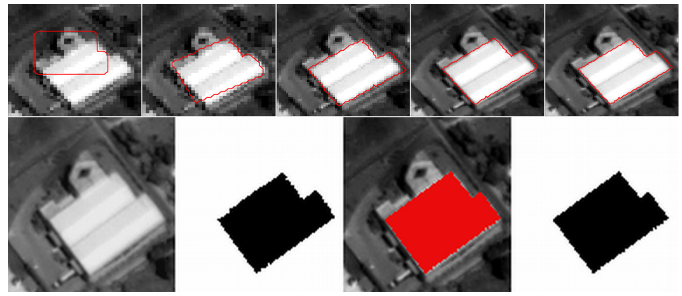
A. *Planar Projective Homography*

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 [36]. 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) [36], [37].

Here, similarly to the formulations of [24] the homograph is calculated directly in its explicit form:

$$\mathcal{T} = r + \frac{1}{d} t n^T \tag{6}$$

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



(a) (b) (c) (d)

Fig. 3. Curve evolution based on the single prior-based segmentation of Equation 8. First row: The different steps until algorithms convergence, are shown, where the resulted boundaries in red (last image) accurately describe the building. Second row: Initial image (a), algorithms binary output (b), the ground truth superimposed in red color (c) and the binary ground truth (d).

from $\mathbf{y}$ to $\mathbf{y}'$ can be expressed as:

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

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:

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \; y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}, \tag{7}$$

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 [38], [37].

In particular, (**i**) the translation in the image is described by the vector $t = (t_x, t_y, t_z)$, (**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).

Rather than relying on point correspondence, the observed contour in the image and the prior model are registered using the calculus of variations. Note that since the recovery of the homography and the segmentation process are jointly addressed, only the prior shape is known in advance. The prior shape is matched to the shape being segmented as part of its detection procedure. In order to minimize the energy functional (Equation 4) one has to simultaneously evolve the level set function ϕ, estimate the region descriptors (r_{obj} and r_{bg}) and recover the transformation $\mathcal{T}(\mathbf{x})$ for a given prior level-set function $\tilde{\phi}$. 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 $\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right)$ is substituted for $\tilde{\phi}$ in Equation 5.

B. Energy Minimization

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

$$E_{prior}(\phi, \mathcal{T}) = \int_{\Omega} \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right))\right)^2 \, d\mathbf{x} \quad (8)$$

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:

$$\frac{\vartheta E_{prior}}{\vartheta u} = 2\mu \int_{\Omega} \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right))\right) \frac{\vartheta \mathcal{T}(u)}{\vartheta u} \, d\mathbf{x} \quad (9)$$

Following such a formulation the optimization procedure can cope with low-level misleading visual information (due to shadows, occlusions, etc) and can produce a successful building boundaries detection (Figure 3). The prior-driven segmentation approach managed to detect correctly building boundaries and overcome data-due limitations. The transformation of the shape prior (Figure 1f) was successfully recovered and the binary result (Figure 3b) highly matches the ground truth one (Figure 3d). Above observations are supported, also, by the quantitative measures shown in Table 1. All measures obtained high values of over 95% (Table 1: fig.3) contrary to the 78% of the overall detection rate that was obtained without the use of a prior (Table 1: fig.2).

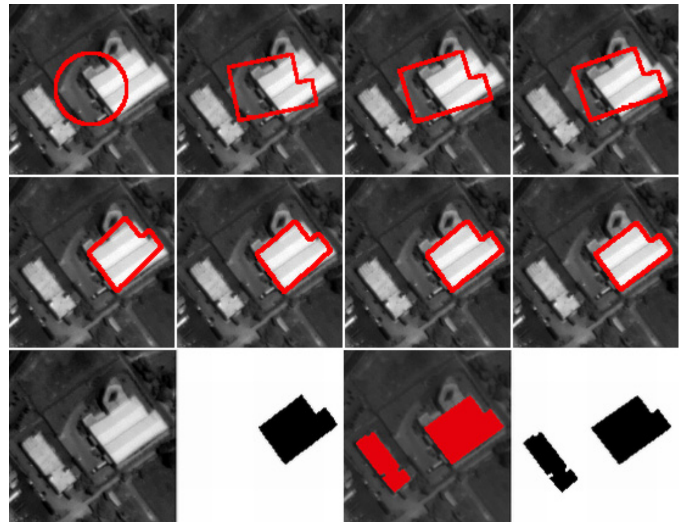
However, such a prior formulation (Equation 8) can not account for multiple buildings as is demonstrated in Figure 4. The optimization allows to reconstruct correctly the building at the right part of the image but all other unfamiliar objects, like the building in the left, are suppressed from the segmentation. The level set process obviously lost its capacity to handle multiple (independent) objects. The binary result (Figure 4b) differs much from the ground truth and the calculated quantitative measurements report a poor overall detection quality lower than 65% (Table 1: fig.4).

IV. 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 $L : \Omega \rightarrow \{-1, +1\}$, 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.

A. The Case With Two Buildings

Let us consider without loss of generality the simple case of two prior models and therefore a binary label process. The



(a) (b) (c) (d)

Fig. 4. Curve evolution based on the single prior-based segmentation of Equation 8. First two rows: The different steps until algorithms convergence, are shown. The final segmentation managed to extract only the building on the right. Third row: Initial image (a), algorithm's binary output (b), the ground truth superimposed in black color (c) and the binary ground truth (d).

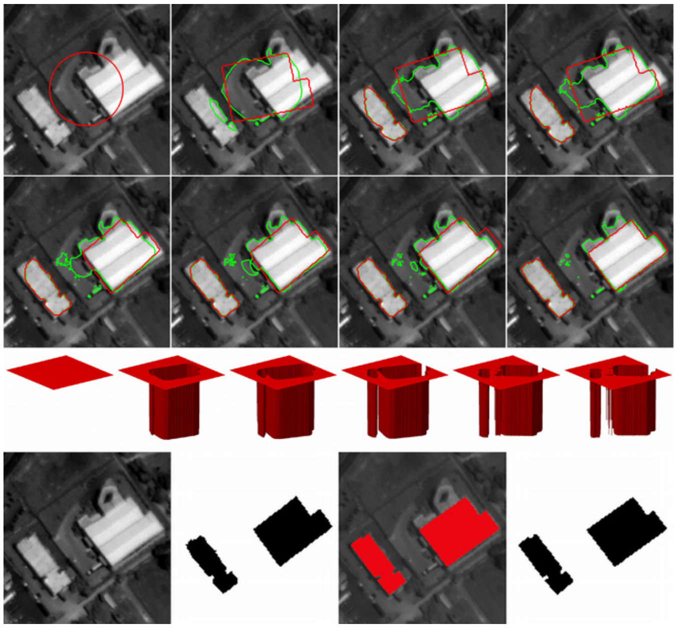
corresponding shape energy is given by:

$$E_{prior}(\phi, \mathcal{T}, L) = \int \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right))\right)^2 (L+1)^2 \, d\mathbf{x} + \int \lambda^2 (L-1)^2 \, d\mathbf{x} + \rho \int |\nabla L| d\mathbf{x} \quad (10)$$

with the two parameters $\lambda, \rho > 0$. For a fixed ϕ, minimizing the first two terms in above Equation 10, oblige the labeling function L to obtain the following qualitative behavior:

$$\begin{aligned} L \rightarrow +1, &\quad \text{if } |H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right))| < \lambda \\ L \rightarrow -1, &\quad \text{if } |H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}\left(\mathcal{T}(\mathbf{x})\right))| > \lambda \end{aligned} \quad (11)$$

Thus, L enforces the shape prior $\tilde{\phi}$ in those image areas where the level set function ϕ is similar to the prior (i.e L = 1). The last term in equation (10), acts as any TV regularization operator and forces the boundary that separates labelling regions to have the minimal length. The main challenge in the context of optimization with respect to the membership function L lies on the fact that is defined in the discrete domain along with variables which are part of the continuous domain. Addressing the optimization of discrete and continuous variables being interdependent is rather challenging, therefore we consider a fully continuous approach for L. One can either replace L with the $sign(L)$, which can be expressed using as $L/|L|$ and then carry on the derivations using this new model. The derivative of the cost function is more complex but can be determined in a straightforward way. However, this approach might suffer from instabilities due to the numerical approximation of the gradient. A more practical solution consists of projecting the continuous solution once convergence has been obtained in the current iteration to the discrete set of values using the formula presented (11) which was considered for the shake of simplicity and stability of the numerical approximation of



(a) (b) (c) (d)

Fig. 5. Curve evolution based on the prior-based dynamic labeling of Equation 10. First two rows: The different steps until algorithms convergence, are shown. The algorithm managed to extract both buildings and resulting contours (in red) describe accurately their boundaries. The evolution of the data-driven term is, also, shown in green. Third row: 3D plots from the evolution of the labeling function. Starting from its initialization ($L(:)$=1) the labeling evolves dynamically controlling the regions where the shape prior is been applied. Fourth row: Initial image (a), algorithm's binary output (b), the ground truth superimposed in red color (c) and the binary ground truth (d).

the gradient. The same concept was used when multiple labels were considered.

Minimizing the energy of Equation 4 with the prior formulation of Equation 10 results to the successful building boundaries detection presented in Figure 5. The shape prior permits to reconstruct the building on the right and in contrast to Figure 4 the process dynamically selects the region where to impose the prior. The selection process and corresponding evolution of the labeling function are shown in Figure 5 (third row). Consequently, the correct detection of the two unknown objects is unaffected by the prior and the resulted binary output (Figure 5b) highly match the ground truth (Figure 5d). All measures from the quantitative evaluation (Table 1: fig.5) reported high scores in correctness, completeness and quality (all were over 91%), contrary to the previous much lower ones (Table 1: fig.4).

Still, a limitation of the above labeling formulation (Equation 10) is that it only accounts for a single shape prior. The following modification can allow the usage of two different priors $\tilde{\phi}_1$ and $\tilde{\phi}_2$:

$$E_{prior}(\phi, \mathcal{T}, L) = \frac{1}{\sigma_1^2} \int \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_1\,(\mathcal{T}_1(\mathbf{x}))) \right)^2 (L+1)^2\, d\mathbf{x} + \frac{1}{\sigma_2^2} \int \left(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_2\,(\mathcal{T}_2(\mathbf{x}))) \right)^2 (L-1)^2\, d\mathbf{x} \;+\; \rho \int |\nabla L|\, d\mathbf{x} \tag{12}$$

Similar to [26], the terms associated with the two objects are normalized with respect to the variance of the respective template: $\sigma_i^2 = \int \phi_i^2 d\mathbf{x} - \int \phi_i d\mathbf{x}^2$.

B. The General Case

Let us now consider the general case of a larger number of building shape priors (like those in 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

$$\mathbf{L} : \Omega \to R^k, \quad \mathbf{L}(\mathbf{x}) = (L_1(\mathbf{x}), ..., L_k(\mathbf{x})) \tag{13}$$

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 [25], [26]:

$$x_i(\mathbf{L}) = \frac{1}{4^k} \prod_{j=1}^{k} (L_j - w_j)^2, \quad \text{with } w_j \in \{-1, +1\} \tag{14}$$

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:

$$E_{total} = E_{seg}(\phi, r_{obj}, r_{bg}) + \mu E_{prior}(\phi, \mathcal{T}, \mathbf{L}) \tag{15}$$

where:

$$E_{prior}(\phi, \mathcal{T}, \mathbf{L}) = \sum_{i=1}^{m-1} \int \left(\frac{H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_i\,(\mathcal{T}_i(\mathbf{x})))}{\sigma_i} \right)^2 x_i(\mathbf{L}) d\mathbf{x} + \int \lambda^2 x_m(\mathbf{L}) d\mathbf{x} + \rho \sum_{i=1}^{m} \int |\nabla L| d\mathbf{x} \tag{16}$$

Contrary to [39] and [26] 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:

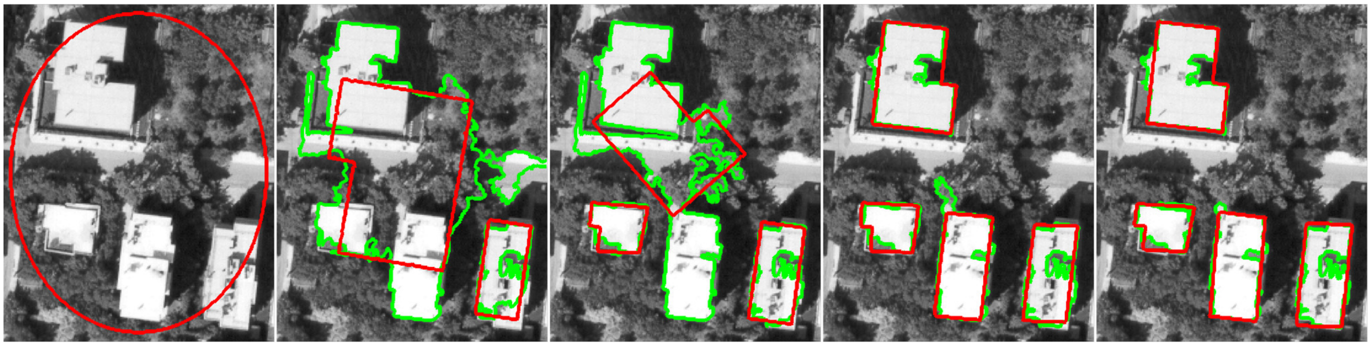
$$k = \lceil \frac{\log(1+q)}{\log 2} \rceil \tag{17}$$

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.

C. Energy Minimization

The prior-based segmentation process is generated by minimizing the functional of Equation 15. Minimization is performed by alternating the update of the region 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 $\tilde{\phi}_i$:

1) Evolution of the Segmentation: For fixed labeling and pose parameters, the level set function ϕ evolves according



(a) (b) (c) (d) (e)

Fig. 6. Curve evolution based on the proposed, here, recognition-driven formulation of Equation 15. The different steps until algorithms convergence, are shown (in red). The algorithm did manage to extract all fourth buildings and resulting contours (in red) describe accurately buildings boundaries. The evolution of the data-driven term is, also, shown in green.

to:

$$\frac{\vartheta E_{total}}{\vartheta \phi} = \frac{\vartheta E_{seg}}{\vartheta \phi} - 2\mu \sum_{i=1}^{m-1} \frac{H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_i\,(\mathcal{T}_i(\mathbf{x})))}{\sigma_i^2} x_i(\mathbf{L}) \tag{18}$$

Apart from the first image-driven component of Equation 3, there is an additional relaxation term towards the prior $\tilde{\phi}_i$ in all image regions where $x_i > 0$. Thus, the segmentation favors the curve propagation in regions indicated by the labeling function, ameliorating the segmentation process.

2) 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_j corresponds to an evolution of the form:

$$\frac{\vartheta E_{total}}{\vartheta L_j} = -\mu \sum_{i=1}^{m-1} \frac{(H_\epsilon(\phi) - H_\epsilon(\tilde{\phi}_i\,(\mathcal{T}_i(\mathbf{x}))))^2}{\sigma_i^2} \frac{\vartheta x_i}{\vartheta L_j} - \mu\,\lambda^2 \frac{\vartheta x_m}{\vartheta L_j} - \mu\,\gamma\,div \frac{\nabla L_j}{\|\nabla L_j\|}, \tag{19}$$

where the derivatives of the indicator functions x_i are calculated from (14). The first two terms in Equation 19 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. For example, in case where $k = 2$, then four regions can be modeled by the indicator functions and L_j takes the following form:

$$x_1(\mathbf{L}) = \frac{1}{16}(L_1-1)^2(L_2-1)^2\,,\; x_2(\mathbf{L}) = \frac{1}{16}(L_1+1)^2(L_2-1)^2\,,$$
$$x_3(\mathbf{L}) = \frac{1}{16}(L_1-1)^2(L_2+1)^2\,,\; x_4(\mathbf{L}) = \frac{1}{16}(L_1+1)^2(L_2+1)^2 \tag{20}$$

3) 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 9. In order, though, to handle both global and local shape deformations a multiscale optimization was introduced. The multiscale approach is implement via a fixed point iteration on both the level set function ϕ and the shape priors $\tilde{\phi}_i$ with a down-sampling 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 $\phi^l, (l = 0, 1, ...)$, 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:

$$\frac{\vartheta E_{total}}{\vartheta u_i^l} = 2\mu\; x_i^l(\mathbf{L}^l) \int_\Omega \left(\frac{H_\epsilon(\phi^l) - H_\epsilon(\tilde{\phi}_i^l\,(\mathcal{T}_i^l(\mathbf{x})))}{\sigma_i^2} \right) \frac{\vartheta \mathcal{T}_i^l(u_i^l)}{\vartheta u_i^l} \tag{21}$$

Above equation is analogous to Equation 9 (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 the shape $\tilde{\phi}_i$, i.e. to the area where $x_i > 0$ and (ii) moreover, is calculated via fixed point iterations l.

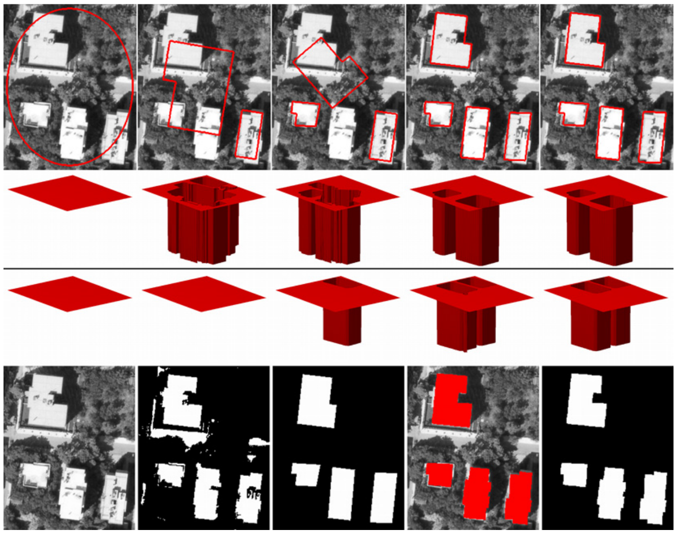
V. Evaluation

A. Experimental Results

Having already discussed the shortcomings of a purely data-driven procedure against a single prior-based one (Figures 2 and 3) and the advantages of using a labeling function in order to retain level sets multi-object segmentation properties (Figures 4 and 5), we proceed by presenting experimental results from the application of the introduced recognition-driven variational framework to high resolution aerial and satellite data sets. Note that in all paper's figures, the width of the (superimposed on the original image) evolving contours has been enlarged for visualization purposes and it is not related with method's detection accuracy. With just a standard interpolation technique, the spatial localization of the detected boundaries is performed with sub-pixel accuracy. Experimental results and videos, which demonstrate method's performance, are also provided in author's web-page[1]

In Figure 6, different optimization steps until algorithm's convergence are shown towards automatic building extraction from a high resolution satellite image. Starting with an arbitrary elliptical curve (Figure 6a) and after a couple of

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



(a) (b) (c) (d) (e)

Fig. 7. Qualitative evaluation after the application of the proposed recognition-driven segmentation to a high resolution aerial image. First row: The algorithm managed to extract all fourth buildings and resulting contours (in red) 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 just 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 of Equation 3 (b), algorithm's binary output (c), the ground truth superimposed in red color (d) and the binary ground truth (e).

iterations the data-driven term (shown in green) resulted into two main segments (Figure 6b). 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 into the boundaries shown in red. Among the eight competing priors from the database (Figure 1), (h) was chosen in order to recover the smaller segment in the bottom right. The competing procedure converged, also, to the prior (d) in order to recover the bigger segment in the middle. The later does not corresponds to a semantic image object. Obviously, this state (Figure 6b) was not the global optimum and the algorithm continued until convergence (Figure 6e). All four building were extracted and their detected boundaries are shown in red. Three shape priors from the database (Figure 1: d, f, and h) were finally chosen for the recovery of the four detected buildings.

In addition, in the top row of Figure 7, the result of the same prior-based contour evolution (in red) is shown superimposed on the original satellite image. The recognition-driven labelling process detected, in an unsupervised manner, image building regions and simultaneously the selected priors managed efficiently to reconstruct 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), just one labeling function was capable to handle the two detected segments. In algorithms convergence the segmentation result obtained with $k = 2$ labeling functions.

TABLE I
QUANTITATIVE EVALUATION

Data set	Quantitative Measures		
	Completeness	Correctness	Quality
Fig.2	0.988	0.784	0.783
Fig.3	0.988	0.971	0.953
Fig.4	0.656	0.974	0.645
Fig.5	0.941	0.975	0.918
Fig.7a	0.868	0.790	0.705
Fig.7b	0.926	0.946	0.879
Fig.8a	0.813	0.918	0.758
Fig.8b	0.877	0.927	0.820
Fig.9a	0.825	0.952	0.797
Fig.9b	0.847	0.977	0.831

Each function controlled which image region was 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 (Figure 7b) from the purely intensity-based segmentation (Equation 3) and the one from the proposed, here, prior-based process (Figure 7c) demonstrates the superior results that were obtained. The resulting output from the developed algorithm highly matches the ground truth one (Figure 7e). Note, also, that the pure intensity-based segmentation (Equation 3) resulted into a different segmentation outcome (Figure 7b) when compared to the data-driven term of Equation 18, whose result is shown in Figure 6 (in green). The later, being influenced by the labeling function, was more robust and managed to surpass the irrelevant non-semantic segments. Thus, the second term of Equation 18 does ameliorate the segmentation process.

Above observations are supported by the quantitative evaluation, which indicated that: (i) the purely intensity-based segmentation scored really low with an overall detection quality at about 70% (Table 1: fig.7a) 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: fig.7b). These quantitative results can be compared with the lower rates reported by other automatic algorithms [29] but not directly since different data were used and apart from buildings the detection was focused on other man-made objects, as well. However, algorithm's efficiency should be emphasized. For the construction of an

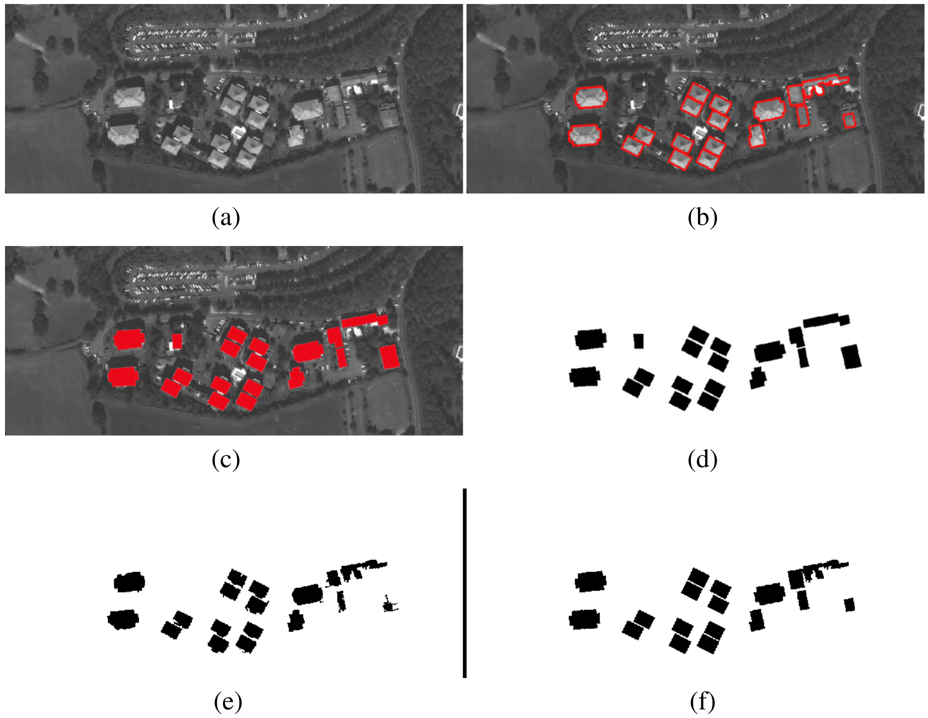


Fig. 8. Results from the application of the proposed recognition-driven segmentation to a high resolution aerial image. First row: Initial image (a) and the detected buildings in red color (b). Second row: the ground truth superimposed in red color (c) and the binary ground truth (d). Third row: the binary output of a pure data-driven segmentation (e) and the binary output after the application of the proposed algorithm (f).

operational system based on the proposed, here, framework a more extensive evaluation, that will be performed on several types of complexities and over larger areas (i.e. images of at least 4000x4000 pixels), should take place.

Furthermore, the developed algorithm was applied for the detection of buildings to another two data sets (aerial imagery with appx. 0.7m ground resolution), which both cover a wider area with a complex terrain, multiple objects of different classes, shadows, occlusions, different texture patterns and an important terrain height variability. In figure 8b, the final detected building boundaries are shown superimposed on the original image. 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, as well. The correctness of the detection was high at appx. 93% with a completeness at 88% (Table 1: fig.8b). The overall quality of algorithm's performance was at 82%, while the detection based, only, on the data-term of Equation 18 was lower than 76% (Table 1: fig.8a). The same quantitative and qualitative results have appx. obtained from the algorithm's evaluation on the second aerial date set. The detected building boundaries, shown in figure 9b superimposed on the original image, described sufficiently-enough scene's buildings. The algorithm managed to overcome the misleading low level information, caused by shadows, occlusions, intensity and texture variations and detect in most cases accurately building boundaries. However, two buildings were not at all detected and in a couple of cases resulting boundaries were not fully accurate. These qualitative results can be confirmed by the performed quantitative evaluation which indicates that the overall algorithms correctness was significantly high at 98% and its completeness due to the aforementioned failures at 85% (Table 1: fig.9b). Its overall performance had a quality of 83%, contrary to the data-driven term of Equation 18, which scored lower than 80% (Table 1: fig.9a).

B. Discussion

As it has been demonstrated in the experimental results, the selection of which image regions are associated with which appropriate priors is generated by the vector-valued dynamic labeling in a recognition-driven manner and thus, multiple buildings can be automatically extracted without any *a priori* information for their exact shape or number. Although, in this way, a recognition process has been elegantly integrated into a variational segmentation framework, in cases where the data term can not detect possible building regions the algorithm naturally fails, since all energy terms are associated with the ϕ function. The cost functional (Equation 15) 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. In all our experiments the parameters μ, λ and ρ were left constant.

Moreover, regarding the recovery of the projective transformation parameters, the multiscale optimization approach can guaranty a better approximation of energy's global optimum. In addition, since images are, often, acquired with controlled conditions, in cases where image plane is perpendicular to the

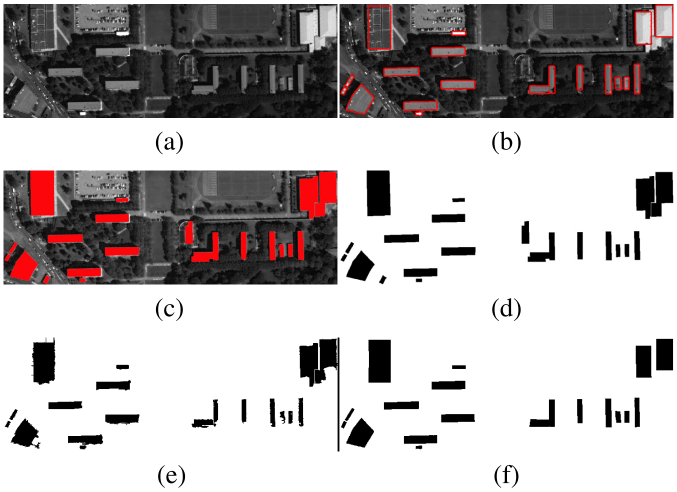


Fig. 9. Results from the application of the proposed recognition-driven segmentation to a high resolution aerial image. First row: Initial image (a) and the detected buildings in red color (b). Second row: the ground truth superimposed in red color (c) and the binary ground truth (d). Third row: the binary output of a pure data-driven segmentation (e) and the binary output after the application of the proposed algorithm (f).

optical axis, the recovery of eight unknowns can be reduced to six by setting the structure parameters ξ and ψ to zero and thus simplify the registration procedure. For non-calibrated cameras the homography should be fully recovered in its implicit form.

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. In all our experiments the eight building templates shown in figure 1 were used, but this database can be updated with other more complex shapes. In cases where the detected building can not be sufficiently described from any shape from the database -under any possible planar projectivity- then the algorithm fails to accurately detect its boundaries. A certain solution is to construct a large database with all the representative shape samples (derived e.g. from cadastral maps) but then the computation time will increase a lot. Searching, in our experiments, in a space of eight possible solutions for every detected segment, the developed algorithm in MATLAB, without an optimized coding, managed to converged approximately after two hours in an ordinary iPentiumM 2GHz,1GB RAM and for an image of approximately half a million pixels. One can imagine that with an efficient C++ implementation the processing time will be decreased by a factor of 500-1000 given prior experience in similar problems. For real-time applications, apart from optimizing the code, its implementation on a parallel system is straightforward by searching in parallel both for all segments and for the best fitted prior shape.

VI. Conclusion

We have introduced a novel recognition-driven variational framework which accounts for automatic and accurate multiple building extraction from aerial and satellite images. We demonstrated how one can integrate prior knowledge on multiple building shapes (like those given from a database) into the segmentation process and introduced the appropriate variational formulations to address multi-object segmentation with multiple competing shape priors. 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 [24] and (ii) without the need of having a priori knowledge for the pose of objects in image's plane, contrary to [26]. In addition, the proposed approach can account for multiple building extraction from single panchromatic images a highly demanding task [2] of fundamental importance in various geoscience and remote sensing applications.

The successful segmentation results, the reliable estimation of the transformation parameters and the adequate performance of the dynamic labeling encourage future research. First, a C++ implementation of the code is in progress towards a significant acceleration of the processing times. Introducing more complex image models towards accounting for other type of satellite images where the Gaussian assumption is not satisfied is a natural extension of our method. A comprehensive solution for general 3D objects would require to extend both the 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 training sets.

Acknowledgment

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

[1] A. Gruen, O. Kuebler, and P. Agouris, *Automatic Extraction of Man-made Objects from Aerial and Space Images I*. Birkhaeuser, Basel, 1995.

[2] E. Baltsavias, "Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, pp. 129–151, 2004.

[3] H. Kaartinen, J. Hyyppa, E. Gulch, H. Hyyppa, L. Matikainen, G. Vosselman, A. Hofmann, U. Mader, A. Persson, U. Soderman, A. Elmqwist, M. Ruiz, M. Dragoja, D. Flamanc, G. Maillet, C. J. Kersten, T., W. E. Hau, R., L. Frederiksen, J. Holmgaard, and K. Vester, "EuroSDR building extraction comparison," in *ISPRS Workshop on High-Resolution Earth Imaging for Geospatial Information*, Hannover, 2005, pp. 17–20.

[4] N. Champion, "Detection of unregistered buildings for updating cadastral databases," ISPRS - EuroSDR Commission 2, Tech. Rep. EuroSDR, http://buildingsdetection.free.fr/ (accessed in 07/07), 2007.

[5] K. Zhang, J. Yan, and S. Chen, "Automatic construction of building footprints from airborne lidar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2523–2533, 2006.

[6] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multi-aspect high-resolution insar data in urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 9, pp. 3583–3593, 2007.

[7] H. Mayer, "Automatic object extraction from aerial imagery-a survey focusing on buildings," *Computer Vision and Image Understanding*, vol. 74, no. 2, pp. 138–149, 1999.

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

[9] G. Cao, X. Yang, and Z. Mao, “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, 2005, pp. 474–479.

[10] K. Karantzalos and D. Argialas, “An image segmentation level set method for building detection,” in *International Symposium of Remote Sensing*, Busan, Korea, 2006, pp. 610–614.

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

[12] J. E. Ball and L. M. Bruce, “Level set segmentation of remotely sensed hyperspectral images,” in *IEEE International Geoscience and Remote Sensing Symposium*, Seoul, Korea, 2005, pp. 5638–5642.

[13] O. Besbes, Z. Belhadj, and N. Boujemaa, “Adaptive satellite images segmentation by level set multiregion competition,” INRIA, Tech. Rep. 5855, 2006.

[14] T. Keaton and J. Brokish, “A level set method for the extraction of roads from multispectral imagery,” in *Workshop on Applied Imagery Pattern Recognition*, 2002, pp. 141–150.

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

[16] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews, Neuroscience*, vol. 2, pp. 194–203, 2001.

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

[18] M. Leventon, E. Grimson, and O. Faugeras, “Statistical Shape Influence in Geodesic Active Controus,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. I:316–322.

[19] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky, “Modelbased curve evolution technique for image segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 463–468.

[20] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. Geiser, “Using shape priors in geometric active contours in a variational framework,” *International Journal of Computer Vision*, vol. 50, pp. 315–328, 2002.

[21] M. Rousson and N. Paragios, “Prior knowledge, level set representations & visual grouping,” *International Journal of Computer Vision*, vol. 76, no. 3, pp. 231–243, 2008.

[22] M. Taron, N. Paragios, and M.-P. Jolly, “Registration with uncertainties and statistical modeling of shapes with variable metric kernels,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, in press.

[23] A. Foulonneau, P. Charbonnier, and F. Heitz, “Affine-invariant geometric shape priors for region-based active contours,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1352–1357, 2006.

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

[25] T. Chan and W. Zhu, “Level set based shape prior segmentation,” Computational Applied Mathematics, UCLA, Los Angeles, Tech. Rep. 03-66, 2003.

[26] D. Cremers, N. Sochen, and C. Schnörr, “A multiphase dynamic labeling model for variational recognition-driven image segmentation,” *International Journal of Computer Vision*, vol. 66, no. 1, pp. 67–81, 2006.

[27] E. Gulch, “Digital systems for automated cartographic feature extraction,” in *International Archives of Photogrammetry and Remote Sensing*, 2000, pp. (B2):241–256.

[28] R. Bonnefon, P. Dherete, and J. Desachy, “Geographic information system updating using remote sensing images,” *Pattern Recognition Letters*, vol. 23, pp. 1073–1083, 2002.

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

[30] M. Irani and P. Anandan, “All about direct methods,” ser. Vision Algorithms: Theory and Practice, Springer-Verlag, A. Z. W. Triggs and R. Szeliski, Eds., 1999.

[31] A. Dervieux and F. Thomasset, “Multifluid incompressible flows by a finite element method,” *Lecture Notes in Physics*, vol. 11, pp. 158–163, 1981.

[32] S. Osher and J. Sethian, “Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation,” *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[33] T. Chan and L. Vese, “Active Contours without Edges,” *IEEE Transactions on Image Processing*, vol. 10, pp. 266–277, 2001.

[34] N. Paragios and R. Deriche, “Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision,” *Journal of Visual Communication and Image Representation*, vol. 13, pp. 249–268, 2002.

[35] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky, “A nonparametric statistical method for image segmentation using information theory and curve evolution,” *IEEE Transactions on Image Processing*, vol. 14, pp. 1486–1502, 2005.

[36] C. Springer, *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

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

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

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

Konstantinos Karantzalos received his engineering diploma from the National Technical University of Athens (NTUA, Greece) in 2000 and his PhD in 2007 from the same school in collaboration with Ecole Nationale de Ponts et Chaussees (France). His thesis was entitled ‘Automatic feature extraction from aerial and satellite imagery with computer vision techniques’. In 2007 he joined the Medical Imaging and Computer Vision Group of the Department of Applied Mathematics at Ecole Centrale de Paris (France) as a postdoctoral researcher.

Dr. Karantzalos has numerous publications in international journals and conferences and in 2006 he received the ‘Best Paper Award’ in the International Symposium of Remote Sensing. His research interests include geoscience and remote sensing, computer vision, pattern recognition, artificial intelligence and underwater photogrammetry. *http://users.ntua.gr/karank/*

Nikos Paragios (http://vision.mas.ecp.fr) received his B.Sc. and M.Sc. from the University of Crete (Greece) in 1994 & 1996 respectively, his Ph.D. from I.N.R.I.A./University of Nice/Sophia Antipolis (France) in 2000 and his HDR (Habilitation a Diriger de Recherches) from the University of Nice/Sophia Antipolis in 2005. He is currently professor at the Ecole Centrale de Paris - one of most exclusive engineering schools "Grange Ecoles" - leading the Medical Imaging and Computer Vision Group at the Applied Mathematics Department.

He is also affiliated with INRIA Saclay Ile-de-France, the French Research Institute in Informatics and Control heading the GALEN group. Prior to that he was professor/research scientist (2004-2005) at the Ecole Nationale de Ponts et Chaussees-France affiliated with Siemens Corporate Research (Princeton, NJ) as a project manager (smart cameras and machine vision group) (2002-2004), senior research scientist (2004) and research scientist (1999-2003). In 2002 he was an adjunct professor at Rutgers University and in 2004 at New York University while N. Paragios was a visiting professor at Yale University in 2007. Professor Paragios has co-edited four books, published more than 100 papers in the most prestigious journals and peer-reviewed conferences of computer vision and medical imaging, has nine US issued patents and more than twenty pending. His work has approx 3,000 citations in googlescholar and his H-number according to the scholar is 25. Professor Paragios is a Senior member of IEEE, and serves as member of the Editorial Board of (i) the International Journal of Computer Vision (IJCV), (ii) the Computer Vision and Image Understanding Journal (CVIU) and (iii) the Medical Image Analysis (MedIA). In 2008 he was the laureate of one of Greece’s highest honor for young academics and scientists, the Bodossaki Foundation Prize in the field of applied sciences while in 2006 he has received the TR35 prize from the MIT technology review journal. Professor Paragios is one of the program chairs of the 11th European Conference in Computer Vision (ECCV’10, Heraklion, Crete).

His research interests include computer vision, medical image analysis, remote sensing and human computer interaction.