A New Region based Active Contour Model for Object Segmentation

**Michela Lecca · Stefano Messelodi ·
Raul Paolo Serapioni**

Abstract We present a novel region based active contour model that segments one or more image regions that are visually similar to an object of interest, said prior. The region evolution equation of our model is defined by a simple heuristic rule and it is not derived by minimizing an energy functional, as in the classic variational approaches. The prior and the evolving region are described by the probability density function (pdf) of a photometric feature, as color or intensity. The heuristic rule enlarges or contracts an initial region of the image in order to equalize pointwise the pdf's of the prior and of the region. Such heuristic rule can be modeled by many mathematical monotonic decreasing functions, each defining an evolution equation for the initial image region. The choice of a particular function is remitted to the user, that in this way can even integrate a priori knowledge possibly useful to break down the computational charge of the method and to increase the detection accuracy. Here we propose two different evolution equations for the general purpose of prior detection without a priori information and we discuss empirically the performances of our model on real-world and synthetic datasets. These experiments show that our model is a valid alternative to the classic models.

Keywords Region-based Active Contour Models · Heuristic Model based on Probability Density Functions · Object Segmentation

M. Lecca
Fondazione Bruno Kessler, Center for Communication and Information Technology, via Sommarive 18, 38123 Trento, Italy
E-mail: lecca@fbk.eu

S. Messelodi
Fondazione Bruno Kessler, Center for Communication and Information Technology, via Sommarive 18, 38123 Trento, Italy
E-mail: messelod@fbk.eu

R. P. Serapioni
University of Science, Department of Mathematics, via Sommarive 14, 38123 Trento, Italy
E-mail: serapion@science.unitn.it

1 Introduction

Active contour models are widely used to segment an object of interest in an image [3], [6], [22], [43], [21], [12], [34], [9], [19], [47], [30], [18].

An active contour is a curve that dynamically evolves in the image domain towards the boundary of the object of interest under the influence of internal and external forces, until the energy associated with these forces is minimized. Internal forces model geometric properties of the curve itself, for instance its smoothness or stiffness. External forces depend on the image features used for the segmentation, and they cooperate with the internal ones to shape the active contour and to make it to adhere to the object boundary. The energy associated with the forces thus depends on the geometric properties of the curve and on the visual features of the image and of the object of interest. It is designed so that the contour with minimum energy lies over the boundary of the object of interest.

The minimization of the energy functional is computationally hard. In general, it ends up in a local minimum, that *hopefully* corresponds to the desired object. Usually, the minimization is implemented by a gradient descent method [45], that iteratively deforms the active contour from an initial curve to different subsequent states, until a minimum is found.
The mathematical expression of the energy functional depends on the visual features used for segmenting the object of interest. When these features describe only geometric cues of the active contour, e.g. its smoothness or stiffness, or when they regard pixelwise properties as edge magnitude or orientation [29], [31],[24], [39], [38], [7], the energy is expressed by a line integral around the active contour. Its minimization is performed by solving the Euler-Lagrange partial differential equations (PDE) of the energy of the initial contour. When the segmentation features describe visual properties of a region, e.g. the mean value or the distribution of the color inside the internal or external part of the active contour, the energy is generally expressed as a double integral over the region [48], [41], [37], [6], [2], [11], [23], [26], [18]. In this case the minimization becomes more involved, because the directional derivative is more naturally defined for curves than for regions [2]. One minimization approach re-writes the energy functional as a boundary integral according to the Green-Riemann Theorem, and then derives the Euler-Lagrange PDE's. Other methods use shape derivation tools [1], [2], [13], that minimize the energy functional with respect to the whole region and not only with respect to its boundary. Historically, region-based active contours have been introduced later than the edge-based active contours, in order to make the image segmentation more robust with respect to weak edges and noise: edge-based active contour models have been developed originally in [25], while the first works on region-based models have been proposed in [10] and [40].

The difficulties in minimizing the energy functional stimulated the pursuit of alternative active contour models. For example, [5], [8] adopt a velocity vector field that deforms an initial curve towards the object of interest without defining any energy functional.
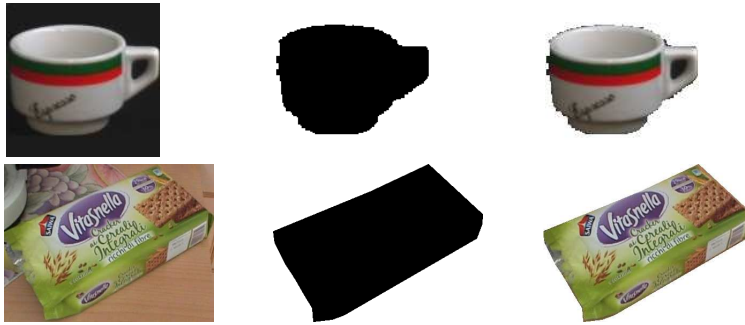
Fig. 1 Some examples of objects of interest (priors) employed in our experiments: the prior is stored as a picture (on left) along with a binary mask (in the middle), whose black pixels specify the part of the picture actually occupied by the object. In this work the mask is used only to compute the pdf of a photometric feature of the interest object depicted in the image (on right). Shape features, that could be extracted from the mask, are not considered here.

In this work we propose a novel region based active contour model that detects in an input image a view of a known object, said *prior*. Examples of prior are a logo, a traffic sign, a face, or a home object as shown in Figure 1. In our procedure we deform iteratively an initial region of the input image until it becomes visually similar to the prior. As proposed by previous works, e.g. [36], [46], [48], [2], [37], [41], [11], [28], we describe the visual appearance of the prior, as well as that of the evolving region, by statistical features. Precisely, we use the probability density function (pdf) of a global photometric feature, as color or intensity. We measure the fidelity of the region to the prior by the extent of the overlapping between their pdf's. We define the region evolution law by a simple heuristic rule, that enlarges or contracts the evolving region by including or excluding image pixels in order to equalize the pdf's of the prior and of the region. Analogously to the approaches in [5], [8], the evolution equation is not derived as the gradient of an energy functional. The main novelty of this work is the heuristic law we introduce. It can be modeled by several different monotonic decreasing functions satisfying three constraints driving the cirve evolution (see Section 4.1). As a consequence, there is not a unique evolution equation of the input region. This is a strength point of the model since it allows the user to design a function tailored for his/her applications. For instance, a priori knowledge about the image and/or the prior can be used to tune the parameters of the evolution equation in order to speed up the convergence of the algorithm or to increase its performances (see Section 4.1). In this work we do not concern with the criteria to appropriately select the function implementing the heuristic upon the application. We rather investigate two different region evolution equations for the detection of a prior in real-world and synthetic images of public datasets without a priori information. For each evolution equation we measured the performances of our active contour model by using the color as visual descriptor. We have choosen color among the other descriptors because of its invariance against many geometric transformations,

like image rescaling or rotation. Moreover, color is widely used in object detection and image retrieval for its high discriminative power [20]. We computed the accuracy on the prior detection by considering two different color representations and by varying the parameters of the model, as for instance the initial region and the metric employed to measure the overlap between the pdf's. Moreover, we also tested the robustness of our algorithm with respect to geometric distortions and Gaussian noise addition. The experiments, that we carried out on synthetic and real-world databases, show very good performances, also in comparison with the classic energy based approaches. This makes our model a valid alternative to the variational methods.

2 Paper Outline

The rest of the paper outlines as follows. In Section 3 we introduce the formalization of the classic active contour model as a reference needed to understand similarities and differences with the non standard one. Section 4 describes our model and provides some implementation details. Section 5 presents our experimental results, and finally Section 6 reports our conclusions and future work.

3 The Classic Approach

In this Section we briefly describe the classic approach in order to introduce some notation and concepts used also in our model.

Let S be the set of the spatial coordinates of the pixels composing an image I, and let P be an object of interest to be segmented in I. An active contour on I is a differentiable parametrized curve

$$\omega : [0, 1] \rightarrow S$$

that dynamically evolves by minimizing an energy functional. The energy is expressed as the sum of the potentials of internal and external forces whose action determines the curve deformation. The forces are chosen so that a (local) minimum of the energy occurs when the active contour lies over the boundary of the object of interest.

Usually, the minimization of the energy functional is performed by a gradient descent method [45], that iteratively deforms the curve ω from an initial contour ω_0 to different subsequent states ω_t, until the equilibrium is reached. Therefore, a real positive parameter t is introduced to describe the family $w : [0, +\infty) \times [0, 1] \rightarrow S$ where $w(t, s) = \omega_t(s)$ ($t \geq 0$, $s \in [0, 1]$) is the evolution of the initial curve ω_0 at the *time t*. The analytic expression of the energy leads to an evolution equation of the general form

$$\frac{\partial w}{\partial t} = F\mathbf{N}, \tag{1}$$

where F is the magnitude of the velocity vector field deforming the curve, and $\mathbf{N}$ is the inward unit normal of ω.

Often, the energy minimization is performed by using the *level-set representation* of the curve, that allows to handle topological changes [32], [44], [4], [35]. For instance, topological changes are necessary when a prior is displayed many times in an image. In this case, an initial contour should split into parts. Describing this splitting and solving the corresponding evolution equation for a curve may be quite hard. To overcome this problem, the active contour $\Gamma_t := \omega_t([0,1])$ is represented by the zero-level set of a smooth function $u : [0, +\infty) \times S \to \mathbf{R}$ depending on the time t. More precisely, the sets $\Omega_t = \{x \in S : u(t,x) > 0\}$ and $S - (\Gamma_t \cup \Omega_t)$ are respectively the regions inside and outside the active contour. If $\omega(t, \cdot)$ is a local representation of Γ_t, then

$$u(t, \omega(t,s)) = 0 \qquad \forall\, s \in [0,1] \tag{2}$$

and differentiating u with respect to t yields

$$0 = \frac{d}{dt} u(t, \omega(t,s)) = \nabla_x u(t, \omega(t,s)) \cdot \frac{\partial \omega}{\partial t}(t,s) + \frac{\partial u}{\partial t}(t, \omega(t,s)).$$

Therefore we have

$$\begin{aligned}
\frac{\partial u}{\partial t}(t, \omega(t,s)) &= -\mathbf{N} \cdot \frac{\partial w}{\partial t}(t,s) \parallel \nabla_x u(t, \omega(t,s)) \parallel \\
&= -F \parallel \nabla_x u(t, \omega(t,s)) \parallel .
\end{aligned} \tag{3}$$

4 The New Region-based Active Contour Model

As mentioned in Section 1, our active contour model detects in an input image I one or more regions which are visually similar to the view of an object of interest P said prior. The prior P is represented by: (i) a picture I_P displaying P and by (ii) a binary mask, whose black pixels specify the part of I_P actually occupied by the prior (see Figure 1). In our framework, this mask is used only to circumscribe the pdf computation to the set of image pixels that actually belong to the object of interest.

The visual appearance of P is described by the probability density function (pdf) q_p of a photometric feature, as color or intensity. The feature over a subset of image pixels $\mathcal{S}$ is usually modeled by a n-dimensional real bounded function h from $\mathcal{S}$ to $\mathbf{R^n}$. In general, a same visual cue (e.g. the color) can be modeled by different functions (see Subsection 5.3, where we consider two different representation of the color). Once the function h representing the feature has been defined, its pdf is computed as

$$pdf(\alpha) = \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \delta(h(x) - \alpha) dx \tag{4}$$

(1) Prior

(2) Input Image

(3) Initial Region

(4) Segmented Object

(5) PDF's of initial region
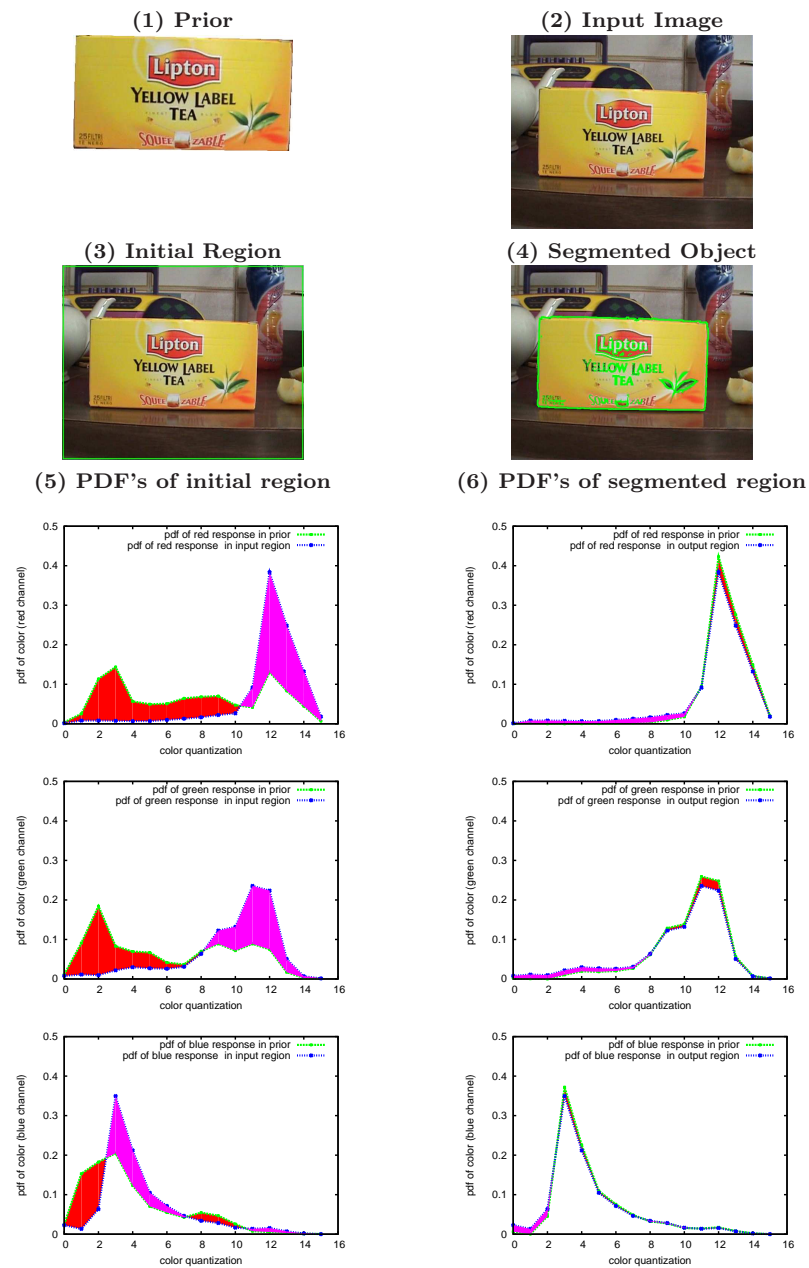
(6) PDF's of segmented region

Fig. 2 (1) A prior; (2) an input image; (3, 4) in green, the boundary of the initial region and that of the region output by our algorithm; (5, 6) the pdf's of the red, green and blue color responses of the prior, of the initial region, and of the output in the RGB-S representation (see text for more details). Purple and red areas in (5) and (6) highlight the differences between the pdf's. The red (purple, resp.) set is upper bounded by the pdf of the prior (of the region. resp.) and lower bounded by the pdf the region (of the prior, resp.).

where $|\cdot|$ indicates the Lebesgue measure of the set between the pipes , δ is the Dirac delta and α ranges over the co-domain of h. In the discrete framework, i.e. when $h(\mathcal{S})$ is discretized, the pdf is approximated by the histogram q given by

$$q(\alpha) = \frac{1}{|\mathcal{S}|}|\{x \in \mathcal{S} : h(x) = \alpha\}| \tag{5}$$

where $|\mathcal{S}|$ is computed as the cardinality of $\mathcal{S}$ and α is defined as above. In the next, with abuse of notation, we will use h to denote the function modeling the visual feature or the visual feature itself.

Previous works, e.g. [48], [42], propose a region-based active contour model for foreground/background separation that use mean and variance as statistical descriptors. However, as pointed out in [41], these features provide often a poor description of the visual appearance of the interest object, hence the pdf is preferred. Notice that differently from other methods ([36], [28]), the pdf we consider in our approach is neither approximated by exponential functions like Gaussian, Poisson or Rayleigh distributions, nor smoothed by density kernel as the Parzen window used in [2] and [41].

Our model deforms iteratively an initial region Ω_0 (which is a subset of S) so that it evolves towards a region Ω whose pdf q_Ω *coincides as much as possible* with the pdf q_p of the prior.
Figure 2 shows an example. The yellow box in (1) is the prior to be detected in the picture in (2). The boundary of the initial region Ω_0 is highlighted in green in (3). The prior and the evolving region are described by the pdf's of their color responses red, green and blue, as shown in (5) and (6) respectively. The algorithm works out the pdf's of the evolving region according to a *heuristic rule*, that excludes (includes, resp.) a pixel from (to, resp.) the evolving region if the region pdf at that color is greater (smaller, resp.) than the prior pdf at the same color. In particular, this heuristic law minimizes the difference with the pdf of the prior, i.e. it minimizes the areas of the red and purple regions highlighted in the graphs of Figure 2 (5) and (6).
Notice that a perfect overlap between the pdf's of the prior and of the evolving region is usually not even possible, because in most real-world scenes, the object of interest imaged in I is similar to the prior P but it is not an exact copy of it. In most instances the pdf's of the prior and of the object of interest images in the input picture differ because of noise due to image compression or blurring, small changes of the camera view point, slight differences of the illuminants under which the prior and the input image have been captured often occur.

This Section continues as follows: Subsection 4.1 describes our region based evolution model, while Subsection 4.2 provides some implementation details.

4.1 The Region Evolution Equation

Let S be the set of the spatial coordinates of the pixels composing an image I. We describe the evolving region as the positivity set of a 2D smooth function

$u : [0, +\infty) \times S \to \mathbf{R}$ that, for each time t, partitions S into the three non-empty sets $\Omega_t := \{x \in S : u(t, x) > 0\}$, $\Gamma_t := \{x \in S : u(t, x) = 0\}$ and $S - (\Omega_t \cup \Gamma_t)$. For $t = 0$, Ω_t coincides with an image region input by the user.

The evolution law of our model is described by

$$\frac{\partial u}{\partial t}(x) = V(x, \Omega_t)\delta_\varepsilon(u(t, x)) \tag{6}$$

where, for $\varepsilon > 0$,

$$\delta_\varepsilon(u(t, x)) = \frac{1}{\pi} \frac{\varepsilon}{u(t, x)^2 + \varepsilon} \qquad \forall\, x \in S \tag{7}$$

and

$$V(x, \Omega_t) = \phi(x, \Omega_t) + \lambda K(t, x) \qquad \forall\, x \in S. \tag{8}$$

The function δ_ε of Equation (7) is a smooth approximation of the Dirac delta function, primed by differentiating the smoothed Heaviside function $H_\varepsilon : \mathbf{R} \to [0, 1]$

$$H_\varepsilon(z) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\frac{z}{\varepsilon}\right).$$

η highly scores each pixel x for which $u(t, x)$ is close to zero, i.e. each pixel near to the border Γ_t of Ω_t. This makes the deformation of Ω_t more remarkable in a neighborhood of the border of Ω_t than in its interior and exterior parts. The parameter ε is a real number strictly positive and it controls the width of the neighborhood around Γ_t.

The function $V(x, \Omega_t)$ of Equation (8) describes mathematically the heuristic rule of our model. In particular, $V(x, \Omega_t)$ is the sum of two terms. The first one is a real bounded function ϕ defined on S. It depends on Ω_t through the ratio

$$Q(x, \Omega_t) = \frac{q_{\Omega_t}(h(x))}{q_p(h(x))} \tag{9}$$

where it is defined, and $Q(x, \Omega_t) := 1$ if $q_{\Omega_t}(h(x)) = q_p(h(x)) = 0$.

The heuristic rule implemented by ϕ attempts to find a region Ω_t such that the ratio in Equation (9) is 1 for any x in Ω_t. We require that the function ϕ is such that:

1. if $Q(x, \Omega_t) > 1$, then $\phi(x, \Omega_t) < 0$: when the percentage $q_{\Omega_t}(h(x))$ of pixels in Ω_t with feature $h(x)$ is greater than the percentage $q_p(h(x))$ in the model, the pixels of Ω_t with value $h(x)$ are negatively scored, and the value of u in these pixels decreases. Eventually, when u becomes strictly negative, the region Ω_t contracts to exclude these points;
2. if $Q(x, \Omega_t) < 1$, then $\phi(x, \Omega_t) > 0$: when the percentage $q_{\Omega_t}(h(x))$ of pixels in Ω_t with feature $h(x)$ is smaller than $q_p(h(x))$, the pixels with value $h(x)$ are positively scored, and the value of u in these points increases. When u becomes strictly positive, Ω_t enlarges to include these points;
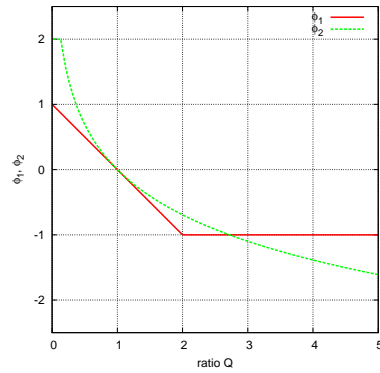
Fig. 3 The functions ϕ_1 and ϕ_2 (see Equations (10) and (11)).

3. if $Q(x, \Omega_t) = 1$, then $\phi(x, \Omega_t) = 0$: the percentages $q_{\Omega_t}(h(x))$ and $q_p(h(x))$ are equal.

Infinite functions satisfy the properties listed above. Any monotonic decreasing function ϕ with $\phi(x, \Omega_t) = 0$ when $Q(x, \Omega_t) = 1$ can be used to define the region evolution equation (6). The function ϕ defines *how much* to score the pixel x with $Q(x, \Omega_t) > 1$ and those with $Q(x, \Omega_t) < 1$. Thus the shape of ϕ determines the speed $\partial u/\partial t$ at any pixel x. The mathematical expression of ϕ is a model parameter to be set up by the user. In this way, the user can adapt the model to his/her specific applications, in order to speed up the convergence of the method and/or to improve the results. Let us give an example. Let us suppose to know that a given prior is effectively present in the input image and that the image region occupied by prior is much smaller than the entire image. Let us use the border of the whole image as initial contour. In order to speed up the convergence of the active contour, the velocity of the region evolution (i.e. $\partial u/\partial t$) should be greater when a pixel has to be removed from the current region than it has to be added to the region. Therefore, the slope of the function ϕ when $Q(x, \Omega_t) > 1$ must be greater than the slope when $Q(x, \Omega_t) < 1$. In this way, background pixels are removed faster from the evolving region than if the same slope is used and this makes the computational time shorter.

As already mentioned in Section 1, here we do not focus on the problem to choose an expression of ϕ upon the application, but rather we propose two different expressions of ϕ, denoted by ϕ_1 and ϕ_2 (see Figure 3), and we use them for the general purpose of prior detection when no a priori information about the image are given. In Section 5 we compare these functions in terms of accuracy on the detection of the prior.

Precisely, $\phi_1 : S \to [-1, 1]$ and $\phi_2 : S \to [-2, 2]$ are defined as follows:

$$\phi_1(x, \Omega_t) = \begin{cases} -1 & \text{if } q_p(h(x)) = 0 \wedge q_{\Omega_t}(h(x)) \neq 0 \\ 1 - \min\left(Q(x, \Omega_t), 2\right) & \text{otherwise} \end{cases} \tag{10}$$

and

$$\phi_2(x, \Omega_t) = \begin{cases} 2 & \text{if } \{(q_p(h(x)) = 0 \wedge q_{\Omega_t}(h(x)) \neq 0) \vee \\ & \log Q(x, \Omega_t) < -2\} \\ \max(-2, -\log Q(x, \Omega_t)) & \text{otherwise.} \end{cases}$$
(11)

The second term in Equation (8) has been introduced to ensure the smoothness of the surface described by u and the stability of the solution. In particular, K is the curvature of the implicit curve $u(t, x) = 0$ and λ is a strictly positive real parameter weighting the contribution of K to the evolution. K can be computed by

$$K(t, x) = \text{div}\left(\frac{\nabla_x u(t, x)}{\| \nabla_x u(t, x) \|}\right).$$
(12)

The term $K(t, x)$ is not defined when the gradient of u is null, i.e. it is not defined in any stationary point. At these points, we set $K(t, x) = 0$. Notice that the function u must be a $\mathcal{C}^2$ function with respect to the spatial variable x in order to ensure the existence of $K(t, x)$.

Finally we observe that the differential equation (6) makes sense only if the function u is (i) of class $\mathcal{C}^2$ with respect to the space variable x, in order to ensure the existence of $K(t, x)$ and (ii) of class $\mathcal{C}^1$ with respect to the time variable t.

4.2 Implementation Details

The algorithm implementing our active contour model requires as input an image I, a prior P, a surface u as specified in Section 4.1, and a set of parameters, listed in Table 1.

In the discrete space, the time derivative of u in Equation (6) is approximated by

$$u^{(n)}(x) = u^{(n-1)}(x) + (\phi^{(n-1)}(x, \Omega^{(n-1)}) + \lambda K^{(n-1)}(x))\eta^{(n-1)}(x)\Delta t, \quad (13)$$

where Δt is a time step, and n is an integer ranging over $\{1, \ldots, N_{max}\}$. When $n = 0$, $u^{(0)}$ is the input surface, while $u^{(n)}$ is the evolution of $u^{(0)}$ after the time $n\Delta t$. $\Omega^{(n-1)}$ indicates the evolving region at time $n - 1$. $K^{(n-1)}$ is defined in Eq. (12) for the function $u^{(n-1)}$, $\phi^{(n-1)}$ is the function ϕ (either ϕ_1 or ϕ_2) evaluated on $\Omega^{(n-1)}$, and $\eta^{(n-1)}$ is the function η evaluated on $u^{(n-1)}$. In this work, the surface u at time $t = 0$ is defined as

$$u^{(0)}(x) = \begin{cases} \mathcal{D}(x, \partial D)^{0.25} & \text{if } x \in D \\ -\mathcal{D}(x, \partial D)^{0.25} & \text{if } x \in S - D \end{cases}$$
(14)

User Parameters	Description
λ	Weight of curvature term
Δt	Time Step
f_{init}	Step for re-initialization of u
f_s	Step for smoothing u
d	Metric to stop the algorithm (L^1, L^2 or EMD)
ϕ	(1) ϕ_1; (2) ϕ_2
alg_type	(1) no smoothing; (2) smoothing with median filter
D	$D = \{x \in S : u^{(0)}(x) > 0\}$ (see Eq. (14))

Table 1 Parameters to be set up by the user in the current implementation of our active contour algorithm. See the text for more details.

where D is a non-empty closed subset of S, ∂D is its boundary, and $\mathcal{D}$ is the Euclidean distance between the point x and the set ∂D. The initial region Ω_0 is thus the positivity set of $u^{(0)}$. The function $u^{(0)}$ is clearly continuous and smooth. During the surface evolution, the function u is re-initialized according to Equation (14) (with frequency f_{init}) in order to preserve its smoothness and thus the stability of the solution. In our algorithm, u can be further smoothed with a frequency f_s: this additional smoothing helps to include (remove resp.) isolated pixels close to (far from resp.) the boundary of the evolving region. In the current implementation, smoothing is performed by using the median filter, that replaces each value $u(x)$ respectively with its median value over a square of 9×9 pixels centered at x. The activation of the smoothing routine is controlled by the parameter alg_type, that is 1 when no smoothing is applied, 2 otherwise.

In principle, the algorithm stops when the region Ω does not evolve any more. However, in order to avoid slight changes of the solution, the algorithm ends when a maximum number N_{max} of iterations is reached or when a distance d measuring the dissimilarity (i.e. overlap) between the pdf's of P and of any region $\Omega^{(m)}$, does not improve for M times consecutively. From some qualitative experiments, we observed that $N_{max} = 1000$ and $M = 20$ provide good thresholds for the termination criterion. Distance d can be chosen by the user. In this work three metrics are considered: the L^1, the L^2 and the Earth-Mover-Distance (EMD).

Finally, in the current implementation the parameter ε in Equation (7) has been fixed to be 1.0.

5 Experiments

In this Section we present the experiments we carried out on synthetic and real-world image databases to measure the accuracy of our approach, also in comparison with the classic variational method. In our experiments, the feature used for the segmentation is the color. The color pdf is often employed in detection and recognition algorithms [20], because color is invariant with respect to translations and changes of size and/or in-plane rotations over a wide range of scale factors and rotation angles, and it also quite robust to

Gaussian noise addition. However, other features can be used, as for instance texture [41], [37].

This Section is organized as follows: in Subsection 5.1 we explain how we measure the accuracy of our model; in Subsection 5.2 we describe the databases used in our experiments; then we study how the accuracy changes by varying the model parameters (Subsection 5.3) and in presence of geometric affine distortions (Subsection 5.4) and noise (Subsection and 5.5). In Subsection 5.6 we analyze the performances of our model on images with complex background. In Subsection 5.7 we give some suggestions to extend our model in order to include edges and/or shape descriptors and to make it robust to object ccclusions. Finally, in Subsection 5.8 we compare our performances with those output by the classic approach presented in [2].

5.1 Performances Evaluation

Let O be an object view to be detected in an input image I, and let G the part of I (if any) actually depicting O.

We measure the accuracy on the detection of O in I by the *overlap index*

$$\mathcal{X}(G, \Omega^\star) = \frac{\mathrm{Area}(G \cap \Omega^\star)}{\mathrm{Area}(G \cup \Omega^\star)} \tag{15}$$

where $\Omega^\star$ is the region of S output by our algorithm.
The overlap index $\mathcal{X}$ ranges over $[0, 1]$, and the closer $\mathcal{X}$ is to 1, the better the detection is.
In addition we also compute the percentage $\mathcal{I}$ of pixels of G belonging to $\Omega^\star$:

$$\mathcal{I}(G, \Omega^\star) = \frac{\mathrm{Area}(G \cap \Omega^\star)}{\mathrm{Area}(G)}. \tag{16}$$

We say that an object view O is *missed* if it is depicted in the input image I but it has not been detected, i.e. $\mathcal{X}(G, \Omega^\star) = 0$. In the next, we denote by $\mathcal{M}$ the percentage of missed object views of any test set.

5.2 Databases

Here we describe the three public databases of color images we used in our experiments: COIL-100 [33], GroundTruth-for-COIL100 [15] and MLDB2013 [17].

5.2.1 COIL-100

The database COIL-100 [33] consists of real-world color pictures of 100 objects with size 128×128 pixels. This database has been built up as follows: each

object of COIL-100 has been placed on a motorized turntable, that was counterclockwise rotated through 360 degrees to vary the object pose. The images of the 100 object of COIL-100 have been taken at pose intervals of 5 degrees, so that the database contains 7200 images.

For each picture, the part actually belonging to the object view portrayed is specified by the black pixels of a binary mask. The masks of the object views are downloadable from [16].

Hereafter we denote the 72 views of any object of COIL-100 by v_0, v_5, ..., v_{355}, where the subscript indicates the angle at which the object has been imaged. Moreover, we denote by I_0, I_5, ..., I_{355} the images portraying the views v_0, v_5, ..., v_{355} respectively.

Figure 4 shows some objects from COIL-100 (frontal views), while Figure 5 displays the first five views of an object (from degree 0 to 20).

Fig. 4 COIL-100: some objects from COIL-100 (frontal views).

Fig. 5 COIL-100: first five views (from 0 to 20 degrees) of an object.

5.2.2 GroundTruth100-for-COIL

The database GroundTruth100-for-COIL [15] contains 100 images depicting some objects of COIL-100 displayed against a non-uniform background. These images have different size (430x400, 375x400, 484x484 pixels). Each image has been generated synthetically: some object views have been randomly chosen from COIL-100, then they have been rotated with a random angle in $[0, 2\pi]$ and re-scaled with a random factor in the range [0.7, 1.3]. Finally, they have been superimposed on a synthetic, non-uniform, colored background in a non-overlapping fashion. Some examples are shown in Figure 6.

Each image is tagged with some data about each object view of COIL-100 present in it: an integer number identifying each object (from 1 to 100), the

Fig. 6 GroundTruth-100-for-COIL: Some examples from GroundTruth-100-for-COIL.

degree corresponding to its view (ranging over $\{0, 5, 10, \ldots, 355\}$), the rotation angle and the scale factor at which this object view appears in the image. A segmentation map provides ground-truth information about the image parts occupied by the object views. In particular, each segment is labeled by an integer number i, that identifies an object view (when i is greater than zero) or the background (when i is zero).

5.2.3 MLDB2013

The database MLDB2013 [17] includes a set of 12 objects and a set of 12 real-world test images where the objects have to be searched for. Each object is modeled by a single view, that is represented as in COIL-100 by a color image and by a binary mask specifying the image part actually belonging to the object view. The size of the object images varies from 291x103 (minimum) to 857 x 1142 (maximum) pixels. The ith test image displays the ith object view of the database ($i = 1, \ldots, 12$). The position of a view in the corresponding test image is defined by a binary map, where the white pixels correspond to the background and the black ones to the object view. Some examples are reported in Figure 7. The size of the test images varies from 288 x 384 (minimum) to 857 x 1142 (maximum) pixels.

Differently from the images of GroundTruth-100-for-COIL, the objects imaged in the test pictures of MLDB2013 differ from the correspondent priors not only in scale and in orientation, but also in pose (see for instance Figure 8). These issues make the detection of the objects of MLDB2013 harder than that of the objects of GroundTruth-100-for-COIL.

Fig. 7 MLDB2013: (a, c) two objects in the database, and (b, d) two images where these objects have been searched for.

Fig. 8 MLDB2013: the prior (on left) appears in the test image (on right) in a different pose.

5.3 Sensitivity Analysis

First of all we analyzed the dependency of our model on the values of the parameters listed in Table 1 and on the color representation. We carried out this study on a subset of COIL-100. In particular, for each object of COIL-100 we considered the two pictures I_0 and I_5, which respectively depicts the views v_0 and v_5, where the subscript denotes at which the object has been imaged.. Then we took v_0 as a prior and we used our model to segment v_5 in the image I_5.

Since two consecutive shots of COIL-100 varies slightly, their pdf's are quite similar. In fact, the mean L^2 distance between the pdf's of v_0 and v_5 averaged over the number of objects of COIL-100 is about $8.23 \ 10^{-5}$. However, in general two different views of an object have different colors and thus different color pdf's. We will systematically analyze the robustness of our method against

changes of pose in Section 5.5. Here we pursue the sensitivity analysis with respect to small changes of views, as described before.

We varied the parameters in Table 1 as follows: $\lambda = 0.02$, 0.05, 0.10; $\Delta t = 10$, 50; $f_{init} = 5$, 10; $f_s = 5$, 10; $d = L^1$, L^2, EMD; $\phi = \phi_1$, ϕ_2, alg_type = 1, 2. We repeated the experiments with three different choices of D: (a) a rectangle D_0 with ∂D_0 coincident with ∂S; (b) a square D_1 positioned in the center of S and with size $s = \frac{1}{3}\min(R, C)$ where R and C are the rows and columns of S respectively; (c) D_2 is a disconnected set composed by 4 rectangles, equally spaced in S and with size $\frac{3R}{16} \times \frac{3C}{16}$.

Finally, we repeated the experiments by using two different color representations, that we name RGB-3D and RGB-S.

In RGB-3D, the color is represented by a 3D vector whose components are the red, green, and blue camera responses. The color pdf is approximated by a 3D histogram of size 8×8×8, that we arrange as a 1D vector with 512 entries, in order to compute the scalar ratio $Q(x, \Omega)$ of Equation (9).

In the RGB-S, the red, green and blue color responses are considered separately. The marginal pdf's q^0, q^1, and q^2 of the red, green and blue color responses are computed and each q^i ($i = 0,1,2$) is approximated with an histogram of $N = 16$ bins. The color pdf is thus represented by a 3×16 matrix, whose ith row is the pdf q^i ($i = 0$, 1,2). Therefore, in order to manage such a vector representation, we replaced the Equation (8) by Equation (17) when $\phi = \phi_1$, and by Equation (18) when $\phi = \phi_2$:

$$V(x) = \lambda K(x) + \phi_1'(x) \qquad \forall\, x \in S \tag{17}$$

$$V(x) = \lambda K(x) + \sum_{i=0}^{2} \phi_2^i(x) \qquad \forall\, x \in S \tag{18}$$

where ϕ_1' is obtained from ϕ_1 by replacing $Q(x, \Omega_t)$ with $Q'(x, \Omega_t) := \prod_{i=0}^{2} \frac{q_{\Omega_t}^i(h(x))}{q_p^i(h(x))}$ and ϕ_2^i is obtained from ϕ_2 by substituting q with q^i.

Both the RGB-3D and the RGB-S histograms have been normalized to sum up to 1.0.

Tables 2 and 3 report the values of the parameters providing the best performances with and without smoothing u respectively. The overlap index reported in these tables has been averaged over the number of searched views, that in our case is 100. The results have been broken down by color representation (RGB-3D and RGB-S) and by $\phi = \phi_1$, ϕ_2. The pdf's of the overlap index output from our sensitivity analysis are reported in Figure 10.

The overlap index $\mathcal{X}$ is generally very high and the differences between the results obtained with and without smoothing are negligible. In some cases, smoothing u makes the accuracy better as in Figure 9(first row), while in other cases it provides worst results, because some pixels close to the boundary of the object to be segmented are penalized and thus missed in the detection process, as in Figure 9(second row). The choice of the parameter alg_type giving the best results depends often on the object and on the context in which

Fig. 9 COIL-100: Sensitivity analysis. (a, e) the frontal views of two objects; (b, f) two images depicting the objects in (a, e); segmentation results by using the color representation RGB-3D, $\phi = \phi_2$ with smoothing (c, g) and without smoothing (d, h).

Color	ϕ	λ	D	Δ_t	f_{init}	d	$\mathcal{X}$
RGB-3D	ϕ_1	0.1	D_0	10	5	L^2	0.895
RGB-3D	ϕ_2	0.1	D_0	10	5	L^2	0.912
RGB-S	ϕ_1	0.1	D_0	10	5	L^1	0.898
RGB-S	ϕ_2	0.1	D_0	10	5	L^1	0.896

Table 2 COIL-100: Sensitivity analysis for alg_type $= 1$ (no smoothing of u is applied). The results are broken down by color representation and by ϕ (RGB-3D and RGB-S, $\phi = \phi_1, \phi_2$). The first columns report the sets of the values of the user parameters giving the best overlap index, that is shown in the last column.

the object appears: for instance, when the object is homogeneous and the background contains isolated pixels that have the same color of the object, smoothing u is recommended; on the contrary, when the object is uniformly colored and it includes thin parts (as in case of starred objects), smoothing could remove the pixels of the wiry parts, making the detection worse.

In our experiments, the best performances are achieved by setting the color representation to RGB-3D, $\phi = \phi_2$, alg_type $= 1$, $\lambda = 0.1$, $\Delta t = 10$, $d = L^2$ and $D = D_0$ (see the last column of Table 2). However, these results are very similar to those obtained with the other sets of parameters.

We note that for any choice of alg_type, color representation and ϕ, the best accuracy is obtained when the initial region D contains almost the whole image and thus the object to be segmented. The worst results are obtained for $D = D_1$, i.e. when D is small and almost entirely included in the image region occupied by object view to be segmented. In this case, the overlap index is generally unsatisfactory because several objects of COIL-100 have an homogeneous color and thus the termination criterion is satisfied before the whole object view is retrieved.

In the experiments described in the next Subsections, we report the results obtained by using the parameter set providing the best overlap index.

Color	ϕ	λ	D	Δ_t	f_{init}	f_s	d	$\mathcal{X}$
RGB-3D	ϕ_1	0.05	D_0	10	5	10	L^2	0.889
RGB-3D	ϕ_2	0.05	D_0	10	5	5	L^2	0.904
RGB-S	ϕ_1	0.02	D_0	10	5	10	EMD	0.876
RGB-S	ϕ_2	0.02	D_0	10	5	10	L^1	0.889

Table 3 COIL-100: Sensitivity analysis for alg_type $= 2$ (smoothing of u is applied). The results are broken down by color representation and by ϕ (RGB-3D and RGB-S, $\phi = \phi_1$, ϕ_2). The first columns reports the sets of the values of the user parameters giving the best overlap index, that is shown in the last column.
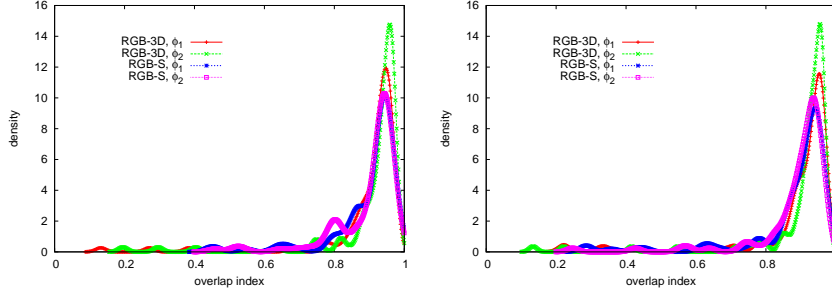

Fig. 10 COIL-100: Sensitivity analysis. These plots show the pdf's of the overlap index obtained by setting alg_type $= 1$(left) and alg_type $= 2$(right). The results are broken down by color representation and by function ϕ. The other parameters are set up as in Tables 2 and 3.

5.4 Robustness against Geometric Distortions

In principle, the description of the color of a region by its pdf is insensitive to many geometric distortions, like changes of size, translations, in-plane rotations. However, in practice, lowering or highering too much the size of the prior makes the recognition hard: in the first case, a lot of visual details are lost, while in the second one artifacts are introduced. Moreover, since the geometric transformations mentioned above do not preserve the discrete nature of the data, this is retrieved in the transformed image by an interpolation procedure and thus the color pdf generally differs from that of the original image version. Therefore, we studied how the accuracy of our model varies when the prior appears in an image at a different scale and with a different orientation.

For this analysis, we considered a synthetic database built up from COIL-100. More precisely, we took each frontal view v_0 of the COIL-100 objects as a prior and we resized and rotated the correspondent image I_0 by linear applications of the form

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \alpha \begin{bmatrix} \cos\theta \ \sin\theta \\ \sin\theta \ \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{19}$$

where $[x, y]$ and $[x', y']$ are the coordinates of the pixels in the image I_0 before and after the transform, α is a scale factor ranging in $\{0.5, 0.75, 1.0, 1.25,$

σ	$L^2[\cdot 10^{-5}]$	$\mathcal{X}$	$\mathcal{I}$
0.50	2.77	0.950	0.983
0.75	4.86	0.945	0.978
1.00	7.31	0.928	0.963
1.25	8.65	0.917	0.953
1.50	9.52	0.910	0.946
1.75	12.5	0.874	0.910

Table 4 COIL-100: Robustness of the model against Gaussian noise. The second column (PDF's dist.) reports the mean value of the L^2 distance between the pdf's of the prior (a frontal view) and of the noised view to be detected in the input image.

1.5} and θ is the rotation angle varying in {0, 45, 90, 135, 180, 225, 270, 315} degrees.

Figure 11 reports the results we obtained. Figure 11 (a) shows the mean L^2 distances between the pdf of the prior and the pdf of its affine transformed versions averaged over the number of views and transforms. We note that these distances range over [0.0015, 0.00040], and they are higher when the image I_0 is rescaled by 0.5. Figures 11(b) and (c) show the overlap index $\mathcal{X}$ and $\mathcal{I}$ for the different transforms: on average $\mathcal{X}$ is 0.94325, while $\mathcal{I}$ is 0.98087. As expected, the values of $\mathcal{X}$ and of $\mathcal{I}$ decrease when the pdf's distance increases.

5.5 Robustness to Noise

We tested the robustness of our approach against Gaussian noise and against changes of viewpoints of the prior.

We measured the robustness against Gaussian noise on a synthetic database, built up from COIL-100. In particular, we took each frontal view v_0 of COIL-100 as prior, and we modified the corresponding image I_0 by adding some Gaussian noise. More precisely, we convolved I_0 by a Gaussian filter with zero mean and variance σ ranging in {0.5, 0.75, 1.0, 1.25, 1.50, 1.75}. Then we used our model to detect the view v_0 in each noisy version of the image I_0.
Table 4 shows the values of $\mathcal{X}$ and $\mathcal{I}$ averaged over the number of tests: as for the geometric distortions, the overlap index decreases when the L^2 distance between the pdf's of the prior and of its noisy version increases. Also in this case, the performances are satisfactory: on average $\mathcal{X}$ and $\mathcal{I}$ are 0.92053 and 0.95528 respectively.

In Subsection 5.3 we studied the performances of our algorithm on a subset of COIL-100, where each frontal view v_0 was a prior and I_5 was the image to be segmented. We observed that the view v_0 is very similar to the view v_5 imaged in I_5, and thus the L^2 distance between the color pdf's of v_0 and v_5 is very small. To study the robustness of our method against changes of pose, we repeated the experiments of 5.3 by considering v_0 as a prior and I_{10}, I_{15} and I_{20} as images to be segmented (see Subsection 5.2.1 for the notation). Table 5 shows the mean overlap index we obtained. This table also reports the accuracy output from the segmentation of the images I_5's already discussed in Subsection 5.3. As expected, the overlap decreases by increasing the distances
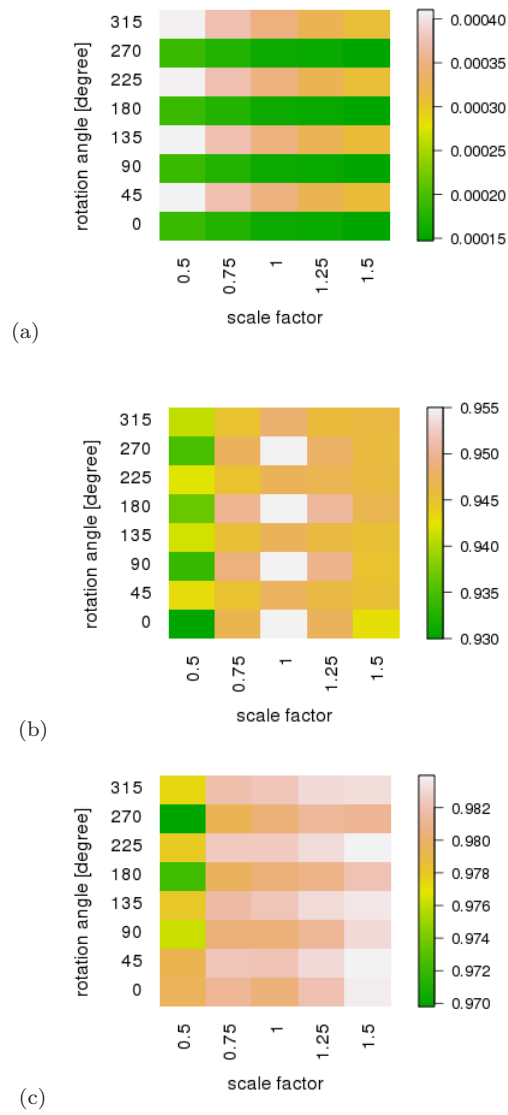
(a)

(b)

(c)

Fig. 11 COIL-100: Robustness against geometric distortions described by Equation (19). (a) Mean L^2 distance between an object view and its geometrical distortions; (b, c) Overlap indices $\mathcal{X}$ and $\mathcal{I}$.

between the pdf's. On average, $\mathcal{X}$ is 0.88466. The last column reports the percentage of object views that have been not detected by our algorithm. This percentage is null for the poses rotated of 5 degrees with respect to the references, while it is less than the 4% for the other cases.

View [degree]	L^2 [$\cdot 10^{-5}$].	$\mathcal{X}$	$\mathcal{I}$	$\mathcal{M}$
5	0.823	0.931	0.959	0
10	1.236	0.903	0.929	0.04
15	1.602	0.869	0.894	0.03
20	1.911	0.835	0.858	0.03

Table 5 COIL-100: Robustness of our model to changes of views. The second column reports the mean value of the L^2 distance between the pdf's of the prior (a frontal view) and of the view to be detected in the input image.

5.6 Results on GroundTruth100-forCOIL and on MLDB2013

The experiments on COIL-100 and on the synthetic databases built up from it showed very good performances. The object views were searched in images with an almost uniform background in order to test our model in an almost ideal framework. Such a configuration occurs in some applications, like in medicine to isolate an anatomic part captured in fMRI or PET images and/or to detect lesions in tissue images [14]. However, in most applications, the object to be detected are displayed against a more complex and cluttered background, as in GroundTruth100-forCOIL and MLDB2013. This makes the object segmentation more difficult, because often the background includes parts visually similar to the object to be detected and therefore false positives can occur. An example is shown in Figure 12, that displays our segmentation results on an image from GroundTruth100-forCOIL. The yellow COIL-100 toy-ship is detected along with some yellow parts of the toy-car and of the box with the white cover. False positives are returned also by segmenting the toy-cat and the toy-car, while the other three objects of COIL-100 (the yellow box, the sandwich, and the terracotta vase) are perfectly segmented. In this example the false positives were generated by the similarity between an object view and parts of other objects displayed in the image. Another example is shown in Figure 13, where the false positives are some background regions with a color similar to the prior.

In these experiments we added to our algorithm an heuristic rule to discard regions with too big or too small area. In particular, let v be a view in the database and let $A(v)$ be its area. Let Ω be an image region with area $A(\Omega)$ detected as an instance of v. Ω is retained as a solution if $0.25A(v) \leq A(\Omega) \leq 2.25A(v)$, i.e. when the minimum and the maximum scale factors at which an object view can appear are 0.5 and 1.5 respectively.

The results are listed in Table 6 for GroundTruth100-forCOIL and in Table 7 for MLDB2013. While the mean value of $\mathcal{I}$ is still high, the overlap index $\mathcal{X}$ is quite low, because many background parts visually similar to the priors have been retained as solution (false positives). In MLDB2013 no object views have been missed, while in GroundTruth100-forCOIL about the 4% has been lost.

Fig. 12 GroundTruth100-for-COIL: Detection of the object views displayed in a picture.

Fig. 13 MLDB2013: the prior (a yellow box, on left) has been correctly detected in the test image (on right), but also some background yellow parts are returned as a solution. This is because these background parts have a color pdf similar to that of the prior. Here the RGB-3D color quantization has been used.

Fig. 14 MLDB2013: The region segmented in the test image (on right) matches the prior in the color pdf, but not in shape.

Measure	Our Approach	INRIA
$\mathcal{X}$	0.529	0.212
$\mathcal{I}$	0.808	0.892
$\mathcal{M}$	0.046	0

Table 6 GroundTruth100-forCOIL: Detection results of our approach and of INRIA [2].

Measure	Our Approach	INRIA
$\mathcal{X}$	0.607	0.519
$\mathcal{I}$	0.692	0.651
$\mathcal{M}$	0	0

Table 7 MLDB2013: Detection results of our approach and of INRIA [2].

5.7 Model Extensions

In this Section we give some hints to extend our model in order to enhance the performances and to make it robust against object partial occlusions. Here we only describe some possible modifications of the model, without going through the details.

As pointed out in Section 1, color is widely used for object detection. However, adding other features to the description of the object of interest might improve the results: for instance, the segmented region shown in Figure 14 matches the color pdf of the prior, but not the shape of the prior. The L^2 distance between the pdf's of the prior and of the output region is very low (6.32262e-05), but about the 20% of the object area in the test image is missed. Imposing constraints on the shape of the object to be segmented may increase the performances. In particular, the active contour model we present here could be extended by adding other terms, taking into account edges as in [4] or other features, in order.

In the current implementation, occluded objects cannot be detected. One way to solve this problem consists in to add a description of the parts of the prior, for instance as suggested in [27]. In this framework, the prior is described by its view v and by parts obtained by occluding v by several half-planes. Then

the whole view v and its parts are described as usual by the pdf of a features (e.g. colors) and the detection is carried out by searching for each view. A test about the ovelap of the solutions is then carried out to retain only the most reliable image regions.

5.8 Comparison with the classic approach

We compared our region based active contour model with the classic variational approach in [2]. Here we briefly describe this method, by using the notation of its authors.
The method in [2], that we shortly call INRIA, segments a prior, in an input image, by minimizing the following energy functional:

$$E(\Omega) = \mathcal{D}(q(\cdot, \Omega), q_p(\cdot)) \tag{20}$$

where q_p is the color pdf of the prior and q is the histogram of the segmentation feature h in the evolving region Ω bounded by the active contour, and $\mathcal{D}(q(\cdot, \Omega), q_p(\cdot))$ is a distance between the pdf's q and q_p.
   The pdf of a feature h over Ω is computed by a Gaussian kernel density estimator as follows:

$$q(\alpha, \Omega) = \frac{1}{|\Omega|} \int_\Omega g_\sigma(h(x) - \alpha) \, dx \tag{21}$$

where $|\Omega|$ is the area of Ω, and g_σ is a Gaussian kernel with variance σ:

$$g_\sigma(h(x) - \alpha) = \frac{1}{(2\pi\sigma^2)^{0.5n}} e^{-\frac{(h(x)-\alpha)^2}{\sigma^2}}.$$

The same formula is used for computing the histogram q_p of the prior.
   Several distance functions $\mathcal{D}$ can be used for matching the pdf's q_p and q. In [2] the Euclidean norm is proposed, i.e.

$$E(\Omega) = \frac{1}{2} \int_{\mathbf{R^n}} (q(\alpha, \Omega) - q_p(\alpha))^2 \, d\alpha. \tag{22}$$

   The evolution equation of the active contour is primed by a shape derivative approach that avoids the conversion of the region functional into a boundary functional:

$$\frac{\partial \omega}{\partial t}(x) = \frac{1}{|\Omega|} \int_{\mathbf{R^n}} (q(\alpha, \Omega) - q_p(\alpha))[g_\sigma(h(x) - \alpha) - q(\alpha, \Omega)] d\alpha \tag{23}$$

Here ω is the active contour bounding Ω.

View [degree]	$L^2[\cdot 10^{-5}]$	$\mathcal{X}$	$\mathcal{M}$
5	0.823	0.857	0
10	1.236	0.830	0
15	1.602	0.804	0
20	1.911	0.776	0

Table 8 COIL-100: Robustness of INRIA model to changes of views. The second column reports the mean value of the L^2 distance between the pdf's of the prior (a frontal view) and of the view to be detected in the input image.

We implemented this approach by using the RGB-S color representation, because using RGB-S instead of RGB-3D allows to break down the computational charge of the convolution in Equation (23). Therefore, we modified the functional in Equation (20) as follows:

$$E(\Omega) = \frac{1}{2} \sum_{i=0}^{2} \int_{\mathbf{R}} (q^i(\alpha, \Omega) - q_p^i(\alpha))^2 \, d\alpha \qquad (24)$$

and thus Equation (23) becomes

$$\frac{\partial \omega}{\partial t}(x) = \frac{1}{|\Omega|} \sum_{i=0}^{2} \int_{\mathbf{R}} (q^i(\alpha, \Omega) - q_p^i(\alpha))[g_\sigma(h^i(x) - \alpha) - q^i(\alpha, \Omega)]d\alpha \quad (25)$$

and $h_i(x)$ is the ith channel response at x.

In Equations (24) and (26), q^i indicates the pdf of a color response and it is computed by the kernel density estimator in Equation (21). In [2] a regularization term is added to guarantee the smoothness of the solution, so that the evolution equation considered is

$$\frac{\partial \omega}{\partial t}(x) = \lambda K(x) + \frac{1}{|\Omega|} \sum_{i=0}^{2} \int_{\mathbf{R}} (q^i(\alpha, \Omega) - q_p^i(\alpha))[g_\sigma(h^i(x) - \alpha) - q^i(\alpha, \Omega)]d\alpha$$

where K is the curvature of the active contour, and λ is a strictly positive parameters weighting the curvature contribution.

In our implementation of INRIA, we used the heuristic termination criterion of our model with $d = L^2$ to stop the curve evolution.

The parameters used by INRIA are σ, λ, D, Δt, f_{init}. We repeated the sensitivity analysis of Subsection 5.3 to set up the values of the parameters giving the best performances. In these experiments, σ varied in $\{0.5, 1.0, 1.5\}$.

The values of parameters providing the best performances are $\sigma = 1.5$, $\lambda = 0.02$, $D = D_0$, $\Delta t = 10$, $f_{init} = 10$. With these values, the best mean overlap index is $\mathcal{X} = 0.8772$.

Tables 6 and 7 show the results obtained on the databases GroundTruth100-forCOIL and MLDB2013, while Table 8 reports the mean overlap index $\mathcal{X}$ obtained on COIL-100 by varying the pose of the imaged object with respect

to the prior. In all the cases, the overlap index of our approach is greater than that of INRIA. INRIA detects in fact more false positives, especially on the images of GroundTruth100-forCOIL. However no object views are missed (see the values of $\mathcal{M}$ in Table 8).

Finally, we note that the computational complexity of the evolution equation of INRIA is greater than that of our model, because of the convolution operation in Equation (25).

6 Conclusions

In this work we proposed a heuristic region based active contour model to detect, in an input image, one or more regions that are visually similar to a 2D object of interest (said prior). The visual appearances of the image regions as well as of the prior are described by the pdf of a photometric feature, in our experiments the color. Our model is *heuristic*, because the evolution equation of the input image region is chosen in order to equalize the pdf of the prior and that of the evolving region, and it is not derived from the minimization of an energy functional. The heuristic law we implemented, works on an initial region by excluding or including image pixels so that the pdf of the output region(s) equalizes pointwise that of the prior. Several mathematical functions can be used to implement such a law. Here we discussed and compared two of them. The experiments carried out on synthetic and real-world image database showed good performances, making our model a good alternative to the classic variational approaches. Our future work will include both the development of an automatic technique to set up the model parameters, and the integration in the model of other visual features, like for instance geometric properties of the prior, as its shape. Clearly, using more features allows a better characterization of the prior and thus a more accurate detection of it.

References

1. Atkinson, K., Han, W.: Theoretical Numerical Analysis. A Functional Analysis Framework, *Texts in Applied Mathematics*, vol. 39. Springer-Verlag (2009)
2. Aubert, G., Barlaud, M., Faugeras, O., Jehan-Besson, S.: Image segmentation using active contours: Calculus of variations or shape gradients? SIAM Applied Mathematics **63**, 2003 (2002)
3. Blake, A., Isard, M.: Active Contours: The Application of Techniques from Graphics,Vision,Control Theory and Statistics to Visual Tracking of Shapes in Motion, 1st edn. Springer-Verlag, Secaucus, NJ, USA (1998)
4. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22**(1), 61–79 (1997)
5. Chakraborty, A., Staib, L., Duncan, J.: Deformable boundary finding in medical images by integrating gradient and region information. Medical Imaging, IEEE Transactions on **15**(6), 859 –870 (1996)
6. Chan, T., Vese, L.: Active contours without edges. Image Processing, IEEE Transactions on **10**(2), 266 –277 (2001)
7. Chen, Y., Tagare, H.D., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K.S., R., Briggs, W., Geiser, E.A.: Using prior shapes in geometric active contours in a variational framework. IJCV **50**, 315–328 (2002)

8. Chesnaud, C., Refregier, P., Boulet, V.: Statistical region snake-based segmentation adapted to different physical noise models. Pattern Analysis and Machine Intelligence, IEEE Transactions on **21**(11), 1145 –1157 (1999)
9. Chuang, C.H., Chao, Y.L., Li, Z.P.: Moving object segmentation and tracking using active contour and color classification models. In: Multimedia (ISM), 2010 IEEE International Symposium on, pp. 73–80 (2010)
10. Cohen, L.D.: On active contour models and balloons. CVGIP: Image Underst. **53**(2), 211–218 (1991)
11. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. Int. J. Comput. Vision **72**(2), 195–215 (2007)
12. Darolti, C., Mertins, A., Bodensteiner, C., Hofmann, U.G.: Local region descriptors for active contours evolution. Trans. Img. Proc. **17**(12), 2275–2288 (2008)
13. Delfour, M., Zolésio, J.: Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization. Advances in Design and Control. Society for Industrial and Applied Mathematics (2011)
14. Derraz, F., Taleb-Ahmed, A., Chikh, A., Bereksi-Reguig, F.: Mr images segmentation based on coupled geometrical active contour model to anisotropic diffusion filtering. In: Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on, pp. 721–724 (2007)
15. FBK-TeV: Technologies of vision - Fondazione Bruno Kessler, groundtruth100-for-COIL (2006). URL http://tev.fbk.eu/DATABASES/objects.html
16. FBK-TeV: Technologies of vision - Fondazione Bruno Kessler, masks of coil-100 dabatabase (2006). URL http://tev.fbk.eu/DATABASES/coil-100-masks.tgz
17. FBK-TeV: Technologies of vision - Fondazione Bruno Kessler, MYDB2013 (2013). URL http://tev.fbk.eu/DATABASES/MLDB2013.html
18. Ge, Q., Xiao, L., Huang, H., Wei, Z.H.: An active contour model driven by anisotropic region fitting energy for image segmentation. Digit. Signal Process. **23**(1), 238–243 (2013)
19. Ge, Q., Xiao, L., Zhang, J., Wei, Z.H.: A robust patch-statistical active contour model for image segmentation. Pattern Recogn. Lett. **33**(12), 1549–1557 (2012)
20. Gevers, T., Gijsenij, A., van de Weijer, J., Geusebroek, J.M.: Color in Computer Vision: Fundamentals and Applications, 1st edn. Wiley Publishing (2012)
21. He, L., Peng, Z., Everding, B., Wang, X., Han, C.Y., Weiss, K.L., Wee, W.G.: A comparative study of deformable contour methods on medical image segmentation. Image and Vision Computing **26**(2), 141 – 163 (2008)
22. Jehan-Besson, S., Barlaud, M., Aubert, G.: Video object segmentation using eulerian region-based active contours. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 1, pp. 353–360 vol.1 (2001)
23. Jehan-Besson, S., Barlaud, M., Aubert, G., Faugeras, O.: Shape gradients for histogram segmentation using active contours. In: Proc. Int. Conf. Computer Vision, pp. 408–415 (2003)
24. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. Journal of Computer Vision **1**(4), 321–331 (1988)
25. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. INTERNATIONAL JOURNAL OF COMPUTER VISION **1**(4), 321–331 (1988)
26. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. Image Processing, IEEE Transactions on **17**(11), 2029 –2039 (2008)
27. Lecca, M., Messelodi, S.: Rotation, Rescaling and Occlusion Invariant Object Retrieval. In: N. Rajpoot, A. Bhalerao (eds.) British Machine Vision Conference - BMVC, pp. 14.1–14.10. BMVA Press (2007)
28. Lecellier, F., Jehan-Besson, S., Fadili, J.: Statistical region-based active contours for segmentation: An overview. IRBM **35**(1), 3 – 10 (2014). Biomedical image segmentation using variational and statistical approaches
29. Leymarie, F., Levine, M.: Tracking deformable objects in the plane using an active contour model. Pattern Analysis and Machine Intelligence, IEEE Transactions on **15**(6), 617–634 (1993)
30. Liu, W., Shang, Y., Yang, X.: Active contour model driven by local histogram fitting energy. Pattern Recognition Letters **34**(6), 655 – 662 (2013)

31. Malladi, R., Sethian, J., Vemuri, B.: Shape modeling with front propagation: a level set approach. Pattern Analysis and Machine Intelligence, IEEE Transactions on **17**(2), 158 −175 (1995)
32. Montagnat, J., Delingette, H., Ayache, N.: A review of deformable surfaces: Topology, geometry and deformation. Image and Vision Computing **19**, 1023–1040 (2001)
33. Nene, S., Nayar, S., Murase, H.: Columbia object image library (COIL-100). Tech. rep., Columbia University (Tech. Rep. 1996). URL http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
34. Ni, K., Bresson, X., Chan, T., Esedoglu, S.: Local histogram based segmentation using the wasserstein distance. Int. J. Comput. Vision **84**(1), 97–111 (2009)
35. Osher, S.J., Fedkiw, R.P.: Level Set Methods and Dynamic Implicit Surfaces, 2003 edn. Springer (2002)
36. Paragios, N., Deriche, R.: Unifying boundary and region-based information for geodesic active tracking. In: Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2, pp. –305 Vol. 2 (1999)
37. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. Int. J. Comput. Vision **46**(3), 223–247 (2002)
38. Paragios, N., Mellina-Gottardo, O., Ramesh, V.: Gradient vector flow fast geodesic active contours. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 1, pp. 67–73 vol.1 (2001)
39. Park, H., Schoepflin, T., Kim, Y.: Active contour model with gradient directional information: directional snake. Circuits and Systems for Video Technology, IEEE Transactions on **11**(2), 252–256 (2001)
40. Ronfard, R.: Region-based strategies for active contour models. International Journal of Computer Vision **13**, 229–251 (1994)
41. Rousson, M., Brox, T., Deriche, R.: Active unsupervised texture segmentation on a diffusion based feature space. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, vol. 2, pp. II–699–704 vol.2 (2003)
42. Rousson, M., Deriche, R.: A variational framework for active and adaptative segmentation of vector valued images. In: Proceedings of the Workshop on Motion and Video Computing, MOTION '02. IEEE Computer Society, Washington, DC, USA (2002)
43. Sang-Mook, L., Abbott, A., Clark, N., Araman, P.: Active contours on statistical manifolds and texture segmentation. In: Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 3, pp. III–828–31 (2005)
44. Sethian, J.A.: Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press (1999)
45. Snyman, J.A.: Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, 2th edn. Applied Optimization, Vol. 97. Springer-Verlag (2005)
46. Vitti, A.: The Mumford-Shah variational model for image segmentation: An overview of the theory, implementation and use. ISPRS Journal of Photogrammetry and Remote Sensing **69**(0), 50 – 64 (2012)
47. Ye, C.L.D., Zheng, Y.: A survey on image segmentation using geometric active contour model. In: Proceedings of the 2012 Second International Conference on Electric Information and Control Engineering - Volume 01, ICEICE '12, pp. 233–236. IEEE Computer Society, Washington, DC, USA (2012)
48. Yezzi A., J., Tsai, A., Willsky, A.: A statistical approach to snakes for bimodal and trimodal imagery. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, pp. 898–903 vol.2 (1999)