Neurocomputing 167 (2015) 658–670

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Likelihood-based feature relevance for figure-ground segmentation in images and videos

Mohand Saïd Allili [a,*], Djemel Ziou [b]

[a] Université du Québec en Outaouais, Département d'informatique et d'ingénierie, Gatineau, QC, Canada
[b] Université de Sherbrooke, Département d'informatique, Sherbrooke, QC, Canada

ARTICLE INFO

Article history:
Received 6 October 2014
Received in revised form
25 December 2014
Accepted 9 April 2015
Communicated by Luming Zhang
Available online 6 May 2015

Keywords:
Figure-ground segmentation
Feature relevance
Positive and negative examples
Gaussian mixture models (GMMs)
Level sets

ABSTRACT

We propose an efficient method for image/video figure-ground segmentation using feature relevance (FR) and active contours. Given a set of positive and negative examples of a specific foreground (an object of interest (OOI) in an image or a tracked objet in a video), we first learn the foreground distribution model and its characteristic features that best discriminate it from its contextual background. For this goal, an objective function based on feature likelihood ratio is proposed for supervised FR computation. FR is then incorporated in foreground segmentation of new images and videos using level sets and energy minimization. We show the effectiveness of our approach on several examples of image/video figure-ground segmentation.

1. Introduction

Object segmentation in images/videos (also called figure-ground segmentation) is important for several applications, such as content-based image/video retrieval (CBIVR) [9,40], automatic image/video annotation [30,36], object-based video coding [41], image/video retargeting [6], robotics and activity recognition [1,34]. In CBIVR and image/video annotation, for example, knowing image/video object content is of prominent importance to enhance the accuracy of semantic labeling of images and videos and answering user queries. Also, newly established multimedia standards for video coding (e.g., MPEG) are based on object content of videos. Therefore, efficient figure-ground segmentation is a critical issue for these applications.

Object segmentation in images is a very challenging problem due to several difficulties such as non-uniform illumination, image clutter and variability within object categories [38]. In the past, approaches have been proposed to tackle these difficulties by using either local or global information (or their fusion) for object segmentation. *Bottom-up* approaches group local cues (e.g., contours, color, texture) to form homogenous regions which can be used to build objects. Popular grouping algorithms are finite mixture models (FMM) [2,9] and graph-cuts [17,37,45]. Based on obtained homogenous regions, some approaches identify foreground objects (resp. backgrounds) through an interactive process exploiting the user's feedback [7,14]. However, these approaches suffer from over/under-segmentation where object parts may be merged with the background and vice versa [2,9]. Also, the need for user interaction with each image limits their usage in large-scale image segmentation. Unlike *bottom-up* approaches which consider local image properties regardless of spatial layout of the segmented object, *top-down* approaches rely on the representation of the global form of objects. These include mainly deformable templates [33], which are also applied as part-based representation models. Templates can be either simple geometrical elements (e.g., ellipses, rectangles, arcs, etc.) [18] or active contours (e.g., the *Snake* model [4]), which are evolved using energy minimization [10,33]. The main difficulty in *top-down* approaches lies in segmenting highly deformable objects, in addition to the need for accurate initialization to compensate for a difficult minimization problem [10,17,33].

Recently, several methods have attempted to combine the advantages of *bottom-up* and *top-down* approaches to achieve better object segmentation. In [8,26], for example, overlaps between segmented images and object fragments are used for object segmentation. The object fragments are usually extracted from a learning set of gray-scale or binary segmented images. However, since each fragment is considered independently, the approach is prone to include object parts in the background. Besides, its computational complexity increases exponentially with the number of fragments and explored image positions. Finally, approaches based on figure-ground color

* Corresponding author.
E-mail addresses: mohandsaid.allili@uqo.ca (M.S. Allili), d.ziou@usherbrooke.ca (D. Ziou).

http://dx.doi.org/10.1016/j.neucom.2015.04.015
0925-2312/

statistical modeling (e.g., using Gaussians [33], mixture of Gaussians [35] or kernel methods [25]) have been proposed for object segmentation. In those approaches, however, segmentation success highly depends on how distinguishable an object is from the background. On the one hand, too many dimensions of noise necessarily overwhelm too few dimensions of signal [28,31]. On the other hand, backgrounds can often be highly correlated with the object (e.g., cars and roads, giraffes and grass, swans and water, etc.) [19]. In videos, objects usually lie against the same background over successive frames. Therefore, accurate figure-ground segmentation can be achievable knowing the most discriminative features that best separate objects from their contextual backgrounds.

In this paper, we propose a new framework combining object/background statistical modeling and *feature relevance* (FR) for efficient figure-ground segmentation in images and videos. For images, FR is computed for each object category by using a set of manually segmented images containing instances of that category (i.e., positive examples) and their contextual backgrounds (i.e., negative examples). Local features are automatically extracted from these images and their figure-ground discrimination power is determined by their likelihood ratio. Our object segmentation approach is formulated as an energy minimization problem and implemented using level sets [32]. An energy functional is proposed to fit figure-ground distributions and encode the contribution of each feature according to its discrimination power. The only assumption of our algorithm is that a segmented object lies in the center of attention of the image. A level set function is evolved from its initial position toward the object boundaries using Euler–Lagrange equations. Finally, an extension of our algorithm to video figure-ground segmentation is proposed. We show the performance of our approach on several figure-ground segmentation on real-world images and videos.

Fig. 1a and b summarizes the two steps composing our approach for figure-ground segmentation in images and videos: (1) *a learning step*: computes FR and figure-ground statistical models using training examples and (2) *a segmentation step*: segments objects in new images (resp. videos) using FR and active contours. Early results of this work have been published in [3]. Herein, we give a more in-depth theoretical analysis of the problem and thorough experiments for validation. We have also added a new section containing complexity analysis of the algorithm and a discussion for future improvements.

This paper is organized as follows: Section 2 presents our approach for FR computation. Section 3 presents our segmentation model with FR. Section 4 presents some experiments that validate the proposed approach. We end the paper with a conclusion and some future work perspectives.

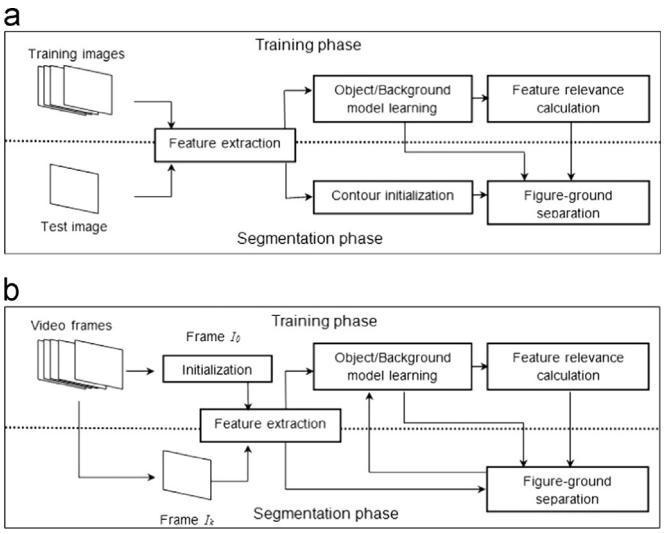

Fig. 1. Outline of our learning-based FR computation and its application to figure-ground segmentation in (a) images and (b) videos.

2. Feature relevance learning for figure-ground segmentation

2.1. Figure-ground distribution models in images

The essence of our approach is to achieve figure-ground segmentation using visual features that best discriminate between objects and their contextual backgrounds. For this goal, we exploit appearance patterns shared between instances of the same object category using a set of training images. These images are chosen to reflect the variety of contextual backgrounds the object may lie against (e.g., tigers in savanna, cars on roads, swans on water, etc.). To build a learning set for FR computation, we manually label locations as object/background in each training image. We call a "positive example" any location chosen in the object; otherwise it is referred to as a "negative example". For each location, we extract color, texture and gradient orientation features from patches centered around the location (see Fig. 2 for illustration). Suppose that we have D features $\{f_1,\ldots,f_D\}$ extracted in each location. Let $\mathcal{C}=\{\mathbf{e}_1,\ldots,\mathbf{e}_{n_1}\}$ and $\overline{\mathcal{C}}=\{\overline{\mathbf{e}}_1,\ldots,\overline{\mathbf{e}}_{n_2}\}$ be two sets of D-dimensional feature vectors extracted from positive and negative examples, respectively, and n_1 and n_2 are their cardinalities. Given that the elements in $\mathcal{C}$ and $\overline{\mathcal{C}}$ can be multi-modal (see for instance Fig. 4), we model feature distributions in $\mathcal{C}$ and $\overline{\mathcal{C}}$ using finite Gaussian mixture models (GMMs) [16].

In a similar way to the naive Bayes classifier [15], we suppose that the features are mutually independent in each class $\mathcal{C}$ and $\overline{\mathcal{C}}$. This is a reasonable assumption since it allows for assessing the discrimination power of each feature individually and reducing complexity and computation time of parameter estimation. Note that to enforce the independence assumption, one could perform *independent component analysis* (ICA) [24] on the features in a pre-processing step. We estimate the GMMs parameters using the *Expectation–Maximization* (EM) algorithm [16], where the number of clusters in each model is automatically determined using the *minimum message length principle* (MML) [39]. We recall that the MML is an information-theoretic principle that gives a good compromise between model complexity and goodness of fit to data [39]. It allows to obtain less complex models which have good fitting to object/background data. In what follows, we denote by $\vec{\theta}_d$ and $\vec{\omega}_d$ the GMM parameters computed for the dth feature using the data in $\mathcal{C}$ and $\overline{\mathcal{C}}$, respectively.

2.2. Feature relevance learning for figure-ground separation

Since we want our segmentation to be driven by the most discriminative features for each object category, we must determine in advance each feature discrimination power. Feature selection methods have been proposed in the past to enhance classification performance [22]. To determine feature subsets ensuring higher discrimination between classes of data, quantitative criteria can be used [28,29]. These criteria can be categorized by whether the evaluation process is data-intrinsic (filters) or classifier-dependent (wrappers). Since exhaustive search of subsets and their evaluation is time-consuming [20], we are constrained to consider simplified and non-exhaustive evaluation strategies to assess about features discrimination. For example, augmented variance ratio (AVR) has been shown to be effective for feature ranking [12]. Similar to Fisher discriminant analysis (FDA) [15], AVR maximizes the ratio between inter-class variance and within-class variance to estimate feature

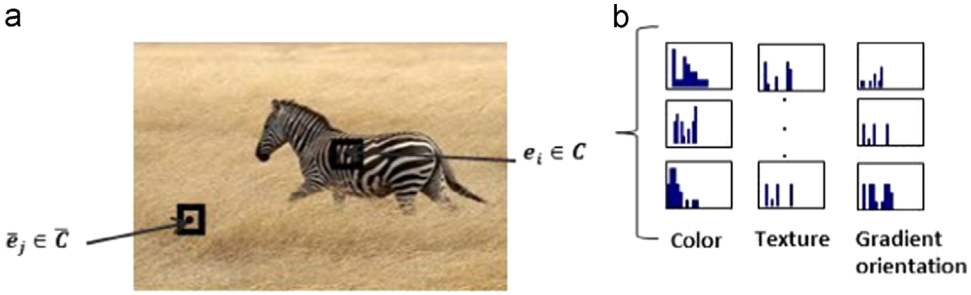

Fig. 2. Illustration of feature extraction from positive and negative examples: (a) represents feature extraction around selected locations, (b) represents features distributions in a location neighborhood.

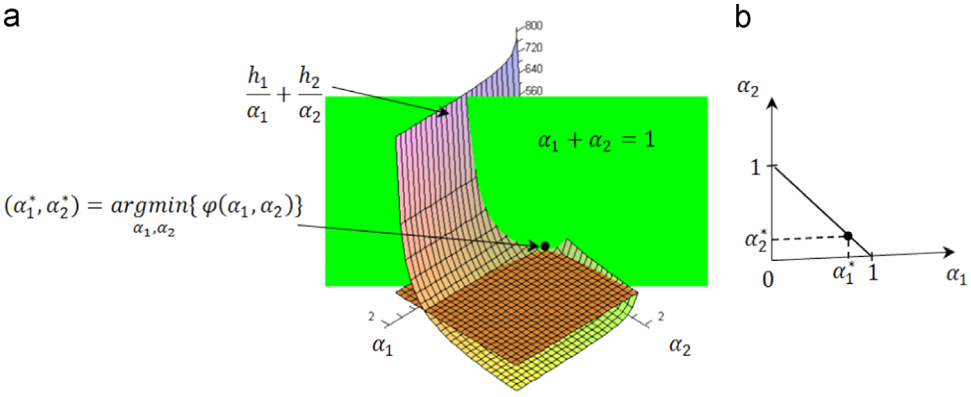

Fig. 3. Illustration of graph of function (1) using an example with $D=2$ and by setting $h_1=70$ and $h_2=10$: (a) represents the graph of function (1) and a plane representing the constraint $\alpha_1+\alpha_2=1$, (b) represents the weight plane coordinates (α_1,α_2) of the graph and the optimal solution (α_1^*,α_2^*). (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)

discrimination power. However, they assume that class data follow a Gaussian distribution.

In segmentation, objects and image backgrounds usually have multi-modal distributions, which violate FDA's Gaussian assumption and invalidate its analytic solution. Since we have the distribution of features modeled as GMMs, we can measure directly feature discrimination using the GMM models. In fact, discriminative features produce likelihood maps where object feature values have higher likelihoods and background feature values have lower likelihoods, and vice versa. Let us define feature weights $\{\alpha_1,\alpha_2,\ldots,\alpha_D\}$ that encode the power of discrimination of each feature, where $\sum_{d=1}^{D}\alpha_d=1$. To estimate these weights, we propose to minimize the following function based on feature log-likelihood ratios:

$$\varphi(\alpha_1,\ldots,\alpha_d)=\sum_{d=1}^{D}\frac{1}{\alpha_d}(A_d-R_d)-\lambda\left(\sum_{d=1}^{D}\alpha_d-1\right), \tag{1}$$

where λ is a Lagrange multiplier ensuring weights summation to 1, and A_d and R_d are defined as follows:

$$\begin{aligned}A_d&=\ln\left[\prod_{i=1}^{n_1}p(\mathbf{e}_{i,d}|\vec{\theta}_d)\prod_{j=1}^{n_2}p(\bar{\mathbf{e}}_{j,d}|\vec{\omega}_d)\right]\\&=\sum_{i=1}^{n_1}\ln\left[p(\mathbf{e}_{i,d}|\vec{\theta}_d)\right]+\sum_{j=1}^{n_2}\ln\left[p(\bar{\mathbf{e}}_{j,d}|\vec{\omega}_d)\right],\end{aligned} \tag{2}$$

and

$$\begin{aligned}R_d&=\ln\left[\prod_{i=1}^{n_1}p(\mathbf{e}_{i,d}|\vec{\omega}_d)\prod_{j=1}^{n_2}p(\bar{\mathbf{e}}_{j,d}|\vec{\theta}_d)\right]\\&=\sum_{i=1}^{n_1}\ln\left[p(\mathbf{e}_{i,d}|\vec{\omega}_d)\right]+\sum_{j=1}^{n_2}\ln\left[p(\bar{\mathbf{e}}_{j,d}|\vec{\theta}_d)\right],\end{aligned} \tag{3}$$

where $\mathbf{e}_{i,d}$ and $\bar{\mathbf{e}}_{j,d}$ denote the dth entry of the vectors $\mathbf{e}_i\in\mathcal{C}$ and $\bar{\mathbf{e}}_j\in\bar{\mathcal{C}}$, respectively;$p(\mathbf{e}_{i,d}|\vec{\theta}_d)$ and $p(e_{i,d}|\vec{\omega}_d)$ are the probabilities of generating the observation $\mathbf{e}_{i,d}$ in $\mathcal{C}$ and $\bar{\mathcal{C}}$, respectively. In the same vein, $p(\bar{\mathbf{e}}_{j,d}|\vec{\theta}_d)$ and $p(\bar{\mathbf{e}}_{j,d}|\vec{\omega}_d)$ are the probabilities of generating the observation $\bar{\mathbf{e}}_{j,d}$ in $\mathcal{C}$ and $\bar{\mathcal{C}}$, respectively. Basically, Eq. (2) represents the sum of the log-likelihoods within the sets of positive and negative sets $\mathcal{C}$ and $\bar{\mathcal{C}}$, respectively, and Eq. (3) represents the sum of the cross-log-likelihoods between these sets.

Since the parameters $\vec{\theta}_d$ and $\vec{\omega}_d$ provide the best fit to the dth feature data in $\mathcal{C}$ and $\bar{\mathcal{C}}$, respectively, the logarithm of the likelihood ratio, denoted by $h_d=A_d-R_d$, is always positive [21]. Therefore, the power of discrimination of the dth feature between $\mathcal{C}$ and $\bar{\mathcal{C}}$ is proportional to the value of h_d. If a feature is not discriminative ($h_d\approx 0$) it has similar distributions in $\mathcal{C}$ and $\bar{\mathcal{C}}$. By opposite, higher values of h_d give the feature more discrimination power between $\mathcal{C}$ and $\bar{\mathcal{C}}$. Note that minimizing (1) according to the parameters α_d gives weights that meet this goal. Indeed, higher weights are needed to penalize higher values of h_d in this equation. After straightforward manipulations, we obtain the following feature weights $\alpha_d, d=1,\ldots,D$ (for more details, see A.1):

$$\alpha_d=\frac{\sqrt{h_d}}{\sum_{k=1}^{D}\sqrt{h_k}}. \tag{4}$$

Eq. (4) assigns a weight to each feature which is proportional to its log-likelihood ratio. This approach is related to distance metric learning (DML) [42,43] where the different features are assigned weights to measure distance between data points. The similarity of the two approaches is more straightforward in the segmentation stage where the contribution of each feature is determined by its power of discrimination. For illustration, Fig. 3a shows the graph of

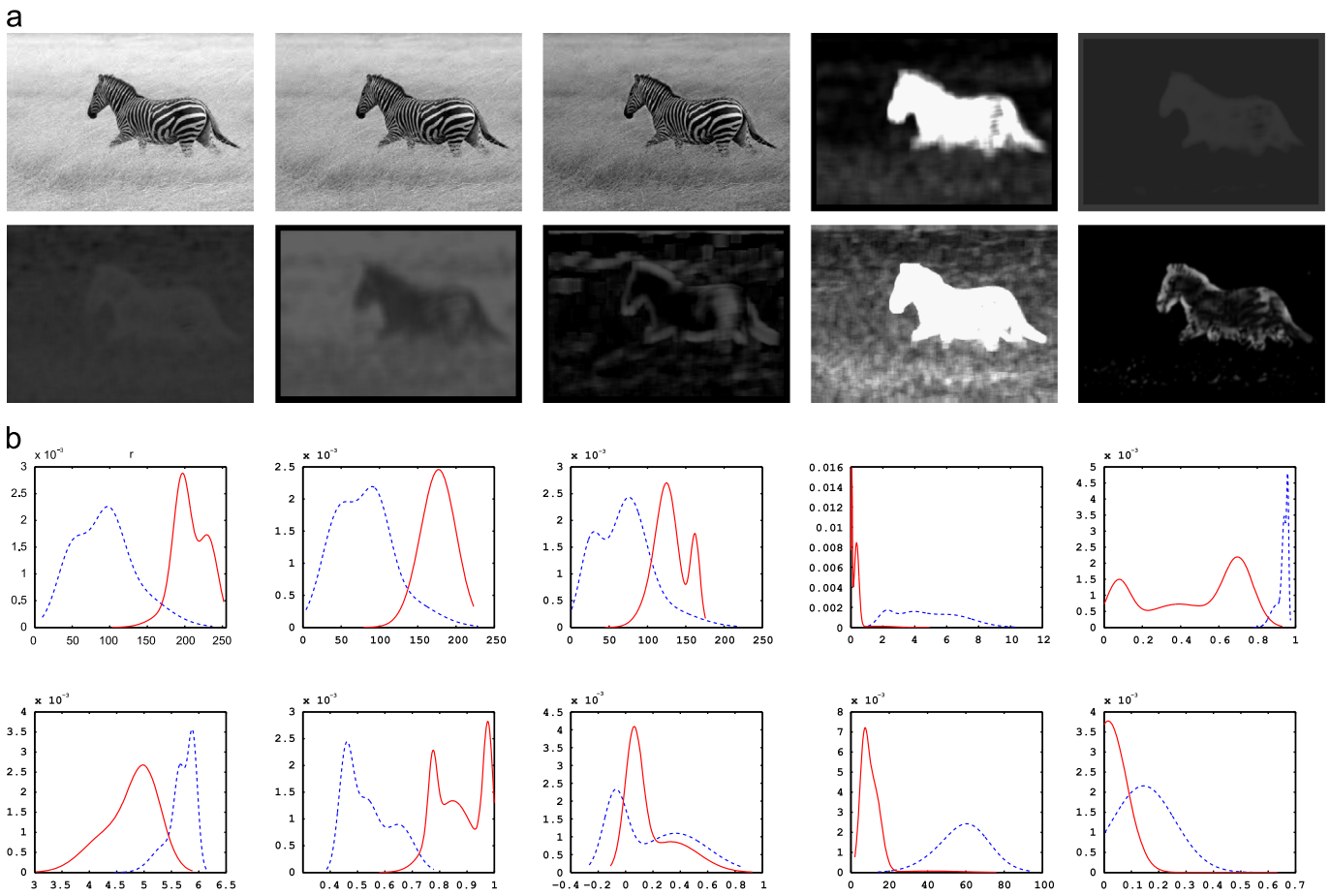

Fig. 4. Illustration of FR calculation for the image in Fig. 2: (a) shows, from left to right and top to bottom, *R*, *G*, *B* color channels, *CT*, *EN*, *ET*, *HM*, *CR*, *VR* and *PL* features, (b) shows, in the same order as (a), GGM models estimated for the features in (a) using the OOI data (dashed blue curve) and background data (continuous red curve). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

function (1) using two features. We set the values of the log-likelihood ratios for these features to 70 and 10, respectively, and we compute α_1 and α_2 using Eq. (4). For more clarity, we separated the graph of the function into two subgraphs: the graph of $h_1/\alpha_1+h_2/\alpha_2$ and the graph expressing the constraint $\alpha_1+\alpha_2=1$ (which is the green plane). The minimum of function (1) is reached for values α_1^* and α_2^* as shown in Fig. 3b. Indeed, function (1) reaches its minimum when the highest weight is assigned to the feature with the highest log-likelihood ratio and vice versa.

Another illustrative example is shown in Fig. 4 where FR is computed for the image in Fig. 2. The considered features are the RGB color channels, average texture feature values of *CT*, *EN*, *ET*, *HM*, *CR* and *VR* calculated for three different orientations and average polarity *PL* calculated for three different scales (see the Experiment section for their definition). Fig. 4 shows values of these features calculated for each image location, as well as the GMMs estimated for their distributions. The obtained feature weights are as follows: α_R : 0.058, α_G : 0.047 and α_B : 0.03, α_{CT} : 0.086, α_{EN} : 0.59, α_{ET} : 0.037, α_{HM} : 0.09, α_{CR} : 0.005, α_{VR} : 0.045 and α_{PL} : 0.014. The calculated weights express exactly the discrimination power of each feature. For instance, the highest weight (0.59) has been assigned to *EN*, which in fact has the strongest discrimination power between the object and the background (see the upper right image). By opposite, the lowest weight (0.005) has been assigned to *CR*, which has the smallest discrimination power (see the third image of the second row).

3. Proposed image figure-ground segmentation algorithm

The previous section introduced our framework to build a feature space that best separate an object category from its contextual background. Given a new image containing an object of interest with known category, we aim to build a partition $\mathcal{P}=\{\mathcal{F},\mathcal{B}\}$ of the image composed of the foreground object $\mathcal{F}$ and the background $\mathcal{B}$. We formulate our segmentation as a classification problem where we make a membership decision for each pixel based on each feature distribution and relevance value.

Using the same set of features as in the learning phase $\{f_1,\ldots,f_D\}$ extracted at each image location, we consider for each feature two parametric GMMs describing its distribution in the object and the background, respectively. Our final segmentation produces an image partition maximizing inner-likelihoods in $\mathcal{F}$ and $\mathcal{B}$ and minimizing their cross-likelihoods, respectively. Without taking into account FR, such segmentation is obtained by maximizing the following function over $\mathcal{F}$ and $\mathcal{B}$:

$$\mathcal{L}(\mathcal{F},\mathcal{B},\vec{\theta},\vec{\omega})=\prod_{d=1}^{D}\left(\prod_{\mathbf{x}\in\mathcal{F}}\frac{p(\mathbf{e}_{\mathbf{x},d}|\vec{\theta}_d)}{p(\mathbf{e}_{\mathbf{x},d}|\vec{\omega}_d)}\prod_{\mathbf{x}\in\mathcal{B}}\frac{p(\bar{\mathbf{e}}_{\mathbf{x},d}|\vec{\omega}_d)}{p(\bar{\mathbf{e}}_{\mathbf{x},d}|\vec{\theta}_d)}\right), \tag{5}$$

where $\vec{\theta}=\{\vec{\theta}_1,\ldots,\vec{\theta}_D\}$ and $\vec{\omega}=\{\vec{\omega}_1,\ldots,\vec{\omega}_D\}$ designate the set of model parameters describing the object and the background of the image, respectively. Here, $\mathbf{e}_{\mathbf{x},d}$ (resp. $\bar{\mathbf{e}}_{\mathbf{x},d}$) represents the dth entry of

the feature vector $\mathbf{e}_{\mathbf{x}}$ (resp. $\overline{\mathbf{e}}_{\mathbf{x}}$) computed at the image location $\mathbf{x}=(x,y)$ inside $\mathcal{F}$ (resp. $\mathcal{B}$). We also have the following mixture probabilities: $p(\mathbf{e}_{\mathbf{x},d}|\vec{\theta}_d)=\sum_{k=1}^{K_d}\pi_{dk}p(\mathbf{e}_{\mathbf{x},d}|\theta_{dk})$ and $p(\overline{\mathbf{e}}_{\mathbf{x},d}|\vec{\omega}_d)=\sum_{h=1}^{\overline{K}_d}\lambda_{dh}p(\overline{\mathbf{e}}_{\mathbf{x},d}|\omega_{dh})$, where the values of K_d and $\overline{K}_d$ are those of the empirical model calculated in the training stage.

Note that function (5) considers all features equally important and, therefore, *discriminative* and *non-discriminative* features have similar contribution to segmentation. A feature f_d is *discriminative* when $r_d^{\mathcal{F}}(\mathbf{x})=p(\mathbf{e}_{\mathbf{x},d}|\vec{\theta}_d)/p(\mathbf{e}_{\mathbf{x},d}|\vec{\omega}_d)\to 1$ and $r_d^{\mathcal{B}}(\mathbf{x})=p(\overline{\mathbf{e}}_{\mathbf{x},d}|\vec{\omega}_d)/p(\overline{\mathbf{e}}_{\mathbf{x},d}|\vec{\theta}_d)\to 1$, which increase function (5). By opposite, f_d is non-discriminative when $r_d^{\mathcal{F}}(\mathbf{x})$ and $r_d^{\mathcal{B}}(\mathbf{x})$ have values less than or equal to 1, which decreases function (5). To emphasize the contribution of *discriminative* features and inhibit *non-discriminative* ones, we introduce the weights α_d in (5) as follows:

$$\mathcal{L}^*(\mathcal{F},\mathcal{B},\vec{\theta},\vec{\omega})=\prod_{d=1}^{D}\prod_{\mathbf{x}\in\mathcal{F}}\{r_d^{\mathcal{F}}(\mathbf{x})\}^{\alpha_d}\prod_{\mathbf{x}\in\mathcal{B}}\{r_d^{\mathcal{B}}(\mathbf{x})\}^{\alpha_d}. \tag{6}$$

The weights in (6) control the influence of each feature according to its power of discrimination. A feature with $\alpha_d\to 0$ will see its contribution reduced to segmentation. By opposite, a feature with greater weight will see its contribution unaltered. By taking the minus-logarithm of the above function and replacing sums with integrals, the maximization problem becomes a minimization of the following functional:

$$\mathbf{J}_r=\sum_{d=1}^{D}\alpha_d\left(\int_{\mathcal{F}}\left[-\ln\left(r_d^{\mathcal{F}}(\mathbf{x})\right)\right]d\mathbf{x}+\int_{\mathcal{B}}\left[-\ln\left(r_d^{\mathcal{B}}(\mathbf{x})\right)\right]d\mathbf{x}\right). \tag{7}$$

This functional has several advantages compared to past methods using energy minimization for segmentation (see for instance [23,35]). First, (7) puts emphasis on the likelihood ratio which encourages pixel labeling by maximizing not only object and background likelihoods, but also by minimizing their cross-likelihoods. Second, (7) incorporates FR which tunes the contribution of each feature according on its power of discrimination. This enables the best discriminative features to drive the segmentation toward better figure-ground separation.

3.1. Segmentation implementation using level sets

Our segmentation is obtained by evolving a closed contour from its initial position toward the object boundaries. The contour is a parametric curve $\Gamma(s): s\in[0,1]\mapsto\mathbf{x}=(x,y)\in\mathbb{R}^2$, where s is the arc-length parameter. We ensure that the final contour is locally smooth by coupling (7) with a boundary potential as suggested in [10]. If L_Γ is the length of the curve Γ, the boundary potential is given as follows:

$$\mathbf{J}_b=\oint_0^{L_\Gamma} g(|P(s)|)\,ds, \tag{8}$$

where $g(|P(s)|)=1/(1+|P(s)|)$ and $P(s)$ is the polarity information defined in [2]. In a nutshell, $P(s)\in[-1,1]$ expresses how the neighborhood of a pixel $\mathbf{x}=(x(s),y(s))$ is textured or uniform. If the neighborhood is homogenous in color and texture, $|P(s)|\approx 0$. In the vicinity of a region contour (i.e., color/texture discontinuity), $|P(s)|\approx 1$. By combining the terms in (7) and (8), we obtain the following energy function:

$$\mathbf{J}_c=\mathbf{J}_r+\vartheta\mathbf{J}_b, \tag{9}$$

where ϑ is a regularization term that controls the contribution of the edge potential. This coefficient is set automatically as suggested in [5]. The energy function (9) is minimized iteratively using the gradient descent method, where each iteration consists of two steps. In the first step, we solve (9) for Γ by keeping the statistical parameters constant. In the second step, we update the object/background statistics using fixed-point iterations.

The implementation of our segmentation is based on the level set method [32] which allows for efficient minimization of (9) and numerical stability. In this method, Γ is embedded as the level zero of a higher dimensional function $\phi(\mathbf{x})$, such that $\Gamma=\{\mathbf{x}=(x,y):\phi(\mathbf{x})=0\}$. Minimization of (9) according to $\phi(\mathbf{x})$ leads to the following Euler–Lagrange equation:

$$\begin{aligned}\frac{\partial\phi(\mathbf{x},t)}{\partial t}=&-\sum_{d=1}^{D}\alpha_d\left[\ln\left(r_d^{\mathcal{F}}(\mathbf{x})\right)-\ln\left(r_d^{\mathcal{B}}(\mathbf{x})\right)\right]|\nabla\phi(\mathbf{x})|\\&+\vartheta\left(P(\mathbf{x})\kappa(\mathbf{x})|\nabla\phi(\mathbf{x})|-\nabla P(\mathbf{x})\cdot\nabla\phi(\mathbf{x})\right),\end{aligned} \tag{10}$$

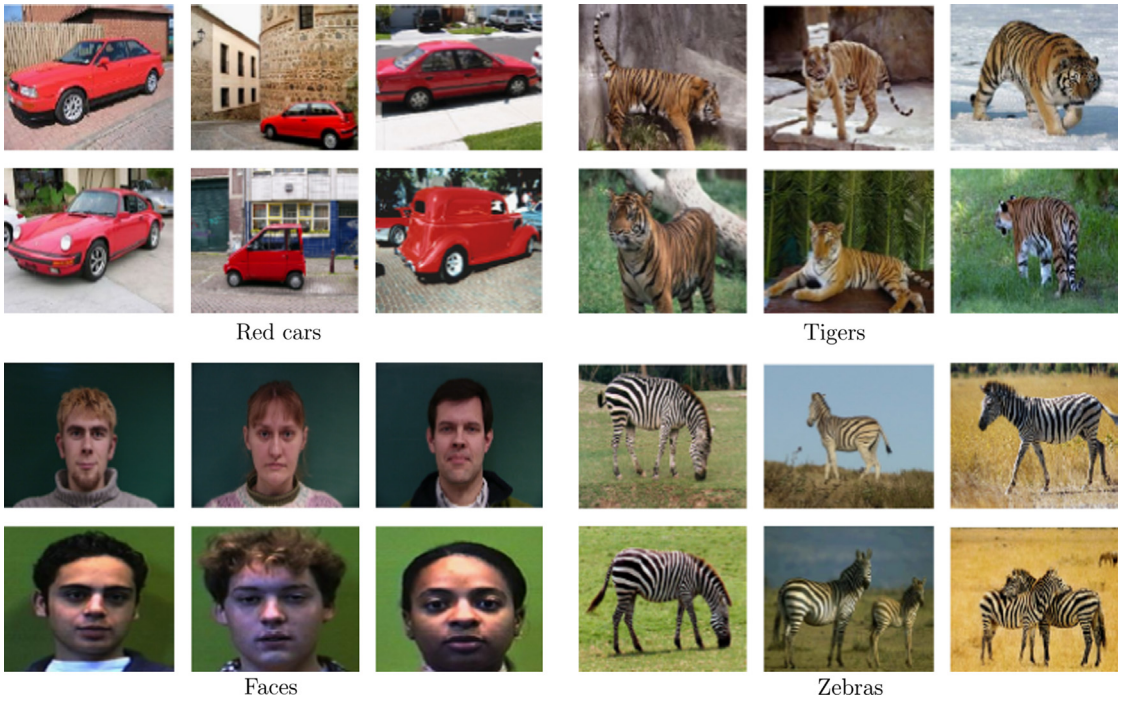

Fig. 5. Sample images used for FR computation in the tested object categories.

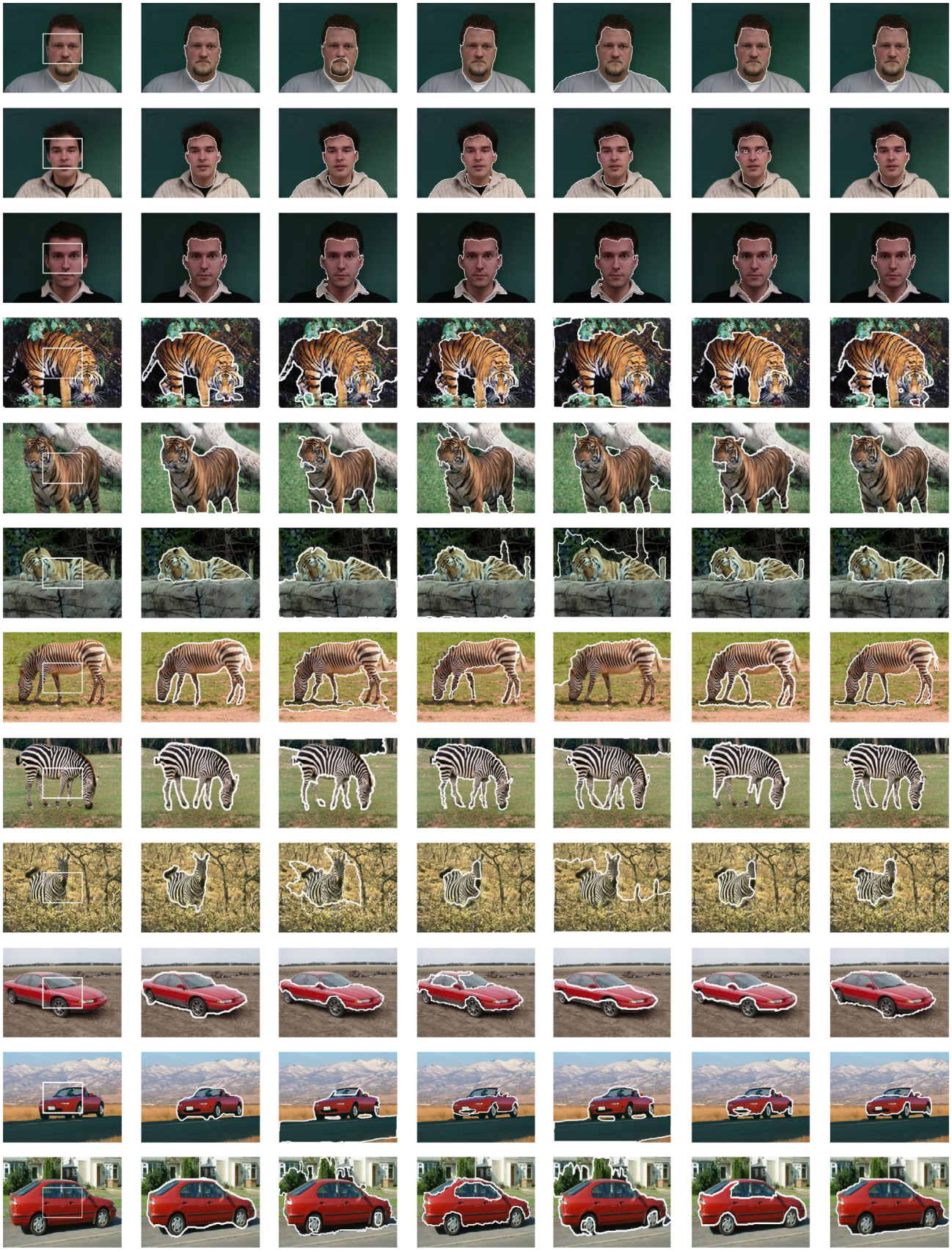

Fig. 6. Examples of figure-ground segmentation using the compared methods: LG-SEG [23], GCUT-SEG [35], FDA-SEG [15] , FBS-SEG [22] and our approach LRR-SEG.

where $\kappa(\mathbf{x})$ stands for the curvature of the zero level set at point $\mathbf{x}$ and t is the time parameter. In practice, all the derivatives in (10) are implemented using finite differences [2].

3.2. Statistical parameter updating

As the contour Γ evolves, data inside and outside Γ change constantly. Therefore, statistical parameters must be updated to fit new data inside and outside the contour. To this end, we minimize (9) w.r.t. to the GMM parameters of each feature in the foreground (resp. background) parts. Each feature generates the following updated parameters $\vec{\theta}^* = \{\pi^*_{dk}, \mu^*_{dk}, \sigma^*_{dk}\}_{k=1}^{K_d}$ for $\mathcal{F}$ and $\vec{\omega}^* = \{\lambda^*_{dh}, \mu^*_{dh}, \sigma^*_{dh}\}_{h=1}^{\overline{K}_d}$ for $\mathcal{B}$, where $\mu^*_{(\cdot)}$ and $\sigma^*_{(\cdot)}$ stand for the Gaussian mean and standard deviation, respectively. In what follows, we develop the formulas for foreground statistics updating. We can obtain the background parameters using similar equations. Using fixed-point iterations (see A.2), we obtain the following updating formulas:

$$\hat{\mu}^*_{dk} = \frac{\int_{\mathcal{F}} t_{dk,\mathbf{x}} \mathbf{e}_{\mathbf{x},d}\, d\mathbf{x} - \int_{\mathcal{B}} \bar{t}_{dk,\mathbf{x}} \bar{\mathbf{e}}_{\mathbf{x},d}\, d\mathbf{x}}{\int_{\mathcal{F}} t_{dk,\mathbf{x}}\, d\mathbf{x} - \int_{\mathcal{B}} \bar{t}_{dk,\mathbf{x}}\, d\mathbf{x}} \tag{11}$$

$$\hat{\sigma}^*_{dk} = \frac{\int_{\mathcal{F}} t_{dk,\mathbf{x}} \left[\mathbf{e}_{\mathbf{x},d} - \mu_{dk}\right]^2 d\mathbf{x} - \int_{\mathcal{B}} \bar{t}_{dk,\mathbf{x}} \left[\bar{\mathbf{e}}_{\mathbf{x},d} - \mu_{dk}\right]^2 d\mathbf{x}}{\int_{\mathcal{F}} t_{dk,\mathbf{x}}\, d\mathbf{x} - \int_{\mathcal{B}} \bar{t}_{dk,\mathbf{x}}\, d\mathbf{x}} \tag{12}$$

$$\hat{\pi}^*_{dk} = \frac{\int_{\mathcal{F}} t_{dk,\mathbf{x}} - \int_{\mathcal{B}} \bar{t}_{dk,\mathbf{x}}\, d\mathbf{x}}{\sum_{l=1}^{K_d} \int_{\mathcal{F}} t_{dl,\mathbf{x}}\, d\mathbf{x} - \int_{\mathcal{B}} \bar{t}_{dl,\mathbf{x}}\, d\mathbf{x}} \tag{13}$$

where $t_{dk,\mathbf{x}} = p(\theta_{dk}|\mathbf{e}_{\mathbf{x},d})$ and $\bar{t}_{dk,\mathbf{x}} = p(\theta_{dk}|\bar{\mathbf{e}}_{\mathbf{x},d})$. In the updating procedure, each component with a weight value smaller than a fixed threshold ϵ is automatically removed from the mixture model (experimental value $\epsilon = 0.01$ gave good results). Note that if the distributions of the object and the background do not overlap for a given feature, Eqs. (11)–(13) boil down to the maximum likelihood estimation in $\mathcal{F}$, since the integrals over the region $\mathcal{B}$ will vanish in these equations. Function (9) is minimized by alternating between Eq. (10) for contour evolution and Eqs. (11)–(13) for statistical updating [2]. Finally, to accelerate convergence of segmentation, all mixture parameters are initialized to those obtained for the empirical models calculated in the training stage.

Table 1
Values of the errors *SD* and *SA* for the compared segmentation methods.

Objects categories	LG-SEG		GCUT-SEG		FDA-SEG		FBS-SEG		LRR-SEG	
	SD_c	SA_c (%)	SD_c	SA_c (%)	SD_c	SA_c (%)	SD_c	SA_c (%)	SD_c	SA_c (%)
Red cars	7.48	51	4.51	68	8.22	61	4.91	76	3.82	77
Faces	4.34	53	3.22	71	5.23	59	3.86	79	2.28	81
Tigers	5.14	48	4.85	67	6.84	52	5.23	68	3.23	83
Zebras	6.33	45	5.05	73	8.12	51	4.67	72	3.12	82

4. Experiments

We conducted experiments on segmentation of four different object categories: (*cars*, *tigers*, *zebras* and *faces*) in real-world still images. We applied also our approach for figure-ground separation in video sequences. Although both applications use the same equations for FR computation and object segmentation, they differ in the way the positive and negative examples are defined. Thus, we develop each application separately.

4.1. Figure-ground segmentation in images

For each object category, we built a database of training images containing instances of objects in the category. We asked four persons to manually select locations in the images that lie in the object or the background, respectively. For each image, we calculate a set of features including color, texture and gradient orientation. Our goal is to segment one or several instances of an object in an unseen image (for example, a herd of zebras is segmented as one instance of the object "zebra"). For color, the perceptually uniform $L^*a^*b^*$ color space is used. For gradient orientation, the polarity information [4] is calculated for each pixel in 3 different scales. For texture, features from the correlogram matrix [2] are used, namely the: *Variance (VR)*, *Contrast (CT)*, *Energy (EN)*, *Entropy (ET)*, *Homogeneity (HM)* and *Correlation (CR)*, calculated for 3 directions $\theta \in \{0°, 45°, 90°\}$, to have a total of $D = 24$. Fig. 5 shows a sample of images used in our dataset. For each object category, we have used 100 different images for training (as described in Section 2) and 200 images for testing (as described in Section 3).

To segment a new image, we initialize for each feature two GMMs associated with the foreground and the background data, respectively. Initial GMM parameters for each category are set to the empirical parameter values obtained in the training stage. More

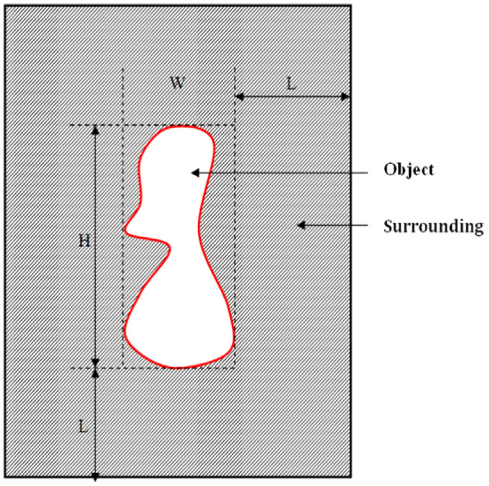

Fig. 8. Illustration of the object surroundings defined using the bounding rectangles B and B′. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)

Fig. 7. Examples of segmentation failure using our method.

specifically, we have obtained the following values for the (object, background) regions: *red cars* ($K_d=3, \overline{K}_d=8$), *faces* ($K_d=3, \overline{K}_d=5$), *tigers* and *zebras* ($K_d=2, \overline{K}_d=6$), $d\in\{1,\ldots,D\}$. We assume that most of our objects lie in the focus of attention; thus, we initialize the object contour using a rectangle with its center equals to the image center. The rectangle width and height in each segmented image are 1/3 of the image width and height, respectively. To measure segmentation quantity for each object category c, $c\in\{tigers, zebras, redcars, faces\}$, we designed two objective criteria, namely: *statistics deviation* (SD_c) and *segmentation accuracy* (SA_c), which are given as follows:

$$SD_c=\frac{1}{N_c}\sum_{i=1}^{N_c}\left(\sum_{d=1}^{D} KL(\hat{q}_d^{(i)}, q_d^{(c)})+KL(q_d^{(c)}, \hat{q}_d^{(i)})\right), \tag{14}$$

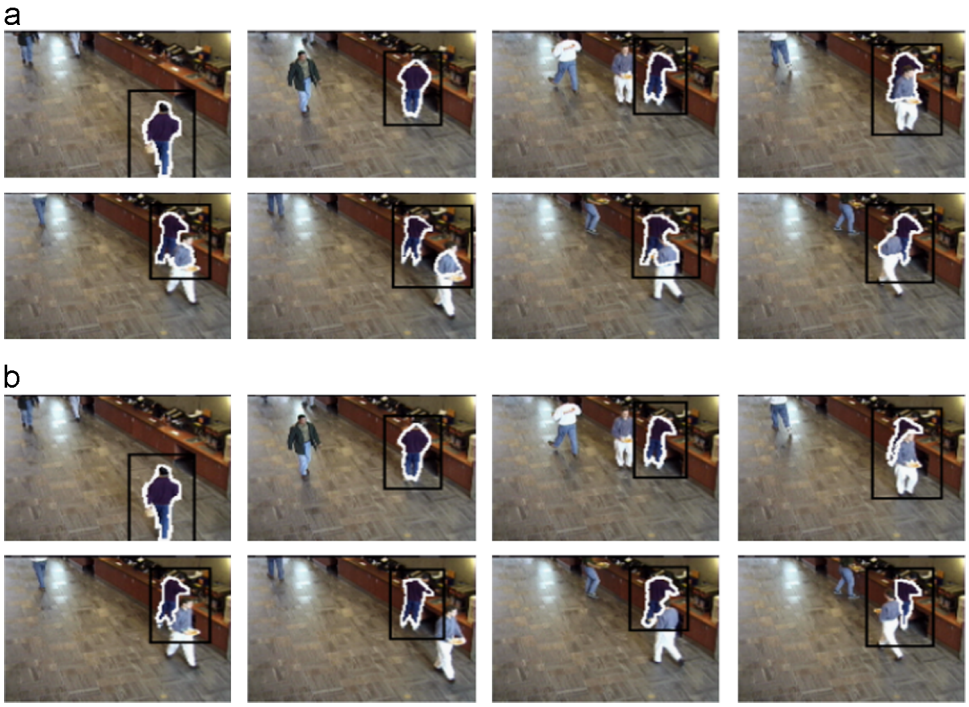

Fig. 9. Part (a) shows tracking results without using FR for frames 04, 12, 76, 81, 82, 84, 110 and 112, from left to right and top to bottom. Part (b) shows tracking results for the same frames using FR.

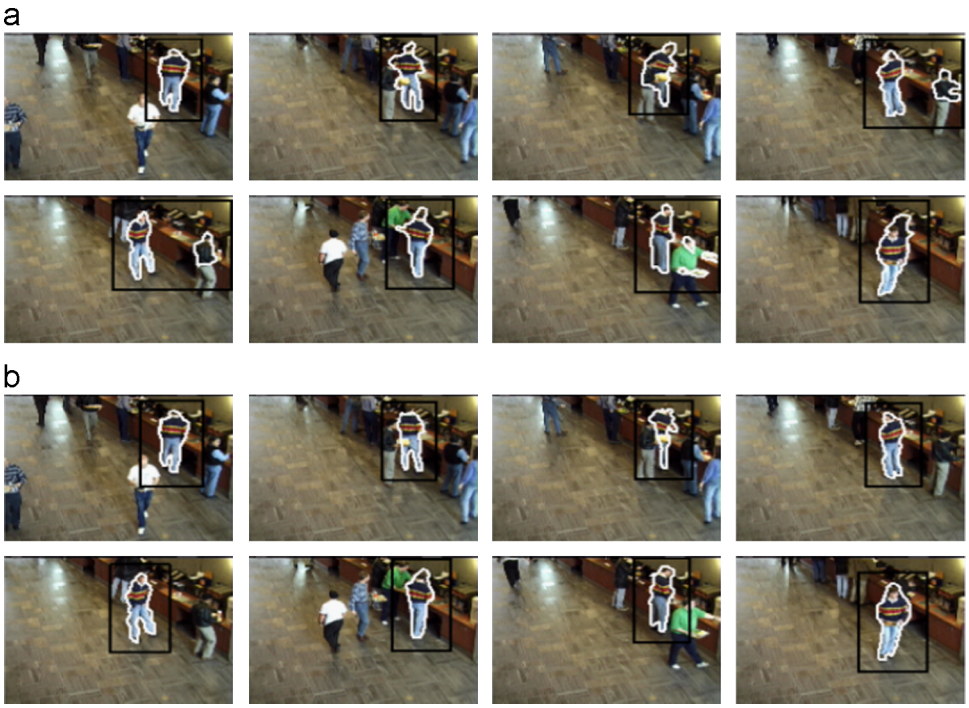

Fig. 10. Part (a) shows tracking results without using FR for frames 01, 20, 30, 63, 80, 127, 143, and 185, from left to right and top to bottom. Part (b) shows tracking results for the same frames using FR.

where N_c is the number of segmented images in the category c, KL designates the *Kullback–Leibler* divergence; $q_d^{(c)}$ and $\hat{q}_d^{(i)}$ are the dth feature distributions for the object category c calculated in the learning step, and the ith segmented image. In other words, SD_c measures average deviation of object statistics from the ground truth after segmenting the images of each category. The second criterion measures the geometric deviation of the object segmentation. Let $\mathcal{F}^{(i)}$ be the output object (i.e., foreground) obtained after segmenting the ith image of category c and $G_{\mathcal{F}}^{(i)}$ is the ground truth, the average segmentation accuracy in category c is measured by the following function:

$$SA_c = \frac{1}{N_c}\sum_{i=1}^{N_c} 1 - \frac{|\mathcal{F}^{(i)} - G_{\mathcal{F}}^{(i)}| + |G_{\mathcal{F}}^{(i)} - \mathcal{F}^{(i)}|}{|\mathcal{F}^{(i)}| + |G_{\mathcal{F}}^{(i)}|}, \tag{15}$$

where "–" and "$|\cdot|$" designate set difference and cardinality, respectively. Note that $SA_c \in [0,1]$ where $SA_c = 1$ corresponds to a perfect segmentation. To assess the performance of our approach, we have compared it with two segmentation methods based on object/background statistical modeling and energy minimization: (1) the method in [23] uses iterative local–global likelihoods for figure-ground segmentation and GrabCut [35] which uses mixture modeling for object/background color distribution, but does not use any FR calculation. We refer to these methods as LG-SEG and GCUT-SEG, respectively. To make GCUT-SEG setting similar to our algorithm, figure-ground seeds are initialized using training images instead of interacting the user directly with the segmented image. Our method using likelihood-ratio for feature relevance is denoted by LRR-SEG. We have also implemented two variants of our method using FDA [15] and forward–backward feature selection [20]. We refer to these variants as FDA-SEG and FBS-SEG, respectively.

Fig. 6 shows segmentation examples for the tested object categories. The first and second columns show contour initialization and the ground truth, respectively. The third to the seventh columns show segmentations using LG-SEG, GCUT-SEG, FDA-SEG, FBS-SEG and LRR-SEG, respectively. From the shown results, we can clearly see that our method has yielded the best segmentation accuracy compared to the other methods. LG-SEG and FDA-SEG yielded the worst results among the compared methods. Indeed, LG-SEG does not use any prior knowledge about the object and background color distribution and feature relevance. On the other hand, FDA assume the data of object/background classes follow Gaussian distributions. This assumption is not realistic for most of the examples and, consequently, some parts of the objects have been absorbed by the background and vice versa. GCUT-SEG and FBS-SEG have the closest performance to our approach. Indeed, GCUT-SEG uses prior knowledge about the object/background distributions to perform segmentation. On the other hand, FBS-SEG selects feature subsets ensuring the best separation between objects and their contextual backgrounds. However, we observed that using FBS is more conservative in that some parts of the objects were segmented to the background. Using FR in our method has yielded more accurate segmentations.

Table 1 shows values of SD_c and SA_c obtained for each object category and the compared methods. For the SA_c criterion, our approach on average outperforms GCUT-SEG by 13% and LG-SEG by 39%, FDA-SEG by 38% and FBS-SEG by 08%. In terms of the SD_c criterion, LRR-SEG also gave the best performance. This is directly related to the quality of segmentation since lesser values of SD_c mean lesser deviation from the ground truth. Note that the worst performance has been obtained for the "red car" category, which can be explained by the high variability in the object/background appearance in this category compared to the other categories. Nonetheless, our approach outperformed the compared methods for this criterion. Finally, Fig. 7 gives some examples where our algorithm failed to accurately separate objects from the background. Indeed, in these images, no feature in our pool has been discriminative enough to separate the objects from their background.

4.2. Figure-ground segmentation in video sequences

Segmenting moving objects from background in video sequences is important for several applications such as video surveillance, video coding and editing [44]. In the past, several methods

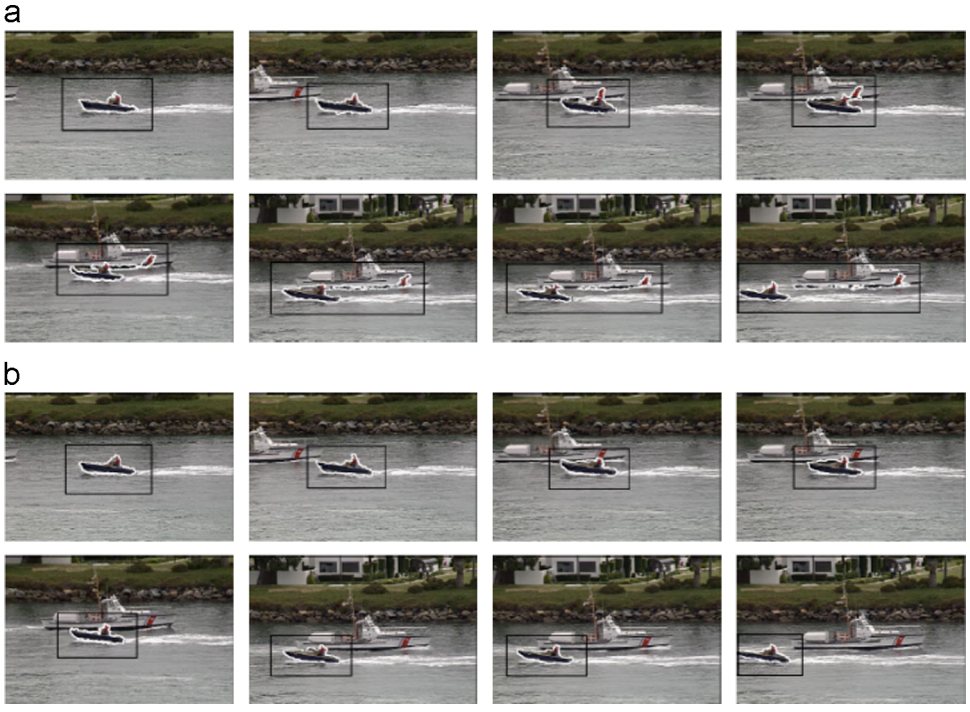

Fig. 11. Part (a) shows tracking results using FR of frames 02, 28, 52, 56, 69, 85, 88 and 103, from left to right and top to bottom. Part (b) shows tracking results for the same frames without using FR.

have been proposed for object tracking using appearance models [44]. Using object histogram, for example, has led to the *mean-shift* method which performs fast and accurate tracking of object positions [13]. Collins et al. [12] have used linear combinations of color channels to build feature sets that enhance *mean-shift* tracking. In general, these methods are capable of localizing object positions in a video sequence. However, they are not efficient in localizing object boundaries. In this paper, we apply our figure-ground segmentation method to track boundaries of moving objects in videos. We apply FR and level sets via Eq. (10) to ensure maximum discrimination between the objects and their surroundings.

Figs. 9–11 show three examples comparing tracking results by using and without using FR. In these examples, neighborhoods of the target objects change constantly due to the movement of other objects which distract tracking. The three sequences are composed of 266, 200 and 130 frames, respectively. For each frame, we calculate the feature weights by extracting the positives and negative examples from its last 4 predecessor frames. The extraction of these examples is performed as follows. Suppose that the bounding rectangle B of the object contour has height H and width W in the frame (see Fig. 8, where the object contour is represented by the red line). The surrounding of the object is the part of the background delimited by the rectangle B′ which has the same center as B and has height: $H+2L$ and width: $W+2L$. The value of L is chosen sufficiently high (60 pixels in our experiments) and the tracking is performed inside B′.

In the first example (see Fig. 9), the immediate background of the tracked object changes due to partial occlusions from other distracting objects. The object has been successfully isolated from

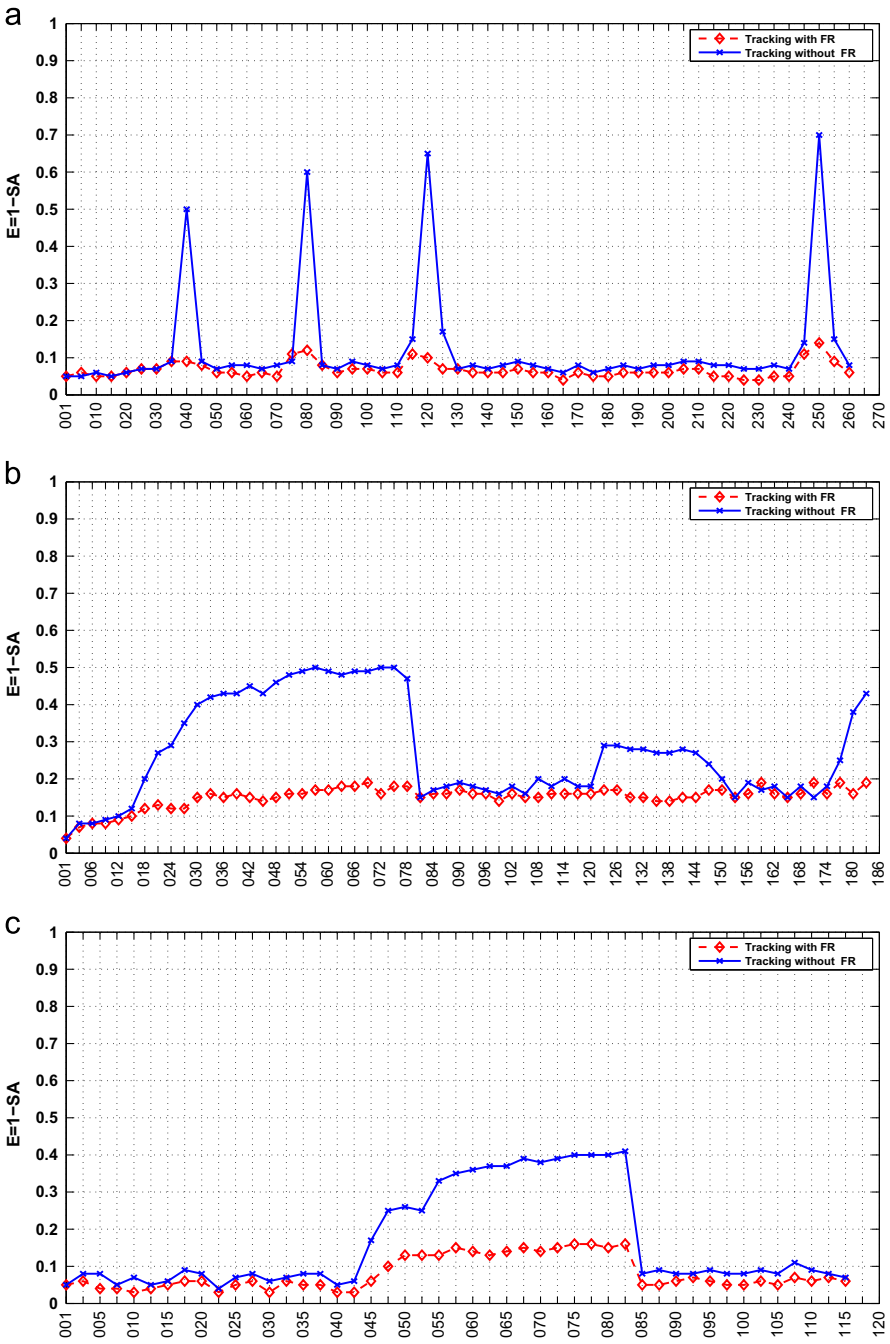

Fig. 12. Evolution of the tracking error $E=1-SA$ w.r.t. the frame number in the: (a) first , (b) second and (c) third sequences, respectively.

the background by using our method. Without using FR, the tracking has been distracted at frames 40, 60, 120 and 250 where parts from contacting/occluding objects have been included in the tracked object. To resume tracking in this case, we had to re-initialize manually the object contour in these frames. Our method successfully avoided these distractions where the level set remained around the real target object boundaries. The same remark holds for the second and third examples (see Figs. 10 and 11), where we let the tracking continue without contour re-initialization when distraction occurs.

Finally, Fig. 12 gives values of $E=1-SA$ (i.e., tracking error) as a function of frame number in the three test sequences. In the first example, where we had to re-initialize tracking at each distraction, we note that tracking is distracted around frames 40, 80, 120 and 250 when FR is not used. The level set includes parts of the background where the target object passes in proximity of other objects. The same remark holds for the second and third examples, where we let run the tracking without re-initialization. Indeed, the tracking recovers itself only when the object causing distraction disappears from the scene. Using FR successfully avoids these distractions which maintains a low error value E in the three examples. These experiments demonstrate the benefit of using FR for figure-ground segmentation in videos.

5. Computational efficiency

For segmenting a new image, suppose that the evolved level set function encloses the object region $\mathcal{F}^{(t)}$ at time t. Then, one step of updating the level set function using (10) has a computation time $O(D|\partial\mathcal{F}^{(t)}|)$, where D is the number of features and $|\partial\mathcal{F}^{(t)}|$ is the length of the contour (in pixels). For statistical parameter updating, the computational complexity is $O(KD|\Omega|)$, where K is the total number of components in the object/background data and $|\Omega|$ is the size of the image. Therefore, the total computational time of each iteration is $O(D|\partial\mathcal{F}^{(t)}|)+O(KD|\Omega|)$. We note that our approach has approximately the same computational complexity as LG-SEG [23] and GCUT-SEG [35], but the latter approaches do not have a pre-training stage. In a Matlab/C environment running on a 2.5 GHz Intel processor, the average time for image segmentation is 4.7 s using our approach, 5.5 s using LG-SEG and 3.4 s using GCUT-SEG . In video sequences, FR is computed dynamically for each frame using previous frames tracking results, which adds some computation time to our algorithm. However, this can be alleviated by using for each frame a contour initialization based on the last tracking results. Thus, a small number of iterations is required for the level set to reach the tracked object boundaries. Currently, it takes about 0.5 s to process a frame using the above computer setting.

6. Conclusions and discussion

We have proposed an approach for figure-ground segmentation in images and videos which efficiently integrate feature relevance (FR), statistical modeling and active contours. Our approach for learning FR is based on maximization of likelihood ratio between positive and negative examples of a given object category. We have incorporated FR in a variational framework for figure-ground segmentation using level sets. Obtained results have demonstrated the advantage of using our approach over recent state-of-the-art methods based on energy minimization. Despite the fact that our approach needs a training stage for FR computation, it does not add too much computation burden to our segmentation since the training can be done in an off-line fashion and once for all segmentations.

Finally, some remarks can be made about the limitations of our method. First, for most of the failed segmentations that we have obtained, we noticed that the backgrounds were either highly textured or contain clutter. In this case, no feature used in our pool could provide enough discrimination between the objects and their background. This problem can be alleviated, for example, by using shape information which can be activated in such cases. Another limitation relates to the assumption that an object lies in the focus of attention in the image (i.e., center of the image), which helps to correctly initialize contours and the algorithm to converge. To relax this constraint, one can add an object localization module to the system to enable automatic contour initialization to the right position. This opens the door to several techniques that can be applied. For example, saliency maps [11] and bag-of-words [27] can be efficiently applied to achieve this goal. In video sequences, our approach can be improved by taking into account correlation between successive frames.

Acknowledgment

The completion of this research has been made possible with the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix A

A.1. Derivation of feature relevance weights

We minimize function (1) according to each parameter α_d by finding the value giving the following equality:

$$\frac{\partial\varphi(\alpha_1,\ldots,\alpha_D)}{\partial\alpha_d}=0\Rightarrow\frac{-1}{\alpha_d^2}(A_d-R_d)-\lambda=0\Rightarrow\lambda=\frac{-(A_d-R_d)}{\alpha_d^2}. \tag{16}$$

By putting $h_d=A_d-R_d$ and using the constraint $\sum_{d=1}^{D}\alpha_d=1$, we get from Eq. (16) the following equation:

$$\frac{h_d}{\alpha_d}+\lambda\alpha_d=0\Rightarrow\sum_{d=1}^{D}\frac{h_d}{\alpha_d}+\lambda\sum_{d=1}^{D}\alpha_d=\sum_{d=1}^{D}\frac{h_d}{\alpha_d}+\lambda=0. \tag{17}$$

$$\Rightarrow\lambda=-\sum_{d=1}^{D}\frac{h_d}{\alpha_d}. \tag{18}$$

Putting together Eqs. (16) and (18)**?** gives

$$\frac{h_d}{\alpha_d^2}=\sum_{j=1}^{D}\frac{h_j}{\alpha_j}. \tag{19}$$

In one side, Eq. (19) gives the equality: $h_1/\alpha_1^2=\cdots=h_d/\alpha_d^2=\cdots=h_D/\alpha_D^2$. Therefore, we can write each parameter α_d as follows:

$$\frac{1}{\alpha_d}=\frac{1}{\alpha_j}\sqrt{\frac{h_j}{h_d}},\quad\forall j=1,\ldots,D. \tag{20}$$

In the other side, by substituting the value of $1/\alpha_d$ of Eq. (20) in Eq. (19), we obtain

$$\frac{h_d}{\alpha_d^2}=\sum_{j=1}^{D}\frac{1}{\alpha_d}\sqrt{\frac{h_d}{h_j}}h_j=\sum_{j=1}^{D}\frac{1}{\alpha_d}\sqrt{h_dh_j}\Rightarrow\alpha_d=\frac{\sqrt{h_d}}{\sum_{j=1}^{D}\sqrt{h_j}} \tag{21}$$

A.2. Derivation of fixed-point iterations for GMM parameter updating

By keeping the terms in (9) containing the foreground parameters, we have

$$B_d = \int_{\mathcal{F}} -\ln\left(\sum_{k=1}^{K_d} \pi_{dk} p(\mathbf{e}_{\mathbf{x},d}|\theta_{dk})\right) d\mathbf{x} + \int_{\mathcal{B}} \ln\left(\sum_{k=1}^{K_d} \pi_{dk} p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dk})\right) d\mathbf{x} \tag{22}$$

Minimizing (9) w.r.t. $\vec{\theta}$ amounts to find the parameters that give the best fit to data in $\mathcal{F}$ and the least fit to data in $\mathcal{B}$, respectively. In other words, the first term of the function yields exactly the maximum likelihood estimation for the parameters. The second term discourages these parameters to fit the background data. The local minima of (9) w.r.t. each parameter θ_{dk} is obtained by setting the first derivative of (9) w.r.t. θ_{ik} equal to zero. Then, we have

$$\frac{\partial B_d}{\partial \theta_{dk}} = -\int_{\mathcal{F}} \left(\frac{\pi_{dk}\frac{\partial p(\mathbf{e}_{\mathbf{x},d}|\theta_{dk})}{\partial \theta_{dk}}}{\sum_{j=1}^{K_d} \pi_{dj} p(\mathbf{e}_{\mathbf{x},d}|\theta_{dj})}\right) d\mathbf{x} + \int_{\mathcal{B}} \left(\frac{\pi_{dk}\frac{\partial p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dk})}{\partial \theta_{dk}}}{\sum_{j=1}^{K_d} \pi_{dj} p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dj})}\right) d\mathbf{x} \tag{23}$$

$$= -\int_{\mathcal{F}} p(\theta_{dk}|\mathbf{e}_{\mathbf{x},d}) \frac{\partial \ln\left(p(\mathbf{e}_{\mathbf{x},d}|\theta_{dk})\right)}{\partial \theta_{dk}} d\mathbf{x} + \int_{\mathcal{B}} p(\theta_{dk}|\bar{\mathbf{e}}_{\mathbf{x},d}) \frac{\partial \ln\left(p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dk})\right)}{\partial \theta_{dk}} d\mathbf{x}. \tag{24}$$

By substituting θ_{dk} with μ_{dk}, then σ_{dk}, we obtain the formulas in Eqs. (11) and (12). To estimate the mixing parameters π_{dk}, we should take into account the constraint $\sum_{k=1}^{K_d} \pi_{dk} = 1$. In doing so, we obtain a new function to minimize using a Lagrange multiplier λ given as

$$C_d = B_d - \lambda\left(\sum_{k=1}^{K_d} \pi_{dk} - 1\right), \tag{25}$$

which is minimized first w.r.t. π_{dk} to obtain

$$\frac{\partial C_d}{\partial \pi_{dk}} = -\int_{\mathcal{F}} \frac{p(\mathbf{e}_{\mathbf{x},d}|\theta_{dk})}{\sum_{j=1}^{K_d} \pi_{dj} p(\mathbf{e}_{\mathbf{x},d}|\theta_{dj})} d\mathbf{x} + \int_{\mathcal{B}} \frac{p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dk})}{\sum_{j=1}^{K_d} \pi_{dj} p(\bar{\mathbf{e}}_{\mathbf{x},d}|\theta_{dj})} d\mathbf{x} - \lambda \tag{26}$$

$$= -\int_{\mathcal{F}} \frac{p(\theta_{dk}|\mathbf{e}_{\mathbf{x},d})}{\pi_{dk}} d\mathbf{x} + \int_{\mathcal{B}} \frac{p(\theta_{dk}|\bar{\mathbf{e}}_{\mathbf{x},d})}{\pi_{dk}} d\mathbf{x} - \lambda. \tag{27}$$

By putting $\partial C_d/\partial \pi_{dk} = 0$, we obtain

$$\pi_{dk} = \frac{\int_{\mathcal{F}} p(\theta_{dk}|\mathbf{e}_{\mathbf{x},d})\, d\mathbf{x} - \int_{\mathcal{B}} p(\theta_{dk}|\bar{\mathbf{e}}_{\mathbf{x},d})\, d\mathbf{x}}{\lambda}. \tag{28}$$

Given that $\sum_{l=1}^{K_d} \pi_{dl} = 1$, we obtain

$$\sum_{l=1}^{K_d} \int_{\mathcal{F}} p(\theta_{dl}|\mathbf{e}_{\mathbf{x},d})\, d\mathbf{x} - \int_{\mathcal{B}} p(\theta_{dl}|\bar{\mathbf{e}}_{\mathbf{x},d})\, d\mathbf{x} = \lambda. \tag{29}$$

By combining Eqs. (28) and (29), we readily derive Eq. (13).

References

[1] J.K. Aggarwal, M.S. Ryoo, Human activity analysis: a review, ACM Comput. Surv. 43 (3) (2011) 16.
[2] M.S. Allili, D. Ziou, Globally adaptive region information for automatic color-texture image segmentation, Pattern Recognit. Lett. 28 (15) (2007) 1946–1956.
[3] M.S. Allili, D. Ziou, Using feature selection for object segmentation and tracking, in: IEEE Canadian Conference on Computer and Robot Vision, 2007, pp. 191–200.
[4] M.S. Allili, D. Ziou, Automatic colour-texture image segmentation using active contours, Int. J. Comput. Math. 84 (9) (2007) 1325–1338.
[5] M.S. Allili, D. Ziou, An approach for dynamic combination of region and boundary information in segmentation, in: IEEE International Conference on Pattern Recognition, 2008, pp. 1–4.
[6] S. Avidan, A. Shamir, Seam carving for content-aware image resizing, ACM Trans. Graph. 26 (3) (2007) Article no. 10.
[7] S. Bagon, O. Boiman, M. Irani, What is a good image segment? A unified approach to segment extraction, in: European Conference on Computer Vision, 2008, pp. 30–44.
[8] E. Borenstein, S. Ullman, Combined top-down/bottom-up segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2109–2125.
[9] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation–maximization and its application to image querying, IEEE Trans. Pattern Anal. Mach. Intell. 24 (8) (2002) 1026–1038.
[10] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, Int. J. Comput. Vis. 22 (1) (1997) 61–79.
[11] K.-Y. Chang, T.-L. Liu, H.-T. Chen, S.-H. Lai, Fusing generic objectness and visual saliency for salient object detection, in: IEEE International Conference on Computer Vision, 2011, pp. 914–921.
[12] R.T. Collins, Y. Liu, Online selection of discriminative tracking features, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1631–1643.
[13] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
[14] A. Delong, L. Gorelick, F.R. Schmidt, O. Veksler, Y. Boykov, Interactive segmentation with super-labels, Energy Minim. Methods Comput. Vis. Pattern Recognit. (2011) 147–162.
[15] O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley, New York, 2002.
[16] M.T. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 381–396.
[17] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vis. 59 (2) (2004) 167–181.
[18] P.F. Felzenszwalb, Representation and detection of deformable shapes, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 208–220.
[19] K. Fragkiadaki, J. Shi, Figure-ground image segmentation helps weakly-supervised learning of objects, in: European Conference on Computer Vision, 2010, pp. 561–574.
[20] F.C. Garcia-Lopez, M. Garcia-Torres, B. Melian, J.A. Moreno-Perez, J.M. Moreno-Vega, Solving feature subset selection problem by a parallel scatter search, Eur. J. Oper. Res. 169 (2) (2006) 477–489.
[21] J.K. Ghosh, M. Delampady, T. Samanta, An Introduction to Bayesian Analysis: Theory and Methods, Springer, New York, 2006.
[22] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[23] G. Hua, Z. Liu, Z. Zhang, Y. Wu, Iterative local–global energy minimization for automatic extraction of objects of interest, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1701–1706.
[24] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, London, 2001.
[25] J. Kim, J.W. Fisher, A.J. Yezzi, M. Cetin, A.S. Willsky, A nonparametric statistical method for image segmentation using information theory and curve evolution, IEEE Trans. Image Process. 14 (10) (2005) 1486–1502.
[26] G. Larivière, M.S. Allili, A learning probabilistic approach for object segmentation, in: IEEE Canadian Conference on Computer and Robot Vision, 2012, pp. 28–94.
[27] D. Larlus, F. Jurie, Latent mixture vocabularies for object categorization and segmentation, Image Vis. Comput. 27 (5) (2009) 523–534.
[28] M. Law, M. Figueiredo, A.K. Jain, Simultaneous feature selection and clustering using a mixture model, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1154–1166.
[29] S. Li, E. Xia, C. Zong, C.-R. Huang, A framework for feature selection methods for text categorization, in: Annual Meeting of the Association for Computational Linguistics, 2009, pp. 692–700.
[30] L.-J. Li, R. Socher, L. Fei-Fei, Towards total scene understanding: classification, annotation and segmentation in an automatic framework, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2036–2043.
[31] B.W. Mel, Seemore: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition, Neural Comput. 9 (4) (1997) 777–804.
[32] S. Osher, J. Sethian, Fronts propagating with curvature-dependant speed: algorithms based on Hamilton–Jacobi formulations, J. Comput. Phys. 79 (1) (1988) 12–49.
[33] N. Paragios, Y. Chen, O. Faugeras, The Handbook of Mathematical Models in Computer Vision, Springer, New York, 2005.
[34] X. Ren, C. Gu, Figure-ground segmentation improves handled object recognition in egocentric video, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3137–3144.
[35] C. Rother, V. Kolmogorov, A. Blake, "GrabCut": interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (3) (2004) 309–314.
[36] H. Sahbi, X. Li, Context based support vector machines for interconnected image annotation, in: Asian Conference on Computer Vision (2010) 214–227.
[37] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[38] R. Szeliski, Computer Vision, Algorithms and Applications, Springer, New York, 2011.
[39] C.S. Wallace, Statistical and inductive inference by minimum message length, Information Science and Statistics, Springer, New York, 2005.
[40] T. Wang, Y. Rui, J.-G. Sun, Constraint based region matching for image retrieval, Int. J. Comput. Vis. 56 (1–2) (2004) 37–45.
[41] Y. Wang, J. Ostermann, Y.-Q. Zhang, Video Processing and Communications, Prentice Hall, New Jersey, 2002.

[42] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest-neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.
[43] L. Yang, R. Jin, Distance Metric Learning: A Comprehensive Survey, Technical Report, 2006.
[44] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Comput. Surv. 38 (4) (2006) Article 13.
[45] R. Zabih, V. Kolmogorov, Spatially coherent clustering using graph cuts, in: IEEE Conference on Computer Vision and Pattern Recognition, II: 2004, pp. 437–444.

Mohand Saïd Allili received the M.Sc. and Ph.D. degrees in computer science from the University of Sherbrooke, Sherbrooke, QC, Canada, in 2004 and 2008, respectively. Since June 2008, he has been an assistant professor of computer science with the Department of Computer Science and Engineering, Université du Québec en Outaouais, Canada. His main research interests include computer vision and graphics, image processing, pattern recognition, and machine learning. Dr. Allili was a recipient of the Best Ph.D. Thesis Award in engineering and natural sciences from the University of Sherbrooke for 2008 and the Best Student Paper and Best Vision Paper awards for two of his papers at the Canadian Conference on Computer and Robot Vision 2007 and 2010, respectively.

Djemel Ziou received the B.Eng. degree in computer science from the University of Annaba (Algeria) in 1984, and Ph.D. in computer science from the Institut National Polytechnique de Lorraine (INPL), France, in 1991. From 1987 to 1993 he served as a lecturer in several universities in France. During the same period, he was a researcher in the Centre de Recherche en Informatique de Nancy (CRIN) and the Institut National de Recherche en Informatique et Automatique (INRIA) in France. Presently, he is a full professor at the Department of Computer Science at the University of Sherbrooke in Canada. He has served on numerous conference committees as a member or chair. He heads the laboratory MOIVRE and the consortium CoRIMedia which he founded. His research interests include image processing, information retrieval, computer vision and pattern recognition.