

# Segmentation by Fusion of Histogram-Based $K$ -Means Clusters in Different Color Spaces

Max Mignotte

**Abstract**—This paper presents a new, simple, and efficient segmentation approach, based on a fusion procedure which aims at combining several segmentation maps associated to simpler partition models in order to finally get a more reliable and accurate segmentation result. The different label fields to be fused in our application are given by the same and simple ( $K$ -means based) clustering technique on an input image expressed in different color spaces. Our fusion strategy aims at combining these segmentation maps with a final clustering procedure using as input features, the local histogram of the class labels, previously estimated and associated to each site and for all these initial partitions. This fusion framework remains simple to implement, fast, general enough to be applied to various computer vision applications (e.g., motion detection and segmentation), and has been successfully applied on the Berkeley image database. The experiments herein reported in this paper illustrate the potential of this approach compared to the state-of-the-art segmentation methods recently proposed in the literature.

**Index Terms**—Berkeley image database, color spaces, fusion of segmentations,  $K$ -means clustering, textured image segmentation.

## I. INTRODUCTION

IMAGE segmentation is a classic inverse problem which consists of achieving a compact region-based description of the image scene by decomposing it into meaningful or spatially coherent regions sharing similar attributes. This low-level vision task is often the preliminary (and also crucial) step in many video and computer vision applications, such as object localization or recognition, data compression, tracking, image retrieval, or understanding.

Because of its simplicity and efficiency, clustering approaches were one of the first techniques used for the segmentation of (textured) natural images [1]. After the selection and the extraction of the image features [usually based on color and/or texture and computed on (possibly) overlapping small windows centered around the pixel to be classified], the feature samples, handled as vectors, are grouped together in compact but well-separated clusters corresponding to each class of the image. The set of connected pixels belonging to each estimated class thus defined the different regions of the scene. The method known as  $K$ -means (or Lloyd's algorithm)

[2] (and its fuzzy version called fuzzy  $C$ -means) are some of the most commonly used techniques in the clustering-based segmentation field, and more generally, “by far, the most popular clustering algorithm used in industrial applications and machine learning” [3].

Many other methods have been proposed and studied in the last decades to solve the textured image segmentation problem. Contrary to clustering algorithms, spatial-based segmentation methods exploit the connectivity information between neighboring pixels and have led to Markov random field (MRF)-based statistical models [4], mean-shift-based techniques [5], graph-based [6], [7] or finally region-based split and merge procedures, sometimes directly expressed by a global energy function to be optimized [8].

Years of research in segmentation have demonstrated that significant improvements on the final segmentation results may be achieved by using notably more sophisticated feature selection procedures, more elaborate clustering techniques (involving sometimes a mixture of different or non Gaussian distributions for the multidimensional texture features [9], [10]), taking into account prior distribution on the labels, region processes, or the number of classes [8], [11], [12], finally, involving (in the case of energy-based segmentation models) more costly optimization techniques.

The segmentation approach, proposed in this paper, is conceptually different and explores a new strategy; in fact, instead of considering an elaborate and *better designed* segmentation model of textured natural image, our technique rather explores the possible alternative of fusing (i.e., efficiently combining) several segmentation maps associated to simpler segmentation models in order to get a final reliable and accurate segmentation result. More precisely, this work proposes a fusion framework which aims at fusing several  $K$ -means clustering results (herein using as simple cues the values of the requantized color histogram estimated around the pixel to be classified) applied on an input image expressed by different color spaces. These different label fields are fused together by a simple  $K$ -means clustering techniques using as input features, the local histogram of the class labels, previously estimated and associated to each initial clustering result.

This paper demonstrates that the proposed fusion method, while being simple and fast performs competitively and often better (in terms of visual evaluations and quantitative performance measures) than the best existing state-of-the-art recent segmentation methods on the Berkeley natural image database (containing also, for quantitative evaluations, ground truth segmentations obtained from human subjects).

Manuscript received September 5, 2007; revised February 6, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

The author is with the Département d'Informatique et de Recherche Opérationnelle (DIRO), Université de Montréal, Faculté des Arts et des Sciences, Montréal H3C 3J7 QC, Canada (e-mail: mignotte@iro.umontreal.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.920761

## II. INITIAL SEGMENTATIONS TO BE FUSED

The initial segmentation maps which will then be fused together by our fusion framework (see Section III) are simply given, in our application, by a  $K$ -means [2] clustering technique, applied on an input image expressed by different color spaces, and using as simple cues (i.e., as input multidimensional feature descriptor) the set of values of the re-quantized color histogram (with equidistant binning) estimated around the pixel to be classified. In our application, this local histogram is equally re-quantized (for each of the three color channels) in a  $N_b = 5 \times 5 \times 5 = 125$  bins descriptor, computed on an overlapping squared fixed-size ( $N_w = 7$ ) neighborhood centered around the pixel to be classified. This estimation can be quickly computed by using a more coarsely requantized color space and then computing the bin index that represents each re-quantized color (see Fig. 1 and Algorithm 1).

Mathematically, let  $b(\mathbf{x}) \in \{0, \dots, N_b - 1\}$  denote the bin index associated with the color vector  $\mathbf{y}(\mathbf{x})$  at pixel location  $\mathbf{x}$  (lying on a pixel grid) and  $\mathcal{N}(\mathbf{x})$  be the set of pixel locations within the squared neighborhood region (of fixed-size  $N_w \times N_w$ ) centered at pixel location  $\mathbf{x}$  (in which local color information will be gathered). An estimate  $\mathbf{h}(\mathbf{x}) = \{h(n; \mathbf{x})\}_{n=0, \dots, N_b-1}$  of 125 bins descriptor, characterizing the color distribution for each pixel to be classified, is given by the following standard bin counting procedure:

$$h(n; \mathbf{x}) = \mathcal{K} \sum_{\mathbf{u} \in \mathcal{N}(\mathbf{x})} \delta[b(\mathbf{u}) - n] \quad (1)$$

where  $\delta$  is the Kronecker delta function and  $\mathcal{K} = 1/(N_w)^2$  is a normalization constant ensuring  $\sum_{n=0}^{N_b-1} h(n; \mathbf{x}) = 1$  (see Fig. 1 and Algorithm 1).

---

**Algorithm 1.** Estimation, for each pixel  $\mathbf{x}$ , of the bins descriptor.

---

### Estimation of the $N_b = q^3$ bins descriptor.

$\mathcal{N}_{\mathbf{x}}$  Set of pixel locations  $\mathbf{x}$  within the  $N_w \times N_w$  neighborhood region centered at  $\mathbf{x}$ .

$h[]$  Bins descriptor: Array of  $N_b$  floats ( $h[0], h[1], \dots, h[N_b - 1]$ ).

$\lfloor \cdot \rfloor$  Integer part of.

**For** each pixel  $\mathbf{x} \in \mathcal{N}_{\mathbf{x}}$  with color value  $R_{\mathbf{x}}, G_{\mathbf{x}}, B_{\mathbf{x}}$

**do**

- $k \leftarrow q^2 \cdot \lfloor q \cdot R_{\mathbf{x}}/256 \rfloor + q \cdot \lfloor q \cdot G_{\mathbf{x}}/256 \rfloor + \lfloor q \cdot B_{\mathbf{x}}/256 \rfloor$ .
  - $h[k] \leftarrow h[k] + 1/(N_w)^2$ .
- 

In this simpler model, a *texton* (i.e., the repetitive character or element of a textured image, also called a texture primitive) is herein characterized by a mixture of colors or more precisely by the values of the re-quantized (local) color histogram. This model is simple to compute, allows significant data reduction while being robust to noise and local image transformations and

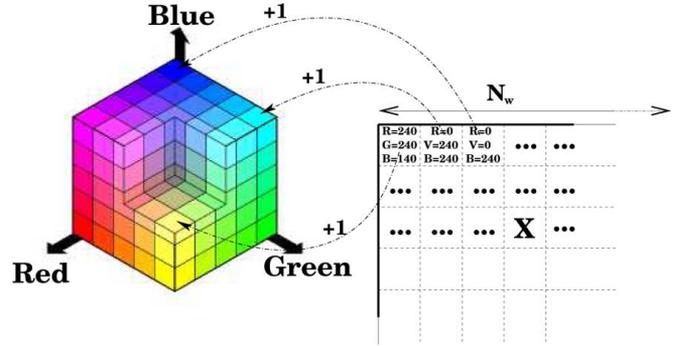


Fig. 1. Estimation, for each pixel  $\mathbf{x}$ , of the  $N_b = q^3$  bins descriptor ( $q = 5$ ) in the RGB color space. The RGB color cube is first divided into  $N_b = q^3$  equal-sized smaller boxes (or bins). Each  $R_{\mathbf{x}}, G_{\mathbf{x}}, B_{\mathbf{x}}$  color value associated to each pixel contained in a (squared) neighborhood region (of size  $N_w \times N_w$ ) centered at  $\mathbf{x}$ , increments (+1) a particular bin. The set of bin values represents the (non-normalized) bin descriptor. We then divide all values of this  $N_b$  bins descriptor by  $(N_w \times N_w)$  in order to ensure that the sum of these values integrates to one.

has already demonstrated all its efficiency for tracking applications [13].

Finally, these (125-bin) descriptors are grouped together into different clusters (corresponding to each class of the image) by the classical  $K$ -means algorithm [2] with the classical Euclidean distance. This simple segmentation strategy of the input image into  $K_1$  classes is repeated for different color spaces which can be viewed as different image channels provided by various sensors or captors (or as a multichannel filtering where the channels are represented by the different color spaces). In our application, we use  $N_s$  segmentations provided by the  $N_s = 6$  color spaces, namely the  $\mathcal{C} = \{\text{RGB, HSV, YIQ, XYZ, LAB, LUV}\}$  color spaces [1], [14]–[16]. Of course, these initial segmentations to be fused can result of the same initial and simple model used on an input image filtered by another filter bank (e.g., a bank of Gabor filters [11], [17] or any other 2-D decomposition of the frequential space) or can also be provided by different segmentation models or different segmentation results provided by different seeds of the same stochastic segmentation model.

Each color space has an interesting property, which can efficiently be taken into account in order to make more reliable the final fusion procedure. For example, RGB is an additive color system based on tri-chromatic theory and nonlinear with visual perception. This space color seems to be the optimal one for tracking applications [18]. The HSV is interesting in order to decouple chromatic information from shading effect [13]. The YIQ color channels have the property to code the luminance and chrominance information which are useful in compression applications (both digital and analogue). Besides, this system is intended to take advantage of human color characteristics. XYZ has the advantage of being more psycho-visually linear, although they are nonlinear in term of linear component color mixing. The LAB color system approximates human vision, and its  $L$  component closely matches human perception of lightness [1]. The LUV components provide an Euclidean color space yielding a perceptually uniform spacing of color approximating a Riemannian space [17]. Each of these properties will be efficiently combined by our fusion technique.

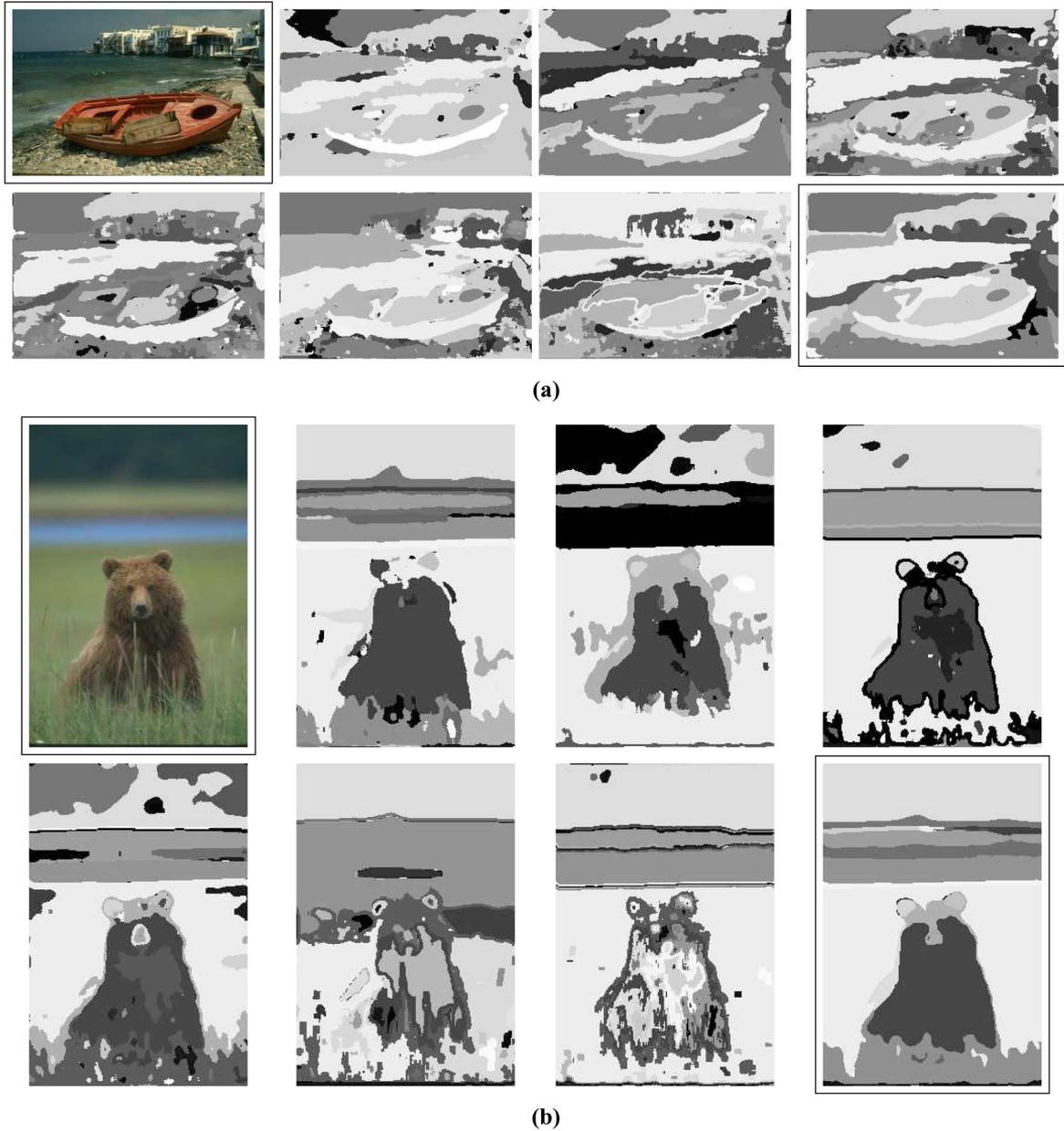


Fig. 2. Examples of fusion results ( $FCR_{[K_1=6|K_2=6|\kappa=0.130]}$ ). From top to bottom and left to right: (top left) input natural image from the Berkeley image database. Six segmentation results (into  $K_1 = 6$  classes) associated to clustering model described in Section II on the top left input image expressed in the RGB, HSV, YIQ, XYZ, LAB, and LUV color spaces and final segmentation map (into  $K_2 = 6$  classes) resulting of the fusion of these six clusterings (bottom right) (see Table I for an objective and quantitative comparison).

### III. FUSION OF SEGMENTATION MAPS

The key idea of the proposed fusion procedure simply consists of considering, for each site (or pixel to be classified), the local histogram of the class (or *texton*) labels of each segmentation to be fused, computed on a squared fixed-size ( $N_w$ ) neighborhood centered around the pixel, as input feature vector of a final clustering procedure. For a fusion of  $N_s$  segmentation with  $K_1$  classes into a segmentation with  $K_2$  classes, the preliminary feature extraction step of this fusion procedure thus yields to  $N_s$  ( $K_1$ -bin) histograms which are then gathered together in order to form, a  $K_1 \times N_s$ -dimensional feature vector or a final ( $K_1 \times N_s$ )-bin histogram which is then normalized to sum to one, so that it is also a probability distribution function. The proposed

fusion procedure is then herein simply considered as a problem of clustering local histograms of (preliminary estimated) class labels computed around and associated to each site. To this end, we use, once again, a  $K$ -means clustering procedure exploiting, for this fusion step, an histogram-based similarity measure derived from the Bhattacharya similarity coefficient. Given a normalized histogram  $\{h(n; \mathbf{x})\}_{n=0, \dots, N_b-1=K_1 N_s-1}$  (at pixel location  $\mathbf{x}$ ) and a reference histogram  $\{h^*(n)\}_{n=0, \dots, N_b-1}$  (representing one of the  $K_2$  cluster centers of each class of a  $K$ -means procedure), the Bhattacharya distance between these two histograms is defined as

$$D_B[h^*, h(\mathbf{x})] = \left( 1 - \sum_{n=0}^{N_b-1} \sqrt{h^*(n)h(n; \mathbf{x})} \right)^{1/2} \quad (2)$$

and a  $K$ -means algorithm based on this distance converges in all tested examples.

The preestimated label fields to be fused (see Section II), along with the fusion procedure can be viewed (and qualitatively explained) as a two-step hierarchical segmentation procedure in which, first, a texton segmentation map (in each color space) is estimated and, second, a final clustering, taking into account this mixture of textons (expressed in the set of color space  $\mathcal{C}$ ), is then used for a final clustering. We recall that a texton, in our framework, is defined by a nonparametric mixture of colors (see Section II).

Consequently, in this final clustering (the fusion procedure), two sites for which the local-class-label histogram (i.e., the mixture of textons in the different color spaces given by the  $K_1 \times N_s$  bins histogram) are not too far away from each other will be in the same class in the resulting fused segmentation. Inversely, two sites associated to different local-class-label histograms will likely belong to different classes in the final segmentation.

Fig. 2 shows an example of the clustering segmentation model presented in Section II (into  $K_1 = 6$  classes) of an input image expressed in the RGB, HSV, YIQ, XYZ, LAB, and LUV color spaces and the final segmentation map (into  $K_2 = 6$  classes) which results of the fusion of these  $N_s$  clusterings. We can notice that none of them can be considered as reliable except the final segmentation result (at bottom right) which visually identify quite faithfully the different objects of the scene.

A final merging step is necessary and is used to avoid over-segmentation for some images. It consists of fusing each region (i.e., set of connected pixels belonging to the same class) of the resulting segmentation map with one of its neighboring region  $\mathcal{R}$  if the distance  $\mathcal{D}_{\text{MERGING}}$  is below a given threshold  $\kappa$  (or if its size is below 50 pixels with the closest region in the  $\mathcal{D}_{\text{MERGING}}$  distance sense)

$$\mathcal{D}_{\text{MERGING}} = \min_{\mathbf{x} \in \mathcal{R}} \left\{ \sum_{\mathcal{C}} D_{\mathcal{B}}[h^{\circ}(n), h^{\ddagger}(n; \mathbf{x})] \right\}. \quad (3)$$

In (3), the first summation is done on the six used color spaces, and  $h^{\circ}(n)$  designates the normalized nonparametric histogram of the set of pixels belonging to the region to be merged and  $h^{\ddagger}(n; \mathbf{x})$  is the normalized histogram, computed on a squared fixed-size ( $N_w$ ) windows (at pixel location  $\mathbf{x}$  and totally included in the region). For this merging procedure, the two histograms are equally re-quantized (for each of the three color channels) with  $4 \times 4 \times 4$  bins (see Fig. 3 where this merging strategy is, for this example, intensively used).

## IV. EXPERIMENTAL RESULTS

### A. Set Up

In all the experiments, we have considered our fusion methods on initial segmentations obtained with the following parameters: the size of the squared window, used to compute the local histogram for the initial segmentations or the fusion procedure is set to  $N_w = 7 \times 7$ . The number of bins for each local re-quantized histogram is set to  $N_b = 5 \times 5 \times 5$ . We use  $N_s = 6$  segmentations provided by the following color

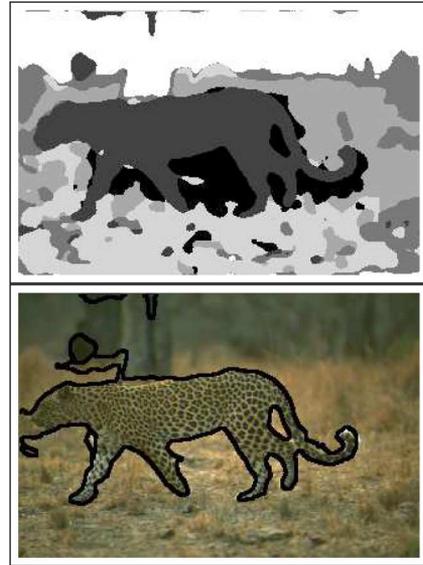


Fig. 3. Example of final merging step using the Bhattacharya distance on different color spaces as merging criterion on a fused segmented image of the Berkeley database.

spaces RGB, HSV, YIQ, XYZ, LAB, and LUV. Several quantitative performance measures will be given for several values (comprised between 6 and 13) of  $K_1$  and  $K_2$ , respectively, the number of classes of the segmentation to be fused and the resulting number of classes of the final fused segmentation map. The optimal value of  $\kappa$  seems to be comprised between 0.10 and 0.15.

We have used the Matlab source code, proposed by A. Y. Yang in order to estimate the quantitative performance measures (i.e., the four image segmentation indices) presented in Section IV-B. This code is kindly available on-line at address [http://www.eecs.berkeley.edu/~yang/software/lossy\\_segmentation/](http://www.eecs.berkeley.edu/~yang/software/lossy_segmentation/).

### B. Comparison With State-of-the-Art Methods

We have replicated the scenario used in the evaluation of state-of-the-art segmentation methods described in [23] and [25]. In these experiments, we have to test our segmentation algorithm on the Berkeley segmentation database [21] consisting of 300 color images of size  $481 \times 321$ . For each color image, a set of benchmark segmentation results, provided by human observers (between 4 and 7), is available and will be used to quantify the reliability of the proposed segmentation algorithm. As proposed in [23]–[25], we have compared our segmentation algorithm (called FCR for fusion of clustering results) against four unsupervised algorithms, available publicly. For each of these algorithms, their internal parameters are set to their optimal value (see [23]) and/or corresponds to the internal values suggested by the authors. These algorithms are namely the mean-shift [5] (with  $h_s = 13$ ,  $h_r = 19$ ), Ncuts [6] (with a number of segments  $K = 20$ , agreeing with the average number of regions found in the segmentation maps given by the human observers [25]), and FH [22] (with a smoothing parameter  $\sigma = 0.5$ , a threshold value  $k = 500$  and a minimal region size equals to 200 pixels), and, finally, the CTM algorithm proposed in [23] and [24] (with  $\eta = 0.1$  and  $\eta = 0.2$ ).

As in [23] and [25], all color images are normalized to have the longest side equals to 320 pixels (in this paper, this operation was done by the Linux command *convert* which is a member of the ImageMagick suite of tools). The comparison is based on the following performance measures, namely a probabilistic measure called PRI (higher probability is better) and three metrics VoI, GCE, and BDE (lower distance is better). The qualitative meaning of these performance measures are recalled as follows.

- 1) The Rand index [19] counts the fraction of pairs of pixels whose labellings are consistent between the computed segmentation and the ground truth. This quantitative measure is easily extended to the probabilistic Rand index (PRI) [26] by averaging the result across all human segmentations of a given image.
- 2) Contrary to the PRI, based on pairwise relationships, the variation of information (VoI) metric [20] is based on relationship between a point and its cluster. It uses mutual information metric and entropy to approximate the distance between two clusterings across the lattice of possible clusterings. More precisely, it measures the amount of information that is lost or gained in changing from one clustering to another (and, thus, can be viewed as representing the amount of randomness in one segmentation which cannot be explained by the other).
- 3) The global consistency measure (GCE) [21] measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. For a perfect match (in this metric sense), every region in one of the segmentations must be identical to, or a refinement (i.e., a subset) of, a region in the other segmentation. Segmentation which are related in this manner are considered to be consistent, since they could represent the same natural image segmented at different levels of detail (as the segmented images produced by several human observers for which a finer level of detail will merge in such a way that they yield the larger regions proposed by a different observer at a coarser level).
- 4) The boundary displacement error (BDE) [22] measures the average displacement error of one boundary pixels and the closest boundary pixels in the other segmentation.

As noticed in [23], PRI seems to be more highly correlated with human hand segmentations. Let us also mention that a inherent problem with the GCE measure is that it does not penalize oversegmentation at all (the highest score is given by assigning each pixel to an individual region). Some of these interesting performance measures thus have degenerate cases (i.e., unrealistic bad segmentations give abnormally high score), these complementary measures have thus to be considered all together in order to quantify the performance of a given segmentation method.

Table I shows the obtained results for the images presented in Fig. 2. Table II shows the obtained results for different values of  $K_1$  and  $K_2$  and  $\kappa$ . Fig. 4 shows the distribution of the different performance measure over the 300 images of the Berkeley image database for  $K_1 = 13$ ,  $K_2 = 6$ ,  $\kappa = 0.135$ . We can notice that the discussed fusion strategy allows to give very competitive results among these four different quantitative performance measures with a relative low variance over the set of images of the Berkeley image database. Fig. 6 displays some

TABLE I  
PERFORMANCE MEASURES FOR, RESPECTIVELY, THE CLUSTERING RESULT EXPRESSED IN EACH COLOR SPACE AND THE FUSION RESULT GIVEN BY OUR ALGORITHM (HIGHER IS BETTER FOR PRI AND LOWER IS BETTER FOR VoI, GCE, AND BDE) ON THE TWO IMAGES PRESENTED IN FIG. 2

	PERFORMANCE MEASURES			
	PRI	VoI	GCE	BDE
HUMANS	0.8754	1.1040	0.0797	4.994
<b>Image (a)</b>				
RGB	0.80510	2.8997	0.36994	<b>5.5826</b>
HSV	0.79946	2.8670	0.42518	5.8838
YIQ	0.87188	2.9994	<b>0.23586</b>	5.6613
XYZ	0.80458	3.3829	0.37047	5.9814
LAB	0.87949	2.9628	0.25179	6.1200
LUV	0.82565	3.2305	0.39333	5.9997
<b>FUSION</b>	<b>0.88711</b>	<b>2.3962</b>	0.26709	5.6369
<b>Image (b)</b>				
RGB	0.80910	2.1282	0.17975	<b>9.3065</b>
HSV	0.79055	2.2627	0.26257	11.1331
YIQ	0.80683	2.0082	0.25385	11.5585
XYZ	0.79904	2.4031	0.16182	9.8310
LAB	0.74751	3.0254	0.33556	10.8777
LUV	0.80549	2.4849	0.22742	9.3800
<b>FUSION</b>	<b>0.83375</b>	<b>1.6808</b>	<b>0.13678</b>	10.2006

TABLE II  
AVERAGE PERFORMANCE OF OUR ALGORITHM FOR SEVERAL VALUES OF ITS INTERNAL PARAMETERS (PARAMETER VECTOR  $[K_1|K_2|\kappa]$  INDICATED FOR EACH EXPERIMENT) AND FOR DIFFERENT PERFORMANCE MEASURES (HIGHER IS BETTER FOR PRI AND LOWER IS BETTER FOR VoI, GCE AND BDE) ON THE BERKELEY IMAGE DATABASE (<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>)

ALGORITHMS	PERFORMANCE MEASURES			
	PRI [19]	VoI [20]	GCE [21]	BDE [22]
<b>Humans</b>	0.8754	1.1040	0.0797	4.994
FCR $_{[K_1=13 K_2=6 \kappa=0.135]}$	<b>0.7882</b>	2.3035	0.2114	8.9951
FCR $_{[K_1=6 K_2=6 \kappa=0.130]}$	0.7842	2.3925	0.2169	9.2463
FCR $_{[K_1=13 K_2=13 \kappa=0.145]}$	0.7849	2.5494	<b>0.1752</b>	<b>8.7754</b>
FCR $_{[K_1=9 K_2=6 \kappa=0.140]}$	0.7835	2.2990	0.2157	9.2627
FCR $_{[K_1=12 K_2=4 \kappa=0.125]}$	0.7789	2.2071	0.2405	9.5758
CTM $_{\eta=0.1}$ [23][24]	0.7561	2.4640	0.1767	9.4211
CTM $_{\eta=0.2}$ [23][24]	0.7617	<b>2.0236</b>	0.1877	9.8962
Mean-Shift [5]	0.7550	2.477	0.2598	9.7001
NCuts [6]	0.7229	2.9329	0.2182	9.6038
FH [7]	0.7841	2.6647	0.1895	9.9497

examples of segmentations obtained by our algorithm. The results for the entire database are available online at <http://www.iro.umontreal.ca/~mignotte/ResearchMaterial>. In short, the proposed algorithm outperforms for several different internal parameters, all the well-known segmentation algorithms presented in Table II in terms of PRI and BDE indices. In term of the VoI

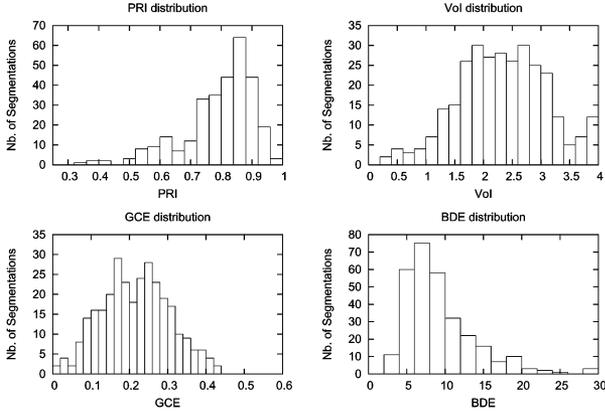


Fig. 4. Distribution of the difference performance measures, respectively from top to bottom; PRI, Vol, GCE, BDE over the 300 images of the Berkeley database for  $\text{FCR}_{[K_1=13|K_2=6|\kappa=0.135]}$ .

TABLE III  
INFLUENCE OF THE DISTANCE CHOICE USED IN THE FINAL FUSION PROCEDURE  
(AVERAGE PERFORMANCE ON THE BERKELEY IMAGE DATABASE)

$\text{FCR}_{[K_1=12 K_2=4 \kappa=0]}$ DISTANCES	PERFORMANCE MEASURES			
	PRI	Vol	GCE	BDE
<b>Bhattacharya</b>	<b>0.7613</b>	<b>2.440</b>	<b>0.2424</b>	<b>10.167</b>
Euclidean	0.7228	2.599	0.2702	11.562
Manhattan	<b>0.7620</b>	2.468	<b>0.2422</b>	10.197
Chord	0.7604	2.441	0.242	10.207
Kolmogorov	0.7388	2.665	0.2619	11.044
Histogram intersect.	0.7619	2.471	0.2423	10.225
Kullback	0.7508	2.535	0.2504	10.533
Shannon-Jensen	0.6966	3.058	0.3037	11.543

indices, only the CTM algorithm performs equivalently or better and for the GCE measure, our algorithm gives, on average, similar results than the others and outperforms them all for a set of parameters (in which  $K_2$  is high, leading to a classical over segmentation).

### C. Sensitivity to Parameters

$\kappa$  allows to refine the final segmentation map and allows, to a certain extent, to avoid some over-segmented partition maps results (especially when  $K_2$  is high). With  $\kappa = 0$  (i.e., without the final fusion step and the other parameters being set to  $K_1 = 12$  and  $K_2 = 4$ ), we obtain  $\text{PRI} = 0.7613$  (to be compared to  $\text{PRI} = 0.7789$  when  $\kappa = 0.125$ ; see Table II).

In order now to quantify the influence of the distance choice, used in the final clustering used as fusion procedure, we have compared the performance measures obtained with our method ( $K_1 = 12$ ,  $K_2 = 4$  and  $\kappa = 0$ ) using a Bhattacharya distance and several other metrics [27] (see Table III). We can notice that several distances are as efficient as the Bhattacharya distance, for example, the Manhattan distance ( $L_1$  norm) or the histogram intersection-based distance. For the same parameters given in Table II ( $\kappa \neq 0$ ), these two metrics allow to obtain similar performance measures comparatively to the Bhattacharya distance.

TABLE IV  
INFLUENCE OF THE SIZE OF THE WINDOW  $N_w$  USED TO ESTIMATE THE LOCAL HISTOGRAMS (AVERAGE PERFORMANCE ON THE BERKELEY IMAGE DATABASE)

$\text{FCR}_{[K_1=12 K_2=4 \kappa=0]}$ SIZES	PERFORMANCE MEASURES			
	PRI	Vol	GCE	BDE
<b><math>7 \times 7</math></b>	<b>0.7613</b>	<b>2.440</b>	<b>0.2424</b>	<b>10.167</b>
$5 \times 5$	0.7621	2.506	0.2349	9.821
$9 \times 9$	0.7604	2.401	0.2515	10.552

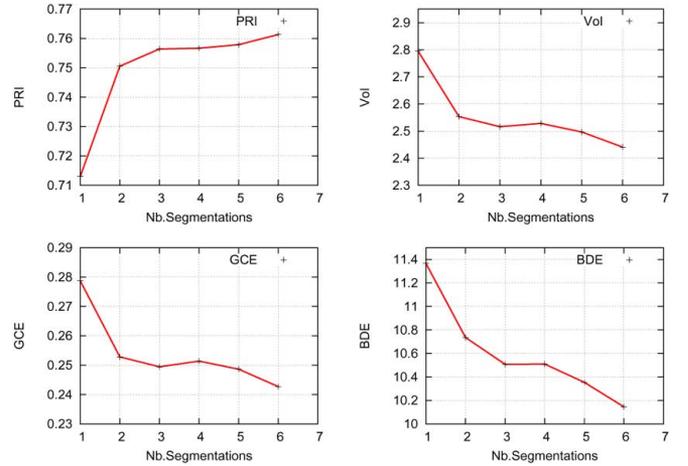


Fig. 5. Evolution of the PRI, Vol, GCE, and BDE measures as a function of the number of segmentations ( $N_s$ ) to be fused for the  $\text{FCR}_{[K_1=12|K_2=4|\kappa=0]}$  algorithm. For  $N_s = 1$ , i.e., without fusion, the segmentation model is the one described in Section II with  $K_1 = K = 4$ .

We have also quantified in Table IV the influence of the size of the window (used to estimate the local color histogram). These tests show that the performance measures are not too much sensitive to this internal parameter.

We can also notice (see Fig. 5) that all the performance measures are all the more better than  $N_s$  (number of segmentation to be fused) is high. This experiment shows the validity of our fusion procedure and also the performance measures obtained by the simple segmentation model presented in Section II.

### D. Discussion

Tests have shown that a higher value for  $k_2$  ( $k_1 \geq k_2 < 15$ ) will induce a lower (consequently better) GCE and BDE performance measure but also a higher (consequently less good) Vol measure. The PRI measure is quite influenced by the value of  $\kappa$  whose optimal value seems to be comprised between 0.10 and 0.15. The fusion method is not too much sensitive to the value of  $N_w$  (size of the window used to estimate the local histograms), and, finally, the performance measures are all the more better than  $N_s$  (number of segmentations to be fused) is high.

The segmentation procedure takes less than one minute for an AMD Athlon 64 Processor 3500+, 2.2 GHz, 4435.67 bogomips and running on Linux. The source code (in C++ language) of our algorithm is available at the following address <http://www.iro.umontreal.ca/~mignotte/ResearchMaterial>.

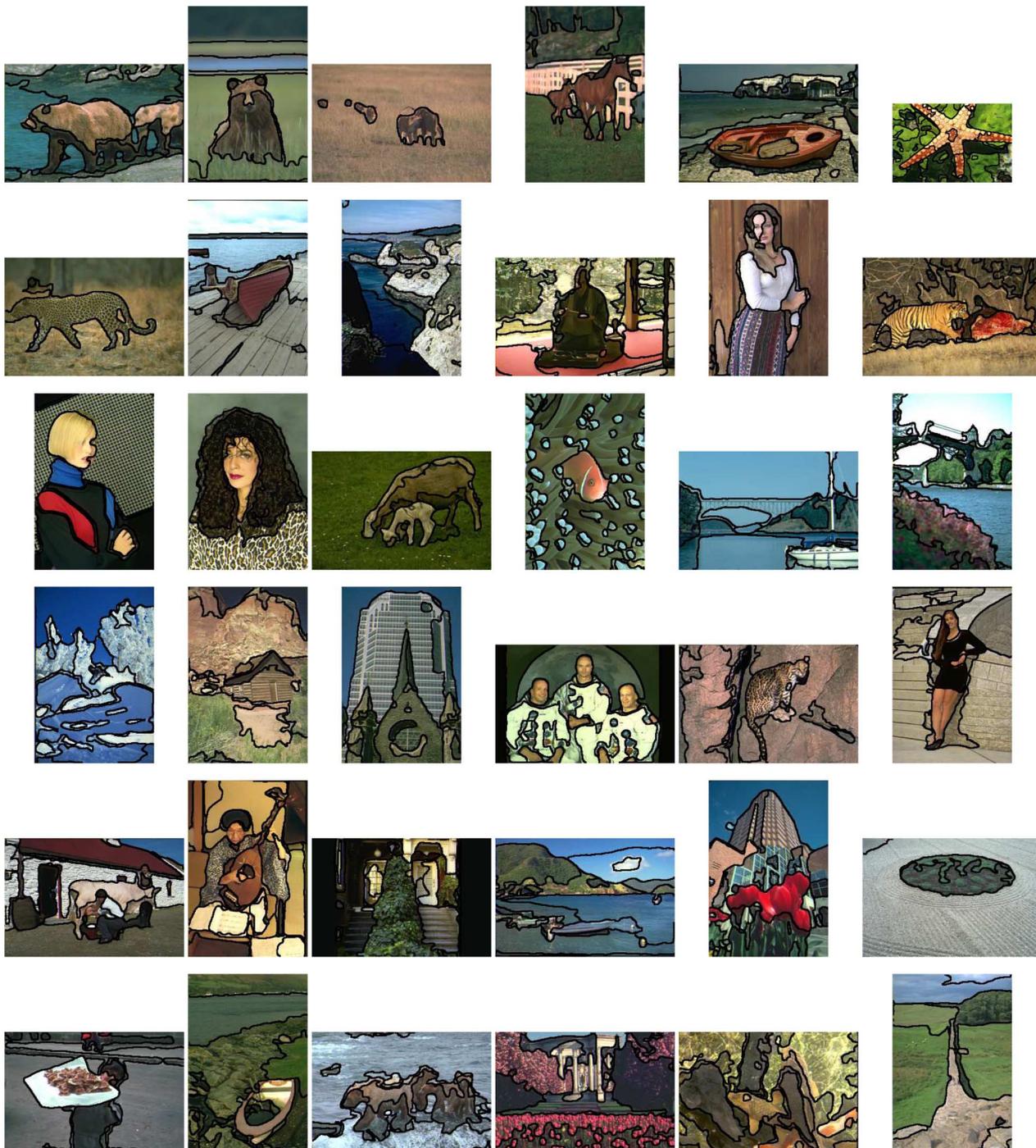


Fig. 6. Example of segmentations obtained by our algorithm  $\text{FCR}_{[K_1=13|K_2=6]|\kappa=0.135}$  on 24 images of the Berkeley image database (see also Table II and Fig. 4 for quantitative performance measures and <http://www.iro.umontreal.ca/~mignotte/ResearchMaterial> for the segmentation results on the entire database).

## V. CONCLUSION

In this paper, we have presented a new segmentation strategy based on a fusion procedure whose goal is to combine several segmentation maps in order to finally get a more reliable and accurate segmentation result. The initial segmentations to be fused can be the output result of the same initial and simple model used on an input image filtered by a given filter bank, or it can also be provided by different segmentation models or different segmentation results provided by different seeds (or different

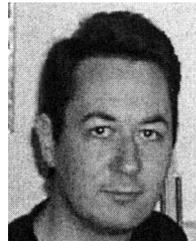
variation of parameters) of the same stochastic segmentation model. This fusion framework remains simple, fast, easily parallelizable, general enough to be applied to various computer vision applications, and performs competitively among the recently reported state-of-the-art segmentation methods.

## ACKNOWLEDGMENT

The author would like to thank the reviewers for their many valuable comments and suggestions that helped to improve both the technical content and the presentation quality of this paper.

## REFERENCES

- [1] S. Banks, *Signal Processing, Image Processing and Pattern Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
- [2] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–136, Mar. 1982.
- [3] P. Berkhin, "Survey of clustering data mining techniques," *Accrue Software*, San Jose, CA, 2002.
- [4] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.
- [5] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [7] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, pp. 167–181, 2004.
- [8] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.
- [9] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using a hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.
- [10] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Three-class Markovian segmentation of high resolution sonar images," *Comput. Vis. Image Understand.*, vol. 76, no. 3, pp. 191–204, 1999.
- [11] F. Destremes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov random field models and its Bayesian estimation," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2920–2935, Oct. 2006.
- [12] Z. Kato, T. C. Pong, and G. Q. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model," in *Proc. Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, pp. 961–964.
- [13] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proc. Eur. Conf. Computer Vision*, Copenhagen, Denmark, Jun. 2002, pp. 661–675.
- [14] J. B. Martinkauppi, M. N. Soriano, and M. H. Laaksonen, "Behavior of skin color under varying illumination seen by different cameras at different color spaces," in *Proc. SPIE, Machine Vision Applications in Industrial Inspection IX*, San Jose, CA, Jan. 2001, pp. 102–113.
- [15] J.-P. Braquelaire and L. Brun, "Comparison and optimization of methods of color image quantization," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 1048–1052, Jul. 1997.
- [16] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 371–381, Mar. 2007.
- [17] Z. Kato, "A Markov random field image segmentation model for color textured images," *Image Vis. Comput.*, vol. 24, no. 10, pp. 1103–1114, 2006.
- [18] E. Maggio and A. Cavallaro, "Multi-part target representation for color tracking," in *Proc. Int. Conf. Image Processing*, Italy, Genova, Sep. 2005, pp. 729–732.
- [19] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop on Empirical Evaluation Methods in Computer Vision*, Jun. 2005, vol. 3, pp. 34–41.
- [20] M. Meila, "Comparing clusterings—An axiomatic view," in *Proc. 22nd Int. Conf. Machine Learning*, 2005, pp. 577–584.
- [21] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.
- [22] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," in *Proc. 7th Eur. Conf. Computer Vision Part III*, Copenhagen, Denmark, May 2002, pp. 408–422, LNCS.
- [23] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, 2007, submitted for publication.
- [24] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.
- [25] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Elect. Eng. Comput. Sci. Dept. Univ. California, Berkeley*, 2006 [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-195.html>
- [26] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.
- [27] S.-H. Cha and S. N. Srihari, "On measuring the distance between histograms," *Pattern Recognit.*, vol. 35, pp. 1355–1370, 2002.



**Max Mignotte** received the D.E.A. (postgraduate degree) in digital signal, image, and speech processing from the INPG University, Grenoble, France, in 1993 and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidentale (UBO) and the Digital Signal Laboratory (GTS) of the French Naval Academy, in 1998.

He was an INRIA Postdoctoral Fellow at the University of Montréal (DIRO), QC, Canada, from 1998 to 1999. He is currently with DIRO at the Computer Vision and Geometric Modeling Lab as an Assistant Professor (Professeur Adjoint). He is also a member of the Laboratoire de recherche en Imagerie et Orthopédie (LIO), Centre de Recherche du CHUM, Hôpital Notre-Dame) and a researcher at CHUM. His current research interests include statistical methods and Bayesian inference for image segmentation (with hierarchical Markovian, statistical templates, or active contour models), hierarchical models for high-dimensional inverse problems from early vision, parameters estimation, tracking, classification, shape recognition, deconvolution, 3-D reconstruction and restoration problems.