

Segmentation of non-natural objects in landscape images using ridgelet transform

Langis Gagnon
Samuel Foucher

Computer Research Institute of Montreal (CRIM)
Research & Development Department
550 Sherbrooke Street West, Suite 100
Montreal, QC, H3A1B9
Canada
E-mail: langis.gagnon@crim.ca

Abstract. *This study reports about the detection of non-natural structures in outdoor natural scenes. In particular, we present a new approach based on ridgelet transform for the segmentation of man-made objects in landscape scenes. Multiscale directional moments of ridgelet coefficients are used as features along with a principal component analysis (PCA) followed by a linear discriminant analysis (LDA), kernel-based LDA (KLDA), or support vector classifier (SVC). The statistical learning is done on about 3,000 image patches that represent natural and artificial content. Performances are measured in terms of image patch type classification (natural versus non-natural) and man-made object segmentation on two different image test sets. Results using ridgelets are compared to Gabor features. Altogether, we compare performance for six different feature/classifier combinations: ridgelets+LDA, ridgelet+KLDA, ridgelets+SVC, Gabor+LDA, Gabor+KLDA, and Gabor+SVC, and various external parameter values. Results show that most of the time, the combinations with ridgelets provide comparable or better performance. © 2009 SPIE and IS&T. [DOI: 10.1117/1.3134137]*

1 Introduction

The goal of this paper is to present a new ridgelet-based procedure for the detection of non-natural (man-made) structures in images of outdoor natural scenes and to report comparative results. The comparison is done with respect to the use of local ridgelet transform (RIT),^{1,2} Gabor transform (GT), and linear discriminant analysis (LDA), kernel-based LDA (KLDA), or a support vector classifier (SVC).

There were two motivations behind this work: (1) a need expressed to us by a local academic organization in landscape and environmental design for a system that could automatically identify natural outdoor scenes with non-natural structures, and (2) a desire to contribute to the development of semantically meaningful segmentation tools for object-based image retrieval (OBIR). The identification of image regions that correspond to objects is a key issue to resolving the semantic gap in OBIR (i.e., the disparity between the content of retrieved information and the user's interpretation of that information³). This is one of the major topics in the development of human-centered OBIR systems.^{4,5}

Paper 08017RR received Jan. 31, 2008; revised manuscript received Mar. 3, 2009; accepted for publication Apr. 6, 2009; published online May 19, 2009.

1017-9909/2009/18(2)/023010/11/\$25.00 © 2009 SPIE and IS&T.

The problem of segmenting non-natural objects in landscape scenes is difficult because natural environments contain a large amount of clutter. Our approach was inspired by a recent work on scene categorization based on the use of GT and spectral shapes.⁶ From these shapes, generic image categories of natural scenes, like “naturalness” or “openness,” can be identified. In a recent paper, we proposed to use the RIT to overcome some drawbacks of the GT when applied to discontinuities along a curve.⁷ We did this in conjunction with a discriminant analysis (DA) for learning spectral shapes. In addition, our scheme was designed to apply to images of nondyadic dimension. A similar idea was reported later by other authors but based on SVC.⁸

The main reason for using ridgelets, and more generally multiscale harmonic analysis like curvelets⁹ and contourlets,¹⁰ is that these approaches provide a more efficient representation of natural images. This goes back to the work of Donoho, who has shown that ridgelet coefficients are highly sensitive to the geometric content of the image (i.e., the presence of straight lines).¹¹ In fact, ridgelets provide a redundant and sparse representation that is more efficient than 2-D wavelet transform (WT) for representing linear singularities.¹¹ In addition, the sparseness of the RIT coefficients changes significantly according to the geometric content of the image (i.e., the presence of straight or curved lines). Other advantages of the RIT include (1) fast runtime due to the intensive use of fast Fourier transform (FFT), (2) simple and fast inverse transform, and (3) fine directional resolution.

Recently, a few publications have directly addressed the problem of man-made object detection in natural images.^{12–14} In Ref. 12, the authors used a multiscale graphical model, known as a tree-structured belief network (TSBN) or multiscale random field (MSRF), to detect man-made structures. This approach was extended in Ref. 13 using dynamic TSBN (DTSBN) to overcome the “blocky” segmentation of the TSBN approach that results from the fixed structure of the nodes. In Ref. 14, a more classical approach was proposed that uses texture descriptors and anisotropic diffusion. The algorithm combines hierarchical splitting, agglomerative merging, and pixelwise classification to obtain rough man-made object segmentation. However, our approach does not build on any of these works.

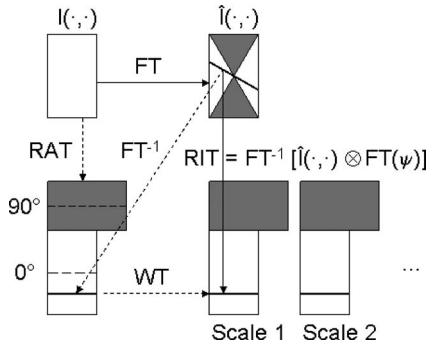


Fig. 1 Schematic description of RAT and RIT. Bold arrows show our numerical implementation. Dotted arrows show the conceptual link between RAT and RIT. All dimension proportions are respected. Since the image is rectangular, there are more quasi-vertical than quasi-horizontal directions; thus, the gray rectangles (quasi-horizontal directions) have a lower height than the white rectangles. See text for more details.

Instead, it is based on our previous work that demonstrated how global outdoor scene classification, using local RIT and DA, could allow localization of artificial objects in natural scenes.⁷ A similar work using RIT and SVC was reported recently but with the important assumption that buildings are known to be present *a priori*.¹⁵ We do not make this assumption in our work.

This paper is organized as follows. Section 2 gives an overview of the continuous RIT and the discrete implementation we reported in Ref. 7. Section 3 describes the methodology and databases used for training and testing. Finally, Sec. 4 presents performance results in terms of detection rate (DR), false alarm rate (FAR), and receiver operating characteristic (ROC) curves.

2 Theoretical Background

This section briefly reviews the theoretical background about RIT and the practical discrete implementation scheme we proposed in Ref. 7. We refer the reader to seminal works for more details on RIT.^{1,2}

The continuous RIT of an image $I(x_1, x_2)$ is the projection of I onto a standard wavelet basis oriented along the lines $x_1 \cos \theta + x_2 \sin \theta = b$. That is,

$$\text{RIT}_I(a, b, \theta) = \iint \psi_{a,b,\theta}(x_1, x_2) I(x_1, x_2) dx_1 dx_2, \quad (1)$$

where

$$\psi_{a,b,\theta}(x_1, x_2) = a^{-1/2} \psi[(x_1 \cos \theta + x_2 \sin \theta - b)/a]$$

$$a > 0, \quad b \in \mathbf{R}, \quad \theta \in [0, 2\pi]. \quad (2)$$

The RIT is similar to a 2-D WT except that the localization parameters are not the point coordinates but rather the line parameters (b, θ) . Lines and points in a plane can be related through a radon transform (RAT),

$$\text{RAT}_I(\theta, t) = \iint I(x_1, x_2) \delta(x_1 \cos \theta + x_2 \sin \theta - t) dx_1 dx_2, \quad (3)$$

So RIT, represented by Eq. (1), can be reformulated in terms of RAT as follows:

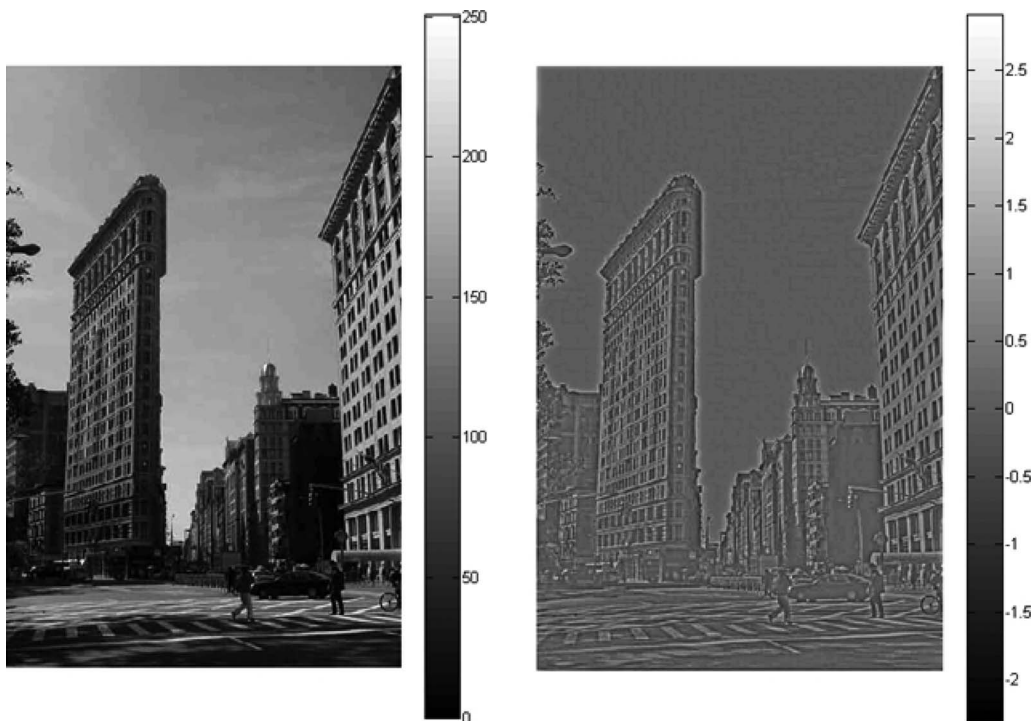


Fig. 2 Effect of the preprocessing filter.

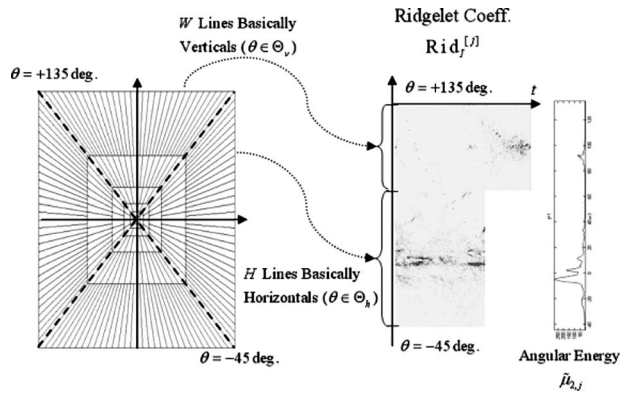


Fig. 3 Schematic representation of the feature vector extraction process.

$$RIT_I(a, b, \theta) = \int RAT_I(\theta, t) \psi[(t - b)/a] dt. \quad (4)$$

Equation (4) simply states that the RIT is the 1-D WT of a RAT.²

For efficient numerical implementation, one expresses Eq. (4) in terms of the Fourier transform (FT) via the Fourier projection-slice theorem. The FT of a RAT can be written as

$$FT[RAT_I(\theta, t)] = \int RAT_I(\theta, t) e^{-\lambda i t} dt = \hat{I}(\lambda \cos \theta, \lambda \sin \theta), \quad (5)$$

where the “hat” denotes the 2-D FT of the image I . Equation (5) states that the RAT for a given orientation can be derived from the inverse 1-D FT of the radial lines of \hat{I} . Thus, Eq. (5) can be rewritten in the frequency domain as

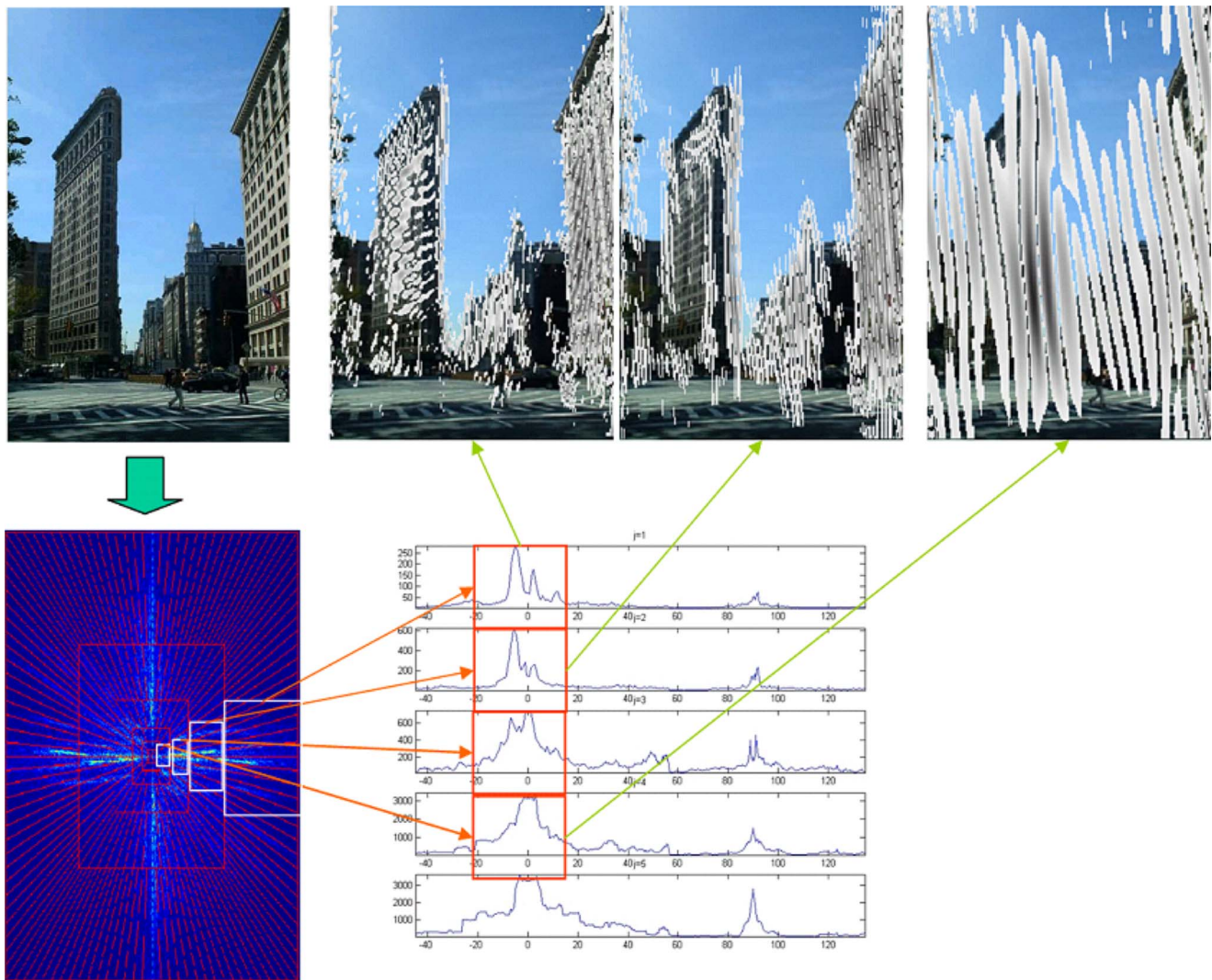


Fig. 4 Example of relation between the energy of the ridgelet coefficients and the geometrical contents of an image. White boxes on the Fourier spectrum (white boxes in bottom part) are regions that generate strong ridgelet coefficients at various scales, corresponding to strong geometric features within the image. The curves show the energy at various angles and scales. The energy maxima correspond to elements in the image representing strong geometrical content.

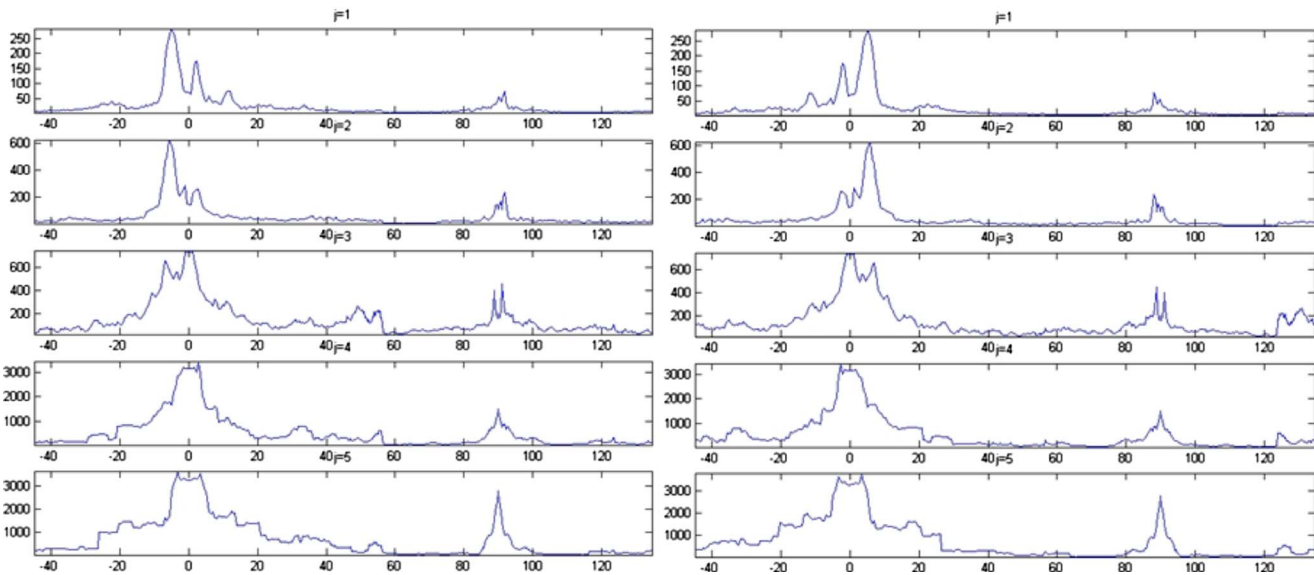


Fig. 5 Effect of a horizontal flip on the angular distribution of the ridgelet coefficients energy.

$$\text{RIT}_f(a, b, \theta) = \text{FT}^{-1}(\hat{I}(\lambda \cos \theta, \lambda \sin \theta) \text{FT}\{\psi[(t-b)/a]\}). \quad (6)$$

For a discrete numerical implementation of Eq. (6), the rectangular grid of the spatial frequency plane must be mapped onto a polar grid. We use the simple approach proposed in Ref. 16 that maps a rectangular grid to a recto-polar grid using a nearest-neighbor assignment. To this aim, the radial lines are divided into two categories: (1) the quasi-horizontal lines corresponding to the angular domain $\Theta_h = [-\pi/4, \pi/4]$, and (2) the quasi-vertical lines for $\Theta_v = [\pi/4, 3\pi/4]$. Each point of the recto-polar grid within the Θ_h (respectively, Θ_v) domain is then assigned the value of the nearest-neighbor point of the rectangular grid along the vertical (or horizontal) axis.

For rectangular images, the angular domains are dependent on the height (H) and width (W) of the image. That is,

$$\Theta_h = [-\text{atan}(H/W), \text{atan}(H/W)],$$

and

$$\Theta_v = [\pi/2 - \text{atan}(W/H), \pi/2 + \text{atan}(W/H)]. \quad (7)$$

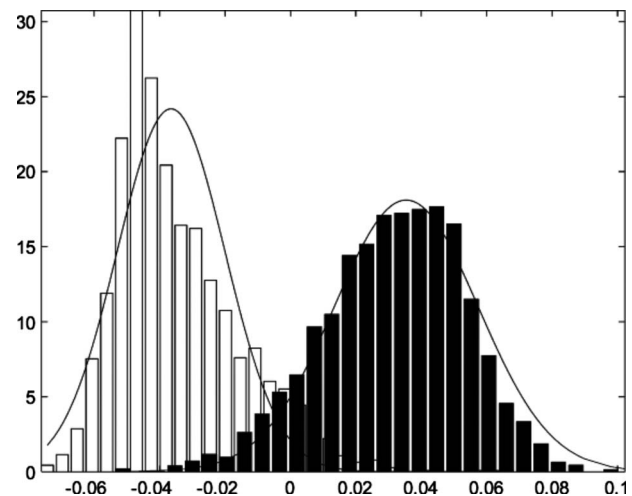


Fig. 6 Distribution of the discriminant function values on the training set.

Table 1 Performance on validation set for block classification.

Method	p	$\Delta\theta$ (deg)	DR (%)	FAR (%)
Ridgelets +sub-LDA	2	0.5	87.6	14.4
	2	1.0	90.2	13.0
	2	2.5	84.9	14.3
	4	1.0	90.0	13.3
Gabor +sub-LDA		7.5	86.7	19.2
Ridgelets +sub-KLDA	2	1.0	96.4	9.4
	4	1.0	95.6	9.1
Gabor +sub-KLDA		7.5	96.6	6.6
Ridgelets +sub-SVC	2	1.0	98.8	0.52
Gabor +sub-SVC		7.5	88.5	4.7

The RIT of an image can thus be decomposed into two images of dimensions $H \times W$ and $W \times H$, which can be noted by $\text{RIT}_I^{[j],h}(m,n)$ and $\text{RIT}_I^{[j],v}(m,n)$, respectively (Fig. 1). The parameter j is the scale index ($j=1,2,3,\dots,J$), and (m,n) are the coordinates of the RIT pixel. The index m labels the orientation angle of the RIT. The RIT thus generates a total of $2 \times W \times H$ coefficients for each scale j . Finally, we use the undecimated version of the 1-D WT to preserve the translation and maximize the RIT redundancy.

Figure 1 gives a schematic representation of the numerical implementation expression by Eq. (6). Given an image I (top left), the RIT (bottom right) is calculated following the clockwise path (arrows) rather than the counter-clockwise path (dashed arrow), which schematically represents Eq. (4). (The full arrow path is the most efficient because it is based on a minimum number of FTs. The dotted arrow path is a less efficient way because it necessitates the calculation of a RAT.) At each scale, two transformed images are cal-

culated for each interval expressed in Eq. (7). A radial line in the FT image \hat{I} (bold line in top-right rectangle) is mapped into a horizontal line in the RIT images. The two horizontal dashed rows in the bottom-left rectangles correspond to the radial angles 90 deg and 0 deg, respectively.

3 Methodology

3.1 Preprocessing

The grayscale version of an image is first filtered to locally enhance the high-frequency information using a technique proposed in Ref. 17 (Fig. 2). The filter response is given by

$$I' = \frac{I * HP}{\varepsilon + \sqrt{(I * HP)^2 * LP}}, \quad (8)$$

where I is the image patch, LP is a low-pass Gaussian filter, and HP is the complementary high-pass filter ($HP=1-LP$). The parameter ε (set to 20 in our implementation) prevents singularity in perfectly homogeneous regions. The filter enhances textural information in low-contrast areas (e.g., shadows) and limits the variability of lighting conditions.

3.2 Feature Extraction

We use the ridgelet coefficients as features. First, we calculate the normalized directional energy and skewness of the ridgelet coefficients (i.e., the normalized directional moments of orders $p=2$ and 4) along various orientations (Fig. 3) and for each scale j . These are given by

$$\tilde{\mu}_{p,j}(m) = \begin{cases} \frac{1}{S_p^{[j]} H} \sum_n [\text{RIT}_I^{[j],h}(m,n)]^p, & m \in \Theta_h \\ \frac{1}{S_p^{[j]} W} \sum_n [\text{RIT}_I^{[j],v}(m,n)]^p, & m \in \Theta_v \end{cases}, \quad (9)$$

where $p=2,4$ and $S_p^{[j]}$ is the equivalent gain of the wavelet transform at scale j . The behavior of the ridgelet coefficients' energy as a function of the image content is illustrated in Fig. 4.

Second, as proposed in Ref. 6, we use a symmetry property of the directional moments to artificially double the size of the training set by duplicating all images and restricting the calculation of the feature vectors within the range $0 \leq \theta \leq 90$ deg. In fact, the effect of a horizontal mirror operation on the image induces a flip on the value of the

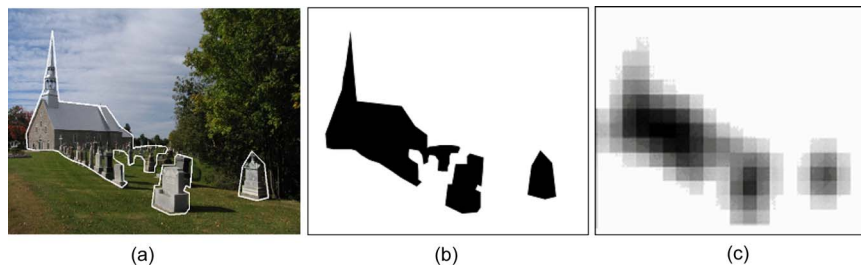


Fig. 7 Example of ground truth for (a) one of the test images, and (b) its corresponding pixel-level mask image. (c) Example of ground truth at the 32×32 block scale. The fraction of artificial pixels in each 32×32 block is an indication of the “non-naturalness” of the block, normalized between 0 (“natural”) and 255 (“non-natural”).

Table 2 Performance on test set I for block classification.

Method	p	$\Delta\theta$ (deg)	DR (%)			FAR (%)		
			$T=50\%$	$T=75\%$	$T=100\%$	$T=50\%$	$T=75\%$	$T=100\%$
Ridgelets +sub-LDA	2	0.5	58.0	67.1	77.6	3.0	4.7	6.8
	2	1.0	61.2	74.0	87.0	2.5	4.1	6.6
	2	2.5	55.8	65.4	81.6	2.8	4.4	6.5
	4	1.0	61.9	75.4	87.0	2.5	4.2	6.7
Gabor +sub-LDA		7.5	66.5	74.3	87.0	5.1	7.1	9.4
Ridgelets +sub-KLDA	2	1.0	69.4	82.0	91.5	3.1	5.0	7.8
	4	1.0	67.8	80.2	89.2	2.8	4.7	7.4
Gabor +sub-KLDA		7.5	64.2	73.3	82.1	3.8	5.7	8.0
Ridgelets +sub-SVC	2	1.0	69.8	82.3	90.6	4.4	6.4	9.1
Gabor +sub-SVC		7.5	58.4	68.9	79.8	2.5	4.2	6.5

directional moments of around 0 deg on the Θ_h domain and around 90 deg on Θ_v (illustrated in Fig. 5). Thus, the components of the feature vector in the range $0 \leq \theta \leq 90$ deg are the same as those of the mirror images in the range $(-45 \leq \theta \leq 0$ deg; $90 \leq \theta \leq 135$ deg).

Finally, since the domains Θ_h and Θ_v depend on the image size, it is necessary to resample the feature vectors at the same resolution angle $\Delta\theta$ in order to obtain the same vector dimension for all images. The final feature vector at scale j is thus of dimension $90 \text{ deg}/\Delta\theta$.

3.3 Feature Reduction

A principal component analysis (PCA) is applied to the feature vectors to compress the feature space. Only the components representing 95% of the total variance are kept. The RIT is performed with four levels ($J=4$) and an angular resolution $\Delta\theta=1$ deg. Once projected in the PCA subspace, the vector dimension is reduced from $360 (=4 \times 90/1)$ to 76. We found in our tests that the PCA step was useful because it removed noise from the feature vectors and prevented overfitting during training, making the classifier (the next step) more robust to slight feature variations.

3.4 Statistical Learning

Finally, statistics of the two-class problem are learned using a LDA or a KLDA with a Gaussian kernel.¹⁸ Both LDA and KLDA were used in our tests to compare classification performances (see Sec. 4). In the remainder of the paper, we use the terms subspace LDA (sub-LDA) and sub-KLDA to refer to the PCA projection.¹⁹

Figure 6 shows the distribution of the discriminant function values in the training set for the LDA classifier with $\Delta\theta=1$ deg, $p=2$, and $J=4$. Values around -0.04 characterize the patches of natural scenes while values around 0.04 describe non-natural structures (the fact that the absolute values are the same is a coincidence).

Once the classifier has been trained to discriminate natural and artificial image patches, it can be applied to outdoor images to locally characterize the “naturalness” of the visual content. We applied an RIT on the entire image over blocks of 64×64 pixels with an overlap of 32 pixels. These values were chosen because they provided a good tradeoff between spatial resolution, localization of the information, and processing time.

4 Dataset Description

4.1 Training Dataset

Our training database was composed of 6,009 image patches of 64×64 pixels with 3,078 elements manually labeled as natural and 2,931 as artificial. Half of this dataset was kept for training and the other half for validation, which served to confirm the optimal choice of parameters and the training convergence (Table 1). These patches were randomly extracted from a set of about 580 images acquired by a research partner to develop a scouting system for the film and tourist industries.²⁰ Images were acquired following a strict procedure (well framed, correct orientation and exposition, etc.). The image set was composed of various types of outdoor natural scenes (urban, suburban, landscape, seascape, etc.), with or without buildings or other man-made structures.

Table 3 Performance on test set II for block classification.

Method	p	$\Delta\theta$ (deg)	DR (%)			FAR (%)		
			$T=50\%$	$T=75\%$	$T=100\%$	$T=50\%$	$T=75\%$	$T=100\%$
Ridgelets +sub-LDA	2	0.5	69.3	76.1	74.3	10.2	12.9	15.7
	2	1.0	62.3	71.4	77.0	7.3	9.7	12.4
	2	2.5	70.4	77.8	77.0	10.0	12.8	15.7
	4	1.0	63.8	74.9	79.7	7.7	10.0	12.8
Gabor +sub-LDA		7.5	78.2	79.9	75.7	14.4	17.7	20.5
Ridgelets +sub-KLDA	2	1.0	73.0	80.7	82.4	6.6	9.7	12.8
	4	1.0	72.8	80.4	81.1	7.2	10.3	13.4
Gabor +sub-KLDA		7.5	69.6	74.5	73.0	9.2	12.1	14.9
Ridgelets +sub-SVC	2	1.0	71.6	78.0	77.0	9.4	12.5	15.6
Gabor +sub-SVC		7.5	65.0	70.6	68.9	6.5	9.5	12.3

4.2 Test Sets and Ground Truth

Our first test set (test set I) was composed of 83 images, of which about 50% come from the same source as the training set (but with no overlap with the training set). The second test set (test set II) was composed of 40 images taken from the Flickr website. The pictures were smaller (512 pixels in width) than those of test set I and had variable quality and a strong JPEG compression.

A ground truth was built by hand for the test sets to assess performance. Figure 7 gives an example of the way the ground truth was manually done on a sample image of the test set. The ground truth was used to generate a binary mask (Fig. 7) where each pixel of the test image was labeled as natural or artificial. However, since the classifier training was done on image patches of 64×64 pixels with a 32-pixel overlap (thus, at a 32×32 scale), the pixel-level mask had to be mapped to that coarser scale. We did this as follows. For each of the 64×64 blocks of the binary mask, we calculated the fraction of artificial pixels in it (the number of pixels belonging to the artificial object divided by 64^2). Second, we merged the fractions from overlapping blocks using a cosine-based weighting function (see Ref. 17 for details) to obtain a fractional value of “naturalness” at the 32×32 resolution. A block (rather than a pixel) was then labeled as artificial if the fraction was equal to or larger than a given threshold T . We set T to 50%, 75%, and 100% for the tests to get various object boundaries (a threshold of 100% would label all the blocks as natural except those that were 100% inside the original mask im-

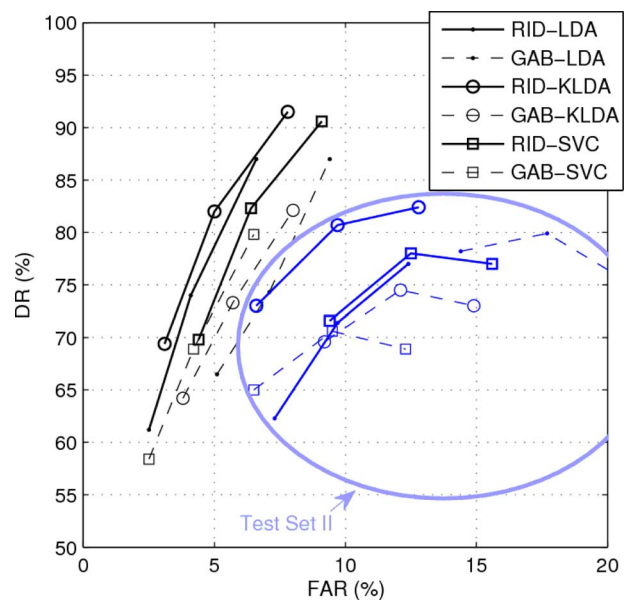


Fig. 8 Performances on the two test sets for the 32×32 block classification. The three markers joined by a single line represent values for three different degrees of naturalness ($T=50\%$, 75% , 100%). For the ridgelet features, only the results for $p=2$ are shown here.

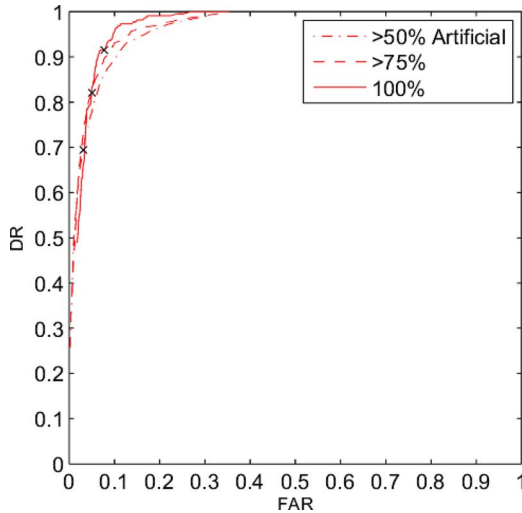


Fig. 9 ROC curves for $p=2$, $\Delta\theta=1$ and sub-KLDA classification for test set I. The “X” gives the points for a Bayesian decision threshold of 0.5.

age). Thus, each threshold T corresponded to a different ground truth image for which the non-natural object border slightly varied.

5 Results

In this section we present performance results of our approach and compare them with other standard feature/classifier combinations. We first give a quick indication of performance on the validation set, then on two different test sets. Our aim here is not to compare different classifiers, but rather the performance of the ridgelet features against Gabor ones for a given classifier.

5.1 Performance on Validation Set

Table 1 shows a typical classification performance obtained on the validation set for the identification of the image patches. Ridgelet parameters were set to $J=4$, $p=2$ and 4, and $\Delta\theta=0.5$ and 1.0. Gabor parameters were the same used in Refs. 6 and 17: 24 orientations and 5 dyadic frequency bands ranging from 0.02 to 0.32 cycles/image, resulting in a feature vectors dimension of 120. DR gives the percentage of “artificial” (non-natural) image patches correctly classified. FAR is the percentage of “natural” image patches labeled as “artificial.”

The ridgelets+sub-KLDA method provided good performances (DR=96.4 and FAR=9.4). Note that when the PCA projection was not done, we got a lower DR (92.0%) and a higher FAR (11.31%) for $p=2$ and $\Delta\theta=1$. This indicates the usefulness of the PCA projection. We also calculated performance with SVC using the same library as in Ref. 8 (see also Ref. 18 for accessing the public library). The two parameters required for SVC (the penalty parameter C and the kernel scale factor γ) were determined separately for the ridgelets and Gabor approaches by following the multigrad procedure described in Ref. 21 so that the RR value was maximized on the validation set. Performances on the validation set were all good, but this set served only to confirm the optimal choice of parameters and track the training convergence. The real performance measures are given in the following test sets.

5.2 Performance on Test Sets

For the test sets, we first calculated the detection performance in terms of image patch (block) classification (natural and man-made), and second, of the man-made object segmentation as a whole. Performances were also evaluated in terms of area under the ROC curve (AUC), because there

Table 4 AUC estimated for the test sets for block classification. n_a and n_n are the number of artificial and natural blocks, respectively. ΔAUC is the approximate confidence interval at the 95% confidence level.²⁴

T	Test Set I			Test Set II			Test Set I+II		
	50%	75%	100%	50%	75%	100%	50%	75%	100%
n_a	2122	1163	223	812	419	74	2934	1582	297
n_n	23577	24536	25476	6651	7044	7389	30228	31580	32865
ΔAUC	0.15	0.20	0.45	0.25	0.33	0.66	0.13	0.17	0.38
Method									
Ridgelets+sub-KLDA	0.963	0.968	0.971	0.926	0.931	0.906	0.955	0.960	0.959
Gabor+sub-KLDA	0.943	0.944	0.952	0.899	0.900	0.880	0.933	0.934	0.938
Ridgelets+sub-SVC	0.947	0.961	0.966	0.892	0.907	0.892	0.935	0.950	0.953
Gabor+sub-SVC	0.934	0.932	0.928	0.914	0.912	0.902	0.930	0.928	0.926

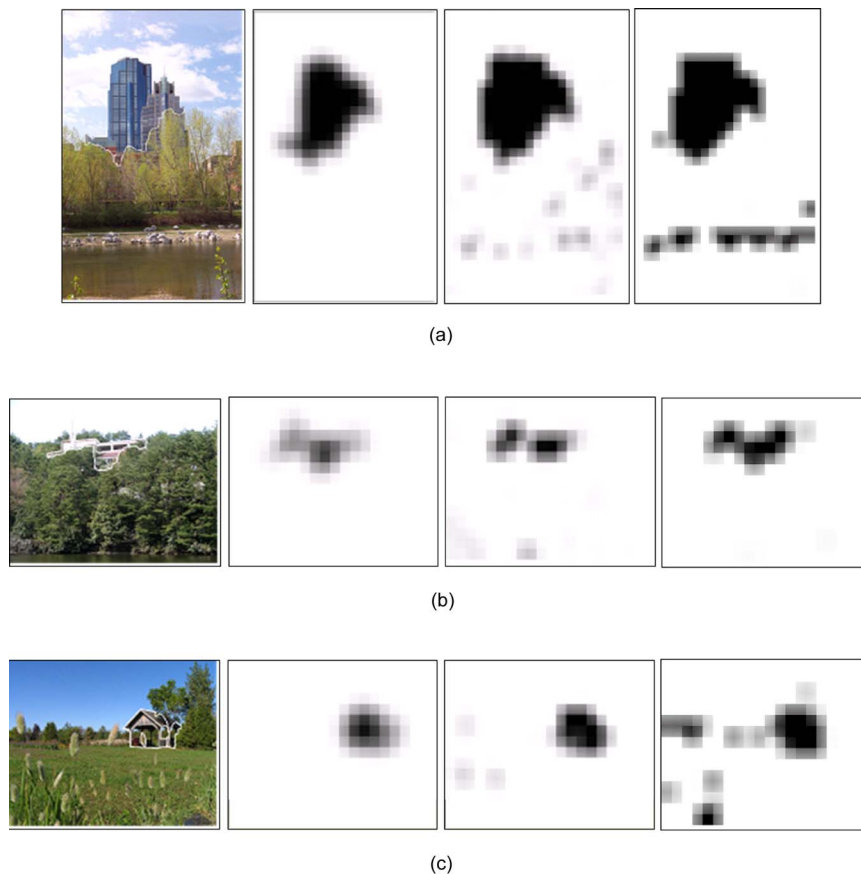


Fig. 10 Examples of man-made object detections on three images of test set I. Each group of four images shows the original image with the overlaid ground truth (first image), the fraction of artificial pixels in the ground truth mask within each 64×64 block (second image), and the *a posteriori* probability map (for $T=50\%$) based on the ridgelet features (third image) and on the Gabor features (fourth image).

is growing evidence in the literature the AUC that constitutes a more consistent and reliable measure for machine learning algorithm evaluation.^{22–24}

5.2.1 Performance of image patch classification on test sets

For each patch, we computed the probability of containing artificial content. The probability map for the entire image was built using a cosine-based weighting function, as in Ref. 17, which provides a good probability map estimate, especially in regions made of blocks with mixed content (e.g., 50% artificial). Because of the overlap, the effective resolution at which the decisions were taken was 32×32 blocks formed by the intersection of the 64×64 tiles.

Tables 2 and 3 show the DR and FAR for the classification of the image patches on test sets I and II, respectively, and for three different thresholds T . As expected, the DR and FAR increased as T increased. The best results for the ridgelets+sub-KLDA method were obtained for $\Delta\theta = 1$ deg. $p=2$ and sub-KLDA. The ridgelet+sub-SVC method had slightly better DR performances on test set I but with a higher FAR. However, the ridgelets+sub-SVC method performed worse than ridgelets+sub-KLDA on the lower-quality images (Table 3). Lower recognition rates on test set II can be explained in part by the fact that the

images were smaller (i.e., smaller objects to detect) and noisier due to the JPEG compression. The blocky effects of the JPEG compression can be amplified by the preprocessing filter within the featureless image regions (e.g., sky).

The performance trend is summarized in Fig. 8, which clearly shows that, overall, classification results improved with ridgelet features instead of Gabor ones, irrespective of the classification method used. This is because RIT provided a much finer directional analysis of the image spectrum compared to the GT.

In Tables 2 and 3, a Bayesian decision threshold of 0.5 was used to construct the probability map of the natural/artificial two-class problem. Other decision thresholds gave different results, as illustrated by the ROC curve in Fig. 9 for $p=2$, $\Delta\theta=1$ and sub-KLDA classification (we obtained similar results for other parameter combinations). Detection performances are dependent on the degree of “artificialness” of the patches, as can be seen in Fig. 9, where the ROC curve degrades when mixed patches are considered. The area values under the ROC curves are given in Table 4, as well as an approximate confidence interval estimate at the 95% confidence proposed by Agarwal *et al.*²⁴ The AUC number was globally larger for ridgelets than Gabor (a larger number is an indication of better performance).

Finally, Fig. 10 shows a typical segmentation result ob-

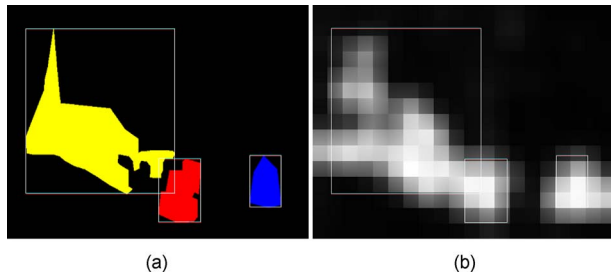


Fig. 11 For the object detection performance, we computed the fraction of pixels well classified on the probability map within the ROI that delimited each object (white boxes). (a) Isolated nonconnected objects on the ground truth image, (b) fraction of pixels well classified as artificial on the probability map for $T=50\%$.

tained from four images of test set I for $p=2$, $\Delta\theta=1$, $T=50\%$, sub-KLDA, and sub-LDA. Overall, the detected man-made object boundaries were quite similar to the ground truth. False alarms were caused mainly by natural objects in the foreground that had some geometric features (e.g., rocks with sharp edges). Increasing T to 100% improved the confidence of the image block classification but reduced the surface of the detected man-made regions.

5.2.2 Performance of object classification on test sets

We also measured the detection performance in terms of the whole non-natural objects. To this aim, we first built a bounding box around each region of interest (ROI) of the ground truth object after applying the threshold T . Within each object ROI (white rectangles in Fig. 11), we computed the fraction of the 32×32 image patches well classified as artificial on the detected probability map inside that ROI. If at least 50% of the blocks within the object were detected, we considered that the object had been correctly detected. If a group of connected 32×32 image blocks were detected and did not overlap an object ROI, it was counted as an object false alarm. Note that a lesser performance in terms of 32×32 block detection does not necessarily translate to a degraded object detection as long as at least 50% of the object was identified. Tables 5 and 6 summarize the object detection performances for various feature/classifier combinations at a Bayesian decision threshold of 0.5 (ridgelet parameters were set to $p=2$ and $\Delta\theta=1$). Note that as T increased, the number of objects considered to be ground truth decreased. Globally, the best DR and FAR combination appears to be for ridgelets. Gabor features gave the worst results in terms of FAR.

6 Conclusion

We have reported about a method based on the local RIT for the detection of non-natural objects in images of landscape scenes. The best test result for the classification of image blocks using ridgelets+sub-KLDA was obtained for an angular sampling rate of 1 deg and directional moments of order 2. This performance was comparable to that of ridgelets+sub-SVC on test set I but was better for lower-quality images (test set II). In terms of object detection, ridgelets gave better results than Gabor features, and with better false-object detection.

Table 5 Performance on test set I for object detection.

DR (% of true objects detected)			Average false objects per images
$T=50\%$ (99 obj.)	$T=75\%$ (63 obj.)	$T=100\%$ (30 obj.)	
Ridgelets+sub-KLDA			
73.7	88.9	90.0	0.40
Ridgelets+sub-SVC			
73.7	87.3	93.3	1.11
Gabor+sub-KLDA			
75.8	82.5	80.0	1.45
Gabor+sub-SVC			
74.8	82.5	83.3	0.89

The Ridgelets+sub-KLDA performances were also compared to other recent published results using different and more complex statistical learning approaches (e.g., Refs. 12 and 14) on similar databases. However, those latter methods are more complex to implement/train (multi-scale Markov random field, dynamic trees, etc.). Here, we concentrated on comparing ridgelets against Gabor features for standard classifiers (LDA, KLDA, or SVC) because they are the most often used in practical applications. We have shown that combining ridgelets features with those standard classifiers provides a simple approach that gives performance results comparable or even better than other features/classifier combinations. The difference is not very large, but the trend is there. In addition, and at least of equal importance, ridgelets have the advantage of a simple feature-extraction procedure.

Table 6 Performance on test set II for object detection.

DR (% of true objects detected)			Average false objects per images
$T=50\%$ (46 obj.)	$T=75\%$ (33 obj.)	$T=100\%$ (15 obj.)	
Ridgelets+sub-KLDA			
84.8	93.9	93.3	0.85
Ridgelets+sub-SVC			
80.4	90.9	93.3	1.51
Gabor+sub-KLDA			
82.6	87.9	93.3	1.75
Gabor+sub-SVC			
87.0	87.9	86.7	0.92

Obviously there is room for improvement and future work. The RIT was implemented in the Fourier space to provide a fast computation time. Typically, the processing task takes less than 2 sec of computer processing time for a 1000×1000 -pixel image using MATLAB on a 2.8-GHz processor. This processing time could easily be reduced by a factor of 10 with a C or C++ implementation. Also, color information could be used as an additional feature input. Finally, tests with kernels other than Gaussian (for the KLDA) could also be explored.

Acknowledgments

This work was supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada. The authors thank Prof. P. Poullaouec-Gonidec from the UNESCO Chair in Landscape and Environmental Design of Université de Montréal (<http://www.unesco-paysage.umontreal.ca>) for providing part of the training and test sets. We also thank the reviewers for their very constructive comments.

References

- E. Candès, "Ridgelets: Theory and applications," Ph.D Thesis, Stanford University (1998).
- M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation," *IEEE Trans. Image Process.* **12**(1), 16–28 (2003).
- A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1349–1380 (2000).
- N. Sebe, M. S. Lew, X. Zhou, T. S. Huang, and E. Bakker, "The state of the art in image and video retrieval," in *International Conference on Image and Video Retrieval (CIVR'03)*, pp. 1–8 (2003).
- A. Smeaton, W. Kraaij, and P. Over, "The TREC video retrieval evaluation (TRECVID): A case study and status report," in *Proc. of RIAO*, pp. 25–37 (2003).
- A. Torralba and A. Oliva, "Statistics of natural image categories," *Network Comput. Neural Syst.* **14**, 391–412 (2003).
- F. Foucher, V. Gouaillier, and L. Gagnon, "Global semantic classification of scenes using ridgelet transform," *Proc. SPIE* **5292**, 402–413 (2004).
- H. Le Borgne and N. O'Conner, "Natural science classification and retrieval using ridgelet-based image signatures," in *Proc. of the Advanced Concepts for Intelligent Vision Systems (ACIVS 2005)*, pp. 116–122 (2005).
- E. Candès and D. L. Donoho, "Curvelets, multiresolution representation, and scaling laws," *Proc. SPIE* **4119**, 1–12 (2000).
- M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.* **14**(12), 2091–2106 (2005).
- D. Donoho and A. G. Flesia, "Can recent innovations in harmonic analysis 'explain' the recent findings in natural image statistics?," *Network Comput. Neural Syst.* **12**, 371–393 (2001).
- S. Kumar and M. Hébert, "Man-made structure detection in natural images using a causal multiscale random field," in *Proc. IEEE Conference on Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 119–126, (2003).
- S. Todorovic and M. C. Nechyba, "Detection of artificial structures in natural-scene images using dynamic trees," in *Proc. 17th International Conference on Pattern Recognition (ICPR 2004)*, Vol. 1, pp. 35–39 (2004).
- X. Dai, Y. Suzuki, and J. Maeda, "Unsupervised rough segmentation of natural images containing man-made objects," *IPSSJ J.* **45**(1), 346–349 (2004).
- J. Malababic, H. Le Borgne, N. Murphy, and N. O'Conner, "Detecting the presence of large buildings in natural images," in *Proc. 4th Int. Workshop on Content-Based Multimedia Indexing (CBMI 2005)*, Riga, Latvia, 12–23 June 2005, IEEE, Piscataway, NJ (2005).
- J.-L. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.* **11**(6), 670–684 (2002).
- A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.* **42**(3), 145–175 (2001).
- S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proc. 1999 IEEE Signal Processing Workshop*, Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds., pp. 41–48, IEEE, Piscataway, NJ (1999).
- A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2), 228–233 (2001).
- V. Gouaillier, L. Gagnon, S. Paquette, and P. Poullaouec-Gonidec, "Use of the MPEG-7 standard as metadata framework for a location scouting system—An evaluation study," in *Proc. International Conference on Dublin Core and Metadata Applications (DC-2005): Vocabularies in Practice*, pp. 157–161 (2005).
- C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," technical report, Department of Computer Science, National Taiwan University (2008).
- C. X. Ling, J. Huang, and H. Zhang, "A statistically consistent and more discriminating measure than accuracy," in *Proc. IJCAI 2003*, pp. 329–341 (2003).
- C. Cortes and M. Mohri, "Confidence interval for the area under the ROC curve," NIPS, http://books.nips.cc/papers/files/nips17/NIPS2004_0197.pdf (2004).
- S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth, "Generalization bounds for the area under the ROC curve," *J. Mach. Learn. Res.* **6**, 393–425 (2005).



Langis Gagnon received a PhD in mathematical physics from Université de Montréal (QC, Canada) in 1988. Until 1995, he was research officer at the Centre d'Optique, Photonique et Laser (COPL) of Université Laval (QC, Canada), the Centre de Recherches Mathématiques (CRM), and the Laboratoire de Physique Nucléaire (LPN) of Université de Montréal. From 1995 to 1998, he was a specialized researcher at Lockheed Martin Canada. He then joined the Computer Research Institute of Montreal (CRIM), where he is now a principal researcher and team leader (vision and imaging). Gagnon has published more than 150 scientific articles related to the fields of image processing, object recognition, and nonlinear optical modeling.



Samuel Foucher has an engineering degree in telecommunications and a PhD in remote sensing from Université de Sherbrooke (QC, Canada) and Université de Rennes I (France). Between 1999 and 2002, he worked in industry as a research scientist on image mining. He then joined the Computer Research Institute of Montreal (CRIM) to work on polarimetric image processing and content-based video technologies. His research interests are in image processing, multiresolution encoding techniques (wavelets), data fusion, belief theory, and Markovian techniques.