Global Semantic Classification of Scenes using Ridgelet Transform

S. Foucher*, V. Gouaillier, L. Gagnon

R&D Department, Computer Research Institute of Montreal,
550 Sherbrooke West, Suite 100, Montreal (QC), Canada, H3A 1B9

ABSTRACT

In recent years, new harmonic analysis tools providing sparse representation in high dimension space have been proposed. In particular, ridgelets and curvelets bases are similar to the sparse components of naturally occurring image data derived empirically by computational neuroscience researchers. Ridgelets take the form of basis elements which exhibit very high directional sensitivity and are highly anisotropic. The ridgelet transform have been shown to provide a sparse representation for smooth objects with straight edges. Independently, for the purpose of scene description, the shape of the Fourier energy spectra has been used as an efficient way to provide a "holistic" description of the scene picture and its semantic category. Similarly, we focus on a simple binary semantic classification (artificial vs. natural) based on various ridgelet features. The learning stage is performed on a large image database using different state of the art Linear Discriminant techniques. Classification results are compared with those resulting from the Gabor representation. Additionally, ridgelet representation provides us with a way to accurately reconstruct the original signal. Using this synthesis step, we filter the ridgelet coefficients with the discriminant vector. The resulting image identifies the elements within the scene contributing to the different perceptual dimensions.

Keywords: ridgelet, semantic extraction, statistical learning, CBIR

1. INTRODUCTION

The purpose of this paper is to investigate the potential of a new harmonic analysis tool called the ridgelet transform applied to landscape classification. This new multiscale representation has been shown to produce more efficient and sparse representations of discontinuities distributed along smooth 2-D curves [1-5]. In order to capture curved edges, different schemes have been proposed to localize the ridgelet transform : a) the *monoscale ridgelet* where the image plane is partitioned into smoothly overlapping blocks [4] and b) the *curvelet transform* which is a multiscale ridgelet transform where ridgelet lengths and widths are related by a scaling law [5,6]. Curvelet and ridgelet basis have a striking resemblance and share similar properties (multiscale, multiorientation, anisotropy,…) with sparse basis of natural images established in the Natural Scene Statistics (NSS) literature [7,8].

Semantic classification of images is a difficult task and could help in the indexation of large database making retrieval more efficient and more meaningful. There are mainly two kinds of approaches in scene categorization: object based and scene based. Where as the former focuses on the recognition of specific objects within the scene (buildings, sky, etc.), the latter deal with the scene as a whole and is sensitive to the organization of objects. Our approach is similar to the work of Oliva and Torralba which used a Gabor representation in order to capture global spectral templates sensitive to the scene organization [9-14]. Semantic information is built directly from a pool of low-level features taking advantage of the regularities found in the statistical distribution of features per scene category [9]. Using learned discriminant spectral templates, superordinate categories of real world images can be distinguished (such as *naturalness* and *openness*) [9,13,14].

In the present work, we focus on the categorization of natural scenes versus artificial scenes. Artificial scenes are composed of man-made objects, having strong geometrical features such as dominant vertical and horizontal edges [14].

* sfoucher@crim.ca; phone 514 840-1235; fax 514 840-1244; crim.ca

Therefore, the ridgelet transform seems to be a suitable representation providing low-level features and defining a semantic naturalness axis on which images could be ranked.

The paper is organized as follows. Section 2 gives some background on the ridgelet transform, its numerical implementation and its application to rectangular non-dyadic images. In section 3, we describe the proposed ridgelet based features used for Artificial-Natural scene classification. In section 4, we present different linear and non linear statistical learning techniques. Classification results for natural versus artificial scene discrimination are presented in section 5 and compared to the results obtained with Gabor filters.

2. RIDGELET TRANSFORMATION

2.1 Background

The ridgelet transform was originally proposed by Candès and Donoho [1,2] and offers a sparse coding of image with edges. A ridgelet is defined as a wavelet distributed along a ridge of orientation $x_1 \cos\theta + x_2 \sin\theta = b$ within the Cartesian plan (x_1, x_2) :

$$\psi_{a,b,\theta}(x_1, x_2) = a^{-1/2}\psi((x_1 \cos\theta + x_2 \sin\theta - b)/a), a > 0, b \in \mathbf{R}, \theta \in [0, 2\pi[\quad (1)$$

The ridgelet coefficients Rid_I of an image *I* are obtained by projection on this base :

$$\mathrm{Rid}_I(a,b,\theta) = \iint \psi_{a,b,\theta}(x_1, x_2) I(x_1, x_2) dx_1 dx_2 \quad (2)$$

This projection is related to the Radon transform which consists in integrating the image along lines of different orientations :

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

Consequently, the ridgelet transform (2) can be seen as 1-D wavelet transform of the Radon transform Rad_I along the translation parameter *t*:

$$\mathrm{Rid}_I(a,b,\theta) = \int \mathrm{Rad}_I(\theta,t)\psi((t-b)/a) dt \quad (4)$$

An important property can be exploited, called the projection-slice therorem, which says that the 1-dimensionnal constant-θ slice of the Radon transform and the 1-dimensionnal radial slice of the Fourier transform makes a Fourier transform pair:

$$\hat{I}(\lambda\cos\theta, \lambda\sin\theta) = TF_t\left[\mathrm{Rad}_I(\theta,t)\right] = \int \mathrm{Rad}_I(\theta,t) e^{-\lambda i t} dt \quad (5)$$

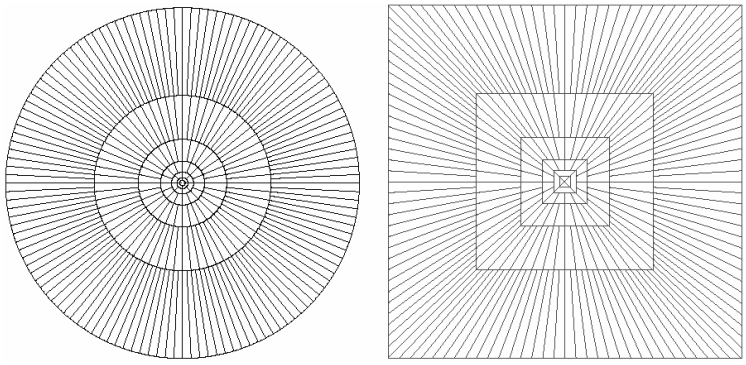
Which means that the Radon transform for a given orientation can be derived from the inverse 1-dimensional Fourier transform of the image Fourier transform performed along radial lines. Consequently, Equation (4) can be simply written in the frequency domain:

$$\mathrm{Rid}_I(a,b,\theta) = TF_\lambda^{-1}\left[\hat{I}(\lambda\cos\theta, \lambda\sin\theta) TF_t\left[\psi(t/a)\right]\right] \quad (6)$$

2.2 Digital implementation

In order to apply equation (6) the frequency domain has to be re-sampled on a polar grid. This task is not easy because radial slices do not always intersect the Cartesian grid. A simple and efficient approximation of the polar grid is the rectopolar grid which consists in concentric square (see Fig. 1). The new frequency samples are distributed along radial lines intersecting the null frequency component of the Fourier spectrum. These new frequency components can be

calculated using various interpolation scheme [3,6]. A simple method, which we apply here, is to use a nearest neighbor rule [6]. The set of radial lines can be divided into two subsets: the *basically horizontal lines* with $\theta \in \Theta_1 = [-\pi/4, \pi/4[$ and the *basically vertical lines* with $\theta \in \Theta_2 = [\pi/4, 3\pi/4[$. An example of basically horizontal lines and nearest neighbor interpolation is shown on Fig. 2.



(a) (b)

Fig. 1. a) Ridgelet tiling of the frequency domain and b) Digital ridgelet tiling (from [3])

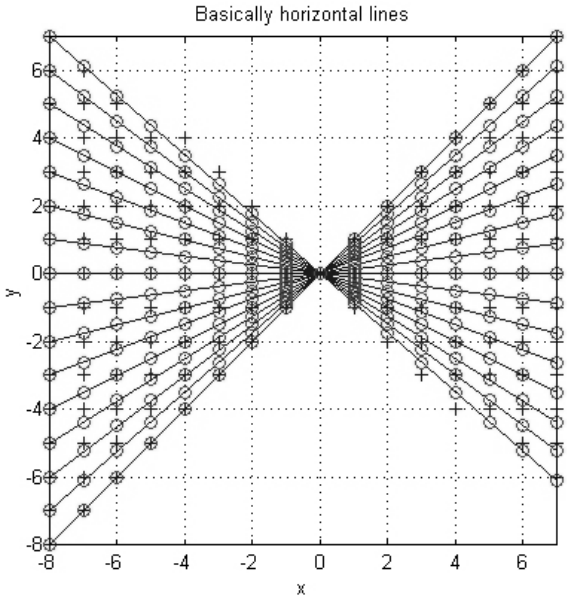


Fig. 2. Cartesian-to-Pseudopolar conversion for the basically horizontal lines for a 16x16 grid, The red circles indicate pseudopolar points and the cross marks indicate nearest neighbor positions on the Cartesian grid.

Our implementation of the ridgelet transform follows the approach proposed by Starck *et al.* [6] for curvelet/ridgelet based denoising. The forward ridgelet transform is summarized on Fig. 3. The wavelet transform is the "à trous" algorithm using B-spline functions [6]. This implementation enjoys many suitable features: a) fast because of the intensive use of the Fast Fourier Transform (FFT); b) rotation and translation invariance due to subband sampling above the Nyquist rate; c) the reconstruction step (inverse transform) is a simple co-addition in the wavelet domain followed by an arithmetic mean of all the spectral values in the rectopolar grid which are nearest neighbors.

2.3 Application to rectangular non-dyadic images

For a rectangular images, the horizontal and vertical lines define the following two angle sets which are dependent on the image dimensions:

$$\Theta_1 = \left[\text{atan}(H/W), \text{atan}(H/W)\right[$$
$$\Theta_2 = \left[\pi/2 - \text{atan}(W/H), \pi/2 + \text{atan}(W/H)\right[\tag{7}$$

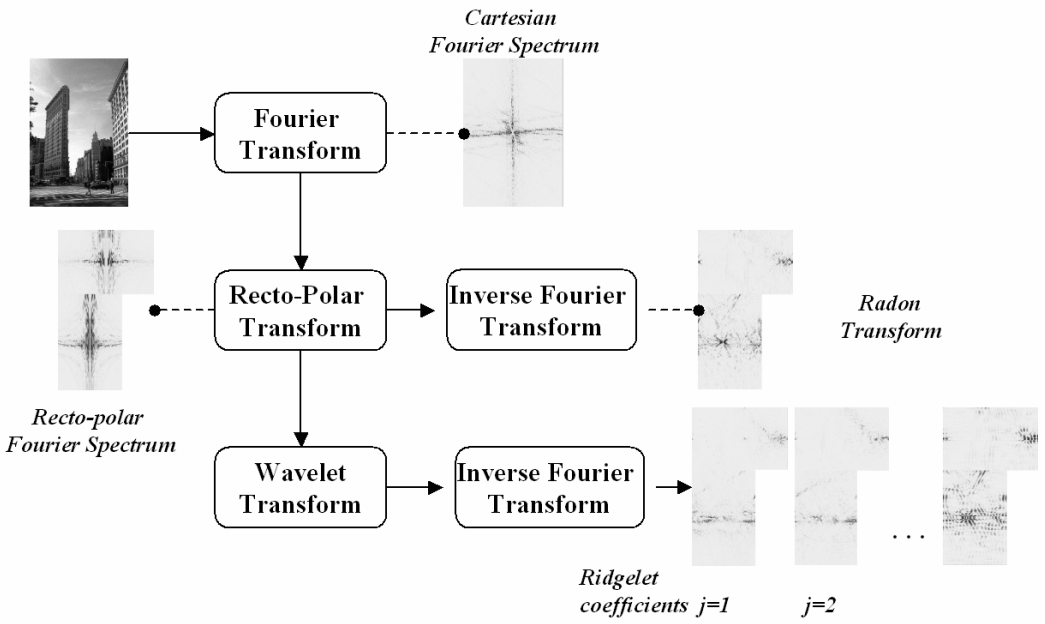


Fig. 3. Block diagram of the ridgelet forward transform

2.4 Statistical behavior

Since the ridgelet transform is a wavelet transform in a geometrical domain, ridgelet statistics are dependent on the image size. We want ridgelet based features of a white noise image for which the p.d.f. is $N(\mu_I, \sigma_I^2)$ to be invariant under image resizing. The Radon transform integrates the image along translated lines with different orientations. Therefore the mean value of the radon transform is related to the image mean $\mu_I = E[I]$ as follows:

$$\mu_{\text{Rad}_I|\theta_m^1} = E_t\left[\text{Rad}_I(m,n) \,|\, \theta_m^1\right] = W\mu_I$$
$$\mu_{\text{Rad}_I|\theta_m^2} = E_t\left[\text{Rad}_I(m,n) \,|\, \theta_m^2\right] = H\mu_I \tag{8}$$

According to the central limit therorem, the p.d.f. of Rad_I is a Gaussian distribution:

$$p_{\text{Rid}_I^{[j]}}(w = \text{Rid}_I^{[j],1}(m,n)) \approx \frac{1}{\sqrt{2\pi S_2^{[j]} W \sigma_I^2}} e^{-\frac{(w)^2}{2S_2^{[j]} W \sigma_I^2}}$$
$$p_{\text{Rid}_I^{[j]}}(w = \text{Rid}_I^{[j],2}(m,n)) \approx \frac{1}{\sqrt{2\pi S_2^{[j]} H \sigma_I^2}} e^{-\frac{(w)^2}{2S_2^{[j]} H \sigma_I^2}} \tag{9}$$

This statistical model is verified on simulated data on Fig. 4.

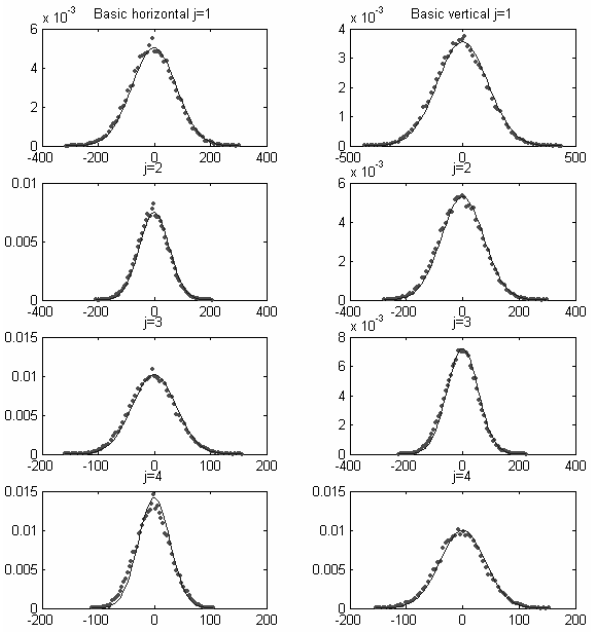


Fig. 4. observed probability density functions (red points) for a white noise image (*H=128, W=256,* $\mu_I = 128, \sigma_I = 10$) at different scales. The continuous curves are given by equation (9).

3. PROPOSED RIDGELET FEATURES

We propose to use the ridgelet coefficients directional moments of order *p* as an image feature. First, variance values have to be normalized by the wavelet power gain ($S_2^{[j]} = \int \left|\Psi^{[j]}(t)\right|^2 dt$) and the image dimensions:

$$\tilde{\mu}^2_{p,\mathrm{Rid}_I^{[j]}|\theta_m}(m) = \begin{cases} \dfrac{E_n\left[\left(\mathrm{Rid}_I^{[j],s}(m,n)\right)^p \mid \theta_m{}^s\right]}{S_2{}^{[j]} H} & , s = 1 \\ \dfrac{E_n\left[\left(\mathrm{Rid}_I^{[j],s}(m,n)\right)^p \mid \theta_m{}^s\right]}{S_2{}^{[j]} W} & , s = 2 \end{cases} \tag{10}$$

Assuming that images are acquired in a standard viewpoint (earth at the bottom, sky at the top), we can exploit the fact that semantic content is invariant under a horizontal symmetry operation:

$$\tilde{\mu}^{[j]}_{p,\mathrm{Rid}_I}(m) = \begin{cases} (\tilde{\mu}_{p,\mathrm{Rid}_I^{[j]}|\theta_m} + \tilde{\mu}_{p,\mathrm{Rid}_I^{[j]}|3\pi/4-\theta_m})/2 & , \theta_m \in \left]\pi/4, 3\pi/4\right] \\ (\tilde{\mu}_{p,\mathrm{Rid}_I^{[j]}|\theta_m} + \tilde{\mu}_{p,\mathrm{Rid}_I^{[j]}|-\theta_m})/2 & , \theta_m \in \left]-\pi/4, \pi/4\right] \end{cases} \tag{11}$$

This assumption is similar to the approach proposed by Oliva and Torralba [14] and leads to a dimension reduction by a factor 2. The angle sets are dependent on the image dimensions so we have to re-sample on a constant angle resolution $\Delta\theta$ in order to obtain the same feature vector for each image. The resulting feature vector based on *J* ridgelet scales is a vector $\hat{\mu}_{p,\mathrm{Rid}_I} = [\hat{\mu}_{p,\mathrm{Rid}_I}{}^{[1]T}, ..., \hat{\mu}_{p,\mathrm{Rid}_I}{}^{[J]T}]^T$ of length $J \times \left\lfloor \dfrac{\pi/2}{\Delta\theta} \right\rfloor$.

4. STATISTICAL LEARNING

4.1 Background

Statistical learning and in particular Fisher's Linear Discriminant Analysis (LDA) techniques is widely used in computer vision problems such as face recognition. We note $X = \{\mathbf{x}_1, ..., \mathbf{x}_l\}$ the training set composed of two subsets $X^1 = \{\mathbf{x}_1^1, ..., \mathbf{x}_{l_1}^1\}$ and $X^2 = \{\mathbf{x}_1^2, ..., \mathbf{x}_{l_2}^2\}$ (where $l = l_1 + l_2$) belonging respectively to the considered classes "Natural" and "Artificial". Recently, enhanced methods have been proposed to improve learning performances when feature dimension is high compared to the training set size and reach non linear solutions [15-18]. Linear discriminant techniques aims at finding a discriminant vector **w** (also called Fisher vector) which maximize the Fisher criterion:

$$\mathbf{w} = \arg\max_{\mathbf{n}} \left\{ \frac{\mathbf{n}^T \mathbf{S}_B \mathbf{n}}{\mathbf{n}^T \mathbf{S}_W \mathbf{n}} \right\} \tag{12}$$

where $\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$ and $\mathbf{S}_W = \sum_{i=1,2} \sum_k (\mathbf{x}_k^i - \mathbf{m}_i)(\mathbf{x}_k^i - \mathbf{m}_i)^T$ are the between and the within class variance respectively. This approach is optimal for two Gaussian distributions with equal covariance structures and leads to a linear solution. The equation (12) is equivalent to solve a generalized eigenvalue problem:

$$\mathbf{S}_B \mathbf{W} = \lambda \mathbf{S}_W \mathbf{W} \tag{13}$$

Usually, this equation is solved by first diagonalizing $\mathbf{S}_W$ before diagonalizing the projection of $\mathbf{S}_B$ on the eigenvectors of $\mathbf{S}_W$. The inverse approach, called the Direct Discriminant Analysis (DDA) has been shown to be more robust when feature space dimension is high and therefore $\mathbf{S}_W$ potentially singular [16]. Many variants have been proposed in order to avoid overfitting and small sample size problems, such as the Subspace LDA (Sub-LDA) where a preliminary PCA analysis is performed in order to reduce feature space dimensionality and reduce overfitting. This technique was employed by Torralba and Oliva [9] to build discriminant spectral templates.

To overcome the linear assumption, the use of kernel functions have been proposed [17,18] where a nonlinear mapping $\Phi : X \to F$ is applied before performing a discriminative analysis leading to the Kernel LDA technique (KLDA). The nonlinear mapping is not directly performed but is implicitly expressed through the use of dot products forming the kernel matrix:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \left(\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \right) \tag{14}$$

For instance, in the case of a gaussian kernel, the projection of a new sample **x** onto the nonlinear discriminant vector **w** is then $\mathbf{w}^T \mathbf{x} = \sum_i \alpha_i \exp(-\|x - x_i\|^2 / c)$. In the new high dimensional space *F*, within and between classes variances become:

$$\begin{aligned} \mathbf{w}^T \mathbf{S}_B \mathbf{w} &= \boldsymbol{\alpha}^T \left(\bar{\mathbf{k}}_1 - \bar{\mathbf{k}}_2 \right) \left(\bar{\mathbf{k}}_1 - \bar{\mathbf{k}}_2 \right)^T \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{I} \boldsymbol{\alpha} \\ \mathbf{w}^T \mathbf{S}_W \mathbf{w} &= \boldsymbol{\alpha}^T \left(\mathbf{K} \left(\mathbf{I} - \mathbf{v}_1 \mathbf{v}_1^T - \mathbf{v}_2 \mathbf{v}_2^T \right) \mathbf{K}^T \right) \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{N} \boldsymbol{\alpha} \end{aligned} \tag{15}$$

where $\mathbf{v}_i$ is an binary indicator vector for the class *i* and $\bar{\mathbf{k}}_i = \frac{1}{l_i}\sum_{n=1}^{l_i}\mathbf{K}(\cdot, \mathbf{x}_n^i)$. The maximization of the Fisher criterion (12) using equations (15) is ill-posed because the matrix **N** is large ($l \times l$) and is possibly singular. Numeric stability can be improve by adding a diagonal regularization term:

$$\mathbf{N}' = \mathbf{N} + \mu\mathbf{I} \tag{16}$$

Non linear mapping can be also applied using a Direct Discriminant technique instead of LDA, this variant has been proposed by Lu *et al.* [16]. We denote KDDA this technique.

4.2 Training and testing sets

We use $\left(\bar{\mu}_{2,\mathrm{Rid}_1}\right)^{1/2}$ with *p=2, J=4* and $\Delta\theta = 0.5^o$ as an image feature vector. Its dimension is then equals to 180. The training step is performed using images from various sources (professional images, digital amateur images, internet). The training set is composed of $l_1 = 1058$ natural scenes and $l_2 = 990$ artificial scenes. Before computing the ridgelet features, images are normalized using a technique proposed by Oliva and Torralba [12] where high frequency information is enhanced within low contrast image regions. On Fig. 5, we give the mean feature vector for the two classes computed over the training set. We also compute Gabor features using 5 dyadic scales and 24 orientations per scale. Non linear techniques are applied using a gaussian kernel $\mathbf{K}(\mathbf{x}_j, \mathbf{x}_i) = \exp(-\left\|\mathbf{x}_j - \mathbf{x}_i\right\|^2 / c)$ with *c* set to one third of the feature space dimension. Discriminant Fisher vectors obtained by Sub-LDA and DDA are represented on Fig. 6. We observe that the discriminative information comes mainly from differentiating diagonal ridgelet energies with vertical and horizontal energies. This is similar to the results obtained by Oliva and Torralba [14]. The testing set is composed of 354 natural scenes and 289 artificial scenes (amateur digital images).

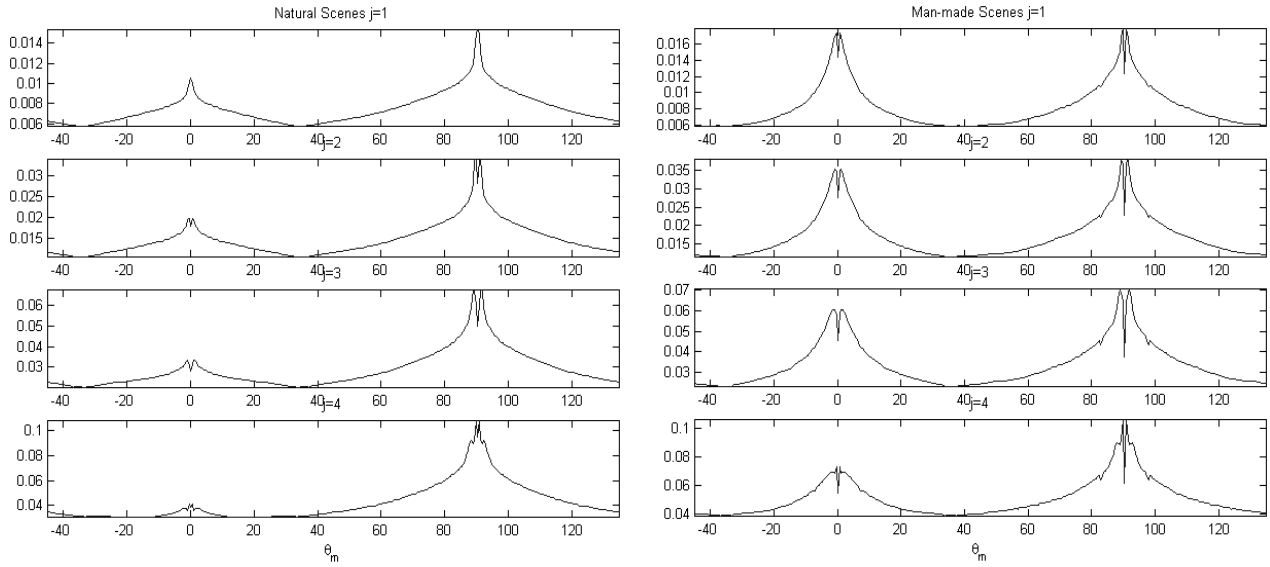


Fig. 5. Mean value for $\left(\bar{\mu}_{2,\mathrm{Rid}_1}\right)^{1/2}$ estimated from 1098 natural images and 990 man-made scenes ($\Delta\theta = 0.5^o$)

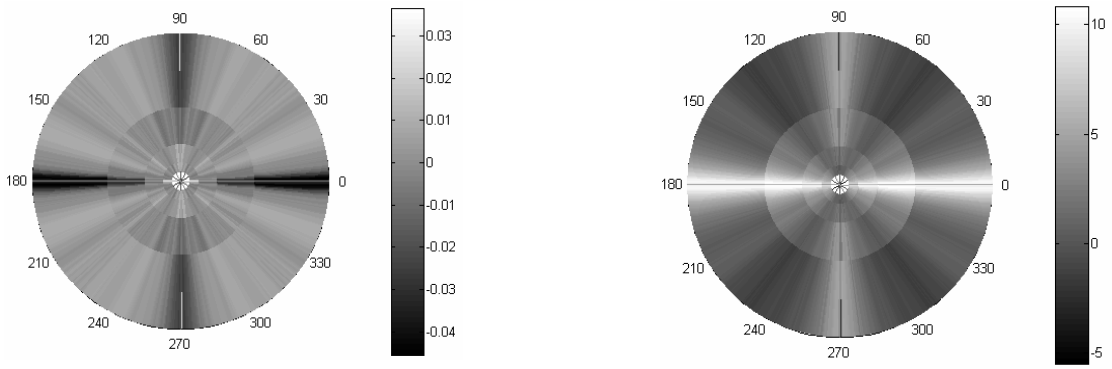


Fig. 6. Radial representation of the Fisher discriminant vector obtained by Sub-LDA (left) and DDA (right).

5 RESULTS

Comparison between the different discriminant techniques is represented on Fig. 7. Classification rate and Kappa coefficient values are given in Table 1. The Kappa coefficient is a classification rate corrected for the probability of the good classification occurring by chance.

	Method	Learning			Testing		
		Cl. Rate (Nat.)	Cl. Rate (Art.)	Kappa	Cl. Rate (Nat.)	Cl. Rate (Art.)	Kappa
Gabor	DDA	76.6	50.8	50.5	65.3	48.8	42.6
	Sub-LDA	80.9	62.5	59.9	70.3	67.1	56.2
	KLDA	90.1	87.9	**81.6**	79.9	84.1	**71.9**
	KDDA	77.2	67.4	60.5	65.5	72.3	56.2
Ridgelet (*p=2*)	DDA	81.8	50.0	53.2	91.0	99.3	**90.4**
	Sub-LDA	92.0	80.1	77.7	90.7	98.3	89.4
	KLDA	99.4	90.5	**91.0**	85.9	99.0	85.7
	KDDA	23.8	27.8	62.6	82.8	100.0	83.9

Table 1. Classification rate and Kappa coefficient values (%) on the training and testing set for the different learning methods. Best results are indicated in bold.

Kernel LDA gives the best result at the training step whereas DDA produces a better Kappa value on the test set. Compared to the Gabor basis, the ridgelet feature reach higher classification rate. Direct discriminant techniques (DDA and KDDA) have very poor performances on the training set but give good results on the testing set.

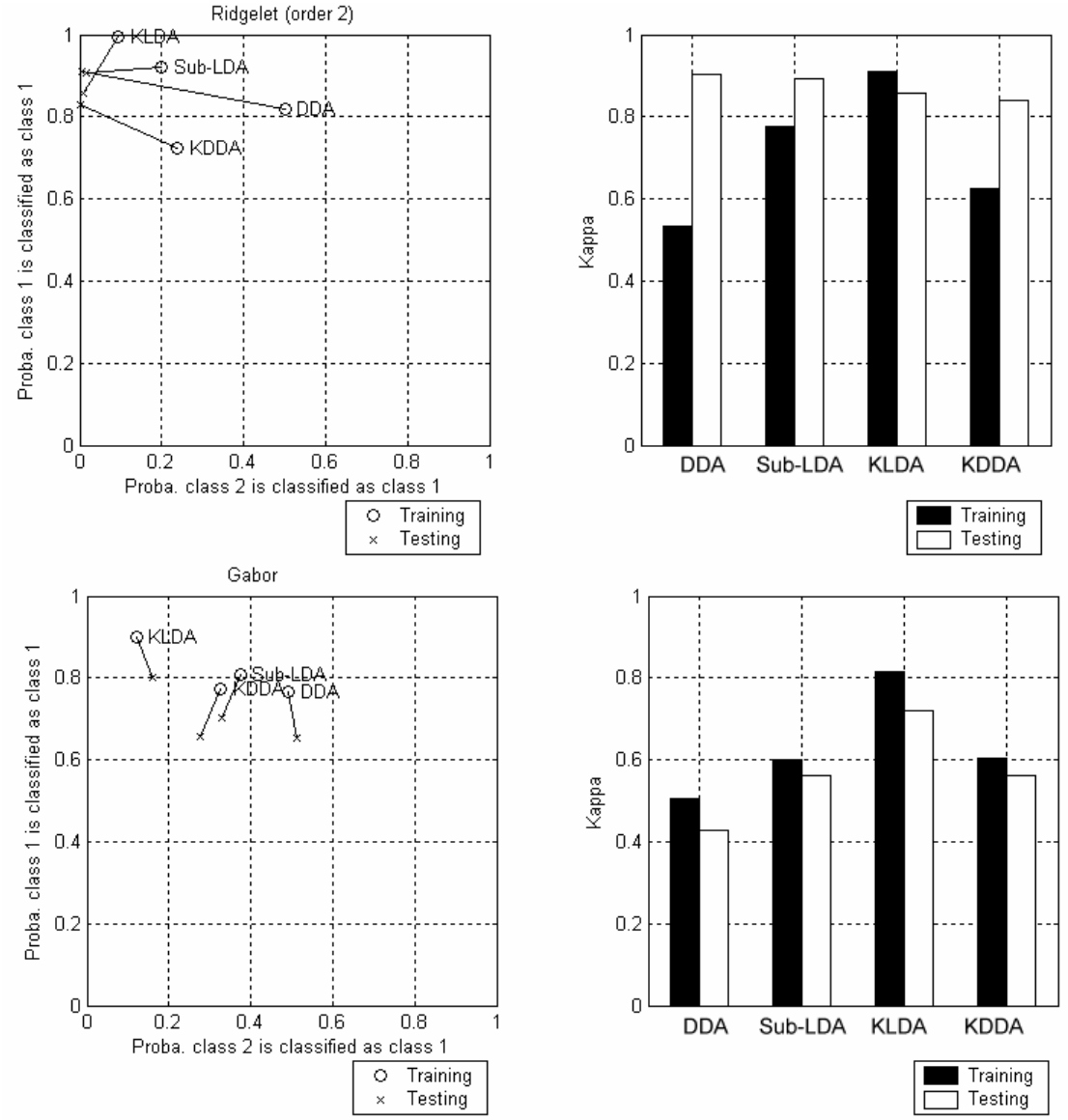
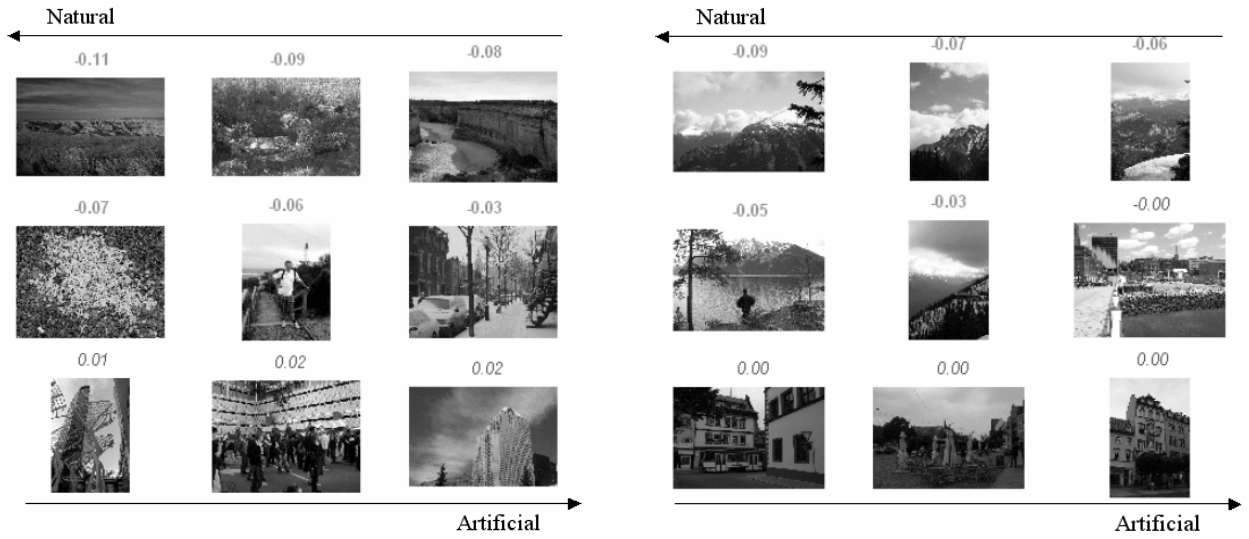


Fig. 7. Figures in the left column represent the training and testing performances for the different methods placed in a detection/false alarm probability plane (ROC curve). The right graph represents Kappa values for the different methods.

Images can be ordered according to the discriminant distance which defines an Naturalness-Artificialness distance (see Fig. 8.).



(a) (b)

Fig. 8. Ordering of images belonging to the training set (a) and to the testing set (b) using the discriminative distance obtained by KLDA (distances in italic (reps. In bold) or red (resp. or green) are labeled as artificial (resp. natural)).

6. FISHER FILTERING

When a linear discriminant technique is applied, such as Sub-LDA or DDA, the resulting Fisher vector can be used as a way to identify elements within the image that contributes to the degree of naturalness or artificialness. This approach was extensively used by Oliva and Torralba [9,13,14] and produced various Discriminant Spectral Templates (DST). We apply the same approach to the ridgelet domain by weighting ridgelet coefficients using the positive and negative part of the Fisher vector. We note $w^{+,[j]}$ and $w^{-,[j]}$ the positive and negative components respectively of the Fisher vector affecting the ridgelet coefficients at level *j*. We filter the ridgelet coefficients using the following relation:

$$\mathrm{Rid}_I{}^{+/-,[j]}(m,n) = \frac{w^{+/-,[j]}(m)}{\gamma \times \sigma^{[j]}_{Rid_I}(m)} \mathrm{Rid}_I{}^{[j]}(m,n) \tag{17}$$

where $\gamma = \sqrt{H}$ or $\gamma = \sqrt{W}$ for basically vertical and horizontal lines respectively. The quantity $\sigma^{[j]}_{\mathrm{Rid}_I}(m)$ is the standard deviation of the feature $\tilde{\mu}^{[j]}_{2,\mathrm{Rid}_I}(m)$ computed on the all training set. After filtering, we obtain two images I^+ and I^- by reconstruction from the filtered ridgelet coefficients $\left\{Rid_I{}^{+,[j]}\right\}_{j=1,..,J}$ and $\left\{Rid_I{}^{-,[j]}\right\}_{j=1,..,J}$ respectively (the low pass image is set to zero). On fig. 9 we give an example on a cityscape scene where I^+ shows scene elements contributing to the scene naturalness such as the trees in the foreground whereas I^- is sensitive to geometric elements on the buildings in the background.

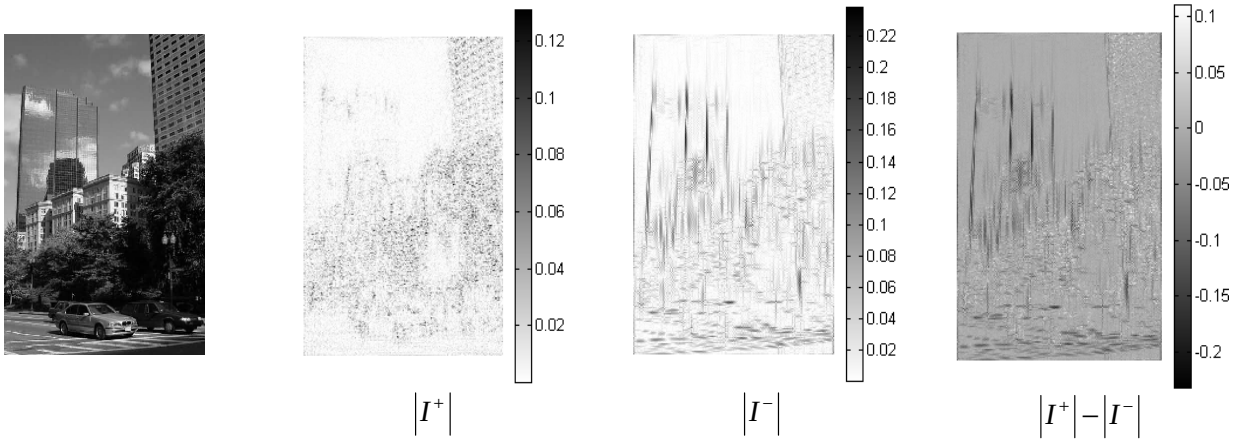


Fig. 9. Example of Fisher filtering using the Sub-LDA discriminant vector.

5. CONCLUSION

In this paper, we have used the ridgelet transform of real world images in order to produce low-levels features. Ridgelet based features seems to capture efficiently man-made elements within the scene. The ridgelet transform is relatively fast compared to the Gabor filters because it does not require a filter response calculation. Further work directions will try to incorporate color information within this framework and study localized ridgelet transform (monoscale ridgelets and curvelets). Higher order moments should also be investigated such as the fourth order moment which is related to the signal kurtosis. However, construction of a representative and exhaustive training set remains a challenge because of the wide varation in scene content. Another challenge is the robustness to variations in the photographic point of view. Different statistical learning methods have been used, the nonlinear methods produced better classification rate but numerical implementation is still an issue. Linear discriminant methods are less performant but are faster and results interpretation is easier.

ACKNOWLEDGEMENTS

This work is supported in part by the Natural Science and Engineering Research Council of Canada (NSERC) and the Ministère du développement économique et régional du gouvernement du Québec.

REFERENCES

1. E. Candès, *Ridgelets: Theory and Applications*, PhD. Thesis, Department of Statistics, Stanford University, 1998.
2. E. Candès and D. L. Donoho, *New Tight Frames of Curvelets and Optimal Representations of Objects with Smooth Singularities*, Technical Report, Stanford University, 2002.
3. D. Donoho and A.G. Flesia, *Digital Ridgelet Transform based on True Ridge Transform*, Beyond Wavelets, Academic Press Inc., pp. 1-33, 2001.
4. E. Candès, *Monoscale Ridgelets for the Representation of Images with Edges*, Technical Report, Department of Statistics, Stanford University, 1999.
5. E. Candès and D. Donoho, *Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges, Curves and Surfaces*, L. L. Schumaker et al. Eds, Vanderbilt University Press, Nashville, TN, 1999.
6. J.-L. Starck, E. Candès and D. Donoho, *The Curvelet Transform for Image Denoising*, IEEE Transactions on Image Processing, Vol. 11, No. 6, pp.670-684, 2002.
7. D. Donoho and A.G. Flesia, *Can recent innovations in harmonic analysis `explain' key findings in natural image statistics?*, Network: Computation in Neural Systems, No. 12, August, pp. 391-412, 2001.
8. B. Olshausen and D. Field, *Emergence of Simple-Cell Receptive Feld Properties by Learning a Sparse Code for Natural Images*, Nature, No 381, pp. 607–609, 1996.
9. A. Torralba and A. Oliva, *Statistics of Natural Image Categories*, Network: Computation in Neural Systems, No 14, August, pp. 391-412, 2003.

10. A. Oliva and A. Torralba, *Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*. International Journal of Computer Vision, Vol. 42, No 3, pp. 145-175, 2001.
11. A. Oliva, and A. Torralba, *Scene-Centered Description from Spatial Envelope Properties*, Lecture Note in Computer Science Serie. Proc. 2nd Workshop on Biologically Motivated Computer Vision, Tuebingen, Germany, 2002.
12. A. Oliva, and A. Torralba, *Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*, International Journal of Computer Vision, Vol. 42, No. 3, pp. 145-175, 2001.
13. A. Torralba and A. Oliva, *Semantic Organization of Scenes using Discriminant Structural Templates*, Proceedings of the International Conference on Computer Vision ,Korfu, Grece, pp. 1253-1258, 1999.
14. A. Oliva, A. Torralba, A. Guerin-Dugue, and J. Herault, *Global semantic classification of scenes using power spectrum templates*, Proceedings of The Challenge of Image Retrieval (CIR99), Springer Verlag BCS Electronic Workshops in Computing series, Newcastle, UK., 1999.
15. T. Cooke, Two Variations on Fisher`s Linear Discriminant for Pattern Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No 2, p. 268-273, 2003.
16. J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, *Face Recognition Using Kernel Direct Discriminant Analysis Algorithms*, IEEE Transactions on Neural Networks, Vol. 14, No. 1, pp. 117-126, January 2003.
17. S. Mika, G. Rätsch, J. Weston, B. Schölkopf and K.-R. Müller. *Fisher Discriminant Analysis with Kernels*. In Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas Eds, Neural Networks for Signal Processing IX, pp. 41-48, 1999.
18. S. Mika, G. Rätsch, J Weston, B. Schölkopf, A. Smola, and K.-R. Müller. *Constructing descriptive and discriminative non-linear features: Rayleigh coefficients in kernel feature spaces*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No 5, pp. 623-628, May 2003.