Parametric probabilistic modeling and information theory tools in textured images analysis

Yannick Berthoumieu

Laboratoire IMS UMR 5218
Université de Bordeaux

November 25, 2011

Contents

Introduction

- *Topic*: Characterizing texture contents for segmentation, classification and indexing.

- *Framework*: Scale space decomposition with wavelet, curvelet etc.
- *Tools*: information theory. (with Lionel Bombrun, Nour-eddine Lasmar, Aurélien Schutz)

Parametric random field

- Statistics (mean, variance, Kurtosis ...) 1980
- Parametric field modeling (Markovian, 2-D Autoregressive model, WOLD ...) 1990
- Scale space and *marginal* probabilistic modeling 2000
- Scale space and *joint* probabilistic modeling 2010

Homogeneous random Field

Definition

A random field $\boldsymbol{F}(s)$, defined on $\mathbf{S} = \mathbb{R}^2$, is a function whose intensities $f(s) \in \mathbb{R}^p$ (color image p=3) are random, for any value of s.

Definition

The homogeneous parametric field is associated to a specific density characterized by a finit set of parameters $\boldsymbol{\theta} \in \mathbb{R}^n$ independant of the pixel position in the field.

Examples

Gaussian, Gamma, Weibull, Uniform, Pareto ...

MGRF (1/2)

Besag , Cross&Jain ...

Definition

Markov-Gibbs random Field (MGRF) - Define a neighborhood $\Delta_i \subset S^i$as the set of all neighboring sites of a site $i \in \mathbf{S}$. A random field is an MRF if for each site $i \in \mathbf{S}$, $p\left(f_i|f^i\right) = p\left(f_i|f_j : j \in \Delta_i\right)$

and a Gibbs distribution if $p\left(f\right) = \frac{1}{Z}e^{\left\{-\sum\limits_{C\in\mathbf{C}} V_C\left(f_i:i\in C\right)\right\}}$with $V\left(f_i : i \in C\right)$ is the interaction function in a clique C for the pixel i over the cliques $\mathbf{C}$ for the image lattice.

MGRF - Pair-wise parametric modeling (2/2)

Definition

The conditional density is the discret pair-potential corresponding to $V_C(f_i : i \in C)$, i.e. $p(f_i|f^i) = p(f_i|f_j : j \in \Delta_i; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$is the parameter set defining the pixel dependance within the clique.

Example

The Gaussian model, or auto-normal model, is

$$p(f_i|f_j : j \in \Delta_i, \boldsymbol{\theta} = [\beta_{ij}, \sigma]) \sim \mathcal{N}\left(f_i - \sum_{j \in ´_i} \beta_{ij} f_j, \sigma\right).$$

Main drawback (and also the strength): the exponential pair-wise separable component (undirect Graph).

Maximum entropy principle (Maxent) 1/2

Definition

The MaxEnt principle suggests to select the density which maximizes the Entropy, i.e.

$$p^* = \underset{p \in \boldsymbol{F}}{arg\,max} H(p)$$

$$\text{s.t. } \boldsymbol{E}_p(L_j) = \boldsymbol{E}_{p^*}(L_j) : \; L_j \in \boldsymbol{L} = \{L_j : j = 1..K\}$$

where

- $\boldsymbol{E}p(.)$ is the expectation operator,
- $H(p) = \int p(f) \log(p(f))\, df$ is the shannon entropy function,
- L_j a set of observed features (mean, correlation, kurtosis ...).

Maximum entropy principle (Maxent) 2/2

Definition

The solution of MaxEnt is a Gibbs distribution (Lagrangian minimizer) as follow

$$p = \frac{1}{Z} \exp\left(\sum_j \lambda_j L_j\right) \text{ with } Z = \sum_f \exp\left(\sum_j \lambda_j L_j\right).$$

See. FRAME modeling [Zhu 1998]

Characterizing texture

- Problems: Segmentation, classification and indexing
  - Local modelling for tractable amount of parameters and for developping iterative process
    $\Longrightarrow p\left(f_i | f_j : j \in \Delta_i ; \boldsymbol{\theta}\right) = p_{\Delta}\left(\mathbf{f_i}, \boldsymbol{\theta}\right)$
- Main issue: Non-Gaussian famillies for random field
  - Wavelet coefficients
- How? Baysian decision based on the parametric form
  - $c^{*}\left(f\right) = \underset{c \in K}{arg\ max}\left[p\left(c | f\right)\right]$

Parametric family

Definition

Let $\mathcal{F}$ denote a parametric family of probability density functions $\mathcal{F} = \{p(f;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \mathbb{R}^n\}$ where the set $\boldsymbol{\theta}$ is assumed not to be redundant, i.e. if $p(f;\theta_1) = p(f;\theta_2)$ then $\theta_1 = \theta_2$.

Examples

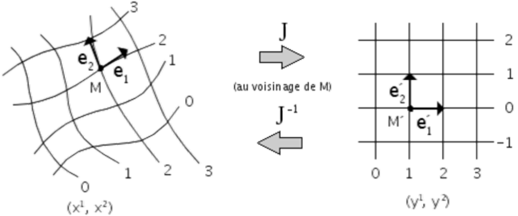
Gaussian law $\theta = (\mu, \sigma)$ with $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$.

Geometric point of view

Definition

Due the definition of a homeomorphism $\varphi : \mathcal{F} \to \mathbb{R}^n$ taking each $p\left(f;\boldsymbol{\theta}\right)$to its coordinates $\boldsymbol{\theta}$, i.e. $\varphi\left(p\left(f;\boldsymbol{\theta}\right)\right) = \boldsymbol{\theta}$, the family is called a *statistical manifold*.

Let $\frac{\partial}{\partial\theta_k}p\left(f;\boldsymbol{\theta}\right)$, for $k = 0, \ldots, n$, be the tangent vector to the manifold, the inner product between two basis vectors is defined by the metric tensor $g_{kl}\left(\boldsymbol{\theta}\right) = E\left(\frac{\partial}{\partial\theta_k}p\left(f;\boldsymbol{\theta}\right)\frac{\partial}{\partial\theta_l}p\left(f;\boldsymbol{\theta}\right)\right)$. The matrix $[g_{kl}]$ is the well known *Fisher information matrix*.

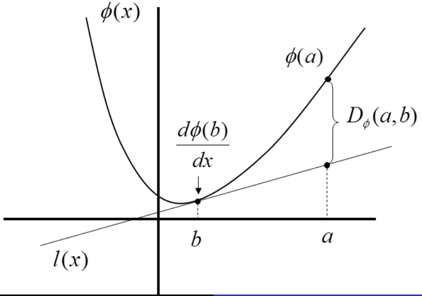


Similarity measure and Divergence (Riemannian manifold)

Bregman 1967, Csiszar 1974, Amari 1984, Tsallis 1998 ...

Definition

The Bregman divergence is defined as follow
$D_{\phi}(p \parallel q) = \phi(p) - \phi(q) + \langle p - q, \nabla\phi(q) \rangle$ for any strictly convex function ϕ.



Bregman divergences

Domain	$\phi(\mathbf{x})$	$d_\phi(\mathbf{x},\mathbf{y})$	Divergence
$\mathbb{R}$	x^2	$(x-y)^2$	Square loss
$\mathbb{R}_+$	$x \log x$	$x \log(\frac{x}{y}) - (x-y)$	
$[0,1]$	$x \log x + (1-x)\log(1-x)$	$x \log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y})$	Logistic loss [3]
$\mathbb{R}_{++}$	$-\log x$	$\frac{x}{y} - \log(\frac{x}{y}) - 1$	Itakura-Saito distance
$\mathbb{R}$	e^x	$e^x - e^y - (x-y)e^y$	
$\mathbb{R}^d$	$\Vert\mathbf{x}\Vert^2$	$\Vert\mathbf{x}-\mathbf{y}\Vert^2$	Squared Euclidean distance
$\mathbb{R}^d$	$\mathbf{x}^T A\mathbf{x}$	$(\mathbf{x}-\mathbf{y})^T A(\mathbf{x}-\mathbf{y})$	Mahalanobis distance [4]
d-Simplex	$\sum_{j=1}^d x_j \log_2 x_j$	$\sum_{j=1}^d x_j \log_2(\frac{x_j}{y_j})$	KL-divergence
$\mathbb{R}^d_+$	$\sum_{j=1}^d x_j \log x_j$	$\sum_{j=1}^d x_j \log(\frac{x_j}{y_j}) - \sum_{j=1}^d (x_j - y_j)$	Generalized I-divergence

Properties of the Bregman divergence

If close-form for the divergence for θ,
$D_\phi(\theta_1 \parallel \theta_1) \geq 0$
$D_\phi(\theta_1 \parallel \theta_2) = 0 \text{ iff } \theta_1 \sim \theta_2$
$D_\phi(\theta + d\theta \parallel \theta) \approx \frac{1}{2} \sum g_{kl}(\theta)\, d\theta_k d\theta_l$

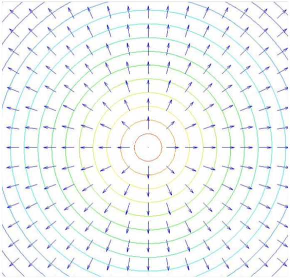
Warning Right-Left divergence: $D_\phi(\theta_1 \parallel \theta_2) \neq D_\phi(\theta_2 \parallel \theta_1)$

In general (not the case for exponential familly with natural parameters), Pythagorean theorem is

$$D_\phi(p \parallel q) \leq D_\phi(p \parallel r) + D_\phi(r \parallel q)$$

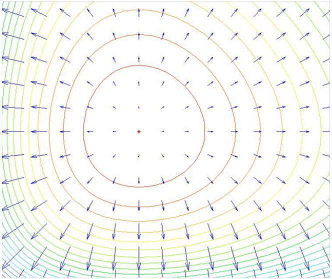
Specific geometry (Fisher Matrix)

Gaussian Distribution



$$d_{\phi}(x,\mu) = \|x-\mu\|^2$$

Exponential Distribution



$$d_{\phi}(x,\mu) = \frac{x}{\mu} - \log\frac{x}{\mu} - 1$$

Geodesic ditance

Remark: The Taylor expansion of the Kullback-Leibler divergence is the geodesic distance.

$$GD\left(\theta_1,\theta_2\right)=\int_{\theta_1}^{\theta_1}\mathrm{d}s=\int_0^1\sqrt{\sum_{\mu,\nu}g_{\mu\nu}\dot{\theta^{\mu}}\dot{\theta^{\nu}}}\mathrm{d}t$$

Maximum Likelihood and KL right side

If p is an empirical distribution (i.e., a set of samples f_i), choosing q that minimizes $KL_R(p||q)$ with q constrained to be a distribution in a parametric model $\boldsymbol{\theta}$ is equivalent to maximum likelihood estimation.

Consequence: in the parametric framework, for classification task we have $c^*(f) = \underset{c \in K}{arg\,max} \left[p\left(f | \theta_c\right)\right] = \underset{c \in K}{arg\,min} \left[KL_R\left(\theta_f \parallel \theta_c\right)\right].$

MaxEnt versus KL_L left side

If L_j is a set of emperical features (moments), choosing p that minimizes $KL_L(p||q)$ with q a specific distribution leads to a close form to the maximum entropie estimate (if q is the uniform is exaclty the MaxEnt).

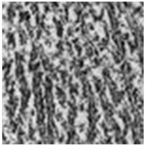
$max\left(H\left(p\right)\right)$ st
$\sum p(f) = 1$
$p(f) > 0$
$\int r_j\left(f\right)p(f)df = L_j$

$$p(f) \sim \exp\left(\sum \lambda_j r_j\left(f\right)\right)$$

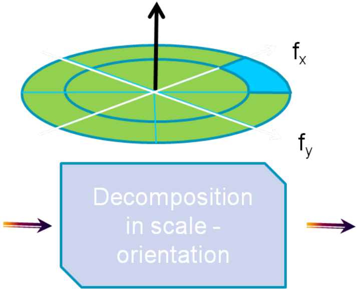
$minD\left(p \parallel q\right)$ st
$\sum p(f) = 1$
$p(f) > 0$
$\int r_j\left(f\right)p(f)df = L_j$

$$p(f) \sim \exp\left(\sum \lambda_j r_j\left(f\right)\right)q\left(f\right)$$

Scale and orientation decompositon



$X_c(m,n)$



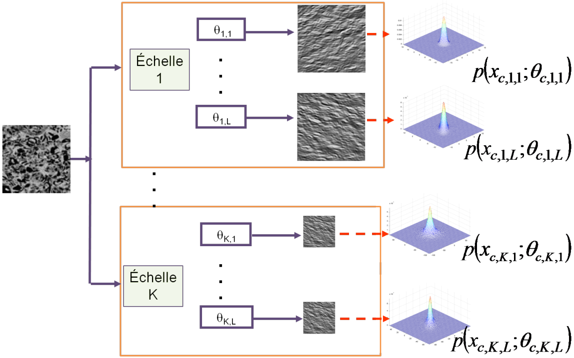
$\{\mathbf{x}_{c,k,l}(m,n)\}$

Wavelets real/complex

- Gabor
- Steerable filters
- Bandelets
- Grouplets
- Dual-Tree

(m,n) spatial indexes
c: color channel
k: scale index
l: orientation index

Texture modeling

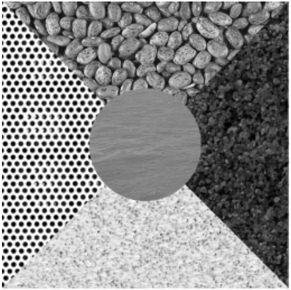


$$\Longrightarrow d\left(p_{\Delta}\left(\mathbf{f}_1, \theta_1\right) \| p_{\Delta}\left(\mathbf{f}_2, \theta_2\right)\right)$$

Classification or indexing texture bases

Commun databases for evalution of proposed modeling (Vistex, Brodatz, Outex ...)

Segmentation issue



Example of test image for evaluating texture segmentation.

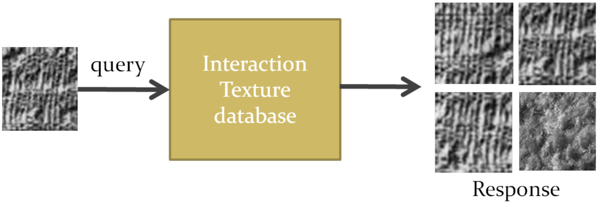
Previous works

Proposed parametric models

- Gaussian [Unser 1995, Manjunath 1996]
- Generalized Gaussian density (GG) [Do 2002]
- Bessel K forms (BKF) [Srivastava 2002]
- Gamma [Mathiassen 2002]
- Weibull [Kwitt 2008, 2010]
- Generalized Gamma [Drissi 2010]

Remark: all of them are not within the exponential family (Natural parameters)

Indexation issue



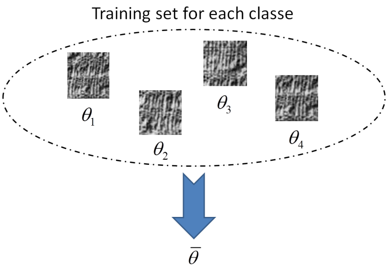
A query = L *best samples* in the database

$$\Longrightarrow [f_1^*, \ldots, f_L^*] = \min_{Database} \left[D\left(p_\Delta\left(\theta_q\right), p_\Delta\left(\boldsymbol{\theta}_{\mathbf{Database}}\right)\right)\right]$$

$$D\left(.\right) = \sum_{ij} KL\left(\theta_f^{ij} \parallel \theta_{DataBase}^{ij}\right) \text{ for}$$

$$i = 1..Nscale, \quad j = 1..Norientation$$

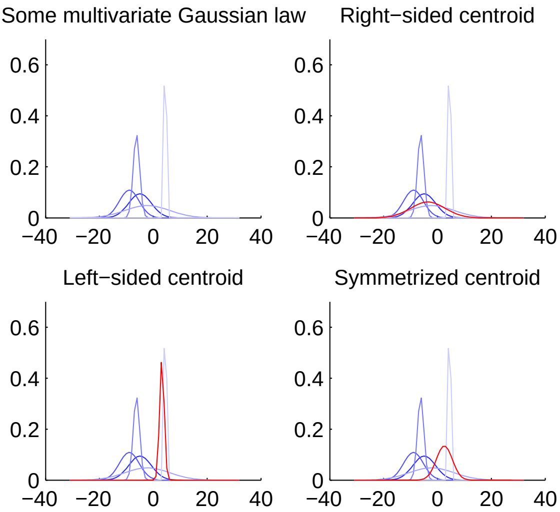
Barycentric law for clustering



The barycenter, i.e. $\overline{\theta}$, must to be conformed to the geometry of the manifold induced by (α, β).

$$\text{Barycenter: } \overline{\theta} = \underset{\theta \in F}{arg\ min} \left[\sum_{j=1..4} D\left(\theta_j, \overline{\theta}\right) \right]$$

Left, Right and symmetrized



Generalized Gaussian density

Mallat, Do&Vetterli, Portilla, Simoncelli ...

$$p(f) \longrightarrow \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{|f|}{\alpha}\right)^{\beta}\right) \text{ with } \boldsymbol{\theta} = (\alpha, \beta)^t \in \mathcal{M} = (\mathbb{R}_+^*)^2$$

- Kullback-Leibler
  $$KL(p_1 \| p_2) = \log\left(\frac{\beta_1 \alpha_2 \Gamma(1/\beta_2)}{\beta_2 \alpha_1 \Gamma(1/\beta_1)}\right) - \frac{1}{\beta_1} + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)}$$
- Estimate based on Maximum Likelihood (Do 2001)

Convex form and Newton approach

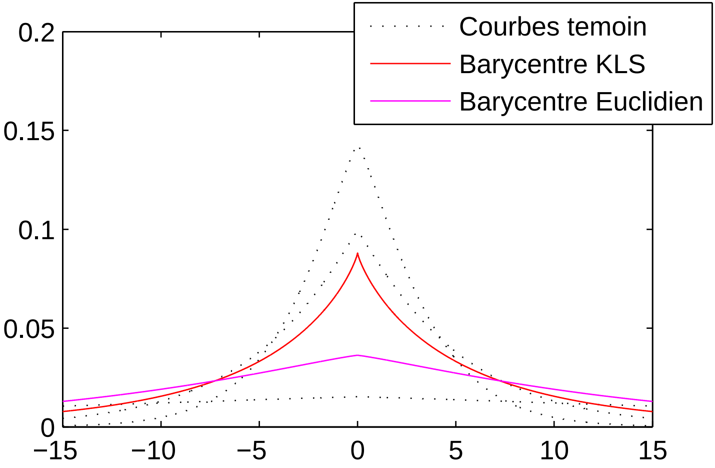
Let $\widetilde{\theta} = \{\theta_1, ..., \theta_K\}$ be the set of K observed models for a given subband, the barycentric model is given by:

$$\overline{\theta} = \underset{\theta \in F}{argmin} \left(\sum_j D\left(\theta_j, \overline{\theta}\right) \right) \text{ with}$$

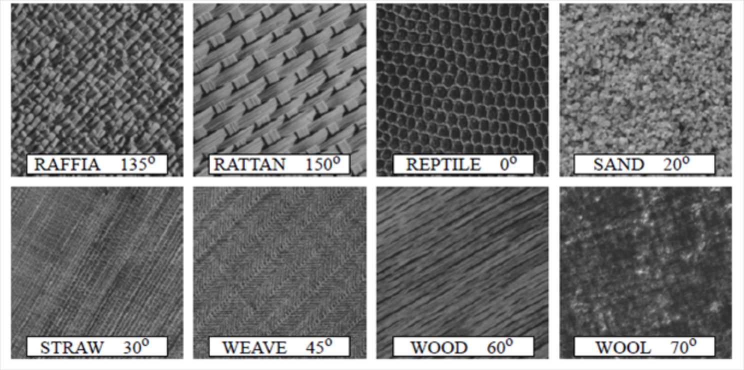
$$D\left(\theta, \overline{\theta}\right) = \frac{1}{2}\left(KL\left(\theta \parallel \overline{\theta}\right) + KL\left(\overline{\theta} \parallel \theta\right)\right)$$

Iterative approach: $\overline{\theta}_{k+1} = \overline{\theta}_k + \varepsilon \left[g_{ij}\right]^{-1} \nabla_\theta \left(D\left(\theta_j, \overline{\theta_k}\right)\right)$

Example


0.2
0.15
0.1
0.05
0
−15
−10
−5
0
5
10
15
Courbes temoin
Barycentre KLS
Barycentre Euclidien

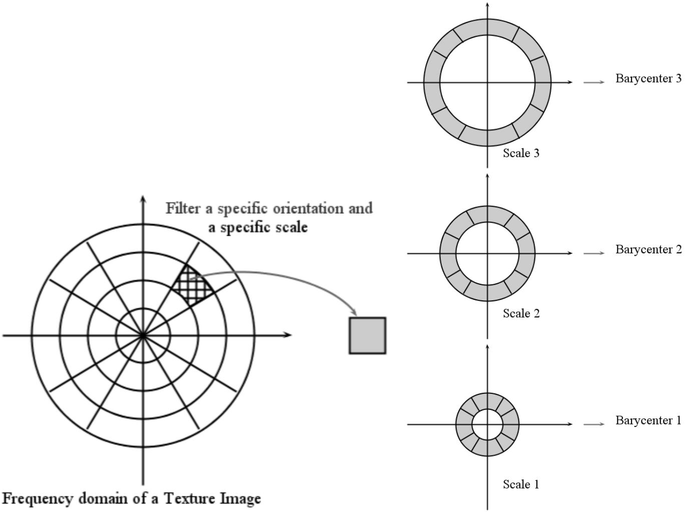
Invariance of rotation



Consider a database with non-rotated and rotated textures

Comparing subband by subband is not invariante.

Barycenter by scale

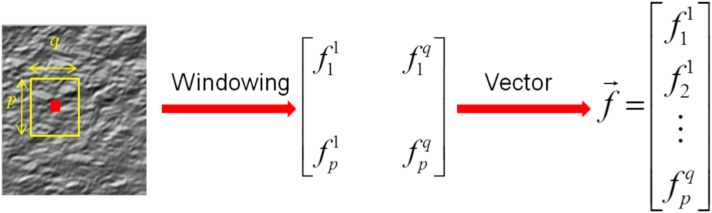


Résultats

Mean percentage of well-classified images

	Brodatz
Indiv. Subband	68%
Right Barycenter	85%
Symmetrized Baryc.	89%

Spatial dependance



Modeling the spatial correlation

Non-Gaussian densities

- Copulas
  - $p(\mathbf{f}, \theta) = c(P_1(f_1), ..., P_{pq}(f_{pq}), \boldsymbol{M}) \prod_{i=1..pq} p(f_i, \boldsymbol{\lambda})$
  - Covariance matrix $\boldsymbol{M}$ and $\boldsymbol{\lambda}$ the marginal parameters
- Elliptical density
  - $p(\mathbf{f}, \theta) = \frac{1}{C} h_{\boldsymbol{\lambda}} \left[(\mathbf{f})^T \boldsymbol{M}^{-1} \mathbf{f} \right]$
  - Covariance matrix $\boldsymbol{M}$ and $\boldsymbol{\lambda}$ the parameters of the elliptical generator

Gaussian copula

Definition

Sklar's Theorem 1959 -

Let $P(f_i)$ be the continuous marginal (cumulative) distributions, there exists a unique pq-copula such that:

$$p(\mathbf{f}, \boldsymbol{\theta}) = c\left(P_1(f_1), ..., P_{pq}(f_{pq}), \boldsymbol{M}\right) \prod_{i=1..pq} p(f_i, \boldsymbol{\lambda}).$$

A Gaussian copula is defined by:

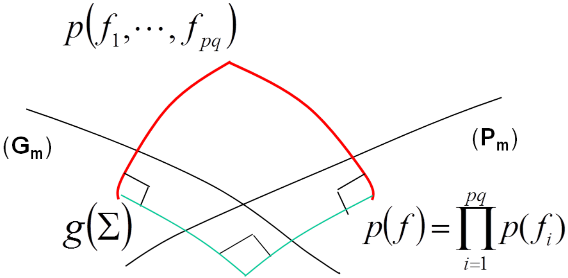
$$c(\mathbf{u}) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{g}^T(\mathbf{M}^{-1} - \mathbf{I_{pq}})\mathbf{g}}{2}\right) \text{ with } g_i = \Phi^{-1}(u_i) \text{ where}$$

$\Phi(.)$ is the cumulative function of the Gaussian density.

Probabilistic discrepancy (Kullback-Leibler)

For the set of orientation and scale subbands, we have:

$$d\left(p_\Delta(\mathbf{f_1},\boldsymbol{\theta_1}) \,\|\, p_\Delta(\mathbf{f_2},\boldsymbol{\theta_2})\right) = \underbrace{\sum_{i=1..pq} KL\left(p_1(f_i), p_2(f_i)\right)}_{\text{Marginal Part}} + \underbrace{\frac{1}{2}\left[trace\left(M_2^{-1}M_1\right) + log\left|\frac{M_2}{M_1}\right| - pq\right]}_{\text{Dependance Part}}$$



Elliptical profiles

- Joint Generalized Gaussian density
  - $p(\mathbf{f}|\mathbf{M}, m, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} h_{m,\beta}\left(\mathbf{f}^T \mathbf{M}^{-1}\mathbf{f}\right)$ with the density generator $h_{m,\beta}(x) = \frac{\beta\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{2\beta}\right) 2^{\frac{p}{2\beta}}} \frac{1}{m^{\frac{p}{2}}} \exp\left(-\frac{|x|^\beta}{2m^\beta}\right)$
- Joint student-t density
  - $p(\mathbf{f}|\mathbf{M}, m, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} h_{m,\beta}\left(\mathbf{f}^T \mathbf{M}^{-1}\mathbf{f}\right)$ with the density generator
  $h_{m,\beta}(x) = \frac{1}{(2\pi)^{\frac{pq}{2}}} \frac{(\beta m)^\beta}{\Gamma(\beta)} \Gamma\left(\frac{pq}{2} + \beta\right) \times \left(\frac{x}{2} + \beta m\right)^{-(\beta + \frac{pq}{2})}$

Parameter estimation

By differentiating the log-likelihood of vectors $(f_1, \dots, f_{pq})$ with respect to $\mathbf{M}$, the maximum likelihood estimator (MLE) of the matrix $\mathbf{M}$ denoted as $\hat{\mathbf{M}}$ satisfies the following fixed point (FP) equation

$$\hat{\mathbf{M}} = \frac{2}{N} \sum_{i=1}^{N} \frac{-g_{m,\beta}(\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i)}{h_{m,\beta}(\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i)} \mathbf{x}_i \mathbf{x}_i^T \text{ with } g_{m,\beta}(y) = \partial h_{m,\beta}(y) / \partial y$$ [1]

No-closed form for this kind of model, we propose Geodesic distance with linear approximation.

[1] Joint work with Frédéric Pascal (Supelec/Orsay) and Jean-Yves Tourneret (IRIT/Toulouse)

Classification results

TABLE I
AVERAGE RETRIEVAL RATES (%) IN THE TOP 16 MATCHES USING ORTHOGONAL WAVELET TRANSFORM WITH DAUBECHIES FILTER DB4 AND DUAL TREE COMPLEX WAVELET TRANSFORM WITH EB1 (VISTEX)

Type of Transform	Models			
	GG	Wbl	GC-MGG	GC-MWbl
1 scale				
OWT, db4	70.5176	69.3652	**79.7754**	75.8105
DT-CWT	72.8906	73.1738	**81.6602**	77.5879
2 scales				
OWT, db4	76.4160	75.9180	**81.9434**	79.6094
DT-CWT	78.7402	79.6289	**83.7012**	82.3633

TABLE III
AVERAGE RETRIEVAL RATES (%) FOR MULTIVARIATE MODELS IN THE TOP 16 MATCHES USING ORTHOGONAL WAVELET TRANSFORM WITH DAUBECHIES FILTER DB5 AND DUAL TREE COMPLEX WAVELET TRANSFORM WITH *EB1*.

Type of Transform	MG	MGmix	GC-MGG	GC-MWbl
1 scale				
OWT, db5	62.3828	72.1387	**79.5703**	75.1758
DT-CWT	65.7129	78.0371	**81.6602**	77.5879
2 scales				
OWT, db5	70.1660	78.7402	**82.0508**	80.0781
DT-CWT	71.2695	81.8262	**83.7012**	82.3633

The segmentation issue

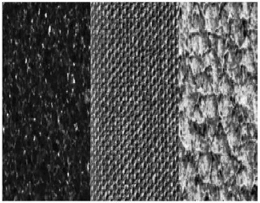
Main ingredients (suppose models associated to the class)

- Label field (Pott's model with K components)
  - $$p\left(x_i = k\right) = \frac{\exp\left(-\sum_{j\in\triangle_i} \beta\delta\left(x_i \neq x_j\right)\right)}{\sum_{k=1..K} \exp\left(-\sum_{j\in\triangle_i} \beta\delta\left(k_i \neq x_j\right)\right)}$$
- Using the SoftMax principle (Sernov's theorem):
  - $$p\left(f_i | x_i = k\right) = \frac{\exp\left(-\sum_{j=1..N_{Subbands}} KLS\left(\theta_j, \theta_k^{Ref}\right)\right)}{\sum_{l=1..K} \exp\left(-\sum_{j=1..N_{Subbands}} KLS\left(\theta_j, \theta_l^{Ref}\right)\right)}$$

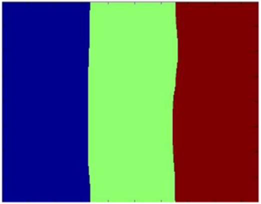
Optimization: Iterative Conditional Mode (ICM)

- $\hat{x}_i \leftarrow \underset{k}{argmax} \left[log\left(p\left(f_i | x_i = k\right)\right) + \lambda log\left(p\left(x_i = k\right)\right)\right]$

Results



Textured image



Segmentation

	GG model	GC-MGG
% Pixel miss-classified	4%	0.97%