Markov random fields for texture recognition with local invariant regions and their geometric relationships

Juliette Blanchet, Florence Forbes & Cordelia Schmid

INRIA Rhône-Alpes (Teams Lear and Mistis), 655, avenue de l'Europe – Montbonnot – 38334 Saint-Ismier Cedex – France

Abstract

We describe a new probabilistic framework for recognizing textures in images. Images are described by local affine-invariant descriptors and their spatial relationships. We introduce a statistical parametric models of the dependence between descriptors. We use Hidden Markov Models (HMM) and estimate the parameters with a recent technique based on the mean field principle. Preliminary results for texture recognition are promising and outperform existing techniques.

I Introduction

Context : **local photometric descriptors** computed for invariant interest regions have proved to be very successful in applications such as object recognition [6], texture classification [3] and texture recognition [4]. They are distinctive, robust to occlusions and invariant to image transformations. Their geometric organization is very informative but **modelling their relative spatial organization** is still an open issue.

Previous work : in [4] neighborhood statistics are modeled by co-occurrence of descriptors and included into the recognition step based on relaxation [7]. It does not use an explicit organizational model of the data during learning. Similarly, [5, 8] use features augmented with spatial information based on a a two-level scheme : (1) intensity-based textons, (2) histograms of textons distributions over local neighborhoods. **No spatial model is explicitly assumed** so that the neighborhood information captured is somewhat weakened.

Our approach : our claim is that there is some gain in assuming that the feature vectors are dependent statistical variables and consequently in using parametric statistical models to account for this dependencies explicitly.
We show that recognition can be improved by using a **Hidden Markov Model (HMM) as organizational model** when learning the texture classes. The parameter estimation of such a model is in this context not trivial. We use recent estimation procedures based on the Expectation-Maximization (EM) algorithm and on the mean field principle of statistical physics [2].

II Hidden Markov Models for textures

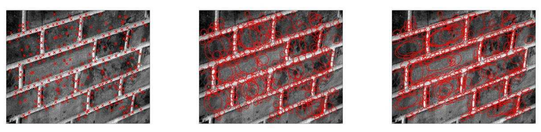
1. Feature extraction

Features are local characteristics : compact and complete image description, information about shape and local structure, invariant to various transformations.

Extraction in 2 steps :

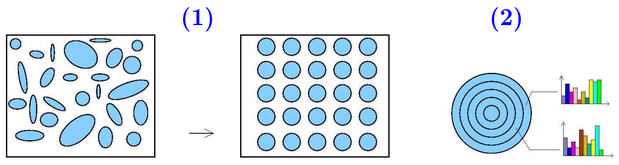
– Interest point detection (Laplacian) at some specific signal changes : **(1)** find locations in scale space where a normalized Laplacian measure attains a local maximum, **(2)** associate to each interest point a circular blob with radius=scale, **(3)** turn circles into ellipses (affine adaptation)

(1) (2) (3)



– Descriptor computation (Spin images) : **(1)** turn ellipses into unit circles, **(2)** two dimensional histogram of the intensity ⇒ affine invariant descriptors : multi-dimensional vectors (80)

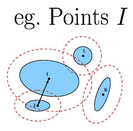
(1) (2)



2. Modelling textures

Spatial dependencies : Descriptors are spatially dependent ⇒ Statistical model that captures spatial relationships between descriptors ⇒ Hidden Markov Random Fields (HMRF)

Neighborhood graph : eg. Points I and J are neighbours



Markov Random Field : $Z = \{Z_i, i \in \mathcal{S}\}$ discrete field, $\mathcal{N}(i)$ neighbors of site i

$$\forall \mathbf{z},\ \begin{cases} P(z_i|\mathbf{z}_{\mathcal{S}\setminus\{i\}}) = P(z_i|\mathbf{z}_{\mathcal{N}(i)}) \\ P(\mathbf{z}) > 0 \end{cases} \Leftrightarrow \forall \mathbf{z},\ \ P(\mathbf{z}) = W^{-1}\exp(-H(\mathbf{z})) \quad \text{(gibbs measure)}$$

where : $H(\mathbf{z}) = \sum_i V_i(z_i) + \sum_{i\sim j} V_{ij}(z_i, z_j)$ (energy function)

$W = \sum_{\mathbf{z}} \exp(-H(\mathbf{z}))$ (normalizing constant, intractable)

Hidden Markov Random Field : taking observations into account

Incomplete data $\begin{cases} \text{Observations} & \mathbf{X} = \{X_i, i \in S\} \text{ measures} \\ \text{Hidden data} & \mathbf{Z} = \{Z_i, i \in S\} \text{ labels, discrete MRF, } Z_i \in \mathcal{L} = [1, ..., K] \end{cases}$

$\longrightarrow$ posterior MRF : $P_G(\mathbf{z}|\mathbf{x}) = W_H^{-1}\exp(-H_x(\mathbf{z}))$ with $H_x(\mathbf{z}) = H(\mathbf{z}) + H(\mathbf{x}|\mathbf{z})$

$H(\mathbf{z})$: regularizing term (prior, constraints satisfaction), $H(\mathbf{x}|\mathbf{z})$: likelihood term

Recovering the unknow labels by $\hat{z}$: MAP solution : $\hat{\mathbf{z}} = \arg\min_{\mathbf{z}\in\mathcal{L}^n} H_x(\mathbf{z})$

HMRF for textures : each texture class is a "mixture" of K sub-classes

For an image of known texture m

$\begin{cases} \text{Descriptors} & \mathbf{X} = \{X_i, i \in S\} \\ \text{Texture } m \text{ sub-classes (hidden)} & \mathbf{Z} = \{Z_i, i \in S\} \text{ discrete MRF, } Z_i \in [c_{m1}, ..., c_{mK}] \\ X_i \text{ conditionnally independent given } & \mathbf{Z} \end{cases}$

$$p(x_i|\mathbf{\Psi}_m) = \sum_{k=1}^{K} P(Z_i = c_{mk}|\mathbf{\Delta}_m) f(x_i|\mathbf{\Theta}_{mk})$$

– $f(x_i|\mathbf{\Theta}_{mk})$ multivariate Gaussian distribution
– Singleton potentials : $V_i(z_i) = -\alpha_m(k)$ if $z_i = c_{mk}$
– Pair-wise potentials : $V_{ij}(z_i, z_j) = -b_m(k, l)$ if $z_i = c_{mk},\ z_j = c_{ml}$
If $(b_m) = \beta \times I$, Potts model of image segmentation
– $\mathbf{\Delta}_m = (\alpha_m(k), b_m(k, l))$ and $\mathbf{\Psi}_m = (\mathbf{\Theta}_{m1}, ..., \mathbf{\Theta}_{mK}, \mathbf{\Delta}_m)$

III Learning the distribution of descriptors and their organization

1. Estimating parameters

Each texture m is associated to a K sub-class HMRF model :
– Parameters $\mathbf{\Psi}_m = (\mathbf{\Theta}_{m1}, ..., \mathbf{\Theta}_{mK}, \mathbf{\Delta}_m)$ need to be estimated

– EM algorithm commonly used in problems with hidden data
Principle : $\max_{\mathbf{\Psi}_m} \log p(\mathbf{x}|\mathbf{\Psi}_m)$ replaced by $\max_{\mathbf{\Psi}_m} I\!E[\log(p(\mathbf{x}, \mathbf{Z}|\mathbf{\Psi}_m))|\mathbf{x}, \mathbf{\Psi}_m]$
Iterative procedure : at iteration (q) estimator $\mathbf{\Psi}_m^{(q)}$ of $\mathbf{\Psi}_m$

– For HMRF, because of the spatial dependencies, EM intractable (W...)
⇒ Mean field approach to approximate by independent variables

2. Mean Field principle for MRF

– Neglecting fluctuations from the mean in the neighborhood of each site
For site $i : \forall j \in \mathcal{N}(i),\ \ z_j$ fixed to $I\!E[Z_j] = \mu_j$
⇒ system of independent variables $P(\mathbf{z}) \approx \prod_i P(z_i|\mu_{\mathcal{N}(i)})$
– Generalization : set the neighbors to constants $\tilde{\mathbf{z}}$ $P(\mathbf{z}) \approx \prod_i P(z_i|\tilde{z}_{\mathcal{N}(i)})$
– Mean field principle for $\mathbf{Z}|\mathbf{x}$ (MRF) $P(\mathbf{z}|\mathbf{x}) \approx \prod_i P(z_i|\tilde{z}_{\mathcal{N}(i)}, \mathbf{x})$

3. EM-like algorithm

Iteration (q) in two steps :
– **(1).** Create $\tilde{\mathbf{z}}^{(q)}$ from $\mathbf{x}$ and $\mathbf{\Psi}^{(q-1)}$

$\tilde{\mathbf{z}}^{(q)} =$	current conditionnal mean	mean field algorithm
$\tilde{\mathbf{z}}^{(q)} =$	current conditionnal mode	mode field algorithm
$\tilde{\mathbf{z}}^{(q)} =$	simulation	simulated field algorithm

– **(2).** Apply EM for the factorized model to get updated $\mathbf{\Psi}^{(q)}$

Learning step : For each texture m, apply an EM-like algorithm
⇒ estimators of $\mathbf{\Psi}_m$

IV Classification and retrieval

1. Modelling an image of unknow texture

An image of unknown texture is a "mixture" of M textures ⇒ $M \times K$ sub-classes

$\begin{cases} \text{Descriptors} & \mathbf{X} = \{X_i, i \in S\} \\ \text{Hidden data} & \mathbf{Z} = \{Z_i, i \in S\} \text{ discrete MRF, } Z_i \in [c_{11}, ..., c_{MK}] \\ X_i \text{ conditionnally independent given } & \mathbf{Z} \end{cases}$

$$p(x_i|\mathbf{\Psi}) = \sum_{m=1}^{M}\sum_{k=1}^{K} P(Z_i = c_{mk}|\mathbf{\Delta}) f(x_i|\mathbf{\Theta}_{mk})$$

– $f(x_i|\mathbf{\Theta}_{mk})$ Gaussian with $\mathbf{\Theta}_{mk}$ learned
– Hidden field : singleton potentials : (α_m), pair-wise potentials : $MK \times MK$ matrix learned

$$\hat{\alpha} = \begin{pmatrix} (\alpha_1) \\ \vdots \\ (\alpha_M) \end{pmatrix} \quad \text{and} \quad \hat{B} = \begin{pmatrix} (b_1) & \dots & -1 \\ -1 & \ddots & -1 \\ -1 & -1 & (b_M) \end{pmatrix}$$

Retrieval : at site i, with all parameters fixed (learned) :

– a site i is assigned to texture m maximizing $\sum_{k=1}^{K} P(Z_i = c_{mk}|x_i)$

– if single-texture, the image is assigned to texture m maximizing $\sum_{i\in\mathcal{S}}\sum_{k=1}^{K} P(Z_i = c_{mk}|x_i)$

But : $P(Z_i = c_{mk}|x_i)$ is intractable. ⇒ EM-like algorithm to estimate $P(Z_i = c_{mk}|x_i) = p_i(m, k)$ for $m = 1, \dots M,\ k = 1, \dots K$

Initialization : textures are equally likely
$p_i^{(0)}(m, k) = \frac{1}{M}\hat{P}_{learn}(c_{mk}|x_i)$ (learned with EM or "spatial" EM, section III)
Updating (our approach) : using the simulated field algorithm :

$$p_i^{(q+1)}(m, k) \propto f(x_i|\hat{\Theta}_{mk}) \exp(-\hat{\alpha}_m(k) - \sum_{j\in\mathcal{N}(i)}\sum_{m'=1}^{M}\sum_{l=1}^{K} \hat{B}(mk, m'l)\tilde{z}_j^{(q)}(m', l))$$

Alternative : using the NEM algorithm :

$$p_i^{(q+1)}(m, k) \propto f(x_i|\hat{\Theta}_{mk})\pi_{mk} \exp(-\sum_{j\in\mathcal{N}(i)}\sum_{m'=1}^{M}\sum_{l=1}^{K} \hat{B}(mk, m'l)\, p_j^{(q)}(m', l))$$

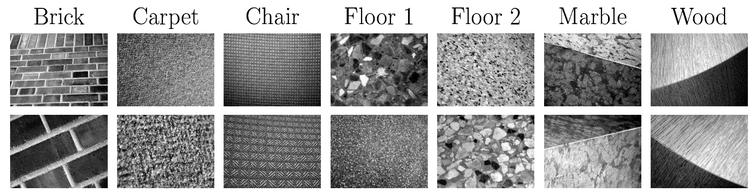
Alternative : using Relaxation (prior energy based, no data, no model !) :

$$p_i^{(q+1)}(m, k) \propto p_i^{(q)}(m, k)\,(1 - \sum_{j\in\mathcal{N}(i)}\sum_{m'=1}^{M}\sum_{l=1}^{K} \hat{B}(mk, m'l)\, p_j^{(q)}(m', l))$$

V Experimental results

1. Sample of 7 texture classes

Brick Carpet Chair Floor 1 Floor 2 Marble Wood



$7 \times 10 = 70$ images

2. Single texture images

10 images $\sim$ 4000 descriptors for each texture

Classification rates in % for individual regions

	Brick	Carpet	Chair	Floor1	Floor2	Marble	Wood
EM	48	77	52	56	50	17	30
Relaxation	78	96	72	86	80	19	42
NEM	82	98	78	88	80	20	43
Simulated F.	81	97	77	80	86	26	46

Significant gain :
– in incorporating spatial relationships (Relaxation, NEM, Simulated Field)
– in using a statistical parametric model (NEM, Simulated Field)

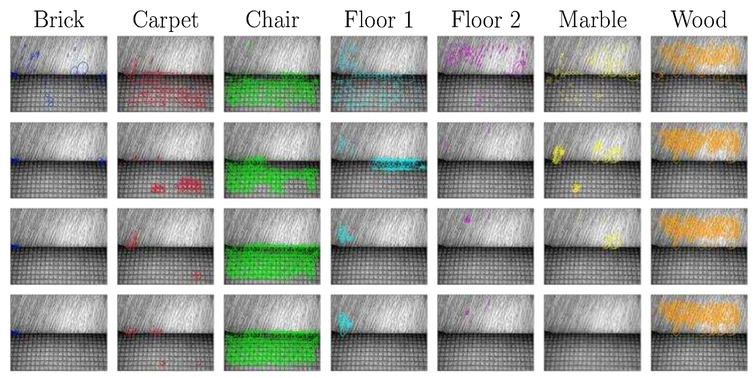
Variant using Delaunay neighborhood graph, image descriptors SIFT and a dimension reduction technique (Simulated Field) [1] :

	Brick	Carpet	Chair	Floor1	Floor2	Marble	Wood
EM	63	35	47	35	36	23	56
Relaxation	93	63	70	74	79	33	95
Simulated F.	99	63	85	88	89	76	98

3. Multi-Texture images

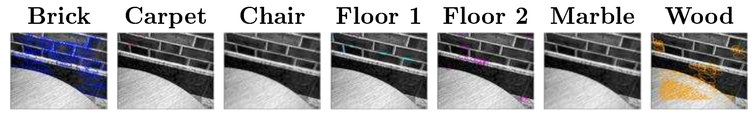
Various algorithms : Maximum likelihood, Relaxation, NEM and Simulated Field for an image composed of Chair and Wood

Brick Carpet Chair Floor 1 Floor 2 Marble Wood



Simulated Field on an image composed of Brick, Marble and Wood

Brick Carpet Chair Floor 1 Floor 2 Marble Wood



– Marble badly learned
– Simulated Field ⇒ homogeneous groups ⇒ classified in brick and wood

VI Conclusion

What we did :
– Hidden Markov Model for feature vectors at irregular locations (Generalization of Potts Model for textures)
– Gain in incorporating spatial relationships between descriptors, in using a statistical parametric model
– Statistically based model selection (BIC) : eg. choice of K

What remains to be done :
– Deviation from the model and data transformation
– Choice of the neighborhood structure and appropriate image descriptors
– Object recognition

Références

[1] C. Bouveyron, S. Girard and C. Schmid, "High Dimensional Discriminant Analysis". *Int. Conf. Applied Stochastic Models and Data Analysis*, pp. 526-534, Brest, France, May 2005.

[2] G. Celeux, F. Forbes and N. Peyrard, "EM procedures using mean-field like approximations for markov-model based image segmentation". *Pattern recognition*, 36 :131–144, 2003.

[3] S. Lazebnik, C. Schmid and J. Ponce, "Sparse Texture Representation Using Affine-Invariant Regions", *Proc. CVPR*, 2003.

[4] S. Lazebnik, C. Schmid and J. Ponce, "Affine-Invariant Local Descriptors and Neighborhood Statistics for Texture Recognition", *Proc. ICCV*, 2003.

[5] J. Malik, S. Belongie, T. Leung and J. Shi, "Contour and Texture Analysis for Image Segmentation", *IJCV*, 43(1), pp. 7-27, 2001.

[6] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector", *Proc. ECCV*, 2002.

[7] A. Rosenfel, R. Hummel and S. Zucker, "Scene Labeling by Relaxation operations", *IEEE Trans. Systems, Man, and Cybernetics*, 6(6), pp.420-433, 1976.

[8] C. Schmid, "Constructing Models for Content-Based Image Retrieval", *Proc. CVPR*, 2001.