Auto-models with mixed states and analysis of motion textures

Patrick Bouthemy, Cécile Hardouin, Gwenaelle Piriou, Jian-Feng Yao

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Auto-models with mixed states and analysis of motion textures

P. Bouthemy, C. Hardouin, G. Piriou, J. Yao

N°5454

Janvier 2005

Systèmes cognitifs

ISSN 0249-6399 ISRN INRIA/RR--5454--FR+ENG

Auto-models with mixed states and analysis of motion textures

P. Bouthemy, C. Hardouin*, G. Piriou, J. Yao†

Systèmes cognitifs
Projet VISTA

Rapport de recherche n˚5454 — Janvier 2005 — 21 pages

Abstract: In image motion analysis as well as for several application fields like daily pluviometry data modeling, observations contain two components of different nature. A first part is made with discrete values accounting for some symbolic information and a second part records a continuous (real-valued) measurement. We call such type of observations “mixed-state observations". In this work we introduce a generalization of Besag’s auto-models to deal with mixed-state observations at each site of a lattice. A careful construction as well as important properties of the model will be given. The performance of the model is then evaluated on the modeling of motion textures from video sequences.

Key-words: Mixed states, auto-models, dynamic textures, motion analysis

(Résumé : tsvp)

* SAMOS/Université de Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex 13, FRANCE
† IRMAR/Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, FRANCE

Auto-modèles à états mixtes et analyse de textures de mouvement

Résumé : Des données de mesure de mouvement dans une séquence d'images, tout comme des données de pluviométrie, comportent deux composantes de nature diffiérente. La première composante contient des valeurs discrètes qui véhiculent une information symbolique. La deuxième composante correspond à de véritables mesures de type continu. Nous appelons ces données *des observations à états mixtes*. Nous introduisons une extension des auto-modèles de Besag afin de modéliser ce type d'observations sur un réseau. La performance des mod èles proposés est ensuite évaluée en analyse de *textures de mouvement* issues des séquences vidéo.

Mots-clé : Etats mixtes, auto-modèles, texture dynamique, analyse du mouvement

1 Introduction

It is of common understanding that the type of any observation data is either continuous or discrete. The situation, where a measurement presents continuous values sometimes and discrete values at other times, is rarely considered in statistical literature. However, such situations are frequent in applications. For examples, daily pluviometry time series at a given site records many zeros when the rain is absent, followed by periods with positive rainfall values (see e.g. (Allcroft and Glasbey, 2003)). Similar phenomena also occur in speech recordings where, interchanges are permanent between absences and presences of the signal. Another example arises in the motion analysis problem from image sequences considered in this paper. Typically, the histograms of local motion measures present a composite picture. An important peak appears at the origin accounting for regions where no motion is present, while a large continuous component encompasses actual motion magnitudes in the images. It then raises the question to find accurate models for this type of data - we shall call them *observations with mixed states*, collected from the image lattice.

From a mathematical point of view, we are searching for models for a random field $\{X_s\}$ with the constraint that the marginal distributions of the X_s's are composed with a discrete component and a continuous component. In its most general form and for the discrete component, we may take any distribution with support on a countable set $\{e_1, \ldots, e_k, \ldots\}$ of symbolic values, while for the continuous component any standard distribution could be considered. However in this work, we will restrict our attention to distributions with one atomic value, typically $\{0\}$, and a continuous component supported on the interval $(0, \infty)$. The state space, called a *mixed-state space* is then $E = \{0\} + (0, \infty)$ with the point 0 playing a special role.

Markov random fields (MRF) models are now a standard tool in image analysis (Chalmond, 2003). However, up to our knowledge, the existing models deal with either continuous variables, or discrete variables, but never with variables that can take continuous as well as discrete values. Furthermore, the discrete component could not be simply neglected, because -as it will becomes clear in the motion analysis application addressed below-, these symbolic values as well as their spatial correlations convey important pixelwise and contextual information. On the other hand, such discrete phenomena are usually taken into account by introducing a label process L_s where, in our case, $L_s = 1$ if no motion is present at pixel s, i.e. $X_s = 0$, and $L_s = 2$ when a positive motion measurement is recorded, i.e. $X_s > 0$. However, the label process is a hidden process and the resulting statistical inference methods need in general a restoration of the hidden process (i.e. segmentation). This classical approach is then possible only upon the cost of a generally huge computation effort.

The approach we propose is different. We aim to give a model which automatically deal with the two types of observations, without introduction (and then the inference) of a hidden process. The main idea is then to introduce mixed-state variables (or distributions) in a random field set-up. More precisely, we will follow J. Besag's construction of auto-models (Besag, 1974) by introducing necessary adaptations for mixed-state variables.

In Section 2, we first recall some backgrounds on motion computation in an image sequence. This will also give a precise description of the data we have at hand, and in particular their mixed-state nature. In Section 3, we introduce a new class of random field models, named *auto-models with mixed states*, for modeling mixed-state observations. The construc-

tion, their basic properties as well as estimation methods will be given. In Section 4 we carry out an application to the analysis of motion textures from video sequences. Finally, concluding remarks are given in Section 5.

2 Motion analysis from image sequences

The data we are dealing with are local motion measurements from video sequences. For the definition and the computation of these measures, we follow the approach developed in (Odobez and Bouthemy,()OB95; ()OB97), where several meaningful applications are also given. Here, we briefly recall the basics of their computation.

2.1 Motion decomposition and the residual motion

The aim of motion computation is to obtain dense motion measures that can be easily and reliably computed from any image sequence (e.g. videos) and that inform on the dynamic content of the depicted scene. The motion information in an image sequence is completely captured by the optical flow. However, methods for estimating optical flow remain complex and time consuming if general video content has to be handled, while not always ensuring accurate and reliable measurements. As a consequence, motion measurements we will analyze in the sequel are related to normal flow. On one hand, they supply a partial motion information only, on the other hand, they can be locally computed in a straightforward way. In case of a static camera, these motion measures are directly related to the scene motion. If the camera is moving, we have first to cancel the camera motion. In that case, we will compute local motion features related to the residual normal flow.

More precisely, we first estimate the dominant image motion which can be assumed (in most cases) due to the camera motion. The following 2D-affine motion model (this is a usual choice, a 8-parameter quadratic model could be considered as well) is considered:

$$\mathbf{w}_\xi(p) = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix}, \tag{1}$$

where $\xi = (a_i, i = 1, \ldots, 6)$ is the model parameter vector and $p = (x, y)$ is an image point. This simple motion model can correctly handle different camera motions such as panning, zooming, tracking. Different methods are available to estimate such a motion model. We use the robust real-time multi-resolution algorithm described in (Odobez and Bouthemy, 1995).

Then, the residual motion measurements $v_{res}(p, t)$ we consider, are defined as the weighted local mean of the normal residual flow magnitudes $|v_n|$, the weights being given by the square of the magnitude of the spatial intensity gradient. This allows us to smooth out the noise attached to the computation of the normal flow and to enforce the reliability of the motion measurements. We get the following expression:

$$v_{res}(p, t) = \frac{\sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2 . |v_n(q, t)|}{\max\left(\eta^2, \sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2\right)}, \tag{2}$$

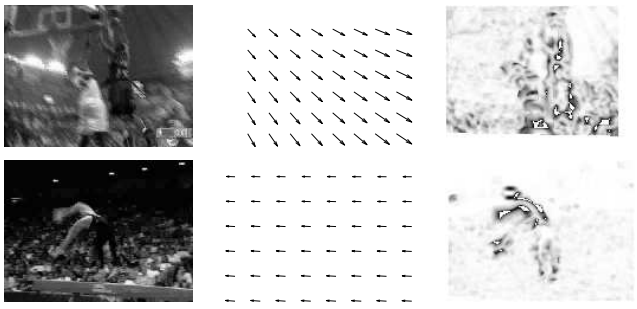

Figure 1: Two images of sport video segments (involving respectively, a zoom combined with an upward-tilt camera motion, and a right panning motion) and their corresponding maps of the estimated dominant image motion fields and of residual motion measurements v_{res} (white = 0; black = maximum value).

with $v_n(q,t) = [I(q,t) - I(q + w_{\hat{\xi}_t}(q), t+1)]/\|\nabla I(q,t)\|$. Here, $\mathcal{F}(p)$ is a local spatial window centered in pixel p (typically a 3×3 window), $\nabla I(q,t)$ is the spatial intensity gradient of pixel q at time t and η^2 is a predetermined constant related to the noise level. This class of local motion measurements have already been proved useful for motion detection (Irani et al., 1992; Odobez and Bouthemy, 1997) and for motion recognition (Fablet and Bouthemy, 2003). Figure 1 displays two images of sport videos with the corresponding maps of the estimated dominant motion vectors and those of residual motion measurements v_{res}. These examples show that the camera motion is reliably captured even in case of multiple moving elements in the scene. It also indicates that the scene motion is correctly accounted by the residual motion measurements. From Eq. (2), it can be straightforwardly noted that we only get information related to motion magnitude, and consequently, we lose the motion direction. However, under the general objective of motion characterization, we aim at addressing issues such as detecting similar motion contents, grouping "qualitative" motion classes, or recognizing predefined motion classes. As demonstrated by the results reported later, these goals can be attained using this type of motion information.

2.2 Motion textures, histograms and mixed states

The above motion computation principle has been applied to various video sequences. In this paper, we are mostly concerned with motions of *natural dynamic scenes*. We will call *motion textures* the resulting local motion measurements $\{v_{res}\}$. Figure 2 gives a set of sample images from six different types of videos including moving escalators, grass, foliage, trees or rivers and sea waves, respectively. The corresponding motion fields $\{v_{res}\}$ are displayed in Figure 3. As a matter of fact, these videos ar made with a static camera; therefore, the $\{v_{res}\}$ are computed with $\mathbf{w}_\xi \equiv 0$ in (1).

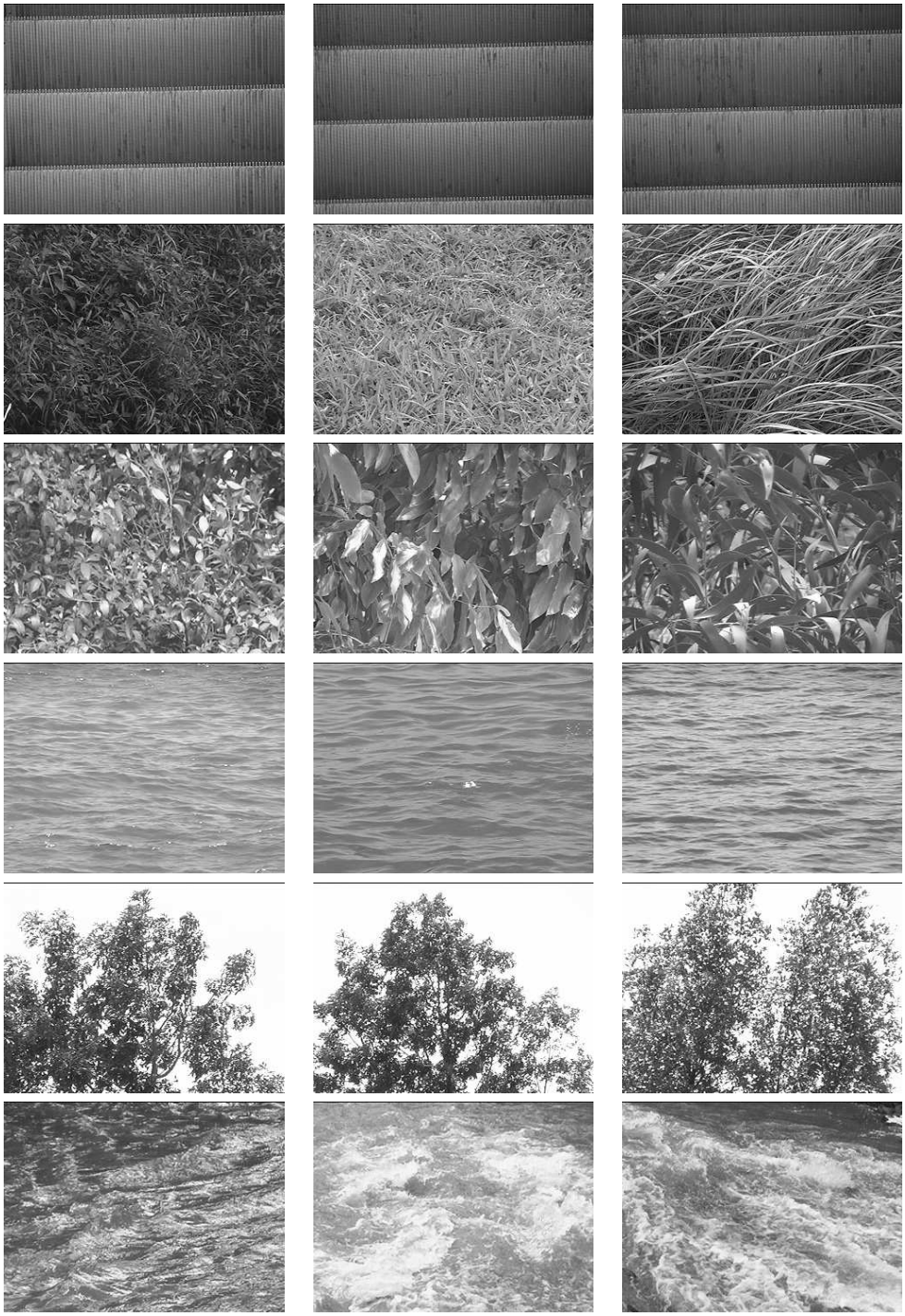

Figure 2: Sample images from different videos. Top to bottom: moving escalators, grass, foliage, sea waves, trees and rivers.

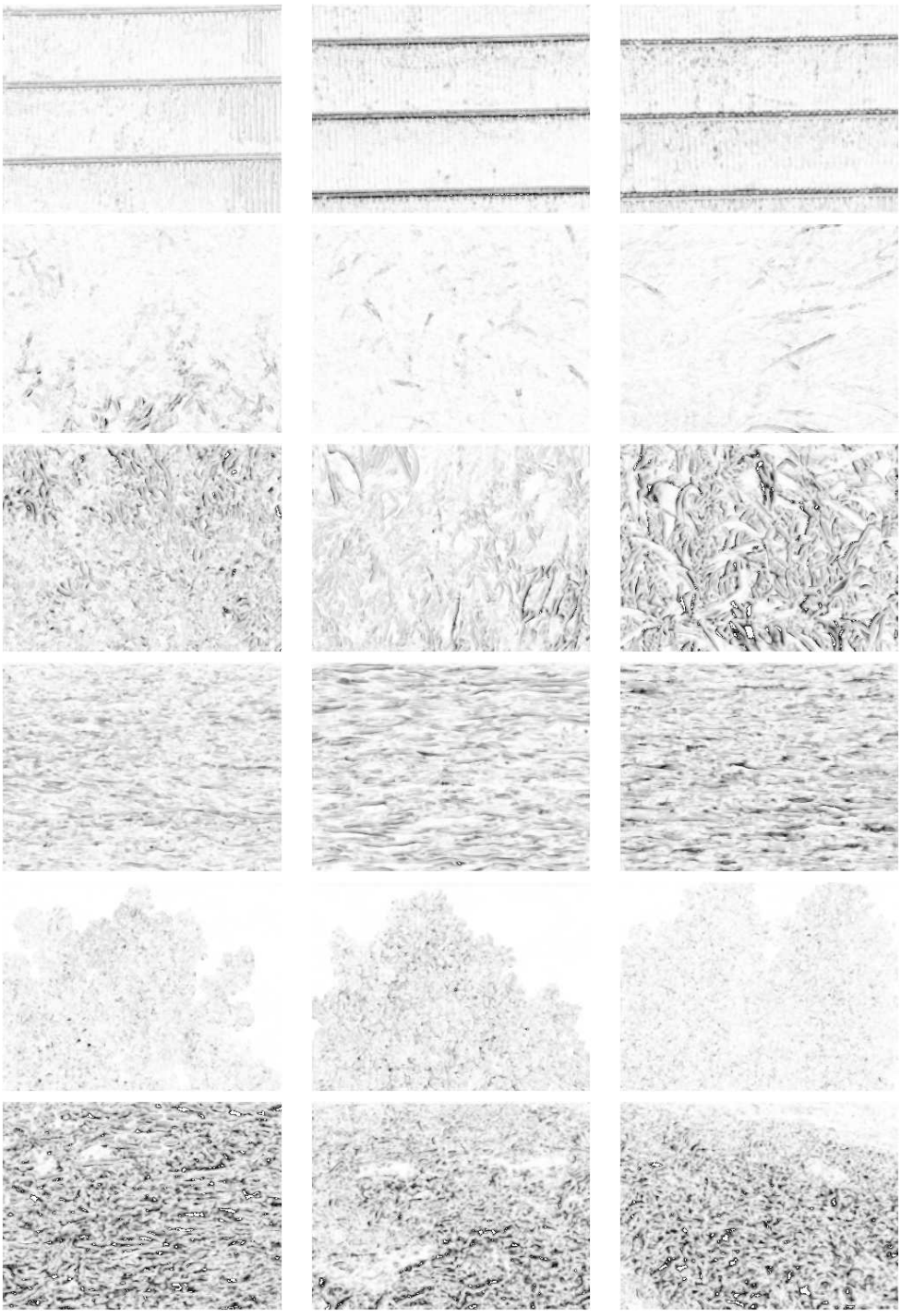

Figure 3: Sample motion measures $\{v_{res}\}$ from the videos of Figure 2. Top to bottom: moving escalators, grass, foliage, sea, trees and rivers (white=0; black=maximum value).

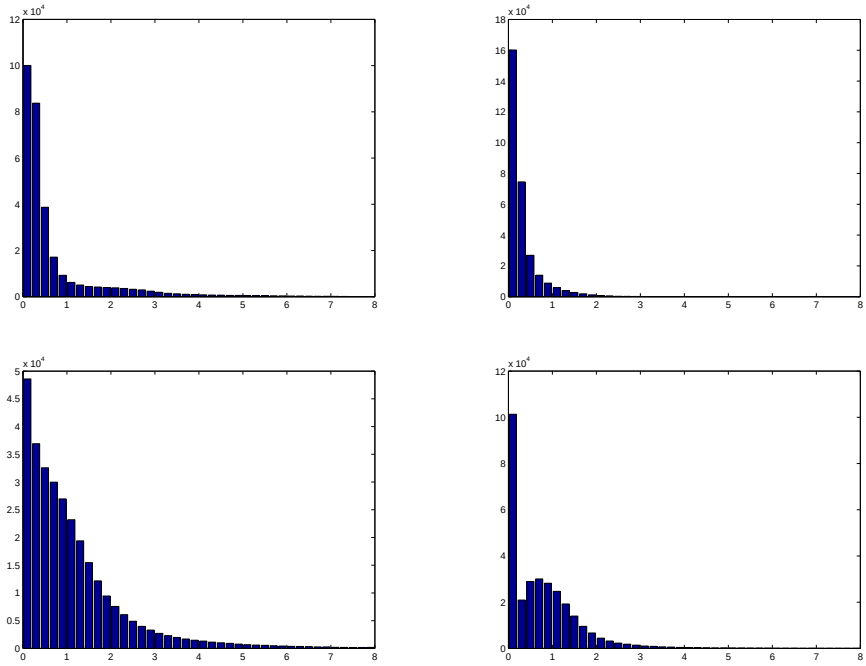

Figure 4: Sample histograms of motion measures $\{v_{res}\}$. Top to bottom, and left to right: moving escalators, grass, foliage and trees.

In Figure 4, we have displayed several typical histograms from motion textures $\{v_{res}\}$. As explained in Introduction, these histograms are of mixed-state type with a prominent peak at the origin accounting for regions where no motion is present, and a continuous component reporting the magnitudes of actual motion in the images.

3 Auto-models with a mixed state space

Our main purpose is to construct a random field model for mixed-state observations on a lattice, like the residual motion field v_{res} introduced in Section 2. Indeed we need first to define a general class of random field models called *multi-parameter auto-models*. This is done in Section 3.1. The mixed-state case is then treated in Section 3.2-3.3. In particular we obtain in Section 3.4 a class of "positive Gaussian" auto-models with the four nearest-neighbours system and whose local conditional distributions are of mixed-state type on the set $\{0\} + (0, \infty)$. Such positive Gaussian auto-models are used in Section 4 to analyze motion textures $\{v_{res}\}$.

3.1 Multi-parameter auto-models

Consider a general system of real random variables $\{X_i, i \in S\}$ indexed by a finite set $S = \{1, \ldots, n\}$. For a *site* i, let

$$\mu_i(x_i|\cdot) = \mu_i(x_i|x_j, j \neq i) \ ,$$

be the probability density of X_i given the event $\{X_j = x_j, j \neq i\}$. An important approach in stochastic modeling consists in specifying the family of all these conditional distributions $\{\mu_i(x_i|\cdot)\}$, and then to determine a joint distribution μ of the system, which is compatible with this family, i.e. the μ_i's are exactly the conditional distributions associated to μ. We refer to the seminar paper (Besag, 1974) which presents general results including a summary of earlier results about the "nearest neighbours systems" from (Whittle, 1963; Bartlett, 1968).

In this paper, we focus our attention on the auto-models introduced by (Besag, 1974). In particular we follow the notations used there. Let us recall that if the joint distribution μ has a everywhere positive density, the Hammersley-Clifford's theorem gives a characterization of μ by an "energy" $Q(x)$ equal to a sum of "potentials" G defined on a set of "cliques". Besag's auto-models are precisely constructed under two assumptions:

[B1] the cliques involve at most two points i.e.

$$Q(x) = \sum_{i \in S} G_i(x_i) + \sum_{\{i,j\}} G_{ij}(x_i, x_j) \ .$$

[B2] for each site i, the conditional distribution $\mu_i(x_i|\cdot)$ belongs to a one-parameter exponential family:

$$\log \mu_i(x_i|\cdot) = A_i(\cdot)B_i(x_i) + C_i(x_i) + D_i(\cdot)$$

Here, the sufficient statistics $B_i(x_i)$ as well as the natural parameters $A_i(\cdot)$ are *real-valued*.

To tackle with mixed-state observations we first need to extend the above one-parameter auto-models to a multi-parameter setting. Let us make more precise some notations. It is understood that we are given a measurable state space $(E, \mathcal{E}, m)$ where E is a subset of $\mathbb{R}^d$. The field X is taking values in a *configuration space* $\Omega = E^S$, equipped with the product structure $(\mathcal{E}, m)^{\otimes S}$. A random field on S is specified by a probability distribution μ on Ω. We will always assume that μ has a *everywhere positive density* P with respect to the product measure $\nu = m^{\otimes S}$. In other words,

$$\mu(dx) = P(x)\nu(dx) \ , \qquad P(x) = Z^{-1} \exp Q(x) \ , \tag{3}$$

where Z is a normalization constant. The positivity condition implies that at each site i, the conditional distribution $(X_i|X_j = x_j, j \neq i)$ has a positive density $\mu_i(x_i|\cdot)$ with respect to $m(dx_i)$.

Throughout the paper, Condition **[B1]** on the size of cliques will be assumed satisfied. To introduce the desired extension, we first need to replace **[B2]** by

[B2'] $\log \mu_i(x_i|\cdot) = \langle A_i(\cdot), B_i(x_i)\rangle + C_i(x_i) + D_i(\cdot) \ , \ A_i(\cdot) \in \mathbb{R}^d, \ B_i(x_i) \in \mathbb{R}^d.$

Secondly the multi-parameter situation requires the following regularity condition

[C] For all $i \in S$, $\mathrm{Span}\{B_i(x_i), x_i \in E\} = \mathbb{R}^d$.

We have then the the following theorem (for a proof, see (Hardouin and Yao, 2004)).

Theorem 1 *Assume the random field probability distribution μ of (3) and its energy function $Q(x)$ satisfy Conditions **[B1]**, **[B2']** and **[C]**. Then there exist for all $i, j \in S,\ i \neq j$, a family of vectors $\alpha_i \in \mathbb{R}^d$ and a family of $d \times d$ matrices β_{ij} such that*

$$A_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B_j(x_j) \ . \tag{4}$$

Consequently the set of potentials is given by

$$G_i(x_i) = \langle \alpha_i, B_i(x_i) \rangle + C_i(x_i) \ , \tag{5}$$

$$G_{ij}(x_i, x_j) = B_i^T(x_i) \beta_{ij} B_j(x_j) \ . \tag{6}$$

A model satisfying the assumptions of the theorem is called a *multi-parameter auto-model*. The results are similar to the one-parameter case defined by Besag under **[B1]** and **[B2]**. The additional condition **[C]** eliminates singular local sufficient statistics $B_i(x_i)$. This condition does not exist for the one-parameter case, since it is automatically satisfied meaning that the B_i's are not identically zero. We will see below that this condition is not restrictive, easily satisfied in most examples. Moreover we notice that Besag's auto-models are originally defined in a more constrained way where the sufficient statistics $B_i(x_i)$ are linear in x_i.

In all the following, we will say that a measurable function $H(x)$ defined on Ω is *admissible*, with respect to ν, if

$$\int_\Omega \exp H(x) \nu(dx) < \infty \ . \tag{7}$$

The following proposition is useful, giving a converse to the theorem above. It also provides a practical way to choose the parameters for a well-defined multi-parameter auto-models (see Section 3.4).

Proposition 1 *Assume that the function Q is defined by **[B1]** with potentials G_i and G_{ij} given in (5)-(6), and that it is admissible. Then the family of conditional distributions $\mu_i(x_i|\cdot)$ belongs to an exponential family of the form **[B2']**, with sufficient statistics $A_i(\cdot)$ verifying (4).*

3.2 Random variables with mixed states

We now consider the mixed-state situation. Let us first define a simple random variable on the mixed-state space $E = \{0\} + (0, \infty)$. This space is equipped with a "mixed" reference measure

$$m(dx) = \delta_0(dx) + \lambda(dx) \ ,$$

where δ_0 is the Dirac measure at 0 and λ the Lebesgue measure on $(0, \infty)$.

Any random variable X taking values in E is called a *mixed-state random variable*, and the associated distribution a *mixed-state distribution*. Such a variable can be constructed as following: with probability $p \in]0, 1[$ we set $X = 0$, and with probability $1 - p$, X follows a exponential family distribution with density

$$g_s(x) = \exp[\langle s, T(x) \rangle - \phi(s)] \ , \quad x \in]0, \infty[\ .$$

Note that $\phi(s)$ is implicitly defined by normalization of the density. This density can be rewritten as

$$g_s(x) = g_s(0)e^{\langle s,\widetilde{T}(x)\rangle}, \quad \text{with } g_s(0) = e^{\langle s,\widetilde{T}(0)\rangle - \phi(s)} ,$$

where we have set $\widetilde{T}(x) = T(x) - T(0)$ such that $\widetilde{T}(0) = 0$. (Note that a priori T is not defined at the point 0, but in practice usual sufficient statistics T can be extended at 0 by continuity).

Let $\delta(x)$ be the Dirac function at 0: $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. Set the complementary function $\delta^*(x) = 1 - \delta(x)$. Then the mixed-state random variable X has the following density function, w.r.t. $m(dx)$,

$$\begin{aligned} f_\theta(x) &= p\delta(x) + (1-p)\delta^*(x)g_s(x) \\ &= (1-p)g_s(0)\exp\left[\delta(x)\log\frac{p}{(1-p)g_s(0)} + \langle s,\widetilde{T}(x)\rangle\right] \\ &= p\exp\left[-\delta^*(x)\log\frac{p}{(1-p)g_s(0)} + \langle s,\widetilde{T}(x)\rangle\right] \\ &= \exp\left[\langle\theta, B(x)\rangle - \psi(\theta)\right] \end{aligned} \tag{8}$$

where we have set

$$\theta = (\theta_1,\theta_2)^T = \left(\log\frac{(1-p)g_s(0)}{p}, s\right)^T , \quad B(x) = (\delta^*(x), \widetilde{T}(x)^T)^T .$$

Here the normalization term $\psi(\theta)$ is defined implicitly.

In other words, X belongs also to an exponential family and the dimension of its parameters θ (or of its sufficient statistic $B(x)$) is one unity bigger than the dimension of the parameter s of its continuous component g_s. We also have the following one-to-one correspondence between the natural parameter θ and the original parameters s and p:

$$s = \theta_2 , \quad p = \frac{g_s(0)}{g_s(0) + e^{\theta_1}} .$$

The following particular mixed-state distribution, called *positive mixed-state Gaussian distribution*, will play a fundamental role in the application developed in Section 4. For this distribution, the continuous component of X is the distribution of the *module* of a zero-mean normal distribution with variance σ^2:

$$g_s(x) = \frac{2}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} = \exp\left[-\frac{x^2}{2\sigma^2} - \phi(s)\right] .$$

We take $s = 2\sigma^{2^{-1}}$ as natural parameter for g_s and $T(x) = \widetilde{T}(x) = -x^2$ as the associated sufficient statistic. Note that $g_s(0) = 2(2\pi\sigma^2)^{-1/2} = 2(s/\pi)^{1/2}$.

The pair $\{\theta, B(x)\}$ of this mixed-state distribution, denoted $\mathcal{G}_m(p,s)$, is then

$$\theta = (\theta_1, s), \quad \theta_1 = \log\left\{\frac{1-p}{p}2(2\pi\sigma^2)^{-1/2}\right\} ,$$

and

$$B(x) = (\delta^*(x), -x^2) .$$

Therefore the parametric dimension is two and the initial parameters (p, s) are recovered from $\theta = (\theta_1, \theta_2)$ by the formula

$$s = \theta_2, \quad p = \frac{2(s/\pi)^{1/2}}{2(s/\pi)^{1/2} + e^{\theta_1}} \ .$$

3.3 Auto-models with mixed states

We are in order to construct auto-models for mixed-state observations on the state space $E = \{0\} + (0, \infty)$. We start by assuming that the family of conditional distributions $\mu_i(x_i|\cdot)$ belongs to the family of mixed-state distribution $f_{\theta_i(\cdot)}(x_i)$ given in (8). Here the parameter $\theta_i(\cdot)$ is a function of neighbouring configuration $(\cdot) = (x_j, j \neq i)$. In other words, we assume that

$$\log \mu_i(x_i|\cdot) = \langle \theta_i(\cdot), B(x_i) \rangle - \psi_i(\cdot) \tag{9}$$

with $B(x) = (\delta^*(x), \widetilde{T}(x))$. Based on Theorem 1, we know that there are a family of vectors $\{\alpha_i\}$ and matrices $\{\beta_{ij}\}$ satisfying $\beta_{ij} = \beta_{ji}^T$, such that

$$\theta_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) \ . \tag{10}$$

Moreover the associated potentials are given in (5)-(6).

3.4 Auto-models with positive Gaussian mixed states

In view of application on motion textures, we now focus our attention on the positive Gaussian case by requiring that at each site i, the conditional distribution μ_i is a mixed-state Gaussian positive distribution $\mathcal{G}_m(p, s)$. Following Eqs.(9)-(10), there are vectors $\alpha_i = (a_i, b_i) \in \mathbb{R}^2$ and 2×2 matrices

$$\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ d_{ij}^* & e_{ij} \end{pmatrix} \ ,$$

such that the energy function is given by

$$Q(x_1, \ldots, x_n) = \sum_{i \in S} \left[a_i \delta^*(x_i) - b_i x_i^2 \right] + \sum_{\{i,j\}} (\delta^*(x_i), -x_i^2) \beta_{ij} (\delta^*(x_j), -x_j^2)^T \ . \tag{11}$$

Let us describe in more details the local conditional distributions $\{\mu_i\}$. By construction at each site i, the conditional distribution $\mu_i(x_i|\cdot)$ is $\mathcal{G}_m(p_i(\cdot), s_i(\cdot))$ with parameters

$$\theta_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) = \left[\log \left(\frac{[1 - p_i(\cdot)] s_i(\cdot)}{p_i(\cdot)} \right), \ s_i(\cdot) \right]^T \ .$$

More explicitly

$$\theta_{i,1}(\cdot) = a_i + \sum_{j \neq i} \left[c_{ij} \delta^*(x_j) - d_{ij} x_j^2 \right] \ , \tag{12}$$

$$\theta_{i,2}(\cdot) = b_i + \sum_{j \neq i} \left[d_{ij}^* \delta^*(x_j) - e_{ij} x_j^2 \right] \ . \tag{13}$$

We have in particular

$$s_i(\cdot) = \frac{1}{2\sigma_i^2(\cdot)} = \theta_{i,2}(\cdot) \ , \qquad p_i(\cdot) = \frac{2[s_i(\cdot)/\pi]^{1/2}}{2[s_i(\cdot)/\pi]^{1/2} + e^{\theta_{i,1}(\cdot)}} \ .$$

It follows that necessarily for all i and its possible neighbouring configuration $(\cdot) = (x_j, j \neq i)$, the variance parameter $s_i(\cdot) = 1/[2\sigma_i^2(\cdot)]$ of the Gaussian component must be positive, i.e.

$$s_i(\cdot) = \frac{1}{2\sigma_i^2(\cdot)} = b_i + \sum_{j \neq i} [d_{ij}^* \delta^*(x_j) - e_{ij} x_j^2] > 0 \ .$$

It can be proved that this is equivalent to require the

Conditions [D]:

(i) for all $\{i, j\}$, $e_{ij} \leq 0$.

(ii) for all i and any subset $A \subset S \backslash \{i\}$, $\quad b_i + \sum_{j \in A} d_{ij}^* > 0$ (in particular $b_i > 0$).

It turns out that these necessary conditions are also sufficient for the admissibility of the energy function Q given in (11) The next proposition is important: together with Proposition 1 it gives the practical choice of parameters for a valid definition of positive Gaussian auto-models. The proof of this proposition, quite technical, is to be found in (Hardouin and Yao, 2004).

Proposition 2 *Under the conditions **[D]**, the energy function Q is admissible.*

3.4.1 A specification for the four nearest-neighbours system

We describe in this section a particular positive Gaussian auto-model using the four nearest neighbours system. Recall that the set of sites is $S = \{1, \ldots, n\} = [1, M] \times [1, N]$ and the neighbour system is to be completed with usual adaptation on the boundary. We denote by $\{i \pm (1, 0),\ i \pm (0, 1)\}$ the four neighbours of i.

Furthermore we assume that the field is homogeneous in space, i.e. the parameters are the same for all sites. Moreover we will allow possible anisotropy between the horizontal and vertical directions. Under all these considerations and by the previous results, there exit a vector $\alpha = (a, b)$ and two 2×2 matrices

$$\beta^{(k)} = \begin{pmatrix} c_k & d_k \\ d_k^* & e_k \end{pmatrix} \ , \quad k = 1, 2$$

such that $\forall i,\ \alpha_i = \alpha$, and for $\forall i, j,\ \beta_{ij} = 0$ unless i and j are neighbours where

$$\beta_{ij} = \beta^{(1)} \quad \text{for } \ j = i \pm (1, 0) \ , \quad \beta_{ij} = \beta^{(2)} \quad \text{for } \ j = i \pm (0, 1) \ .$$

The model has then 10 parameters. Moreover for the application developed in Section 4, we need further constrained the parameters $d_1^*,\ d_2^*,\ e_1$ and e_2 to be zero, since otherwise with $d_k^* > 0$ or $e_k < 0$, the correlation between neighbouring sites becomes negative, i.e. the

field is repulsive and neighbouring sites are "in competition". This is clearly not suited for homogeneous motion textures we intent to analyze here.

Finally, this auto-model with four neighbours system has six parameters, namely $\phi = (a, b, c_1, c_2, d_1, d_2)$. The admissibility condition **[D]** is reduced in the present case to the unique simple condition

$$b > 0 \; .$$

It is useful to note that the parameters of the local conditional distributions $\mu_i = \mathcal{G}_m(p_i(\cdot), s_i(\cdot))$ at a site i take the form

$$\begin{aligned} \theta_{i,1}(\cdot) &= a + \sum_{j=i\pm(1,0)} \left[c_1\delta^*(x_j) - d_1 x_j^2\right] + \sum_{j=i\pm(0,1)} \left[c_2\delta^*(x_j) - d_2 x_j^2\right] \; , \qquad (14) \\ \theta_{i,2}(\cdot) &= b \; . \end{aligned}$$

In case we impose the equality $c_1 = c_2 = c$ and $d_1 = d_2 = d$, we get an isotropic model with four parameters $\phi = (a, b, c, d)$. This model will also be used in Section 4 to test the existence or not of a spatial isotropy.

As for the estimation of the parameter ϕ, we use the pseudo-likelihood method by maximizing the pseudo-likelihood (in fact its logarithm)

$$L(x; \phi) = \sum_{i \in S} \log \mu_i(x_i | x_j, \; j \neq i) \; . \qquad (15)$$

This method has good consistency properties for classical one-parameter auto-models, see e.g. (Guyon, 1995). We conjecture that it is still the case for multi-parameter auto-models considered here, although we are not aware of any proof of such consistency.

4 Applications to motion texture analysis

Temporal textures (or dynamic textures) designate video contents involving natural (almost stationary) dynamic phenomena such as rivers, sea-waves, smokes, steams, fires, fountains, moving grass or foliage, etc. No tractable 3D kinematic models can be exhibited to account for these motions and to allow the derivation of relevant and efficient image motion models. Therefore, the analysis of dynamic textures is a challenging issue while of practical interest for various applications. This problem has been mainly investigated by considering these image sequences as visual signals only, that is by modeling the time-varying intensity function only. In the early work by (Szummer and Picard, 1996) on temporal texture modeling, a spatio-temporal auto-regressive model was introduced which was also causal in the spatial domain and could handle a restricted range of motion contents only. A significant extension has been then designed by (Doretto et al., 2003) exploiting ARMA models and system identification tools. Issues of modeling, learning, recognizing, compressing or synthesizing dynamic textures were addressed with this modeling framework. Recently, multi-scale AR models have been applied to this problem by (Doretto et al., 2004), along with a closed-loop linear dynamic system by (Yuan et al., 2004). However, these methods present two main limitations: they consider linear models only and they operate on the pixel intensities.

Mixed-state auto-models allow us to specify non-linear models, to take into account the spatial context, and to introduce both symbolic information (no motion) and continuous motion values, which is of great interest to handle dynamic textures. Furthermore, we do not model the time-varying intensity function but the motion measurements themselves. Thus, the designed models are intrinsic to the motion content of the video. Consequently, we prefer to use the term "motion texture" in that context.

We report here a set of preliminary results on the modeling of motion textures. All experiments of this section are made with the positive Gaussian mixed-state auto-model introduced in Section 3.4.1.

4.1 Isotropy and anisotropy

In order to evaluate the performance of the proposed modeling, we first examine if the introduced auto-models can realize two fundamental characteristics of a homogeneous texture, namely spatial isotropy and spatial anisotropy. For the positive Gaussian auto-models used here, see Eq.(14) for instance, the isotropy occurs if (and only if) $c_1 = c_2$ and $d_1 = d_2$.

Test experiment 1:

The first dynamic texture we consider represents motion from trees. Such a sample image and the associated motion texture $\{v_{res}\}$ are shown in Figure 5. Moreover this test motion texture from trees is believed to be spatially isotropic.

We have estmated the anisotropic positive Gaussian model using three consecutive $\{v_{res}\}$ maps of the tree sequence. The pseudo-likelihood estimates of the 6 parameters $\phi = (a, b, c_1, c_2, d_1, d_2)$, as defined in (15), are reported in Table I (top row).

From Table I, we see that the parameters c_1 and c_2 on one hand, and the parameters d_1 and d_2 on the other hand, are almost identical (with regard to standard deviations of these estimates we have computed from similar motion texture maps from the same tree sequence). Therefore, spatial isotropy is well reflected here by the equality between the parameters $\{c_k\}$ and $\{d_k\}$. This statement is further confirmed by the estimated *isotropic* positive Gaussian model with four parameters $\phi = (a, b, c, d)$. The corresponding estimates are given in the bottom row of Table I. Clearly the values of c and d are respectively, very close to the c_k and d_k's found above for the anisotropic model.

Test experiment 2:

We next consider a motion texture from a video sequence involving close-up shots of a moving escalator. Since the motion of the moving escalator is a vertical one, we have clearly anistropic motion texture in presence. A sample image and the associated motion measures $\{v_{res}\}$ are given in Figure 6.

As in Experiment 1, we have estimated the 6-parameter positive Gaussian auto-model. The parameter estimates are given in Table II. Clearly, the differences between c_1 and c_2, as well as between d_1 and d_2 are significant. Therefore, by these differences, the fitted model is able to reflect the spatial anisotropy of the considered motion texture.

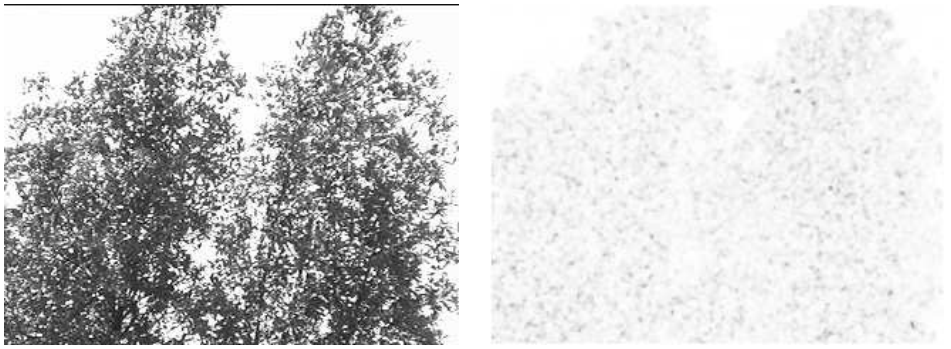

Figure 5: Sample image of "tree" sequence and the associated map of motion measures (white=0; black=maximum value).

anisotropic model	a	b	c_1	c_2	d_1	d_2
	-5.235	2.629	2.105	2.226	-14.045	-14.506

isotropic model	a	b	c	d
	-5.243	2.629	2.156	-14.600

Table 1: Parameter estimates: the anisotropic model (top row) and the isotropic model (bottom row)

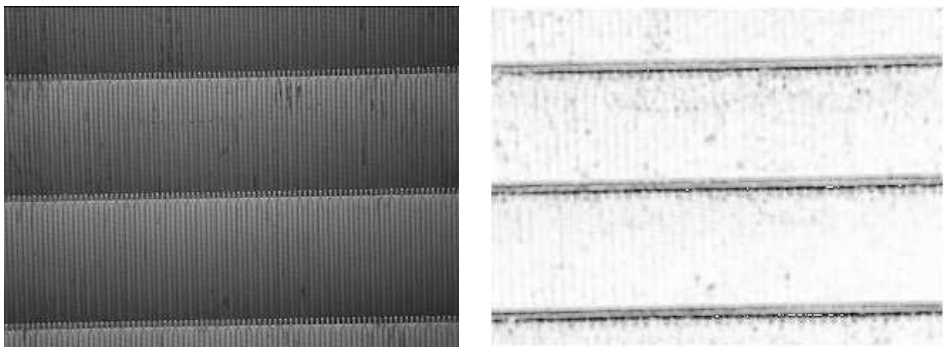

Figure 6: Sample image of the "escalator" sequence and the associated residual motion measures.

anisotropic model	a	b	c_1	c_2	d_1	d_2
	-6.012	0.283	3.143	1.857	-2.032	-4.799

Table 2: Parameters of the anisotropic model for motion from escalators.

4.2 Spatial stationarity

Here we propose to analyze another characteristic feature of motion textures, namely the spatial stationarity. To this end and for a given texture, we divided the motion map into 12 blocks of size 65×65 pixels each. We then fit a anisotropic positive Gaussian auto-model to each of the blocks.

Test experiment 3:

We process here sea-waves images. Figure 7 shows the blocks $B_1 \dots B_{12}$ extracted from a motion map $\{v_{res}\}$ at a given time of the video sequence.

The estimates of the parameters for these blocks are given in Table III, where the standard deviations from these 12 sets of estimates are also computed (bottom row).

From Table III, we can see that the 12 sets of parameters are nearly the same, taking into account the associated standard deviations (however, we should point out the particularly severe instability of the parameter d_2). This then confirms the spatial stationarity believed in motion textures from sea-waves.

Test experiment 4:

We have conducted a similar experiment with a motion texture from a video depicting a river. Sample blocks $B_1 \dots B_{12}$ of motion textures are illustrated in Figure 8.

The estimates of the parameters for these blocks are given in Table IV, where the standard deviations from these 12 sets of estimates are also computed (bottom row).

An overall impression from Table IV is that these 12 sets of parameters are significantly different, resulting in a much bigger standard deviation value as compared to Table IV in the previous experiment. This difference is particularly clear between 3 block lines (of 4 blocks each). Therefore, we can reasonably assert that the non-stationarity from these blocks of river motion texture are correctly reflected by different parameter estimates of the proposed positive Gaussian mixed-state auto-model.

5 Conclusion

For analysis of mixed-state observations such as motion measurements in an image sequence, we have introduced a new class of random field models, namely mixed-state auto-models. This approach is made possible by extending Besag's one-parameter auto-models to the multi-parameter case. We have provided a careful construction of these models as well as a detailed discussion about their basic properties.

The performance of such mixed-state auto-models has been experimentally tested by analyzing different motion textures from video sequences. Although the adopted positive Gaussian model (with the four nearest-neighbours system) of Section 3.4.1 seems crude, our experiments proved that important texture characteristics like spatial isotropy or anisotropy, as well as spatial stationarity or non-stationarity can be corrected reflected by the estimated parameter values.

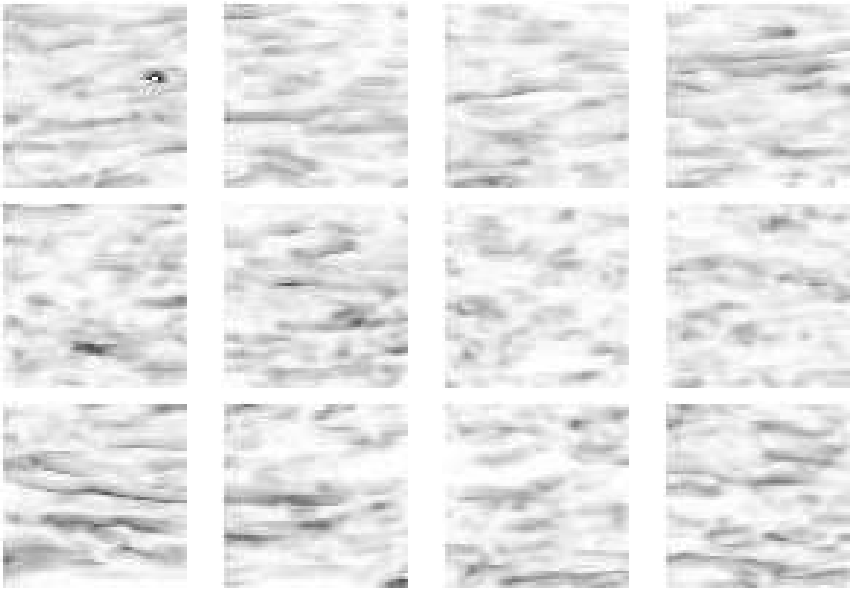

Figure 7: **Sea-waves sequence:** Blocks of residual motion measures of size 65×65 at a giben time of the sequence. Top to bottom and left to right: $B_1 \ldots B_{12}$.

	a	b	c_1	c_2	d_1	d_2
B_1	-7.223	0.315	2.037	3.594	-0.263	-10.792
B_2	-6.083	0.379	1.451	2.793	-1.584	-21.231
B_3	-6.532	0.359	0.947	4.251	-0.561	-14.070
B_4	-6.853	0.328	1.238	3.789	-0.943	-22.252
B_5	-5.729	0.337	0.868	3.593	-0.877	-11.589
B_6	-6.518	0.367	1.437	3.798	-0.578	-10.964
B_7	-5.492	0.475	1.173	3.374	-0.364	-10.463
B_8	-5.638	0.577	1.651	2.549	-0.914	-28.809
B_9	-5.942	0.280	0.508	4.268	-0.394	-14.632
B_{10}	-5.813	0.268	0.944	3.571	-0.949	-12.249
B_{11}	-5.560	0.297	0.797	3.741	-0.754	-8.288
B_{12}	-5.933	0.278	1.111	3.387	-1.086	-14.04
st. d.	0.550	0.090	0.416	0.504	0.369	6.051

Table 3: **Sea-waves:** Estimates of the 6-parameter anisotropic model for 12 blocks of a motion texture from a sea-waves video. The standard deviation from these 12 sets of estimates are given in the bottom row.

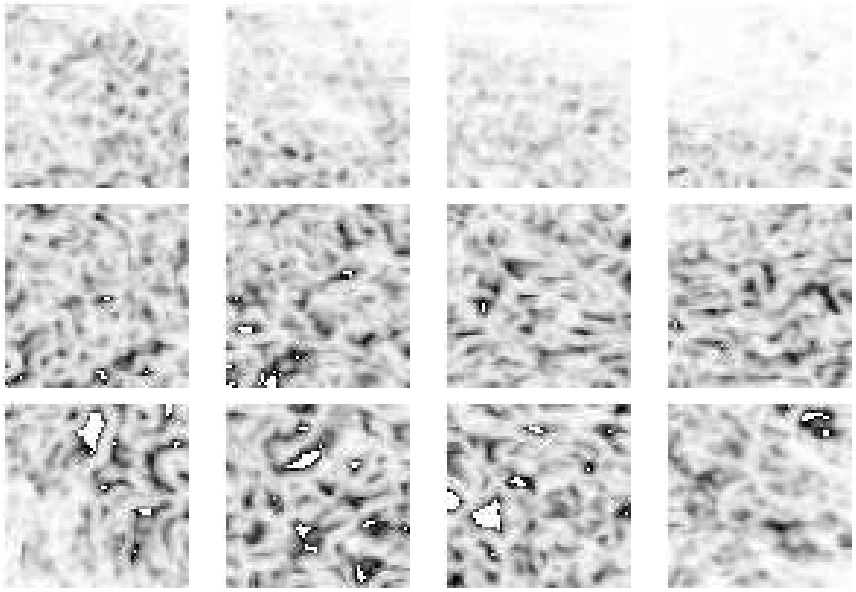

Figure 8: **River sequence:** Blocks of residual motion measures of size 65×65 at a given instant of the sequence. Top to bottom and left to right: $B_1 \ldots B_{12}$.

	a	b	c_1	c_2	d_1	d_2
B_1	-6.671	0.292	2.238	2.979	0.750	-7.882
B_2	-6.751	0.412	2.349	3.184	0.396	-8.820
B_3	-5.216	0.643	1.954	2.187	-8.360	-19.151
B_4	-5.476	0.765	2.147	2.704	-2.704	-11.540
B_5	-0.250	0.112	0.073	5.854	0.113	-0.071
B_6	-0.614	0.097	-0.316	5.469	0.181	-0.554
B_7	0.657	0.107	2.690	1.630	0.228	-1.193
B_8	-0.199	0.109	0.116	5.894	0.099	-0.057
B_9	-7.873	0.104	3.584	2.342	0.518	-5.266
B_{10}	1.472	0.072	1.302	2.279	-0.693	0.112
B_{11}	-2.728	0.094	-2.921	7.528	-6.053	0.477
B_{12}	-5.912	0.110	1.268	3.167	-0.985	-4.284
st. d.	3.361	0.238	1.745	1.899	2.918	6.070

Table 4: **River-sequence:** Estimates of the 6-parameter anisotropic model for 12 blocks of a motion texture from a river video. The standard deviation from these 12 sets of estimates are given in the bottom row.

There are several unanswered questions which need further investigations. First, from a theoretic point of view, the convergence of the pseudo-likelihood estimators has to be established. Also, some efficient Monte-Carlo simulation algorithms of mixed-state auto-models have to be designed in the vein of the theory of Markov random fields.

Secondly, as for the analysis of motion dynamic textures we proposed here, it is clear that the adopted positive Gaussian model need to be improved. A first possibility is to incorporate larger neighbour systems for modeling long-distance spatial correlations. A second possibility that we are currently considering is to include non-Gaussian distributions for the continuous component of motion measurements, like beta distributions. This richer family of distributions would probably improve model fits for various motion textures.

Bibliography

Allcroft, D. J. and Glasbey, C. A. (2003). A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *J. Roy. Statist. Soc. Ser. C*, 52(4):487–498.

Bartlett, M. (1968). A further note on nearest neighbour models. *J. Roy. Statist. Soc.*, A(131):579–580.

Besag, J. (1974). Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.*, B(148):1–36.

Chalmond, B. (2003). *Modeling and Inverse Problems in Image Analysis*, volume 155 of *Applied Mathematical Sciences*. Springer-Verlag, New York.

Doretto, G., Chiuso, A., Wu, Y., and Soatto, S. (2003). Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109.

Doretto, G., Jones, E., and Soatto, S. (2004). Spatially homogeneous dynamic textures. In *8th European Conf. on Computer Vision*, LNCS 3022, pages 591–602. Springer, Prague.

Fablet, R. and Bouthemy, P. (2003). Motion recognition using non parametric image motion models estimated from temporal and multiscale cooccurrence statistics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(122):1619–1624.

Fablet, R., Bouthemy, P., and Pérez, P. (2002). Non-parametric motion characterization using causal probabilistic models for video indexing and retrieval. *IEEE Trans. on Image Processing*, 11(4):393–407.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer-Verlag, New York.

Hardouin, C. and Yao, J. (2004). Markovian auto-models with mixed states. Technical report, IRMAR/Université de Rennes 1.

Irani, M., Rousso, B., and Peleg, S. (1992). Detecting and tracking multiple moving objects using temporal integration. In *Proc. of European Conf. on Computer Vision, ECCV'92*, LNCS 588, pages 282–287, Santa Margherita. Springer-Verlag.

Ngo, C.-W., Pong, T.-C., and Zhang, H.-J. (2002). On clustering and retrieval of video shots through temporal slices analysis. *IEEE Trans. Multimedia*, 4(4):446–458.

Odobez, J. and Bouthemy, P. (1997). *Separation of moving regions from background in an image sequence acquired with a mobile camera*, chapter 8, pages 283–311. Video Data Compression for Multimedia Computing (Eds. H.H. Li,S. Sun and H. Derin). Kluwer Academic Publisher.

Odobez, J.-M. and Bouthemy, P. (1995). Robust multiresolution estimation of parametric motion models. *J. of Visual Comm. and Image Repr.*, 6(4):348–365.

Rui, Y. and Anandan, P. (2000). Segmenting visual actions based on spatio-temporal motion patterns. In *CVPR'2000*, Hilton Head, SC.

Szummer, M. and Picard, R. (1996). Temporal texture modeling. In *IEEE Int. Conf. on Image Processing, ICIP'96*, Lausanne.

Vasconcelos, N. and Lippman, A. (2000). Statistical models of video structure for content analysis and characterization. *IEEE Trans. on Image Processing*, 9(1):3–19.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Statist. Inst.*, 40:974–994.

Yacoob, Y. and Black, J. (1998). Parametrized modeling and recognition of activities. In *Sixth IEEE Int. Conf. on Computer Vision*, Bombay.

Yuan, L., Wen, F., Liu, C., and Shum, H.-Y. (2004). Synthesizing dynamic texture with closed-loop linear dynamic system. In *8th European Conf. on Computer Vision*, LNCS 3022, pages 603–616. Springer, Prague.

Zelnik-Manor, L. and Irani, M. (2001). Event-based video analysis. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifi que,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399