Segmentation of Motion Textures using Mixed-state Markov Random Fields

T. Crivelli[a], B. Cernuschi-Frías[ab], P. Bouthemy[c] and J.F. Yao[d]

[a]Faculty of Engineering, University of Buenos Aires
[b] CONICET, Argentina
[c] IRISA/INRIA Campus de Beaulieu 35042 Rennes Cedex, France
[d] IRMAR/Univ. of Rennes 1. Campus de Beaulieu 35042 Rennes Cedex, France

ABSTRACT

The aim of this work is to model the apparent motion in image sequences depicting natural dynamic scenes (rivers, sea-waves, smoke, fire, grass etc) where some sort of stationarity and homogeneity of motion is present. We adopt the mixed-state Markov Random Fields models recently introduced to represent so-called motion textures. The approach consists in describing the distribution of some motion measurements which exhibit a mixed nature: a discrete component related to absence of motion and a continuous part for measurements different from zero. We propose several extensions on the spatial schemes. In this context, Gibbs distributions are analyzed, and a deep study of the associated partition functions is addressed. Our approach is valid for general Gibbs distributions. Some particular cases of interest for motion texture modeling are analyzed. This is crucial for problems of segmentation, detection and classification. Then, we propose an original approach for image motion segmentation based on these models, where normalization factors are properly handled. Results for motion textures on real natural sequences demonstrate the accuracy and efficiency of our method.

Keywords: Image texture analysis, image motion analysis, segmentation, Markov random fields.

1. INTRODUCTION

Dynamic content analysis from image sequences has became a fundamental subject of study in computer vision, with the objective of representing complex real situations in an efficient and accurate way. Usually, the great amount of image data to be processed requires a modeling framework able to extract a compact representation of the information.

In this context, this work focuses on natural dynamic scenes (such as rivers, sea waves, smoke, fire, moving trees, grass, etc.), where some type of stationarity and homogeneity of motion is present. This problem was recently analyzed in the context of the so-called *dynamic textures*[1,2] Generally speaking, efforts have been mainly dedicated to describe the evolution of the image intensity function over time using linear models[3,4] We will instead consider the image motion information.

We adopt the *mixed-state* Markov Random Fields (MRF) models recently introduced in Ref. 5 to represent the so-called *motion textures*. The approach consists in describing the spatial distribution of local *motion measurements* which exhibit values of two types: a discrete component related to the absence of motion and a continuous part for actual measurements. The former accounts for symbolic information that is far beyond the value of motion itself, providing crucial information on the dynamic content of the scene.

We propose several significative extensions to the first model introduced in Ref. 5 in order to capture more properties of the analyzed motion textures. First, we consider a 8 nearest-neighbor system, which leads to define a 11-parameter mixed-state MRF. Second, we consider a non-zero mean Gaussian distribution for the continuous part which allows us to express spatial correlation between continuous motion values. Also, we have designed a new way for calculating the partition function in Gibbs distributions.

Then, we have defined a motion texture segmentation method exploiting this modeling. One important original aspect is that we do not assume conditional independence of the observations for each texture. Results on real examples will demonstrate the accuracy and efficiency of our method.

2. MOTION TEXTURES

Let $I(i,t)$ be a scalar function that represents the image intensity at image point $i \in S = \{1...n\}$ for time t. We define a Motion Texture as the result of a feature extraction process applied on $I(i,t)$ over a time interval. The definition is general for any motion measure on the image sequence. This accounts for scalar motion measurements, but also for vectorial observations, as well.

Once the motion measure is obtained, a sequence of intensity images is mapped to a sequence of *motion maps or fields*. These fields are in essence functions of image bidimensional location and then the problem is reduced to modeling a spatial distribution of motion values.

2.1. Scalar motion measurements

Obtaining reliable and fast motion information from image sequences is essential in our formulation, which aims at analyzing great amounts of data and describing it with few parameters. We want to be clear in that the objective is not motion estimation by itself, but dynamic content analysis.

The Optical Flow Constraint[6] is a well known condition over intensity images from which velocity fields can be accurately estimated. Locally, it gives a valuable quantitative information about the scene motion. However, the aperture problem allows us to measure only the component of the velocity of an image point in the direction of the spatial intensity gradient, i.e. *normal flow*, defined as $\mathbf{V}_n(i,t) = -\frac{I_t(i,t)}{\|\boldsymbol{\nabla} I(i,t)\|} \frac{\boldsymbol{\nabla} I(i,t)}{\|\boldsymbol{\nabla} I(i,t)\|}$, where $I_t(i,t)$ is the temporal derivative of intensity.

We follow the approach of Ref. 7, for obtaining reliable scalar motion observations. It computes the magnitude of normal flow and applies a weighted average over a small window around an image location, to smooth out noisy measurements and enforce reliability. The result is a smoothed measure of local motion, which is always positive. However, and in contrast to Ref. 7, we introduce a weighted vectorial average of normal flow in order to keep direction information:

$$\tilde{\mathbf{V}}_n(i) = \frac{\sum_{j \in W} \mathbf{V}_n(j) \parallel \boldsymbol{\nabla} I(j) \parallel^2}{max(\sum_{j \in W} \parallel \boldsymbol{\nabla} I(j) \parallel^2, \eta^2)}, \tag{1}$$

where η^2 is a constant related to noise, and W is a small window centered in location i. This average results in a local estimation of normal flow. The projection of this quantity over the intensity gradient direction gives rise to the following scalar motion observation:

$$v_{obs}(i) \quad = \quad \tilde{\mathbf{V}}_n(i) \cdot \frac{\boldsymbol{\nabla} I(i)}{\parallel \boldsymbol{\nabla} I(i) \parallel} \tag{2}$$

with $v_{obs} \in (-\infty, +\infty)$. As already said, the definition of the measurement process, determines to some extent, the type of motion textures that we will be dealing with.

For this case, it is important to emphasize a fundamental property of the resulting field. Real world observations can be represented as continuous random variables within a statistical setting. However, discrete (or symbolic) information is also of interest. The underlying discrete property of no-motion for a point in the image, is represented as a null observation $v_{obs} = 0$. By the way, it is not the value by itself that matters, but the binary property of what we call *mobility*: the absence or presence of motion. Thus, the null motion value in this case, has a peculiar place in the sample space, and consequently, has to be modeled accordingly. We call this type of fields *Mixed-state random fields* as the corresponding random variables take their values in a disrete-continuous space.

3. MIXED-STATE MARKOV RANDOM FIELDS

Generally speaking, we are interested in modeling data that have a mixed nature, i.e., our observations take values from a mixed discrete-continuous space, $E = \{0\} \cup (-\infty, 0) \cup (0, \infty)$. Let $p\left(x_i \mid \mathcal{N}_i\right)$ be the conditional distribution of the mixed scalar observation, where x_i is a mixed-state random variable, w.r.t neighborhood $\mathcal{N}_i$. We are interested in the distribution of motion, and then $x_i = v_{obs}(i)$. The discrete (or symbolic) information is represented by $x_i = 0$ (absence of motion, in our case), and thus we can write:

$$p\left(x_i \mid \mathcal{N}_i\right) \quad = \quad \delta_0(x_i)P_i + p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right)\left(1 - P_i\right) \tag{3}$$

where we define, $P_i = P\left(x_i = 0 \mid \mathcal{N}_i\right)$. In the first term we note that $p\left(x_i \mid \mathcal{N}_i, x_i = 0\right) = \delta_0(x_i)P_i$ where $\delta_0(\cdot)$ is the Dirac functional. Equation (3) is the expression for a mixed-state conditional model where the observation x_i lies in a mixed-state space.

3.1. Gibbs Distribution

As the seminal theorem of Hammersley-Clifford states, Markov Random Fields are equivalent to nearest neighbor Gibbs distributions, that is, the joint distribution of the random variables that compose the field has the form, $p(\mathbf{x}) = exp[Q(\mathbf{x})]/Z$ where $Q(\mathbf{x})$ is an energy function, and Z is called the partition function or normalizing factor of the distribution.

It is a well-known result that the energy $Q(\mathbf{x})$ can be expressed as a sum of potential functions, $Q(\mathbf{x}) = \sum_{\mathcal{C} \subset S} V_{\mathcal{C}}(\mathbf{x})$, defined over all subsets of the lattice space S.[8] The advantage lies in specifying non-null potentials only for specific subsets $\mathcal{C}$ of some desired structure. It is easy to show that two random variables x_i, x_j of the lattice are conditionally independent, if all $\{\mathcal{C} : i, j \in \mathcal{C}\}$ have associated potentials $V_{\mathcal{C}} = 0$. Any $\mathcal{C}$ for which $V_{\mathcal{C}} \neq 0$ is called a clique.

The non-trivial potentials define, as a consequence, a neighborhood structure $\mathcal{N}_i$, which is defined as $\mathcal{N}_i = \{x_j \mid \exists\, \mathcal{C} : i, j \in \mathcal{C} \ \wedge \ V_{\mathcal{C}} \neq 0\}$. The equivalence of this neighborhood description of Markov Random Fields and undirected graphs is straightforward.

We follow the result of Ref. 5 that gives an expression for the potential functions in the case of conditional distributions belonging to the Multiparameter Exponential family, i.e.,

$$\log\left[p\left(x_i \mid \mathcal{N}_i\right)\right] = \mathbf{\Theta}_i^T(\mathcal{N}_i)\mathbf{S}_i(x_i) + C_i(x_i) + D_i(\mathcal{N}_i) \tag{4}$$

under the hypothesis that the corresponding Gibbs energy has non-null cliques with at most two elements. The latter condition corresponds to a class of random fields which are called *auto-models*. Here, $\mathbf{\Theta}_i$ is the parameter vector for the d-dimensional exponential distribution, $\mathbf{S}_i(x_i)$ is a sufficient statistic upon the data, and C_i and D_i complete the expression for the probability density. Then we have,

$$Q(\mathbf{x}) \quad = \quad \sum_{i \in S} V_i(x_i) + \sum_{<i,j>} V_{ij}(x_i, x_j) \tag{5}$$

From these two conditions it can be demonstrated that

$$\mathbf{\Theta}_i(\mathcal{N}_i) = \boldsymbol{\alpha}_i + \sum_{j : x_j \in \mathcal{N}_i} \boldsymbol{\beta}_{ij}\mathbf{S}(x_j) \tag{6}$$

with $\boldsymbol{\beta}_{ij} \in \mathbb{R}^{d \times d}$ and $\boldsymbol{\alpha}_i \in \mathbb{R}^d$. Consequently, the parameter vector that defines the exponential conditional distribution is a function of the neighbors of a particular point i. Note that the conditional dependence between neighbors cannot be arbitrary under the two mentioned hypotheses, having a particular shape as seen in (6).

Finally the expressions for the potentials are given by,

$$V_i(x_i) = \boldsymbol{\alpha}_i^T \cdot \mathbf{S}(x_i) + C_i'(x_i) \quad (7)$$
$$V_{ij}(x_i, x_j) = \mathbf{S}(x_i)^T \boldsymbol{\beta}_{ij} \mathbf{S}(x_j) \quad (8)$$

3.2. Exponential Mixed-state Auto-model

In order to express equation (3) in exponential form, we first introduce the null-argument indicator function $w_0(x)$ which takes the unity value for $x = 0$ and zero elsewhere. Then, we can rewrite the conditional distribution as a sum of excluding terms (up to a zero-measure set):

$$p\left(x_i \mid \mathcal{N}_i\right) = \delta_0(x_i) P_i w_0(x_i) + p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right)(1 - P_i) w_0^*(x_i) \quad (9)$$

with $w_0^*(x) = 1 - w_0(x)$. Strictly speaking, $w_0(x)$ is considered as a very small unitary window centered around zero. This allow us to express the logarithm of the left-hand side as the sum of the logarithms of the two terms on the right-hand side. Some calculations and rearrangements yield:

$$\log\left[p\left(x_i \mid \mathcal{N}_i\right)\right] = \log[\delta_0(x_i)] w_0(x_i) + \log[P_i] + \log\left[p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right) \frac{(1 - P_i)}{P_i}\right] w_0^*(x_i) \quad (10)$$

To avoid inconsistencies, $\delta_0(x)$ should be interpreted as the limit in distribution of a sequence of zero-mean Gaussian pdf's with variance going to zero.

It is easy to show that if $p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right)$ belongs to the exponential family defined by $\tilde{\boldsymbol{\Theta}}_i, \tilde{\mathbf{S}}_i, \tilde{C}_i, \tilde{D}_i$, then the mixed-state conditional distribution is exponential with:

$$\boldsymbol{\Theta}_i^T(\mathcal{N}_i) = \left[\tilde{D}_i(\mathcal{N}_i) + \log\left[\frac{(1 - P_i)}{P_i}\right] \ , \ \tilde{\boldsymbol{\Theta}}_i^T(\mathcal{N}_i)\right] \quad (11)$$
$$\mathbf{S}_i^T(x_i) = \left[w_0^*(x_i) \ , \ \tilde{\mathbf{S}}_i^T(x_i) w_0^*(x_i)\right] \quad (12)$$
$$C_i(x_i) = \tilde{C}_i(x_i) w_0^*(x_i) + \log[\delta_0(x_i)] w_0(x_i) \quad (13)$$
$$D_i(\mathcal{N}_i) = \log[P_i] \quad (14)$$

Here, we suppose that $p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right)$ is Gaussian for every image point, as experiments suggest Gaussian-like motion histograms. In the next section we focus on this particular case and arrive to the complete definition of the mixed-state motion texture model.

3.3. Gaussian Mixed-state Model

We assume that the continuous component of the mixed-state model, i.e. $p\left(x_i \mid \mathcal{N}_i, x_i \neq 0\right)$, follows a Gaussian law with mean m_i and variance σ_i^2. The local characteristics are then expressed as:

$$p\left(x_i \mid \mathcal{N}_i\right) = \delta_0(x_i) P_i + \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(x_i - m_i)^2}{2\sigma_i^2}} (1 - P_i) \quad (15)$$

Considering the Gaussian density, we get the following parameters for the mixed-state model:

$$
\begin{aligned}
\boldsymbol{\Theta}_i^T(\mathcal{N}_i) &= [\theta_{1,i}\ ,\ \theta_{2,i}\ ,\ \theta_{3,i}] \\
&= \left[-\frac{m_i}{2\sigma_i^2} + \log\frac{1}{\sigma_i\sqrt{2\pi}} + \log\left[\frac{1-P_i}{P_i}\right]\ ,\ \frac{1}{2\sigma_i^2}\ ,\ \frac{m_i}{2\sigma_i^2}\right] \\
\mathbf{S}_i^T(x_i) &= \left[w_0^*(x_i)\ ,\ -x_i^2\ ,\ x_i\right] \\
C_i(x_i) &= \log\left[\delta_0(x_i)\right]w_0(x_i) \\
D_i(\mathcal{N}_i) &= \log\left[P_i\right]
\end{aligned} \tag{16}
$$

The parametrization of the conditional distribution in terms of $\boldsymbol{\Theta}_i$ allows us to express the dependence of a point on its neighbors through equation (6). Moreover, the parameters of the original parametrization, P_i, m_i, σ_i, are also functions of the neighborhood and can be obtained easily from the first line of equation (16), resulting in the following expressions:

$$
P_i = \frac{(\sigma_i\sqrt{2\pi})^{-1}}{(\sigma_i\sqrt{2\pi})^{-1} + e^{\theta_{1,i} + \frac{m_i^2}{2\sigma_i^2}}}, \qquad \sigma_i^2 = \frac{1}{2\theta_{2,i}}, \qquad m_i = \frac{\theta_{3,i}}{2\theta_{2,i}} \tag{17}
$$

Finally, les us point that we have a third parametrization of the model, corresponding to $\boldsymbol{\beta}_{ij}$ and $\boldsymbol{\alpha}_i$. These parameters are not a function of the neighbors and, although they can be different for each point, they can describe the whole field in a compact way, under some assumptions as we will see in the next section.

4. A MOTION TEXTURE MODEL

As mentioned before, the definition of motion texture is related to the measurement process. The scalar measure proposed here gives rise to a motion field with mixed-states, where the discrete property corresponds to the absence/presence of motion.

We have seen how the Gaussian mixed-state model is described by a set of parameters in its general form. Now, we want to define those parameters in accordance to a real physical meaning.

First, let us expand the parametrization through $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_{ij}$ knowing, as observed in equation (16), that the conditional mixed-state distributions are exponential with three parameters. Then, we may define:

$$
\beta_{ij} = \begin{pmatrix} d_{ij} & e_{ij} & f_{ij} \\ e'_{ij} & g_{ij} & q_{ij} \\ f'_{ij} & q'_{ij} & h_{ij} \end{pmatrix} \qquad \alpha_i = \begin{bmatrix} a_i & b_i & c_i \end{bmatrix}^T \tag{18}
$$

and from equation (6) we obtain:

$$
\begin{aligned}
\theta_{1,i} &= a_i + \sum_{j:x_j\in\mathcal{N}_i}\left[d_{ij}w_0^*(x_j) - e_{ij}x_j^2 + f_{ij}x_j\right] \\
\theta_{2,i} &= b_i + \sum_{j:x_j\in\mathcal{N}_i}\left[e'_{ij}w_0^*(x_j) - g_{ij}x_j^2 + q_{ij}x_j\right] \\
\theta_{3,i} &= c_i + \sum_{j:x_j\in\mathcal{N}_i}\left[f'_{ij}w_0^*(x_j) - q'_{ij}x_j^2 + h_{ij}x_j\right]
\end{aligned} \tag{19}
$$

This general expressions for Gaussian fields can be reduced considerably. We express two main assumptions regarding the characteristics of the field, that will determine the parameter structure of the model. First we

will assume a homogeneous model, that is, the statistical properties of the random variables are invariant w.r.t. translations over the lattice. This makes the model parameters $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta_{ij}}$ independent of the location i. Secondly, we suppose a cooperative model for the motion texture, as we are interested in textures that tend to build mostly homogeneous regions.

From these hypotheses we can start defining the parameters of the model, leading to some simplifications. First, as we already mentioned, we will have $\boldsymbol{\alpha}_i = \boldsymbol{\alpha} = [a, b, c]^T$ and $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_j$.

More conditions arise from equation (8), knowing that $V_{ij}(x_i, x_j) = V_{ji}(x_j, x_i)$ from which we get $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_{ji}^T$. This is a structural property of Markov random fields as the cliques are not-ordered pairs of points.

Let us note from equation (17) that P_i depends on $\theta_{1,i}$ in an exponential way which in fact is not symmetric with respect to motion values (see first line of equation (19)), if we let the term $f_j x_j$ to be distinct of zero. In other words, if $f_j \neq 0$, the arbitrary positive and negative values of x_j would produce an asymmetry on P_i w.r.t. observations of same magnitude but different sign, leading to a physical non-sense: that a negative motion value influences differently the probability of no motion than a positive one. Then, we need to have $f_j = 0$ and so, $f'_j = 0$ (from $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_{ji}^T$). We also assume that the variance of the continuous part of the model, σ_i^2, is constant for every point and then, $e'_i = g_i = q_i = 0$ and consequently, $e_i = q'_i = 0$. Finally, the model results in:

$$\boldsymbol{\beta}_j = \begin{pmatrix} d_j & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & h_j \end{pmatrix} \qquad \boldsymbol{\alpha} = \begin{bmatrix} a & b & c \end{bmatrix}^T \tag{20}$$

Additionally, we see that we can impose $h_j \geq 0$, to have a cooperation scheme as these coefficients are related to the mean m_i. Obviously $b > 0$ as it defines the conditional variance. We note that the coefficients d_j and h_j corresponding to symmetric neighbors have to be the same so as to obtain an homogeneous process.

Regarding the neighborhood structure, we consider the vertical (v), horizontal (h), main diagonal (d) and anti-diagonal (ad) directions as the interacting ones, for a total of 8 neighbors, and consequently, this leads to the set of 11 parameters which characterizes the motion-texture model:

$$\boldsymbol{\phi} = \{a, b, c, d_h, h_h, d_v, h_v, d_d, h_d, d_{ad}, h_{ad}\} \tag{21}$$

We can now write the complete joint distribution for the mixed-state model starting from the expression for the potential functions:

$$\begin{aligned} Q(\mathbf{x}) &= \sum_{i \in S} \left[\boldsymbol{\alpha}^T \mathbf{S}(x_i) + \log\left[\delta_0(x_i)\right] w_0(x_i)\right] + \sum_{i \in S} \sum_{j: x_j \in \mathcal{N}_i} \mathbf{S}(x_i) \boldsymbol{\beta}_j \mathbf{S}(x_j) \\ &= \sum_{i \in S} \left[a w_0^*(x_i) - b x_i^2 + c x_i + \log\left[\delta_0(x_i)\right] w_0(x_i)\right] + \sum_{i \in S} \sum_{j: x_j \in \mathcal{N}_i} d_j w_0^*(x_i) w_0^*(x_j) + h_j x_i x_j \\ &= \sum_{i \in S} \log(\delta_0^{w_0(x_i)}(x_i)) + \sum_{i \in S} \left[-b x_i^2 + c x_i + \sum_{j: x_j \in \mathcal{N}_i} h_j x_i x_j\right] + \sum_{i \in S} \left[a w_0^*(x_i) + \sum_{j: x_j \in \mathcal{N}_i} d_j w_0^*(x_i) w_0^*(x_j)\right] \end{aligned}$$

Note that the second and third terms can be written as a quadratic form to yield:

$$Q(\mathbf{x}) = \sum_{i \in S} \log(\delta_0^{w_0(x_i)}(x_i)) + (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{B} (\mathbf{x} - \boldsymbol{\mu}) + (\mathbf{w_0} - \mathbf{u})^T \mathbf{K} (\mathbf{w_0} - \mathbf{u}) + \textit{constant} \tag{22}$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$, $\mathbf{u} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$, $\mathbf{w_0} = \begin{bmatrix} w_0(x_1) & w_0(x_2) & \dots & w_0(x_n) \end{bmatrix}^T$

$$\mathbf{B} = \begin{pmatrix} -b & & \mathbf{\frac{h_j}{2}} \\ & -b & \\ & \cdots & \\ \mathbf{\frac{h_j}{2}} & & b \end{pmatrix} \quad \mathbf{K} = \begin{pmatrix} a & & \mathbf{\frac{d_j}{2}} \\ & a & \\ & \cdots & \\ \mathbf{\frac{d_j}{2}} & & a \end{pmatrix}$$

The elements (i, j) of $\mathbf{B}$ and $\mathbf{K}$ that are out of the main diagonal are the corresponding coefficients between neighbors x_i,x_j. Finally, $\boldsymbol{\mu}$ is given by $-\boldsymbol{\mu}\mathbf{B} = [c\ c\ ....\ c]$: observe that the matrix $\mathbf{B}$ has the property that all columns (or rows) add up to the same value $-b + \sum_j \frac{h_j}{2}$ due to the homogeneity of the model, implying that the vector $\mathbf{u}$ is an eigenvector with eigenvalue $\lambda_u = -b + \sum_j \frac{h_j}{2}$. Then $\boldsymbol{\mu}$ is also an eigenvector with eigenvalue $\lambda_\mu = \frac{c}{b - \sum_j \frac{h_j}{2}}$ and consequently, $\boldsymbol{\mu} = \frac{c}{b - \sum_j \frac{h_j}{2}}\mathbf{u}$.

The constant term in equation (22) can be included as part of the normalization factor Z and knowing that $\delta_0^{w_0(x_i)}(x_i) = 1 + \delta_0(x_i)$, the final expression for the joint distribution is:

$$p(\mathbf{x}) = \frac{1}{Z}\left[\prod_{i \in S}(1 + \delta_0(x_i))\right] e^{(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{B}(\mathbf{x}-\boldsymbol{\mu}) + (\mathbf{w_0}-\mathbf{u})^T\mathbf{K}(\mathbf{w_0}-\mathbf{u})} \tag{23}$$

4.1. Parameter Estimation

In order to estimate the parameters of the motion-texture model from motion measurements, we adopt the pseudo-likelihood maximization criterion.[8] Therefore, we search the set of parameters $\hat{\phi}$ that maximizes the function $L(\phi) = \prod_{i \in S} p(x_i \mid \mathcal{N}_i, \phi)$. We use a gradient descent technique for the optimization as the derivatives of L w.r.t ϕ are known in closed form.

5. PARTITION FUNCTION CALCULATION

In this section we propose a new approach for solving the problem of partition function calculation for Gibbs distributions. This issue is crucial as will enable to properly handle the optimization of the global energy function which will be defined for the motion texture segmentation problem.

In the case of image processing applications, it is usual to analyze distributions of different classes in a decision-theory framework, as it occurs in detection, segmentation and classification problems. Thus, it is fundamental to accurately know the partition function in order to allow the comparison of different models.

For a general Gibbs distribution, the expression for the normalizing factor is:

$$Z = \int_\Omega e^{Q(\mathbf{x})} dx_1 ... dx_n \tag{24}$$

where $\mathbf{x} = [x_1 .... x_n]^T$ is the vector of random variables that form the field and Ω is the sample space. Let $\Delta Q(\mathbf{x}_k)$ be a variation on the energy function, not necessarily small, due to an arbitrary variation in the field and being a function of a subset of $\mathbf{x}$, namely $\mathbf{x}_k$. Then we can write the resulting partition function, Z', as a function of the former one, Z:

$$\begin{aligned} Z' &= \int_\Omega e^{Q'(\mathbf{x})} dx_1 ... dx_n = \int_\Omega e^{Q(\mathbf{x}) + \Delta Q(\mathbf{x}_k)} dx_1 ... dx_n \\ &= \int_\Omega \frac{Z}{Z} e^{Q(\mathbf{x})} e^{\Delta Q(\mathbf{x}_k)} dx_1 ... dx_n = \int_\Omega Z p(\mathbf{x}) e^{\Delta Q(\mathbf{x}_k)} dx_1 ... dx_n \\ &= Z \int_{\mathbf{x}_k} p(\mathbf{x}_k) e^{\Delta Q(\mathbf{x}_k)} d\mathbf{x}_k = Z E_{\mathbf{x}_k}\left[e^{\Delta Q(\mathbf{x}_k)}\right] \end{aligned} \tag{25}$$

where the expectation operator is applied w.r.t. the marginal probability distribution of $\mathbf{x}_k$. Conceptually, the approach consists in estimating the unknown partition function from reference values which are either known in closed form or can be easily estimated.

In c-class segmentation applications, it is common to formulate the problem as an energy optimization one. For model-based segmentation of images, the partition function for each class must be known. Available optimization methods are mostly based on iteratively computing energy changes as a result of adding or taking out points from a class. Defining ΔQ appropriately, we then have an expression for the change on the normalizing factor. For example, for the case of the auto-models we define here, taking out a point x_i from a class is equivalent to extracting the cliques corresponding to that point, and for the sake of obtaining the partition function, the integral must be calculated with respect to one less variable. This change in Z can be expressed setting:

$$\Delta Q(\mathbf{x}_k) = \Delta Q(x_i, \mathcal{N}_i) = -\left(V_i(x_i) + \sum_{j:x_j \in \mathcal{N}_i} V_{i,j}(x_i, x_j) \right) + \log \delta_0(x_i) \tag{26}$$

where $\log \delta_0(x_i)$ is a term that allows integrating also with respect to x_i without changing the value of the integral. Observe in the last equation the expression for $-V_i(x_i) + \log \delta_0(x_i)$:

$$-V_i(x_i) + \log \delta(x_i) = -\alpha_i^T \cdot \mathbf{S}(x_i) - \log\left[\delta_0(x_i)\right] w_0(x_i) + \log \delta_0(x_i) = -\alpha_i^T \cdot \mathbf{S}(x_i) + \log\left[\delta_0(x_i)\right] w_0^*(x_i) \tag{27}$$

Remember that $w_0^*(x_i)$ is a small unitary window around zero and thus, $e^{\log\left[\delta_0(x_i)\right] w_0^*(x_i)} \cong w_0(x_i)$. Then, we can write:

$$e^{\Delta Q(\mathbf{x}_k)} = w_0(x_i) e^{\left[-\alpha_i^T \cdot \mathbf{S}(x_i) - \sum_{j:x_j \in \mathcal{N}_i} V_{i,j}(x_i, x_j)\right]} \tag{28}$$

Knowing that $V_{i,j}(0, x_j) = 0$ and $\mathbf{S}(0) = 0$, the exponent in equation (28) is null for $x_j = 0$, and consequently, $e^{\Delta Q(\mathbf{x}_k)} = w_0(x_i)$. Finally:

$$Z' = Z E_{\mathbf{x}_k}\left[e^{\Delta Q(\mathbf{x}_k)}\right] = Z E_{\mathbf{x}_k}\left[w_0(x_i)\right] = Z P(x_i = 0) \tag{29}$$

We have come to a relation for the change on the partition function due to an extraction of a point from a mixed-state motion texture. Let us point out that if the observations would be independent, the partition function would take the form $Z = Z_0^N$ for homogeneous fields, where Z_0 is the partition function for a single point, taking the value $Z_0 = 1/P(x_i = 0)$, and N is the total number of points. This is consistent with equation (29).

6. MOTION TEXTURE SEGMENTATION

The representation of a motion texture with a relatively small set of parameters allows for a parsimonious characterization of the different parts of a natural dynamic scene. One important aspect that the model should fulfill for dynamic content analysis, is the ability of discrimination. In this context, the problem of segmentation, closely related to classification, aims at determining and locating regions in the image that correspond to a same motion texture or class. As we deal with discrete spatial schemes, the problem is equivalent to assign a label to each point in the image grid, indicating that it belongs to a certain motion texture.

Here, we follow a Bayesian approach for determining in an optimal way the distribution of the motion texture labels, uniquely related to a segmentation of the field, with the motion map as input data. Thus, we search for a label realization $\mathbf{l} = \{l_i\}$, where $l_i \in \{1, 2, ..., c\}$ is the class label value, that maximizes $p(\mathbf{l} \mid \mathbf{x}) \propto p(\mathbf{x} \mid \mathbf{l}) p(\mathbf{l})$,

where $\mathbf{x}$ represents the motion map v_{obs}. This corresponds to a MAP (maximum-a-posteriori) estimation of the label field $\mathbf{l}$. If we suppose that the c different motion textures come from independent dynamic phenomena, given the label field, we can write:

$$p(\mathbf{x} \mid \mathbf{l}) = \prod_{k=1}^{c} p(\mathbf{x}_k) = \prod_{k=1}^{c} \frac{e^{Q_k(\mathbf{x}_k)}}{Z_k(\mathbf{l})} \tag{30}$$

We introduce the following notation: we call Q_k the energy function corresponding to the texture class k; $\mathbf{x}_k$ is the vector of motion random variables that belong to texture k, and is a subset of $\mathbf{x}$; $Z_k(\mathbf{l})$ is the corresponding partition function, and depends on the distribution of a texture on the lattice, through the label field $\mathbf{l}$. That is, each texture has its Gibbs distribution, that depends on the parameters $\boldsymbol{\phi}$ of that class, and on the region that it occupies.

This approach allows us to account for conditional dependence, with the only supposition that there is no interaction between different textures.

For the *a priori* information on the segmentation label field, $p(\mathbf{l})$, we introduce another 8 nearest-neighbour Markov random field that behaves as a regularization term for the labeling process, so $p(\mathbf{l}) \propto exp[Q_S(\mathbf{l})]$ with:

$$Q_S(\mathbf{l}) = \sum_{i} \sum_{j:x_j \in \mathcal{N}_i} \gamma(\mathbb{I}_0(l_i - l_j) - 1) \tag{31}$$

where $\mathbb{I}_0(z)$ is the null argument indicator function. $p(\mathbf{l})$ penalizes the differences of labeling between adjacent neighbors, smoothing the segmentation output. The complete formulation can be stated as maximizing the energy:

$$E(\mathbf{l}) = \sum_{k=1}^{c} Q_k(\mathbf{x}_k) - \sum_{k=1}^{c} \log(Z_k(\mathbf{l})) + Q_S(\mathbf{l}) \tag{32}$$

6.1. Initialization

The formulation of the segmentation problem we are dealing with, does not assume that the motion texture model parameters are known. Then, it is necessary to correctly estimate the mixed-state motion texture parameters for each class.

As an initialization of the label field, we divide the motion map in square blocks of some fixed size, and for each block a set of motion-texture model parameters is estimated. Then, we apply a clustering technique to obtain a first splitting of blocks.

In this step, the election of the distance between clusters is crucial. An appropriate distance is the symmetrized Kullback-Leibler (KL) divergence between probability densities. The ideal situation would be to calculate it between the joint probability distributions (the global distribution on each initial block) of the textures we want to segment. This implies knowing these functions (specially the normalizing factor) and also integrating them, resulting in an intricate problem by itself.

Here, we adopt a simplified approach. We calculate the KL distance between marginal distributions $p(x_i)$ for single points: supposing homogeneity and that the distribution for a pixel is a mixed-state Gaussian density, one can estimate easily its parameters (probability of null value, mean and variance of the Gaussian density). Having the parameters is also easy to obtain the divergence expression in closed form.

For the marginal distribution, we then write $p(x) = P(x = 0)\delta_0(x) + (1 - P(x = 0))\mathcal{N}(\mu, \sigma)$, where $\mathcal{N}(\mu, \sigma)$ represents a Gaussian density with mean μ and variance σ^2. This results in the following expression for the Kullback-Leibler divergence of a density p_1 with respect to a density p_2, due to the difference in model parameters:

$$
\begin{aligned}
KL(p_1(x)\|p_2(x)) &= \int_{-\infty}^{\infty} \log\left[\frac{p_1(x)}{p_2(x)}\right] p_1(x)dx \\
&= P_1 \log\left[\frac{P_1}{P_2}\right] + (1-P_1)\left[\log\left[\frac{\sigma_2(1-P_1)}{\sigma_1(1-P_2)}\right] + \frac{1}{2}\left[\frac{\sigma_1^2}{\sigma_2^2} + \frac{(\mu_2-\mu_1)^2}{\sigma_2^2} - 1\right]\right] \qquad (33)
\end{aligned}
$$

In fact this expression is not strictly a distance as it is not symmetric, so we use the symmetrized version $KL(p_1(x), p_2(x)) = \frac{1}{2}(KL(p_1(x)\|p_2(x)) + KL(p_2(x)\|p_1(x)))$.

Once we have a first segmentation of the field by clustering, we re-estimate the parameters for each final cluster in order to obtain a new set of motion-texture model parameters that represent each class.

6.2. Energy Maximization method

Maximizing equation (32) is performed using the technique of *Graph cuts*[9,10] for assigning labels to points in the image grid. In this framework, one seeks the labeling $\mathbf{l}$ that optimizes energy functions of the type:

$$
E(\mathbf{l}) = \sum_{i \in S} D_i(l_i) + \sum_{<i,j>} E^{ij}(l_i, l_j) \qquad (34)
$$

where $D_i(l_i)$ is a function that takes into account the energy associated to assigning the label l_i to image point x_i and that is derived from the observed data. $E^{ij}(l_i, l_j)$ is a term that measures the cost of assigning labels l_i, l_j to neighboring points and is related to spatial smoothness.

The method is based on constructing a directed graph $G = (V, A)$, with vertices V and edges A. The set of vertices include the images points and two special vertices, namely, *terminal vertices*. Then, we define a set $C \in A$ of edges of the graph, which we refer as a *cut*, that splits G into two disjoint sets in such a way that each set contains one of the terminal vertices. The result is an induced graph $G' = (V, A - C)$. From all the possible partitions of this kind, the minimum cut is defined as the one that has the smallest cost $|C|$, i.e., the sum of costs of all edges in C.

Finally, it is shown that there is a one to one correspondence between a partition of G and a complete labeling of the image points. Moreover, assigning appropriate edge weights, obtaining a minimum cut is equivalent to optimizing $E(\mathbf{l})$. Thus, the formulation reduces to computing a min-cut/max-flow problem. In our case we have to rewrite equation (32) in the form of (34). For a two-class problem we have:

$$
E(\mathbf{l}) = Q_1(\mathbf{x}_1) + Q_2(\mathbf{x}_2) - \log(Z_1(\mathbf{l})) - \log(Z_2(\mathbf{l})) + Q_S(\mathbf{l}) \qquad (35)
$$

where for each class k, $Q_k = \sum_{i \in S_k} V_i^k(x_i) + \sum_{<i,j> \in S_k} V_{ij}^k(x_i, x_j)$, V_i^k and V_{ij}^k are the corresponding potentials for each motion texture model and $S_k \in S$ is the subset of N_k points that belong to texture k. Supposing that $Z_k = (1/P(x_i = 0))^{N_k}$ as described in section 5, we define:

$$
\begin{aligned}
D_i(l_i) &= V_i^{l_i}(x_i) + \sum_{j \in \mathcal{N}_i} \frac{V_{ij}^{l_i}(x_i, x_j)}{2} + \log P(x_i = 0 \mid l_i) \\
E^{ij}(l_i, l_j) &= [\mathbb{I}_0(l_i - l_j) - 1]\left[\gamma + (\frac{V_{ij}^{l_i}(x_i, x_j)}{2} + \frac{V_{ij}^{l_j}(x_i, x_j)}{2})\right] \qquad (36)
\end{aligned}
$$

Thus, we have come to a formulation of the energy function that allows the direct application of graph cut algorithms for energy optimization.

7. RESULTS

We have applied our motion-texture segmentation method to natural scenes consisting of at most two moving textures ($c = 2$). In Fig.1 we analyze a situation where two real motion textures (steam and ocean) are combined artificially. We display in Fig.1b the motion measurement map v_{obs}. In Fig.1c we see that segmentation results are quite satisfactory. The results on the Fountain sequence in Fig.2 also show the accuracy of the method.

In Fig.3 we consider another real natural scene. It consists of two overlapping trees moved by the wind. The segmentation result in Fig.3c shows that the two main regions are well separated. This is a complex scene since the trees have not only similar intensity textures, but also similar motion textures.

8. CONCLUSIONS AND FUTURE WORK

We have proposed a new mixed-state motion texture model that has shown to be a powerful non-linear representation for describing complex dynamic content with only a few parameters. An original segmentation framework has also been designed which does not assume conditional independence of the observations for each of the textures, and fully exploits the mixed-state motion texture model. In this context, new results on partition function calculations have been obtained. As demonstrated by the reported experiments, segmentation results were quite satisfactory for two-class problems.

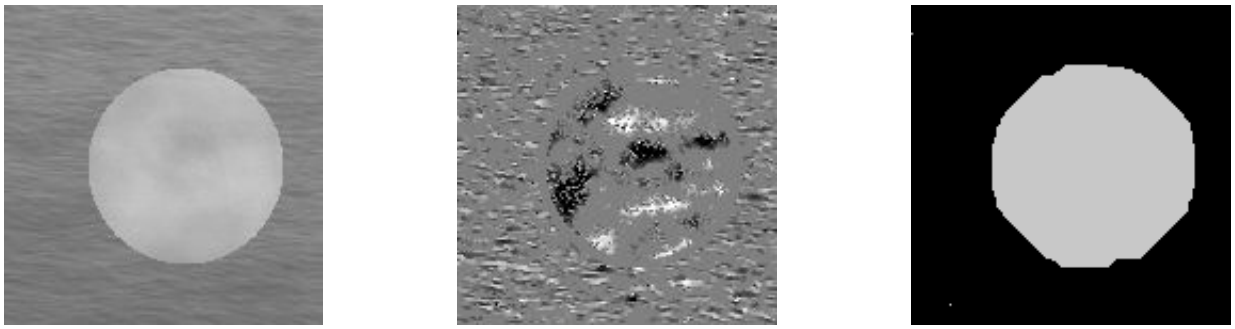
Currently, other issues are being investigated: extensions to include the temporal evolution of the motion measurements, a deeper study of the partition function, introduction of contextual information through the utilization of Conditional Markov Random Fields.

ACKNOWLEDGMENTS

This work was partially supported by the FIM project funded by INRIA and by an INRIA internship, the University of Buenos Aires, and CONICET, Argentina. The authors want to thank Stefano Soatto and Daniel Cremers for supplying the Ocean-Steam sequence.

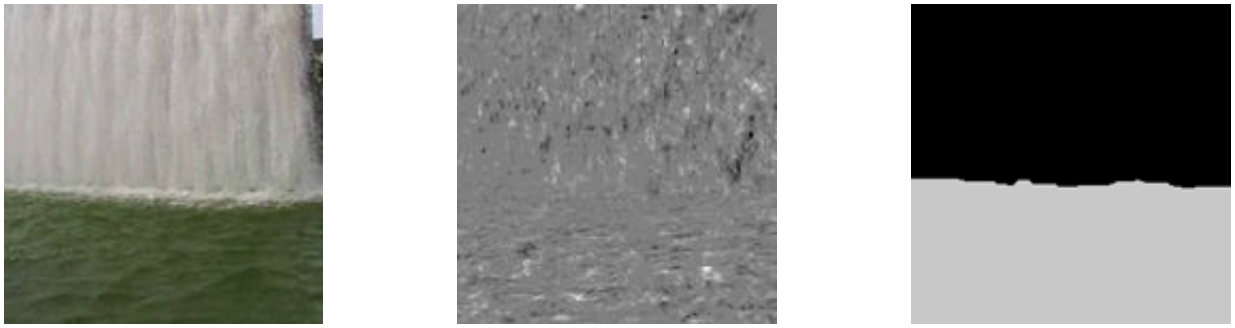
REFERENCES

1. G. Doretto, D. Cremers, P. Favano, and S. Soatto, "Dynamic texture segmentation," in *Proc. of 9th Int. Conf. on Computer Vision. ICCV'03, Nice*, pp. 1236–1242, Oct 2003.
2. A. Chan and N. Vasconcelos, "Mixtures of dynamic textures," in *in Proc. of the 10th IEEE Int. Conf. on Computer Vision, ICCV'05, Beijing*, pp. 641–647, Oct 2005.
3. M. Szummer and R. Picard, "Temporal texture modelling," in *Proc. of the 3rd IEEE Int. Conf. on Image Processing, ICIP'96, Lausane*, pp. 823–826, Sept. 1995.
4. L. Yuan, F. Wen, C. Liu, and H. Shum, "Synthesizing dynamic textures with closed-loop linear dynamic systems," in *Proc. of the 8th European Conf. on Computer Vision, ECCV'04, Prague*, **LNCS 3022**, pp. 603–616, Springer, May 2004.
5. P. Bouthemy, C. Hardouin, G. Piriou, and J. Yao, "Mixed-state auto-models and motion texture modeling." To appear, 2006.
6. B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence* **17**, pp. 185–203, Aug. 1981.
7. R. Fablet and P. Bouthemy, "Motion recognition using non-parametric image motion models estimated from temporal and multiscale co-ocurrence statistics," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**, pp. 1619–1624, Dec 2003.
8. J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B* **36**, pp. 192–236, 1974.
9. Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intellingence* **20**, pp. 1222–1239, Nov. 2001.
10. V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE. Trans. Pattern Analysis and Machine Intelligence* **26**, pp. 147–159, Feb. 2004.



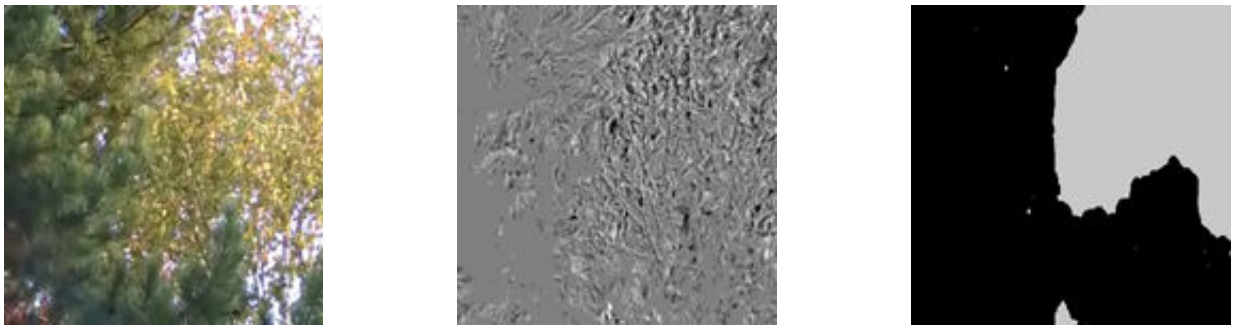
(a) (b) (c)

Figure 1. Ocean-Steam sequence : a) image from the original sequence, b) motion map, c) result of the segmentation process.



(a) (b) (c)

Figure 2. Fountain sequence: a) image from the original sequence, b) motion map, c) result of the segmentation process.



(a) (b) (c)

Figure 3. Tree sequence: a) image from the original sequence, b) motion map, c) result of the segmentation process.