TEMPORAL MODELING OF MOTION TEXTURES WITH MIXED-SATES MARKOV CHAINS

*T. Crivelli *, B. Cernuschi Frías*† , P. Bouthemy ** and J.F. Yao ‡*

* University of Buenos Aires and † CONICET, Argentina
** IRISA/INRIA and ‡ IRMAR/University of Rennes 1, France

ABSTRACT

Dynamic textures are time-varying visual patterns that exhibit certain spatio-temporal stationarity properties and are displayed mostly by natural scene elements. In this paper, we present new statistical models for the characterization of motion in this type of sequences. First we observe that motion measurements present values of two types: a discrete component at zero expressing the absence of motion and a continuous distribution for the rest of the motion values. Thus, we define random variables with mixed-states and propose to model a sequence of motion maps as a Markov chain, where the transition densities are mixed-state probability densities. Based on this approach, we propose a method for dynamic texture segmentation in real sequences showing the efficiency of the proposal in dynamic content analysis applications.

Index Terms— Markov processes, motion analysis, image segmentation, stochastic fields.

1. INTRODUCTION

Dynamic textures are time-varying visual patterns that exhibit certain spatio-temporal stationarity properties and are displayed mostly by natural scene elements. Typical examples can be found in: smoke, rain, moving foliage, etc. Its study in the context of computer vision is relatively recent and the main efforts in the modeling of this type of image sequences was mainly devoted to define linear dynamical systems that describe the evolution of image intensity over time [1, 2]. However, normal and optical flow have been pointed out as a very efficient and natural way of characterizing the local spatio-temporal dynamics of a dynamic texture [3], and thus, its modeling is an issue to explore, and exploit.

When we analyze motion measurements extracted from dynamic texture sequences from a statistical point of view, we observe that motion variables display values of two types: a discrete component at the null motion value, and a continuous distribution for the remaining values. Then, it can be helpful to explicitly define random variables that take either a discrete or symbolic state expressing the absence of motion, or continuous real values accounting for actual measurements.

Compact representations of these types of information have been recently introduced in the form of *mixed-state Markov random fields* [4, 5] for purely spatial schemes.

In this paper, we introduce new mixed-state models for the temporal modeling of dynamic textures, also called *motion textures*. A mixed state Markov chain framework is defined herein, assuming causal dependence as a first approach to studying their evolution over time. An application to motion texture image segmentation is given together with results on real image sequences.

2. MOTION TEXTURES

Let $I_i(t)$ be a scalar function that represents the image intensity at image location i for time t. Here, we consider the normal flow as local motion measurements, defined as $\mathbf{V}_i^n(t) = -\frac{\frac{\partial I_i(t)}{\partial t}}{\|\nabla I_i(t)\|}\frac{\nabla I_i(t)}{\|\nabla I_i(t)\|}$, where $\frac{\partial I_i(t)}{\partial t}$ is the temporal derivative of the image intensity. We follow the approach of [5], introducing a weighted vectorial average of normal flow in order to keep direction information:

$$\tilde{\mathbf{V}}_i^n(t) = \frac{\sum\limits_{j \in W} \mathbf{V}_j^n(t) \parallel \nabla I_j(t) \parallel^2}{max(\sum\limits_{j \in W} \parallel \nabla I_j(t) \parallel^2, \eta^2)}, \tag{1}$$

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:

$$v_i(t) = \tilde{\mathbf{V}}_i^n(t) \cdot \frac{\nabla I_i(t)}{\parallel \nabla I_i(t) \parallel}, \tag{2}$$

with $v_i \in (-\infty, +\infty)$. Once the motion measure is obtained, a sequence of intensity images is mapped to a sequence of *motion maps or fields* called *motion textures*. The resulting motion measurements are of a mixed nature. The underlying discrete property of no-motion for a point in the image, is represented as a null observation $v_i = 0$, and acts as the symbolic component of the model. Thus, the null motion value in this case, has a peculiar place in the sample space, and consequently, has to be modeled accordingly.

3. MIXED-STATE PROBABILITY FRAMEWORK

We first outline the theoretical framework attached to mixed-state random variables. Define $E = \{r\} \cup \mathbb{R}^*$ where $\mathbb{R}^* = \mathbb{R} \setminus \{r\}$, with r a possible "discrete" value. A random variable X defined on this space, called *mixed-state variable*, is constructed as follows: with probability $\rho \in (0, 1)$, set $X = r$, and with probability $= 1-\rho$, X is continuously distributed in $\mathbb{R}^*$. Consequently, the distribution function of X can be expressed as a monotone increasing function with a "step jump" at $X = r$.

In order to compute the probability density function of the mixed-state variable X, E is equipped with a "mixed" reference measure, $m(dx) = \nu_r(dx) + \lambda(dx)$, where ν_r is the discrete measure at r and λ the Lebesgue measure on $\mathbb{R}^*$. Hereafter, we will consider that $r = 0$ without loss of generality, expresing the property of no-motion in the context of motion texture modeling. Let us define the indicator function of the null value $\mathbf{1}_0(x)$ and its complementary function $\mathbf{1}_0^*(x) = 1 - \mathbf{1}_0(x)$. Then, the above random variable X has the following density function, w.r.t. $m(dx)$,

$$p(x) \quad = \quad \rho\mathbf{1}_0(x) + (1-\rho)\mathbf{1}_0^*(x)g(x), \qquad (3)$$

where $g(x)$ is a continuous pdf defined on $\mathbb{R}$. These definitions can be extended easily to multi-dimensional mixed-state variables.

4. TEMPORAL MODELING OF MOTION TEXTURES

We propose modeling the motion textures as a Markov chain of random fields, $\mathbf{X} = \{\mathbf{X}_t\}_{t:0...T}$, defined over a state space E^S, where $\mathbf{X}_t = \{x_i(t)\}_{i\in S}$ represents a motion texture or field computed at time t, and $S = \{1, 2, ...N\}$ is a spatial lattice of sites or image locations. This random field is assumed to have a probability distribution function with an everywhere positive density p, w.r.t the mixed-state product measure $\mu = m^{\otimes S}$. Therefore, the transition kernel for the mixed-state Markov chain (MS-MC) is defined as,

$$P(\mathbf{x}_{t-1}, \mathbf{x}_t) = p(\mathbf{X}_t = \mathbf{x}_t \mid \mathbf{X}_{t-1} = \mathbf{x}_{t-1})\mu(d\mathbf{x}_t). \qquad (4)$$

A similar model but in another context is proposed in [6]. For a first order Markov chain, we can write:

$$p(\mathbf{X}) = p(\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_T) = p(\mathbf{X}_0)\prod_{t=1}^{T} p(\mathbf{X}_t \mid \mathbf{X}_{t-1}). \quad (5)$$

For the purely causal temporal model that we study here, a first assumption to consider is spatial conditional independence within a motion texture for time t. This simplifies considerably the formulation of the problem w.r.t. spatial models [5], for which equation (5) can, rarely, be completely known [7]. Consequently, given the previous instant, and assuming a local dependency on a neighborhood $\chi_{i,t-1}$ of 'past' random variables,

$$p(\mathbf{X}_t \mid \mathbf{X}_{t-1}) = \prod_{i\in S} p(x_{i,t} \mid \mathbf{X}_{\chi_{i,t-1}}), \qquad (6)$$

where $\mathbf{X}_{\chi_{i,t-1}}$ is the subset of $\mathbf{X}_{t-1}$ restricted to a neighborhood of locations $\chi_{i,t-1}$ and $x_{i,t} = x_i(t)$. In our case, we will assume that the temporal neighborhood is a 9-point set which includes the previous (at $t-1$) center, diagonal, anti-diagonal, horizontal and vertical motion variables for a point at time t.

4.1. Gaussian mixed-state Markov Chains (MS-MC)

A mixed-state Markov Chain is defined as a Markov chain where the transition densities are mixed-state probability densities. Particularly, for a Gaussian MS-MC the continuous part of the corresponding conditional mixed-state density (3) follows a Gaussian law with mean $m_{i,t} \equiv m(\mathbf{X}_{\chi_{i,t-1}})$ and variance $\sigma^2_{i,t} \equiv \sigma^2(\mathbf{X}_{\chi_{i,t-1}})$. Then, the local conditional mixed-state densities are defined as

$$p(x_{i,t} \mid \mathbf{X}_{\chi_{i,t-1}}) = \rho_{i,t}\mathbf{1}_0(x_{i,t}) + \rho^*_{i,t}\mathbf{1}^*_0(x_{i,t})\frac{e^{-\frac{(x_{i,t}-m_{i,t})^2}{2\sigma^2_{i,t}}}}{\sqrt{2\pi}\sigma_{i,t}}, \qquad (7)$$

where $\rho_{i,t} = P(x_{i,t} = 0 \mid \mathbf{X}_{\chi_{i,t-1}})$ is now a function of $\mathbf{X}_{\chi_{i,t-1}}$ and $\rho^*_{i,t} = 1 - \rho_{i,t}$.

An interesting case for motion texture modeling is given, when the mean $m_{i,t}$ is a sort of weighted average of its neighbors,

$$m_{i,t} = c + \sum_{j\in\chi_{i,t-1}} h_j x_{j,t-1}, \qquad (8)$$

and $\sigma^2_{i,t} = \sigma^2$ is a constant for every point. This enforces local correlation, captures important properties as the orientation of the texture, and at the same time, keeps the model simple and with a limited number of parameters. Following [8], for such a mixed-state conditional density, we can write the corresponding joint probability density (5) as a Gibbs distribution, i.e. $p(\mathbf{X}) = \exp Q(\mathbf{X})/Z$, with Z the partition function or normalizing factor, and for which $Q(\mathbf{X}) = Q^d(\mathbf{X}) + Q^c(\mathbf{X})$ and,

$$Q^d(\mathbf{X}) = \sum_{i,t}\alpha\mathbf{1}^*(x_{i,t}) + \sum_t \sum_{\langle i,j\rangle : j\in\chi_{i,t-1}} \beta_j\mathbf{1}^*(x_{i,t})\mathbf{1}^*(x_{j,t-1}), \qquad (9)$$

$$Q^c(\mathbf{X}) = -\frac{1}{2\sigma^2}\sum_{i,t}[cx_{i,t} - x^2_{i,t} + \sum_{j\in\chi_{i,t-1}} h_j x_{j,t}x_{j,t-1}]. \qquad (10)$$

Moreover, the probability of the null value is given by $\rho_{i,t} = \left[1 + \sqrt{2\pi}\sigma e^{\alpha + \sum\beta_j\mathbf{1}^*(x_{j,t-1}) + \frac{m^2_{i,t}}{2\sigma^2}}\right]^{-1}$.

Finally, the parameters that define the Gaussian MS-MC model are $\varphi = \{\sigma^2, \alpha, \{\beta_j\}, c, \{h_j\}\}$. Well known estimation techniques can be applied in order to estimate the set of parameters that defines the conditional distributions of equation (7) and the transitions densities, as well. A Maximum Likelihood formulation w.r.t. those parameters is straightforward to obtain.

5. APPLICATION TO TEXTURE SEGMENTATION

The problem of segmentation is equivalent to assign a label to each point in the image grid for every time t, indicating that it belongs to a certain motion texture class. Following a Bayesian approach, we search for a label realization $\mathbf{l} = \{l_{i,t}\}$, where $l_{i,t} \in \{1, 2, ..., c\}$ is the class label value of $x_{i,t}$, that maximizes $p(\mathbf{l} \mid \mathbf{X}) \propto p(\mathbf{X} \mid \mathbf{l})p(\mathbf{l})$, where $\mathbf{X}$ represents the motion sequence. 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:

$$p(\mathbf{X} \mid \mathbf{l}) = \prod_{k=1}^{c} p(\mathbf{X}_k; \varphi_k) = \prod_{k=1}^{c} \prod_{i,t:l_{i,t}=k} p(x_{i,t} \mid \mathbf{X}_{\chi_{i,t-1}}; \varphi_k) \quad (11)$$

$\mathbf{X}_k$ is the vector of motion random variables that belong to texture k, and is a subset of $\mathbf{X}$. For the *a priori* information on the segmentation label field, $p(\mathbf{l})$, we introduce another Markov field on the same temporal neighborhood as for the MS-MC, that behaves as a regularization term for the labeling process, so $p(\mathbf{l}) \propto exp[Q_S(\mathbf{l})]$ with:

$$Q_S(\mathbf{l}) = \sum_{i,t} \sum_{j \in \chi_{i,t-1}} \gamma \mathbb{I}_0(l_{i,t} - l_{j,t-1}), \quad (12)$$

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:

$$E(\mathbf{l}) = \sum_{k=1}^{c} \sum_{i,t:l_{i,t}=k} \log p(x_{i,t} \mid \mathbf{X}_{\chi_{i,t-1}}; \varphi_k) + Q_S(\mathbf{l}), \quad (13)$$

where from (7), (9) and (10),

$$\log p(x_{i,t} \mid \cdot) = \alpha \mathbf{1}^*(x_{i,t}) - \frac{c}{2\sigma^2} x_{i,t} + \frac{x_{i,t}^2}{2\sigma^2} +$$
$$+ \sum_{j \in \chi_{i,t-1}} [\beta_j \mathbf{1}^*(x_{i,t}) \mathbf{1}^*(x_{j,t-1}) + \frac{h_j}{2\sigma^2} x_{i,t} x_{j,t-1}] + \log \rho_{i,t}. \quad (14)$$

The maximization of equation (13) w.r.t $\mathbf{l}$ is performed using fast optimization algorithms based on the technique of *Graph cuts* [9] for assigning labels to points in the image grid.

5.1. Experimental results

We have applied our motion-texture segmentation method to natural scenes consisting of at most two moving textures ($c = 2$). The sequences are processed using non-overlapping temporal windows with a length of $L = 5$ frames. For each window, the texture temporal models are estimated and the segmentation algorithm is applied.

We have first to estimate the parameters of the different motion textures involved in the processed image sequence. After computing the normal flow field from consecutive images within the temporal window, we divide the images of the sequence in blocks of 50x50 pixels and for each block, assuming that neither of the motion textures are moving in that interval, a set of motion-texture model parameters is estimated. Then, we apply a clustering technique to obtain a first splitting of blocks in two classes (composite blocks can be easily discarded at this stage). After this step, we recompute the parameters for each cluster and we obtain an estimate for both motion-texture models.

In Fig. 1 a) and b), we analyze a situation where two real motion textures (steam and ocean) are combined artificially. In Fig. 1 c) and d), we present another combined sequence where we have two separated regions of moving leaves over a motion texture of grass. Finally, in Fig. 1 e) and f), we have another complex real sequence that corresponds to a fountain over a static background. In this last case, we observe that the segmented regions vary across time as the fountain changes its shape. It is worthy to note that, in fact, the static background is a motion texture for our model, as it can be represented with the symbolic value, “absence of motion”, corresponding to null motion measurements.

6. CONCLUSIONS

We have presented new mixed-state models for motion texture temporal modeling. It has been shown that the discrete-continuous nature of the proposed motion measurements, settles the necessity for an appropriate modeling framework, which considers, not only the distribution of motion values, but also the distribution of the symbolic information. As we illustrate in the experiments, these models are very efficient in challenging problems as segmentation. With only a few parameters, they have shown to be a very powerful non-linear representation for describing complex dynamic content.

7. REFERENCES

[1] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto, “Dynamic textures,” *Intl. Journal of Comp. Vision*, vol. 51, no. 2, pp. 91–109, 2003.

[2] R. Vidal and A. Ravichandran, “Optical flow estimation and segmentation of multiple moving dynamic textures,” in *Proc. of CVPR’05*, San Diego, 2005, vol. 2, pp. 516–521.

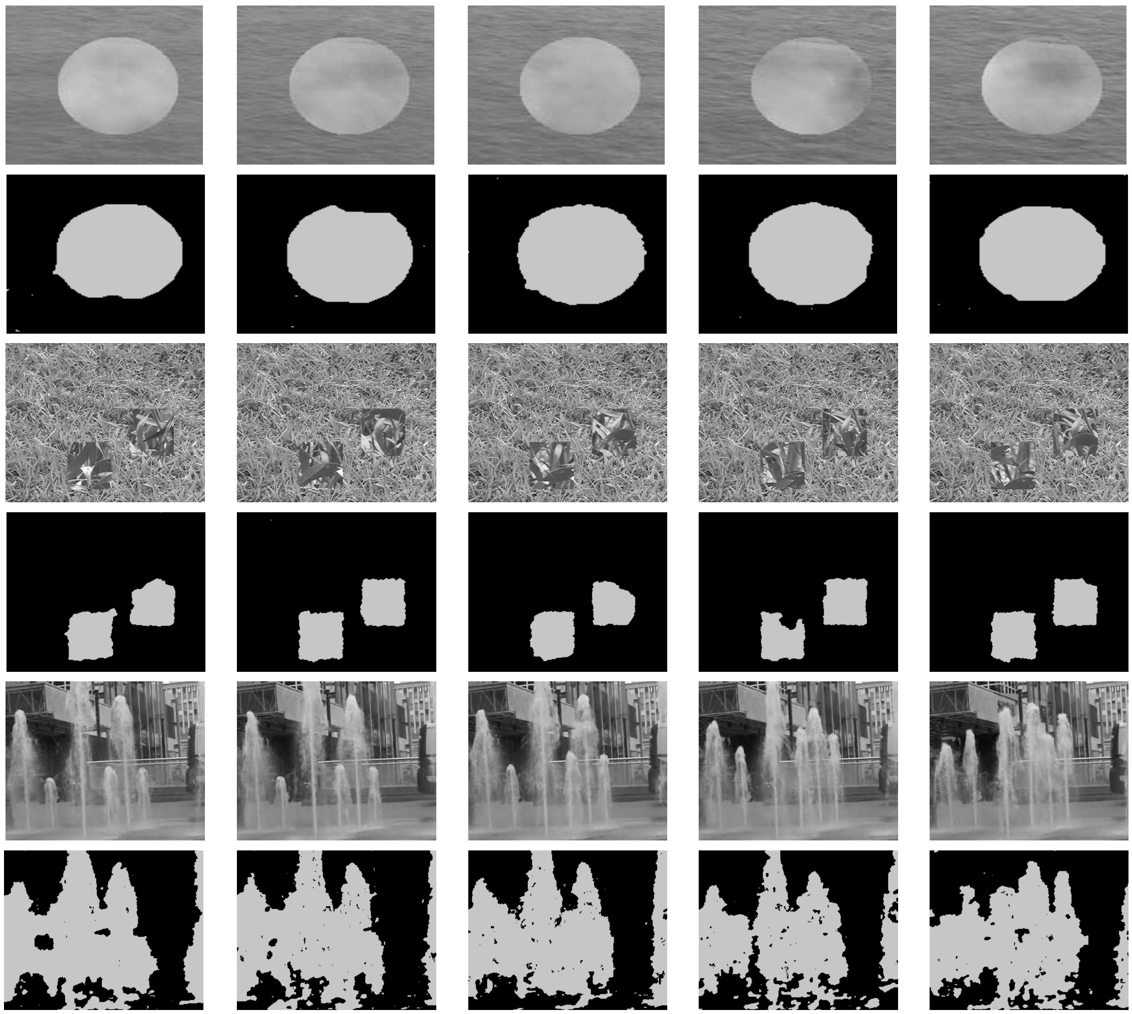

Fig. 1. a) Ocean-steam sequence (frames 5-20-35-50-65), b) segmentation using the proposed model, c) and d) grass-leaves sequence, e) and f) fountain sequence.

[3] S. Fazekas and D. Chetverikov, "Normal versus complete flow in dynamic texture recognition: a comparative study," in *Texture 2005: 4th international workshop on texture analysis and synthesis. ICCV'05*, Beijing, 2005, pp. 37–42.

[4] P Bouthemy, C Hardouin, G Piriou, and J.F. Yao, "Mixed-state auto-models and motion texture modeling," *Journal of Mathematical Imaging and Vision*, vol. 25, pp. 387–402, 2006.

[5] T. Crivelli, B. Cernuschi, P. Bouthemy, and J.F. Yao, "Mixed-state Markov random fields for motion texture modeling and segmentation," in *Proceedings of IEEE International Conference on Image Processing, ICIP'06*, Atlanta, 2006, pp. 1857–1860.

[6] F. Salzenstein and C. Collet, "Fuzzy markov random fields versus chains for multispectral image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1753–1767, November 2006.

[7] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B*, vol. 36, pp. 192–236, 1974.

[8] Bruno Cernuschi-Frias, "Mixed-states markov random fields with symbolic labels and multidimensional real values," *Rapport de Recherche INRIA.*, 2007, Http://hal.inria.fr/docs/00/16/59/37/PDF/RR-6255.pdf.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intellingence*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.