RECOGNITION OF DYNAMIC VIDEO CONTENTS BASED ON MOTION TEXTURE STATISTICAL MODELS

∗ Tomas Crivelli, ∗† Bruno Cernuschi-Frias
∗ Faculty of Engineering, University of Buenos Aires, Buenos Aires, Argentina. †CONICET, Argentina

Patrick Bouthemy
IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

Jian-Feng Yao
IRMAR/Univ. of Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France

Keywords: Motion analysis, Markov random fields, image content classification, dynamic textures.

Abstract: The aim of this work is to model, learn and recognize, dynamic contents in video sequences, displayed mostly by natural scene elements, such as rivers, smoke, moving foliage, fire, etc. We adopt the *mixed-state* Markov random fields modeling recently introduced to represent the so-called motion textures. The approach consists in describing the spatial distribution of some motion measurements which exhibit values of two types: a discrete component related to the absence of motion and a continuous part for measurements different from zero. Based on this, we present a method for recognition and classification of real motion textures using the generative statistical models that can be learned for each motion texture class. Experiments on sequences from the DynTex dynamic texture database demonstrate the performance of this novel approach.

1 INTRODUCTION

In the context of visual motion analysis, *motion textures* refer to dynamic video contents displayed mostly by natural scene elements. They are closely related to *temporal* or *dynamic textures* (Doretto et al., 2003). Different from *activities* (walking, climbing, playing) and *events* (open a door, answer the phone), *temporal textures* show some type of stationarity and homogeneity, both in space and time. Typical examples can be found in nature scenes: rivers, smoke, rain, moving foliage, etc.

When analyzing a complex scene, the three types of dynamic visual information (activities, events and temporal textures) may be present. However, their dissimilar nature leads to considering substantially different approaches for each one in tasks as detection, segmentation, and recognition.

The aim of this work is to model the apparent motion contained in dynamic textures, with special interest in dynamic content recognition. Generally speaking, model-based approaches (Doretto et al., 2003; Saisan et al., 2001; Yuan et al., 2004) have been mainly dedicated to describe the evolution of intensity over time, while motion-based methods (Fazekas and Chetverikov, 2005; Lu et al., 2005; Peteri and Chetverikov, 2005; Vidal and Ravichandran, 2005) propose the use of motion measurements (mainly based on optical and normal flow) as input features for a classification step.

We adopt the *mixed-state* Markov random fields (MS-MRF) model introduced in (Bouthemy et al., 2006), to represent the so-called motion textures. The approach consists in describing the spatial distribution of some motion measurements which exhibit values of two types: a discrete component related to the absence of motion and a continuous part for real measurements.

Based on this, we present a method for recognition and classification of real motion textures using the MS-MRF generative statistical models that can be learned for each motion texture class. Experiments on sequences from the DynTex (Peteri et al., 2005) dynamic texture database demonstrate the performance of this novel approach.

Crivelli T., Cernuschi-Frias B., Bouthemy P. and Yao J. (2008).
RECOGNITION OF DYNAMIC VIDEO CONTENTS BASED ON MOTION TEXTURE STATISTICAL MODELS.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 283-289
DOI: 10.5220/0001078102830289
Copyright © SciTePress

2 LOCAL MOTION MEASUREMENTS

In (Fazekas and Chetverikov, 2005), the effectiveness of normal flow versus complete optical flow measurements, in the context of dynamic texture recognition, is analyzed. They conclude that for small data sets, normal flow is an adequate description of temporal texture dynamics, while complete flow performs better when the number of classes grows. However, they use motion descriptors directly as input features for the classification step. No underlying statistical model is proposed.

Our approach consists in defining a statistical model for motion measurements that allows us to rely on the representativeness of the model, more than on the accuracy of the motion measurements. At the same time, normal flow can be directly computed locally, avoiding the computational burden associated to dense flow estimation. Finally, our method is based on modeling the instantaneous motion maps as spatial random fields, where the amount of spatial statistical interaction between motion variables will intervene in the motion texture recognition.

We consider the normal flow as local motion measurements. However, in contrast to (Fablet and Bouthemy, 2003), we do not consider its magnitude only, but its vectorial expression defined by $\mathbf{V}_n(p) = -\frac{I_t(p)}{\|\nabla I(p)\|}\frac{\nabla I(p)}{\|\nabla I(p)\|}$, where p is a location in the image I. Then, we introduce the following weighted averaging of the normal flow vectors to smooth out noisy measurements and enforce reliability:

$$\tilde{\mathbf{V}}_n(p) = \frac{\sum_{q \in W} \mathbf{V}_n(q) \parallel \nabla I(q) \parallel^2}{max(\sum_{q \in W} \parallel \nabla I(q) \parallel^2, \eta^2)}, \quad (1)$$

where η^2 is a constant, as in (Fablet and Bouthemy, 2003), and W is a small window centered in p. Finally, we consider the following scalar expression:

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

which projects the smoothed normal motion over the spatial intensity gradient direction, resulting in $v_{obs} \in (-\infty, +\infty)$.

In Fig. 1 we observe the result of applying the proposed motion measurements to two pairs of consecutive images for two different sequences. We observe in the motion histograms that the statistical distribution of the motion measurements has two elements: a discrete component at the null value $v_{obs} = 0$, and a continuous distribution for the rest of the motion values. The underlying discrete property of no-motion

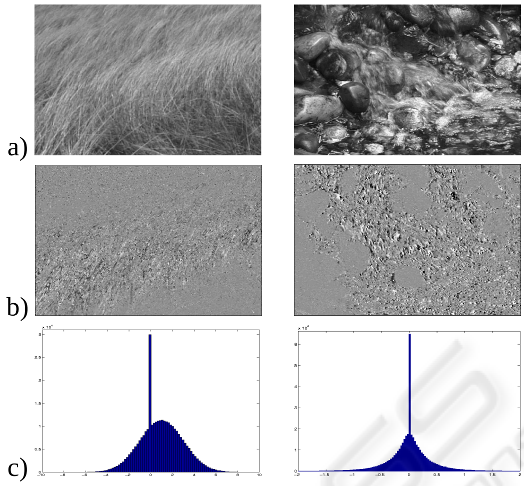

Figure 1: a) Images from original sequences (left: straw, right: water-rocks) obtained from the DynTex texture database and their corresponding, b) motion textures, and c) motion histograms.

for a point in the image, is represented as a null observation, and acts as a symbolic component in the model. It is not the value by itself that matters, but the binary property of what is called *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.

3 MIXED-STATE MRF MODEL

The key observation made in the previous section about the statistical properties of motion measurements, settles the necessity for an adequate representation of the associated random variables. In a first approach as in (Crivelli et al., 2006), observing the histograms, the problem can be formulated defining a probability density for the motion values, that is composed by two terms, i.e.

$$p(x) = \rho\delta_0(x) + (1-\rho)f(x). \quad (3)$$

where $\delta_0(x)$ is the Dirac impulse function centered at zero. This density is well-defined and corresponds to a random variable that has a discrete value with probability mass concentrated at zero. In fact, it is easy to observe that $P(x = 0) = \rho$.

3.1 Measure Theoretic Approach

From a probability theory point of view, it is more natural to redefine the probability space w.r.t. a new probability measure, avoiding to deal with a functional or distribution as the Dirac delta, that may complicate the strict definition of the corresponding density, allowing also to generalize the case of a discrete real value (e.g., $x = 0$) to a generic symbolic value or abstract label, that may lie on an arbitrary label set.

In this section, we outline the theoretical framework attached to mixed-state random variables. Let us define $\mathcal{M} = \{r\} \cup \mathbb{R}^*$ where $\mathbb{R}^* = \mathbb{R} \setminus \{r\}$, with r a possible "discrete" value, sometimes called *ground* 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 $\mathbb{R}^*$. Hereafter, we will assume that $r = 0$, without loss of generality and all the results are inmediately extended to abstract symbolic values for r. This is the main advantage of the measure theoretic approach.

Consequently, the distribution function of X can be expressed as a monotone increasing function with a "step jump" at $X = 0$. In order to compute the probability density function of the mixed-state variable X, $\mathcal{M}$ is equipped with a "mixed" reference measure:

$$m(dx) = \nu_0(dx) + \lambda(dx), \quad (4)$$

where ν_0 is the discrete measure for the value 0 and λ the Lebesgue measure on $\mathbb{R}^*$. Such a measure has already been used in (Salzenstein and Pieczynski, 1997) for simultaneous fuzzy-hard image segmentation. Let us define the indicator function of the discrete value 0 as $\mathbf{1}_0(x)$ and its complementary function $\mathbf{1}^*_0(x) = \mathbf{1}_{\{0\}^c}(x) = 1 - \mathbf{1}_0(x)$. Then, the above random variable X has the following density function w.r.t. $m(dx)$:

$$p(x) \quad = \quad \rho \mathbf{1}_0(x) + (1-\rho)\mathbf{1}^*_0(x) f(x), \quad (5)$$

where $f(x)$ is a continuous pdf from an absolutely continuous distribution w.r.t. λ, defined on $\mathbb{R}$. Equation (5) corresponds to a mixed-state probability density.

4 MIXED-STATE SPATIAL MARKOV MODELS FOR MOTION TEXTURES

Let $\mathbf{X} = \{x_i\}_{i \in 1...N}$ be a motion field or motion texture obtained as in section 2. We define the neighborhood $\mathcal{N}_i$ of any image point i, as the 8-point nearest neighbors set, and $\mathbf{X}_{\mathcal{N}_i}$ as the subset of random variables restricted to $\mathcal{N}_i$. Then,

$$\mathcal{N}_i = \{i_E, i_W, i_N, i_S, i_{NW}, i_{SE}, i_{NE}, i_{SW}\}, \quad (6)$$

where, for example, i_E is the east neighbor of i in the image grid, i_{NW} the north-west neighbor, etc.

The following mixed-state conditional model is considered:

$$p(x_i \mid \mathbf{X}_{\mathcal{N}_i}) = \rho_i \mathbf{1}_0(x_i) + (1-\rho_i)\mathbf{1}^*_0(x_i) f(x_i \mid \mathbf{X}_{\mathcal{N}_i}, x_i \neq 0) \quad (7)$$

where $x_i \in \mathbf{X}$ is the motion information measurement at point $i \in S = \{1,..,N\}$ of the image grid and $\rho_i = P(x_i = 0 \mid \mathbf{X}_{\mathcal{N}_i})$. Consequently, we have a distribution consisting of two parts: a discrete component for $x_i = 0$ and a continuous one for $x_i \neq 0$. This model gives a specific attention to the null value of the random variable x_i, which corresponds to the property of no motion for a point. For the continuous part, f, we assume a Gaussian density with variance σ_i^2 and mean m_i depending on $\mathbf{X}_{\mathcal{N}_i}$. The motion histograms suggest the use of this distribution.

In contrast to (Bouthemy et al., 2006), where a truncated zero-mean Gaussian distribution is assumed, the proposed extension takes into account a strongest correlation between continuous motion values, as we will see shortly. This is crucial in our formulation as it allows describing more realistic motion textures. Then, after some rearrangements, we may write:

$$p(x_i \mid \mathbf{X}_{\mathcal{N}_i}) \quad = \quad exp\left[\boldsymbol{\Theta_i}^T(\mathbf{X}_{\mathcal{N}_i}) \cdot \mathbf{S}(x_i) + \log \rho_i\right] \quad (8)$$

where $\boldsymbol{\Theta_i}^T(\mathbf{X}_{\mathcal{N}_i}) \in \mathbb{R}^d$ and $\mathbf{S}(x_i) \in \mathbb{R}^d$, with $d = 3$ for our case and,

$$\boldsymbol{\Theta}_i(\mathbf{X}_{\mathcal{N}_i}) \quad = \quad \left[\ \log\left(\frac{1-\rho_i}{\sigma_i\sqrt{2\pi}\rho_i}\right) - \frac{m_i^2}{2\sigma_i^2}, \quad \frac{1}{2\sigma_i^2}, \quad \frac{m_i}{\sigma_i^2}\ \right]^T \quad (9)$$

$$\mathbf{S}(x_i) \quad = \quad \left[\ \mathbf{1}^*_0(x_i), \quad -x_i^2, \quad x_i\ \right] \quad (10)$$

As the seminal theorem of Hammersley-Clifford states, Markov random fields with an everywhere positive density function, are equivalent to nearest neighbor Gibbs distributions. The joint p.d.f. 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 in the Markov random fields theory that the energy $Q(\mathbf{X})$ can be expressed as a sum of potential functions, $Q(\mathbf{X}) = \sum_{C \subset S} V_C(\mathbf{X}_C)$, defined over all subsets $\mathcal{C}$ of the lattice space S (Besag, 1974).

For a one-parameter conditional model ($d = 1$) satisfying the assumption that it belongs to the exponential family and that it depends only on cliques that contain no more than two sites, i.e. *auto-models*, the expression for the parameter is known w.r.t. the sufficient statistics $\mathbf{S}(\cdot)$ of the neighbors (Besag, 1974). In the case of multi-parameter auto-models ($d > 1$), the result of (Besag, 1974) was extended in (Bouthemy et al., 2006) and (Cernuschi-Frias, 2007) showing that:

$$\boldsymbol{\Theta}_i(\mathbf{X}_{\mathcal{N}_i}) = \boldsymbol{\alpha}_i + \sum_{j \in \mathcal{N}_i} \boldsymbol{\beta}_{ij} \mathbf{S}(x_j) \tag{11}$$

Moreover, they give the expression of the first and second order potentials of the expansion of the energy function $Q(\mathbf{X})$:

$$V_i(x_i) = \boldsymbol{\alpha}_i^T \cdot \mathbf{S}(x_i) \tag{12}$$

$$V_{ij}(x_i, x_j) = \mathbf{S}^T(x_i) \boldsymbol{\beta}_{ij} \mathbf{S}(x_j) \tag{13}$$

We thus define a mixed-state Markov Random Field (MS-MRF) auto-model for the motion texture, where the local conditional probability densities are mixed-state densities.

4.1 Motion Texture Model Parameters

From equation (11) it is clear that $\boldsymbol{\beta}_{ij} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{\alpha}_i \in \mathbb{R}^3$. First, we propose that, for the Gaussian continuous density, f, the conditional mean is given by,

$$m_i(\mathbf{X}_{\mathcal{N}_i}) = \frac{c_i}{2b_i} + \sum_{j \in \mathcal{N}_i} \frac{h_{ij}}{2b_i} x_j \tag{14}$$

That is, the mean motion value of the continuous part for a point is a sort of weighted average of the values of its neighbors. Second, we assume that the variance is constant for every point, i.e. $\sigma_i^2(\mathbf{X}_{\mathcal{N}_i}) = \sigma_i^2$. From this assumptions, several coefficients of the matrix $\boldsymbol{\beta}_{ij}$ are null. Additionally, note that the second order potentials are defined over non-ordered pairs of points and thus, a symmetry condition arises since $V_{ij} = V_{ji}$. Consequently, from equation (13), $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_{ji}^T$. Finally, we have

$$\boldsymbol{\beta}_{ij} = \begin{pmatrix} d_{ij} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & h_{ij} \end{pmatrix} \qquad \boldsymbol{\alpha}_i = [\ a_i \quad b_i \quad c_i\ ]^T \tag{15}$$

In general terms, the proposed conditional models could be defined by a different set of parameters for each location of the image. This would give rise to a motion texture model with a number of parameters proportional to the image size. Unfortunately, such high-dimensional representation of the associated dynamic information is unfeasible in practice and does not constitutes a compact description of motion textures. Moreover, an increasingly number of frames would be necessary for the estimation process. This conspires against a formulation oriented to efficient content recognition and retrieval. However, if needed, our framework could deal with spatially non-stationary motion textures. Then, we propose to use an *homogeneous* model where $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$ and $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_k$, and k is an index that indicates the position of the neighbor w.r.t the point, and is independent of location i. Moreover, a necessary condition in order to define a stationary spatial process, is that the parameters related to symmetric neighbors (E-W, N-S, NW-SE, NE-SW) must be the same. This is a consequence of the symmetry of the potentials. Then, $k \in \{H, V, D, AD\}$, for the Horizontal, Vertical, Diagonal, and Anti-Diagonal interacting directions. Finally, we have the set of 11 parameters which characterizes the motion-texture model, $\phi = \{a, b, c, d_H, h_H, d_V, h_V, d_D, h_D, d_{AD}, h_{AD}\}$.

We now write the expression of the resulting Gibbs energy for a Gaussian mixed-state Markov random field,

$$\begin{aligned} Q(\mathbf{X}) &= \sum_i a\mathbf{1}^*(x_i) - bx_i^2 + cx_i \\ &+ \sum_{<i,j>} d_{ij}\mathbf{1}^*(x_i)\mathbf{1}^*(x_j) + h_{ij}x_ix_j \end{aligned} \tag{16}$$

Here, we distinguish two main parts: a set of terms related to the interaction between the discrete components of the neighbors, and terms related to a continuous Gaussian Markov random field. Although the energy is functionally decomposed in two parts, this does not mean that the two types of values (discrete-continuous) are independent in the Gibbs field, and effectively, the estimation of the parameters has to be done jointly.

4.2 Model-parameter Estimation

In order to estimate the eleven parameters of the motion-texture model from motion measurements, we adopt the pseudo-likelihood maximization criterion (Besag, 1974), since the partition function for exact maximum likelihood formulation is in general intractable. Therefore, we search the set of parameters $\hat{\phi}$ that maximizes the function $L(\phi) = \prod_{i \in S} p(x_i \mid \mathbf{X}_{\mathcal{N}_i}, \phi)$. We use a gradient descent technique for the optimization as the derivatives of L w.r.t ϕ are known in closed form.

5 MOTION TEXTURE RECOGNITION

One of the key aspects of a model oriented to dynamic content recognition, as the one proposed here, is the ability to define a way of computing some similarity measure between models, in order to embed it in a decision-theory-based application. In this context, the Kullback-Leibler (KL) divergence is a well-known distance (more precisely, a pseudo-distance) between statistical models. Here we present a result for computing this quantity between general Gibbs distributions and obtain an expression for the case of mixed-state models, that will allow us to classify a set of real motion textures.

5.1 A Simmilarity Measure Between Mixed-state Models

The KL divergence from a density $p_1(\mathbf{X})$ to $p_2(\mathbf{X})$ is defined as

$$KL(p_1(\mathbf{X}) \mid p_2(\mathbf{X})) = \int_\Omega \log \frac{p_1(\mathbf{X})}{p_2(\mathbf{X})} p_1(\mathbf{X}) m(d\mathbf{X}), \quad (17)$$

which is independent of the measure m. This is not strictly a distance as it is not symmetric; define then the symmetrized KL divergence as $d_{KL}(p_1(\mathbf{X}), p_2(\mathbf{X})) = \frac{1}{2}\left[KL(p_1(\mathbf{X}) \mid p_2(\mathbf{X})) + KL(p_2(\mathbf{X}) \mid p_1(\mathbf{X}))\right]$.

Now, assume that $p_1(\mathbf{X})$ and $p_2(\mathbf{X})$ are Markov random fields, i.e. $p_1(\mathbf{X}) = Z_1^{-1} \exp Q_1(\mathbf{X})$ and $p_2(\mathbf{X}) = Z_2^{-1} \exp Q_2(\mathbf{X})$. Define $\Delta Q(\mathbf{X}) = Q_2(\mathbf{X}) - Q_1(\mathbf{X})$. Then $\log \frac{p_1(\mathbf{X})}{p_2(\mathbf{X})} = -\Delta Q(\mathbf{X}) + \log \frac{Z_2}{Z_1}$, and

$$d_{KL}(p_1(\mathbf{X}), p_2(\mathbf{X})) = \frac{1}{2}\left(E_{p_2}\left[\Delta Q(\mathbf{X})\right] - E_{p_1}\left[\Delta Q(\mathbf{X})\right]\right) \quad (18)$$

We observe from this general equation, that we do not need to have knowledge of the partition functions of the Gibbs distributions which simplifies enormously the handling of this expression. Now, let $p_1(\mathbf{X})$ and $p_2(\mathbf{X})$ be two Gaussian MS-MRF. Then,

$$\Delta Q(\mathbf{X}) = \sum_i \Delta\alpha \mathbf{S}(x_i) + \sum_{<i,j>} \mathbf{S}(x_i) \Delta\beta_{ij} \mathbf{S}(x_j) \quad (19)$$

where $\Delta\alpha = \alpha^{(2)} - \alpha^{(1)}$ and $\Delta\beta_{ij} = \beta_{ij}^{(2)} - \beta_{ij}^{(1)}$. From this definition:

$$\begin{aligned} E_{p_1}\left[\Delta Q(\mathbf{X})\right] &= \sum_i \Delta\alpha E_{p_1}\left[\mathbf{S}(x_i)\right] \\ &+ \sum_{<i,j>} E_{p_1}\left[\mathbf{S}(x_i)\Delta\beta_{ij}\mathbf{S}(x_j)\right] \quad (20) \end{aligned}$$

As we have an homogeneous model, the expectations in the last equation are equal for each site of the motion field. The same result applies to $E_{p_2}\left[\Delta Q(\mathbf{X})\right]$ and then it is straightforward to compute equation (18). In a practical application, the idea is to use the parameters of the two models that we want to compare, to generate synthetic fields using a Gibbs sampler (Geman and Geman, 1984) from which we can estimate the involved expectations and finally calculate the divergence.

5.2 Experiments

The recognition performance of mixed-state motion texture models was tested with real sequences extracted from the DynTex dynamic texture database (Peteri et al., 2005). We first took motion textures where the homogeneity assumption was mostly valid and divided them in 10 different classes (Figure 2): steam, straw, traffic, water, candles, shower, flags, water-rocks, waves, fountain. A total of 30 different sequences were considered, and for each one, 5 pairs of consecutive images were selected at frames 1,20,40,60,80, for a total of 150 samples. Each motion texture class parameter set was learned from a single pair of images picked from only one of the sequences belonging to each type of motion texture. All sequences were composed by gray scale images with a resolution of 720x576 pixels, given at a rate of 25 frames per second. In order to reduce computation time, the original images were filtered and subsampled to a resolution of 180x144 pixels.

Having estimated the reference model parameters for each of the 10 training samples, we then estimated ϕ for each test sample and computed the distance with each learned parameter vector, as explained in the last section. The recoginition was based on assigning the class of motion texture that was closer to the test sample.

In Table 1 we show the confusion matrix for the 10 motion texture classes. A correct recognition is considered when both, the test sample and the closest reference parameter vector, belong to the same class.

A promising overall classification rate of 90.7% was achieved. As for the confusion matrix, let us note that it is likely that waves are classified as water or water-rocks as they correspond to similar dynamic contents, straw is confused with shower, they have similar vertical orientation, and candles can be classified as traffic, as both classes show a motion pattern consisting of isolated blobs. The non-symmetry of the confusion matrix is associated to the nature of the tested data set, where for some classes, the tested sequences have a closer resemblance to the training sample, while for others, there are notorious variations, that may lead to a misclassification.

Reported experiments for dynamic texture recognition using a model-based approach as the one presented in (Saisan et al., 2001), have shown a classification rate of 89.5%, on similar data sets. Their method is based on computing a subspace distance between the linear models learned for each class, that describe the evolution of image intensity over time. Although the efectiveness of this approach is similar to ours, the method proposed here has a big advan-

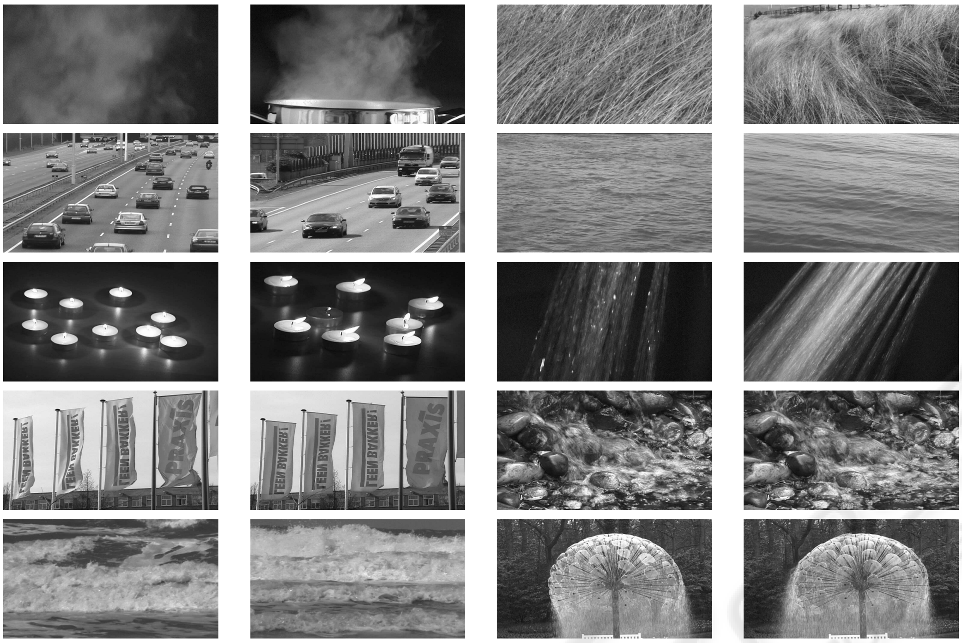

Figure 2: Sample images from the 10 motion texture classes used for the recognition experiments.

Table 1: Motion texture class confusion matrix. Each row indicates how the samples for a class were classified.

	steam	straw	traffic	water	candles	shower	flags	water-rocks	waves	fountain
steam	100%	-	-	-	-	-	-	-	-	-
straw	-	93.3%	-	-	-	6.7%	-	-	-	-
traffic	-	-	86.7%	-	-	-	-	13.3%	-	-
water	-	-	-	100%	-	-	-	-	-	-
candles	-	-	13.3%	-	73.4%	-	-	13.3%	-	-
shower	20%	-	-	-	-	80%	-	-	-	-
flags	-	-	-	-	-	-	100%	-	-	-
water-rocks	-	-	-	-	-	-	-	93.3%	-	6.7%
waves	-	-	-	6.7%	-	-	-	13.3%	80%	-
fountain	-	-	-	-	-	-	-	-	-	100%

tage, that is, we only need two consecutive frames to estimate and recognize the mixed-state models, while in (Saisan et al., 2001) subsequences of 75 frames are used. This is a consequence of modeling the spatial structure of motion rather than the time evolution of the image.

Motion-based methods for dynamic texture classification (Fazekas and Chetverikov, 2005; Peteri and Chetverikov, 2005) have shown an improved performance with recognition rates of over 95% using invariant flow statistics. Although they are the most accurate reported results for addressing this problem, they do not provide a general characterization of dynamic textures, as model-based approaches do. Consequently, more complex scenarios with combined problems, as simultaneous detection, segmentation and recognition of these type of sequences are not directly addressable. Our framework provides an unified statistical representation suitable of beeing applicable to other complex problems, as well (Crivelli et al., 2006; Bouthemy et al., 2006).

6 CONCLUSIONS

We have presented a novel comprehensive motion modeling framework for motion texture recognition. Our approach appropriately handles the mixed-state nature of motion measurements where the parametric representation of the statistical models showed quite satisfactory results on motion texture recognition. In our method, we do not need to process the whole sequence to obtain a reliable estimate of the model in order to achieve an accurate classification rate. The approach is entirely valid, with minor modifications, for motion texture detection as well. Moreover, the reference classes can be learned from a single instantaneous motion map allowing, eventually, to define an adaptive scheme for recognition and classification.

Future prospects are based on considering other dissimilarity measures between statistical models, combining the classification and detection approach with existing motion or dynamic texture segmentation methods and considering the introduction of contextual information through discriminative models, possibly in the form of Conditional Markov Random Fields (CMRF).

REFERENCES

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, 36:192–236.

Bouthemy, P., Hardouin, C., Piriou, G., and Yao, J. (2006). Mixed-state auto-models and motion texture modeling. *Journal of Mathematical Imaging and Vision*, 25:387–402.

Cernuschi-Frias, B. (2007). Mixed-states markov random fields with symbolic labels and multidimensional real values. *Rapport de Recherche INRIA*.

Crivelli, T., Cernuschi, B., Bouthemy, P., and Yao, J. (2006). Segmentation of motion textures using mixed-state markov random fields. In *Proceedings of SPIE*, volume 6315, 63150J.

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

Fablet, R. and Bouthemy, P. (2003). 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(12):1619–1624.

Fazekas, S. and Chetverikov, D. (2005). 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*, pages 37–42.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Lu, Z., Xie, W., Pei, J., and Huang, J. (2005). Dynamic texture recognition by spatio-temporal multiresolution histograms. In *IEEE Workshop on Motion and Video Computing (WACV/MOTION)*, pages 241–246.

Peteri, R. and Chetverikov, D. (2005). Dynamic texture recognition using normal flow and texture regularity. In *Proc. of IbPRIA, Estoril*, pages 223–230.

Peteri, R., Huiskes, M., and Fazekas, S. (2005). Dyntex: A comprehensive database of dynamic textures. http://www.cwi.nl/projects/dyntex/index.html.

Saisan, P., Doretto, G., Wu, Y., and Soatto, S. (2001). Dynamic texture recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. CVPR'01, Hawaii*, pages 58–63.

Salzenstein, F. and Pieczynski, W. (1997). Parameter estimation in hidden fuzzy markov random fields and image segmentation. *Graph. Models Image Process.*, 59(4):205–220.

Vidal, R. and Ravichandran, A. (2005). Optical flow estimation and segmentation of multiple moving dynamic textures. In *Proc. of CVPR'05, SanDiego*, volume 2, pages 516–521.

Yuan, L., Wen, F., Liu, C., and Shum, H. (2004). Synthesizing dynamic textures with closed-loop linear dynamic systems. In *Proc. of the 8th European Conf. on Computer Vision, ECCV'04, Prague*, volume LNCS 3022, pages 603–616.

SciTePress Science and Technology Publications