Dynamic Texture Segmentation

Gianfranco Doretto[†]   Daniel Cremers[†]   Paolo Favaro[‡†]   Stefano Soatto[†]

[†] Dept. of Computer Science, UCLA, Los Angeles, CA 90095, {doretto, cremers, soatto}@cs.ucla.edu
[‡] Dept. of Electrical Eng., Washington University, St. Louis, MO 63130, fava@ee.wustl.edu

Abstract

We address the problem of segmenting a sequence of images of natural scenes into disjoint regions that are characterized by constant spatio-temporal statistics. We model the spatio-temporal dynamics in each region by Gauss-Markov models, and infer the model parameters as well as the boundary of the regions in a variational optimization framework. Numerical results demonstrate that – in contrast to purely texture-based segmentation schemes – our method is effective in segmenting regions that differ in their dynamics even when spatial statistics are identical.

1. Introduction

Consider the following problem, in relation to Fig. 1: An autonomous vehicle must decide what is traversable terrain (e.g. grass) and what is not (e.g. water). This problem can be addressed by classifying portions of the image into a number of categories, for instance grass, dirt, bushes or water. For the most part, such a classification can be accomplished successfully by looking at simple image statistics, such as color or intensity. However, in many situations these are not sufficient, and therefore it may be beneficial to look at *spatio-temporal* statistics, and attempt to classify different portions of the scene based *not* on the statistics of one single image, but on how the statistics of an image change over time during a sequence. Modeling the (global) spatio-temporal statistics of the entire image can be a daunting task due to the complexity of natural scenes. An alternative consists of choosing a simple class of models, and simultaneously estimating regions and their model parameters in such a way that the data in each region is optimally modeled by the estimated parameters. This naturally results in a *segmentation* problem.

In this paper we study the problem of segmenting a sequence of images based on a simple model of its spatio-temporal statistics.

Before we proceed with formalizing the problem, we

Figure 1. *A typical outdoor scene: an autonomous vehicle trying to classify the terrain based on spatial image statistics fails to distinguish water from grass, since the latter reflects on the former and therefore their spatial statistics are very similar (courtesy of Google Image Search).*

would like to point out that segmentation, in this context, is entirely dependent on the *class of models* chosen. Different models result in different partitions of the scene, and there is no "right" or "wrong" result. Ultimately, the usefulness of a statistical segmentation method depends on how well the chosen model captures the phenomenology of the physical scene, but unless one has a physical model to start with, this correspondence cannot be guaranteed. Therefore, in Sect. 1.1 we describe the model we use, which implicitly defines what we mean by "segmentation". It is a Gauss-Markov model of the intensity of the pixels which is known as a *dynamic texture*.

Another issue that we would like to raise at the outset is that what we model is *not a point process*, but rather statistical distributions both in space and in time. Therefore, there will be a "minimum region of integration" in order to

capture such statistics, similarly to what is done in texture segmentation. To segment the image plane into regions of homogeneous dynamic textures, we revert to a region-based segmentation approach pioneered by Mumford and Shah [13]. The Mumford-Shah functional has been extended to the segmentation of color, spatial texture [26] and motion [4]. This work generalizes these approaches to the segmentation of spatio-temporal textures.

1.1. Formalization of the problem

Let $\Omega \subset \mathbb{R}^2$ be the domain of an image and $\{\Omega_i\}_{i=1,\ldots,N}$ be a partition of Ω into N (unknown) regions[1]. We assume that the pixels contained in the region Ω_i are a Gauss-Markov process[2]. In other words, there exist (unknown) parameters $A_i \in \mathbb{R}^{n \times n}, C_i \in \mathbb{R}^{m_i \times n}$, covariance matrices $Q_i \in \mathbb{R}^{n \times n}, R_i \in \mathbb{R}^{m_i \times m_i}$, white, zero-mean Gaussian processes $\{v(t)\} \in \mathbb{R}^n$, $\{w_i(t)\} \in \mathbb{R}^{m_i}$ and a process $\{x(t)\} \in \mathbb{R}^n$ such that the pixels at each region Ω_i at each instant of time are given by

$$\begin{cases} x(t+1) = A_i x(t) + \sqrt{Q_i} v(t); \ x(t_0) = x_{i,0} \\ y_i(t) = C_i x(t) + \sqrt{R_i} w_i(t) \end{cases} \quad (1)$$

where $y(t) \in \mathbb{R}^{m_i}$ represents the vector of intensities at time t of the pixels belonging to Ω_i. Note that we allow the number of pixels m_i to be different in each region, as long as $\sum_{i=1}^{N} m_i = m$, the size of the entire image, and that we require that neither the regions nor the parameters change over time, $\Omega_i, A_i, C_i, Q_i, R_i, x_{i,0} = \text{const}$.

Given this generative model, one way to formalize the problem of segmenting a sequence of images is the following: *Given a sequence of images $\{y(t) \in \mathbb{R}^m, \ t = 1, \ldots, T\}$ with two or more distinct regions $\Omega_i, \ i = 1, \ldots, N \geq 2$ that satisfy the model (1), estimate both the regions Ω_i, and the model parameters of each region, namely the matrices A_i, C_i, the initial state $x_{i,0}$, and the covariance of the driving process Q_i (the covariance R_i is not a parameter of interest since it does not have significant discriminative power).*

1.2. Relation to prior work

The model for the spatio-temporal statistics of one region, that we proposed in the previous section, was first used in [5]. Similar statistical models were also used by others [21, 6, 24, 9]. For synthesis of spatio-temporal textures, statistical generative models can be replaced by procedural techniques such as [19, 25].

The analytical tools we use to infer the model parameters are borrowed from the literature of subspace system

identification [15], and the discrepancy measure between different models is inspired by [3]. Since textures can be considered as a degenerate case of dynamic textures, our work also relates to texture segmentation. For some recent work on texture segmentation see [11, 8, 17] and references therein.

We cast the problem of region segmentation in a variational framework following [13]. In particular, we revert to a level set framework of the Mumford-Shah functional introduced in [2]. The algorithm we propose allows to partition the image domain of a video sequence into regions with constant spatio-temporal statistics. To the best of our knowledge, work in this area is new.

In the next section, we review existing work on how one can infer the parameters of a dynamic texture (1) for each region, assuming the region is known. In the following Section 3.1 we address the dual problem of inferring the regions if the models to which each pixel belongs are known. Since neither is known, we address this chicken-and-egg problem in Sect. 3.

2. Dynamic texture learning and comparison

If the regions $\Omega_i, \ i = 1, \ldots, N$ were known, one would just be left with two problems: one is the learning of the model parameters, which we review in Sect. 2.1, the other is the computation of a discrepancy measure between different dynamic textures, which we discuss in Sect. 2.2.

2.1. Learning

It is well known [10] that a positive-definite covariance sequence with rational spectrum corresponds to an equivalence class of second-order stationary processes. It is then possible to choose as a representative of each class a Gauss-Markov model with the given covariance. In other words, for a given region Ω_i, we can assume that there exist a positive integer n, a process $\{x(t)\}$ with realizations in $\mathbb{R}^n$ (the "state") with initial condition $x(t_0)$, some matrices A_i and C_i, and a symmetric positive semi-definite matrix $\begin{bmatrix} Q_i & S_i \\ S_i^T & R_i \end{bmatrix} \geq 0$, where $S_i = E[w(t)v^T(t)]$, such that $\{y(t)\}$ is the output of model (1). Since we assume that the noise affecting the state $v(t)$ and the noise affecting the output $w(t)$ are independent, we have that $S_i = 0$.

The choice of matrices A_i, C_i, Q_i, R_i is not unique, in the sense that there are infinitely many models that give rise to exactly the same measurement covariance sequence starting from suitable initial conditions[3]. In other words, any given process has *not* a unique model, but an *equivalence class* of models. In order to identify a unique model of

---

[1]That is, $\Omega = \cup_{i=1}^{N} \Omega_i$ and $\Omega_i \cap \Omega_j = \emptyset, i \neq j$.

[2]Although this may seem a restrictive assumption, it has been shown in [20, 6] that sequences such as foliage, water, smoke, and steam are well captured by this model.

[3]Substituting A_i with TA_iT^{-1}, C_i with C_iT^{-1}, Q_i with TQ_iT^{-1}, and choosing the initial condition $Tx(t_0)$, where $T \in GL(n)$ is any invertible $n \times n$ matrix generates the same output covariance sequence.

COMPUTER SOCIETY

the type (1) from a sample path $y(t)$, we choose a representative of each equivalence class as suggested in [5], i.e. we will make the following assumptions: $m_i >> n$, and $\text{rank}(C_i) = n$, and choose a model realization that makes the columns of C_i orthonormal, i.e. $C_i^T C_i = I_n$. This guarantees that the model corresponding to a given dataset is uniquely determined. This model corresponds to a *canonical realization* [7].

The problem of going from data to models can be formulated as follows: *given* measurements of a sample path of the process: $y(1), \ldots, y(T)$; $T >> n$, estimate $\hat{A}_i, \hat{C}_i, \hat{Q}_i$, a canonical model realization of the process $\{y(t)\}$. Ideally, we would want the maximum likelihood solution from the finite sample, that is

$$
\begin{aligned}
&\hat{A}_i, \hat{C}_i, \hat{Q}_i, \hat{x}_{i,0} \qquad\qquad\qquad\qquad\qquad (2)\\
&= \arg \max_{\substack{A_i, C_i \\ Q_i, x_{i,0}}} \log p\big(y(1), \ldots, y(T) | A_i, C_i, Q_i, x_{i,0}\big).
\end{aligned}
$$

Notice that, as we said in Section 1.1, we do not model the covariance of the measurement noise R_i, since that carries no information on the underlying process. The asymptotically efficient solution for the estimation problem (2), as $T \to \infty$, can be found in [23], while accurate description of its implementation, named N4SID, can be found in [15]. In practice, for computational efficiency, we settle for the suboptimal solution described in [5].

2.2. Discrepancy between dynamic textures

Assuming that the parameters $A_i, C_i, Q_i, x_{i,0}$ have been inferred for each region, in order to set the stage for a segmentation procedure, one has to define a discrepancy measure among regions. The difficulty in doing so is that each region is described not only by the parameters above, but by an equivalence class of such parameters, obtained by changes of basis of the state-space $\{x(t)\}$ in model (1). Therefore, a suitable discrepancy measure has to compare *not* the parameters directly, but their equivalence classes.

One technique for doing so has been recently proposed in [3]. It consists of building infinite observability matrices, whose columns span the vector space generated by the measurements $y(t)$ of the model (1), and that represent the high-dimensional subspace of the infinite-dimensional space of all possible measurements. Then one can compute the geometric angles between such subspaces through their embedding.

More formally, let $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{m \times q}$ be two matrices with full column rank, and suppose that $p \geq q$. The q principal angles $\theta_k \in \left[0, \frac{\pi}{2}\right]$ between range(A) and

range(B) are recursively defined for $k = 1, 2, \ldots q$ as

$$
\begin{aligned}
\cos \theta_1 &= \max_{\substack{x \in \mathbb{R}^p \\ y \in \mathbb{R}^q}} \frac{\left|x^T A^T B y\right|}{\|Ax\|_2 \, \|By\|_2} \qquad\qquad (3)\\
&= \frac{\left|x_1^T A^T B y_1\right|}{\|Ax_1\|_2 \, \|By_1\|_2}, \\
\cos \theta_k &= \max_{\substack{x \in \mathbb{R}^p \\ y \in \mathbb{R}^q}} \frac{\left|x^T A^T B y\right|}{\|Ax\|_2 \, \|By\|_2} \\
&= \frac{\left|x_k^T A^T B y_k\right|}{\|Ax_k\|_2 \, \|By_k\|_2}, \quad \text{for } k = 2, \ldots, q
\end{aligned}
$$

subject to $x_i^T A^T A x = 0$ and $y_i^T B^T B y = 0$,
for $i = 1, 2, \ldots, k - 1$.

Now, let M_1 and M_2 be two models of the type (1), with the same output dimensionality, which are characterized in state space terms by their system matrices A_1 and A_2 and output matrix C_1 and C_2 respectively. Their infinite observability matrices $\mathcal{O}_i$, for $i = 1, 2$, are defined as

$$
\mathcal{O}_i \doteq \begin{bmatrix} C_i^T & A_i^T C_i^T & \cdots & (A_i^T)^n C_i^T & \cdots \end{bmatrix}^T \in \mathbb{R}^{\infty \times n}, \tag{4}
$$

and the principal angles between the ranges of $\mathcal{O}_1$, and $\mathcal{O}_2$ are referred to as *subspace angles*. Their computation can be carried out in closed form, and entails the computation of the eigenvalues of the solution of a discrete-time Lyapunov equation [3].

While more than one distance for single-input single-output (SISO) linear dynamical systems have been defined based on subspace angles [3, 12], the extension to the multiple-input multiple-output (MIMO) case is not trivial given the lack of the concept of the inverse of a MIMO system. However, it has been shown in [18] that subspace angles between infinite observability matrices have very high discriminative power under the hypothesis of stability and observability of the compared systems. With this in mind, we measure the discrepancy between different spatio-temporal statistics associated to different models by comparing either the set of subspace angles or their combination via Martin's distance [3] defined as

$$
d_M^2(M_1, M_2) = \ln \prod_{k=1}^n \frac{1}{\cos^2 \theta_k}. \tag{5}
$$

3. Dynamic texture segmentation

In Sect. 2.1 we have seen that, if the boundaries of each region were known, one could easily estimate a simple model of the spatio-temporal statistics within each region. Unfortunately, in general one does not know the boundaries,

which are instead part of the inference process. If the dynamic texture associated with each pixel were known, then one could easily determine the regions by thresholding or by other grouping or segmentation techniques. However, a dynamic texture associated with a certain pixel ξ, as defined in equation (1), depends on the whole region Ω_i containing the pixel ξ. Therefore, we have a classic "chicken-and-egg" problem: If we knew the regions, we could easily identify the dynamical models, and if we knew the dynamical models we could easily segment the regions. Unfortunately, we know neither.

One may be tempted to address this problem by setting up an alternating minimization procedure, starting with an initial guess of the regions, $\hat{\Omega}_i(0)$, estimating the model parameters within each region, $\hat{A}_i(0), \hat{C}_i(0), \hat{Q}_i(0)), \hat{x}_{i,0}$, and then at any given time t seeking for regions $\hat{\Omega}_i(t)$, and parameters $\hat{A}_i(t), \hat{C}_i(t), \hat{Q}_i(t)), \hat{x}_{i,0}$ that minimize a chosen cost functional. Unfortunately, this approach would not lead to a well-posed problem, since one can always explain the image with a few high-order models with large support regions (the entire image in the limit), or with many low-order models with small support regions (individual pixels in the limit). Therefore, a model complexity cost needs to be added, for instance the description length of the model parameters [16] and the boundaries of each region. This significantly complicates the algorithms and the derivation.

Rather than seeking for an estimate of the regions and the model parameters in one shot, we can instead adopt a two-stage algorithm to circumvent the model-complexity issue: We first associate a local *signature* to each pixel, by integrating visual information on a fixed spatial neighborhood of that pixel; then we group together pixels with similar signatures in a region-based segmentation approach. The signatures are computed from the subspace angles relative to a reference model, following the ideas outlined in previous sections. We describe this simple and yet effective approach in the following subsections.

3.1. A geometric approach

We start by considering the neighborhoods $B(\xi) \subset \Omega$ around each pixel $\xi \in \Omega$. We then associate to each pixel location ξ the dynamics of the spatio-temporal region by computing $\mathcal{O}(\xi)$ from $A(\xi), C(\xi) = N4SID\{y(\tilde{\xi}, t) \mid \tilde{\xi} \in B(\xi), t = 1, \ldots, T\}$. For each pixel ξ we generate a local spatio-temporal *signature* given by the cosines of the subspace angles $\{\theta_j(\xi)\}$ between $\mathcal{O}(\xi)$ and a reference model, $\mathcal{O}(\xi_0)$:

$$s(\xi) = \big(\cos\theta_1(\xi), \ldots, \cos\theta_n(\xi)\big). \quad (6)$$

We call this approach "geometric" since the signatures are constructed using subspace angles, rather than responses of banks of filter as is more common in static texture segmentation.

With the above representation, the problem of dynamic texture segmentation can be formulated as one of grouping regions of similar spatio-temporal signatures. We propose to perform this grouping by reverting to the Mumford-Shah functional [13]. A segmentation of the image plane Ω into a set of pairwise disjoint regions Ω_i of constant signature $s_i \in \mathbb{R}^n$ is obtained by minimizing the cost functional

$$E(\Gamma, \{s_i\}) = \sum_i \int_{\Omega_i} \big(s(\xi) - s_i\big)^2 d\xi + \nu |\Gamma|, \quad (7)$$

simultaneously with respect to the region descriptors $\{s_i\}$ modeling the average signature of each region, and with respect to the boundary Γ separating these regions (an appropriate representation of which will be introduced in the next section). The first term in the functional (7) aims at maximizing the homogeneity with respect to the signatures in each region Ω_i, whereas the second term aims at minimizing the length $|\Gamma|$ of the separating boundary.

3.2. A level set formulation

In the following, we will restrict the class of possible solutions of the proposed variational problem to two-phase solutions, i.e. solutions in which each pixel is associated with one of two dynamic texture models. All results do, however, extend to the case of multiple phases. For the implementation of the boundary Γ in the functional (7) we revert to the implicit level set based representation proposed by Chan, Sandberg and Vese [2, 22].

Compared to explicit contour representations, the level set based representations [14] have several favorable properties. Firstly, they do not restrict the topology of the evolving boundary, thereby facilitating splitting and merging during the evolution. And secondly, one does not need to take care of a regridding of control or marker points.

Let the boundary Γ in (7) be given by the zero level set of a function $\phi : \Omega \to \mathbb{R}$:

$$\Gamma = \{\xi \in \Omega \mid \phi(\xi) = 0\}. \quad (8)$$

With the Heaviside function

$$H(\phi) = \left\{ \begin{array}{ll} 1 & \text{if } \phi \geq 0 \\ 0 & \text{if } \phi < 0 \end{array} \right., \quad (9)$$

the functional (7) can be replaced by a functional on the level set function ϕ:

$$
\begin{aligned}
E(\phi, \{s_i\}) &= \int_{\Omega} \big(s(\xi) - s_1\big)^2 H(\phi) \, d\xi \\
&+ \int_{\Omega} \big(s(\xi) - s_2\big)^2 \big(1 - H(\phi)\big) \, d\xi \\
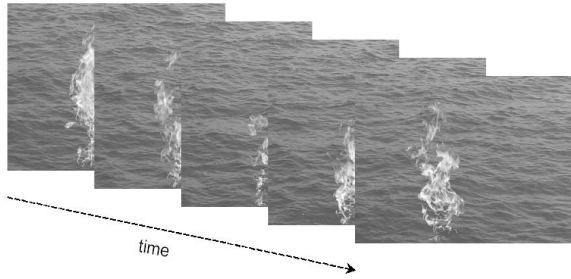&+ \nu |\Gamma|. \quad (10)
\end{aligned}
$$

Figure 2. *Example of a composition of dynamic textures:* fire on ocean waves. The time evolution of the sequence is rendered by overlapping a few snapshots corresponding to different image frames.

3.3. Energy minimization

We minimize the functional (10) by alternating the two fractional steps of:

- Estimating the mean signatures.

  For fixed ϕ, minimization with respect to the region signatures s_1 and s_2 amounts to averaging the signatures over each phase:

  $$s_1 = \frac{\int s\, H(\phi)\, d\xi}{\int H(\phi)\, d\xi}, \quad s_2 = \frac{\int s\, (1 - H(\phi))\, d\xi}{\int (1 - H(\phi))\, d\xi}. \quad (11)$$

- Boundary evolution.

  For fixed region signatures $\{s_i\}$, minimization with respect to the embedding function ϕ can be implemented by a gradient descent given by:

  $$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\nu \nabla \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + (s - s_2)^2 - (s - s_1)^2 \right],$$

4. Experiments

The following experiments demonstrate various aspects of dynamic texture segmentation. In all cases we use sequences of natural phenomena like ocean waves, smoke or fire. Figure 2 shows a few snapshots from a sequence of fire combined with the ocean waves.

In Section 4.1 we show experiments on a sequence where two regions have both different texture and different dynamics. In Section 4.2 we instead keep the dynamics identical, but we change the texture. On the contrary, in Section 4.3 we keep the texture identical, but we change the dynamics. As a last experiment, we test our algorithm on a very challenging sequence: we superimpose a flame on ocean waves. In this case, the region occupied by the flame is changing in time, as opposed to our assumptions (see Section 1.1) of static regions. The contour evolutions are also available on-line [1].

4.1. Smoke on the water...

In this experiment we generate a synthetic sequence by superimposing two sequences: one with ocean waves, and one with smoke. We select a disc in the middle of the images and overlap the sequence of smoke over the sequence of the waves only on this disc. All the sequences are composed of 120 frames and each frame is of 220×220 pixels. In this case both texture and dynamics are very different, resulting in a very strong discrimination between the regions. We identify the local dynamical systems at each pixel by considering neighborhoods of 11×11 pixels. The state of each local system is of dimension 10. Then, we compute the subspace angles between these local dynamical systems and a reference dynamical system. Finally, we segment the signatures with the Mumford-Shah minimization scheme as explained in Sections 3.1-3.3. In Figure 3 we show a few snapshots of the contour evolution, starting from a circle.

4.2. Segmentation by spatial orientation

We generate a synthetic sequence by superimposing two identical sequences of ocean waves, one of which has been rotated of 90 degrees. We select a disc and a square at the opposite corners of the images and overlap the sequence of rotated waves on the basic sequence only within the disc and the square. All the sequences are composed of 120 frames and each frame is of 221×321 pixels.

The texture on the disc and the square regions is similar to the texture in the background region in both gray-scale values and dynamics. However, it can be distinguished on the basis of the different orientation. As in the previous experiment we identify the local dynamical systems at each pixel by considering neighborhoods of 11×11 pixels. The state of each local system is again of dimension 10. Figure 4 shows a few snapshots of the contour evolution, starting from a circle.

4.3. Segmentation by temporal properties

This experiment is complementary to the previous one. To generate a sequence containing regions which only differ with respect to their dynamics, we overlap the ocean sequence in the regions corresponding to the disc and square over an ocean sequence slowed down by a factor of 2.

The evolution of the contour during the cost functional minimization is shown in Figure 5, starting from a circle. We found this experiment to be one of the most compelling ones, because the segmentation is obtained exclusively on the basis of temporal properties of the dynamic texture – this is one of the novelties of our approach.

4.4. ... and fire in the sky

The following experiment is a very challenging one, since we use an input sequence where the regions associated
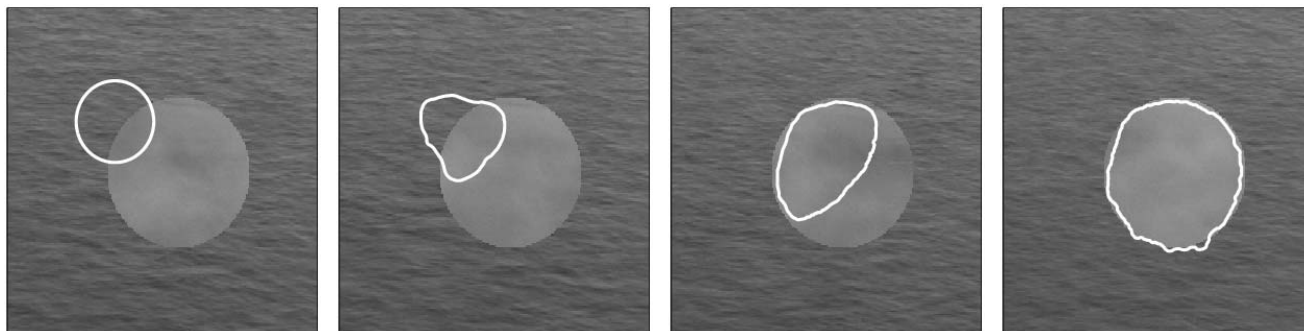
Figure 3. *Smoke on the water:* *in this first experiment the two dynamic textures (the smoke and the ocean waves) are very different both in the dynamics and in the appearance.*
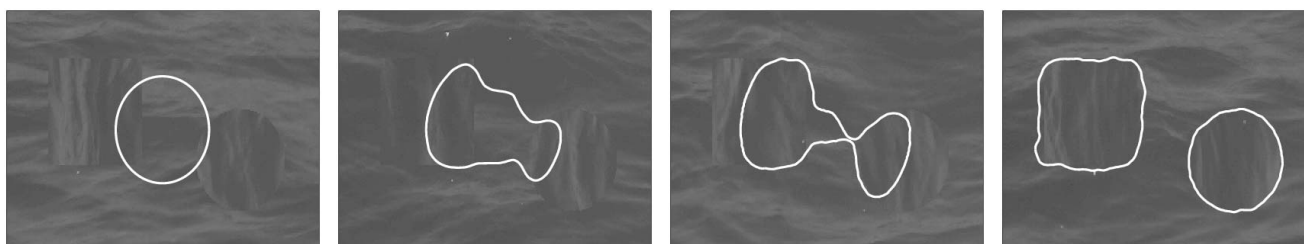

Figure 4. *Segmentation by changing texture:* *in this experiment we segment two dynamic textures that differ only for the texture orientation, but that share the same dynamics and general appearance (grayscale values).*
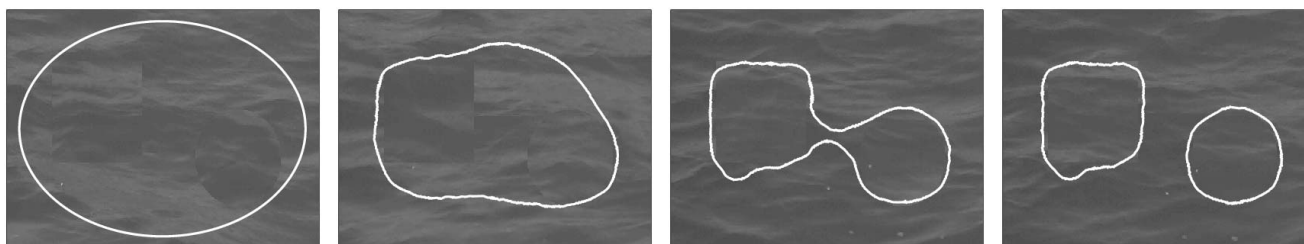

Figure 5. *Segmentation by changing dynamics:* *in this experiment we segment two dynamic textures that are identical in appearance, but differ in the dynamics. Note that this particular segmentation problem is quite difficult, even for human observers. Segmentation is obtained exclusively on the basis of the temporal properties of the dynamic texture. This demonstrates one of the novelties of our approach.*
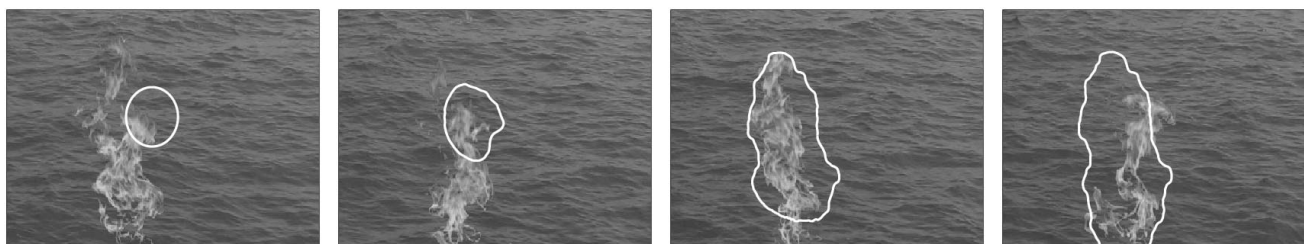

Figure 6. *Fire in the sky:* *this segmentation problem is very challenging, since – contrary to our model assumption – the regions where the dynamic textures are defined (in particular, the flame texture), are changing in time. The segmentation returns an estimate of the average location of the flame, showing that our approach is robust to deviations from the assumption of spatial stationarity.*

with different dynamic textures are also moving in time. We generate a synthetic sequence by superimposing a sequence with fire to a sequence with ocean waves. The flame is continuously changing position in time, thereby mixing in some regions both the fire dynamic texture and the ocean dynamic texture.

In Figure 6 we show a few snapshots of the contour evolution, starting from a circle. Notice that the final contour is the contour of an "average" region obtained by combining the different regions in time. Therefore, our approach shows robustness also to changes in the original hypotheses that dynamic textures were spatially stationary.

5. Conclusions

We have presented a technique to segment an image sequence into regions based on their spatio-temporal statistics. The assumption is that the regions are constant in time or change slowly relative to the dynamics of the irradiance within each region, which in turn is modeled as a second-order stationary process with constant Markov parameters. Our approach draws from the literature on region-based segmentation as well as that on system identification; however, it is not a straightforward concatenation of the two since that would result in either having to compare state space models of different dimensions, or in un-necessary blurring of the region boundaries.

As we illustrate in the experiments, our technique is effective in segmenting regions that differ in their dynamic behavior even when their spatial statistics are identical, and it can be useful in outdoor autonomous navigation and recognition of dynamic visual processes.

References

[1] http://www.cs.ucla.edu/~doretto/projects/dynamic-segmentation.html.

[2] T. Chan, B. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2):130–131, 2000.

[3] K. D. Cock and B. D. Moor. Subspace angles between linear stochastic models. In *Proc. Int. Conf. on Decision and Control*, volume 2, pages 1561–1566, Dec 2000.

[4] D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 53–58, June 2003.

[5] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *Int. Journal of Computer Vision*, 51(2):91–109, February 2003.

[6] A. W. Fitzgibbon. Stochastic rigidity: image registration for nowhere-static scenes. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 662–669, July 2001.

[7] T. Kailath. *Linear Systems*. Prentice Hall,Englewood Cliffs, NJ., 1980.

[8] J. Keuchel, C. Schnöerr, C. Schellewald, and D. Cremers. Binary partitioning, perceptual grouping, and restoration with semidefinite programming. *To appear in PAMI*, 2003.

[9] Y. Li, T. Wang, and H.-Y. Shum. Motion texture: a two-level statistical model for character motion synthesis. *Proc. SIGGRAPH '02*, pages 465–472, July 2002.

[10] L. Ljung. *System Identification: theory for the user*. Prentice Hall, 1987.

[11] J. Malik, S. Belongie, T. Leung, and S. Jianbo. Contour and texture analysis for image segmentation. *Int. Journal of Computer Vision*, 43:7–27, 2001.

[12] R. Martin. A metric for arma processes. *IEEE Trans. on Signal Processing*, 48:1164–7, Apr 2000.

[13] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[14] S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.

[15] P. V. Overschee and B. D. Moor. N4sid: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, Jan 1994.

[16] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[17] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. Technical report, INRIA, Sophia Antipolis, January 2003. RR n. 4695.

[18] P. Saisan, G. Doretto, Y. N. Wu, and S. Soatto. Dynamic texture recognition. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 58–63, December 2001.

[19] A. Schödl, R. Szeliski, D. H. Salesin, and I. Essa. Video textures. In *Proc. SIGGRAPH '00*, pages 489–498, July 2000.

[20] S. Soatto, G. Doretto, and Y.-N. Wu. Dynamic textures. In *Proc. Int. Conf. on Computer Vision*, volume 2, pages 439–446, July 2001.

[21] M. Szummer and R. W. Picard. Temporal texture modeling. In *Proc. Int. Conf. on Image Processing*, volume 3, pages 823–826, 1996.

[22] A. Tsai, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.

[23] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.

[24] Y. Z. Wang and S. C. Zhu. A generative method for textured motion analysis and synthesis. In *Proc. European Conf. on Computer Vision*, volume 1, pages 583–597, June 2002.

[25] L. Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. SIGGRAPH '00*, pages 479–488, 2000.

[26] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.

IEEE
COMPUTER
SOCIETY