Pattern Recognition 48 (2015) 3239–3248

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Classifying dynamic textures via spatiotemporal fractal analysis ☆

Yong Xu [a], Yuhui Quan [a,*], Zhuming Zhang [b], Haibin Ling [c], Hui Ji [d]

[a] School of Computer Science & Engineering, South China University of Technology, Guangzhou 510006, China
[b] Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong, China
[c] Department of Computer & Information Sciences, Temple University, Philadelphia, PA 19122, USA
[d] Department of Mathematics, National University of Singapore, Singapore 117542, Singapore

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form
15 March 2015
Accepted 14 April 2015
Available online 23 April 2015

Keywords:
Dynamic texture
Dynamic fractal spectrum
Fractal dimension
Volumetric analysis
Multi-slice analysis

ABSTRACT

The large-scale images and videos are one kind of the main source of big data. Dynamic texture (DT) is essential for understanding the video sequences with spatio-temporal similarities. This paper presents a powerful tool called dynamic fractal analysis to DT description and classification, which integrates rich description of DT with strong robustness to environmental changes. The proposed dynamic fractal spectrum (DFS) for DT sequences is composed of two components. The first one is a volumetric dynamic fractal spectrum component (V-DFS) that captures the stochastic self-similarities of DT sequences by treating them as 3D volumes; the second one is a multi-slice dynamic fractal spectrum component (S-DFS) that encodes fractal structures of repetitive DT patterns on 2D slices along different views of the 3D volume. To fully exploit various types of dynamic patterns in DT, five measurements of DT pixels are collected for the analysis on DT sequences from different perspectives. We evaluated our method on four publicly available benchmark datasets. All the experimental results have demonstrated the excellent performance of our method in comparison with state-of-the-art approaches.

1. Introduction

The explosive growth in the amount of data makes big data processing and analytics one of the hottest research topics. Roughly speaking, big data analytics aims at examining a large amount of data from various sources to uncover hidden patterns, unknown correlations as well as other useful information. One of the most-visible sources of big data is video, which is being generated pervasively by billions of sensors embedded in various types of devices like surveillance cameras and mobile phones. For analyzing the underlying patterns captured by videos, a fundamental issue is the feature extraction and description of dynamic patterns which are often in the form of dynamic texture.

Dynamic textures (DTs) are often regarded as video sequences of moving scenes that possess certain stationary properties in both space domain and time domain [1,2]. Such video sequences are ubiquitous in real world, like video clips of boiling water, rivers, sea waves, fountains, clouds, smoke, fire, swarm of birds, traffic flow, pedestrians in crowds, whirligig, facial expressions, etc. There are many applications concerning DT, such as video compression, video quality assessment, surveillance, detection of the onset of emergencies, foreground/background separation, and human–computer interaction; see e.g. [3–6]. In recent years, the related topics of DT in computer vision community have ranged from DT modeling and synthesis to recognition and classification. In this paper, we focus on the development of effective DT description and classification techniques, which can be integrated to many recognition systems that involve the characterization of dynamics, e.g. vision sensor based fire detection, DT segmentation based dynamic scene retrieval, real-time facial expression analysis, biometrics, etc.

Compared to static textures, dynamic textures vary not only on the spatial distribution of texture elements, but also on the organization and dynamics over time. One main challenge in the study of DT classification is how to reliably capture the motion behaviors of texture elements, i.e. the properties of dynamics of texture elements over time. Many existing approaches model the dynamics either by treating videos as samples of stochastic dynamical systems or by directly measuring the motion field of

☆ We thank Drs. G. Doretto, B. Ghanem, and G. Zhao for their help with the datasets. Yong Xu would like to thank the supports by National Nature Science Foundations of China (61273255 and 61070091), Engineering & Technology Research Center of Guangdong Province for Big Data Analysis and Processing ([2013]1589-1-11), Project of High Level Talents in Higher Institution of Guangdong Province (2013-2050205-47) and Guangdong Technological Innovation Project (2013KJCX0010). Yuhui Quan would like to thank the partial support by Singapore MOE Research Grant (R-146-000-178-112).

* Corresponding author.
E-mail addresses: yxu@scut.edu.cn (Y. Xu), yuhui.quan@mail.scut.edu.cn (Y. Quan), zhangzm@cse.cuhk.edu.hk (Z. Zhang), hbling@temple.edu (H. Ling), matjh@nus.edu.sg (H. Ji).

http://dx.doi.org/10.1016/j.patcog.2015.04.015
0031-3203/

videos, which are suitable for dynamic textures with regular motions. However, the effectiveness of existing approaches is not satisfactory for dynamic textures with complex motions driven by non-linear stochastic dynamic systems with certain chaos, e.g. turbulent water and bursting fire. This inspired us to develop an effective DT descriptor for classifying DT sequences with complex dynamic behaviors.

1.1. Related work

There are many DT classification approaches, which could be roughly categorized as either generative or discriminative methods. The generative methods attempt to quantitatively model the underlying physical dynamic system that generates DT sequences, and classify DT sequences based on the system parameters of the corresponding physical model. The main difference of these methods lies in the models they build up, e.g. the spatio-temporal autoregressive model [7] and its multi-scale version [8], the linear dynamical systems [9,10], the kernel-based model [11], and the phase-based model [12]. The main drawback of the generative methods is the inflexibility to describe the DT sequences generated by nonlinear physical systems with complex motion irregularities.

In contrast to the generative methods, the discriminative methods are able to describe DT effectively without explicitly modeling the underlying dynamic system. The basic idea of the discriminative methods is to characterize the distribution of local DT patterns. To efficiently extract local DT patterns, many methods have been proposed, e.g. the spatio-temporal filtering for specific motion patterns [13,14], the spatiotemporal extensions of local binary pattern (LBP) encoding [6], the wavelet pattern extraction [5,15], the optical flow based pattern estimation [1,16,17], the space–time oriented pattern analysis [18–20], etc. In practice, the discriminative methods exhibit better performance than the generative methods in DT classification and show advantages in the robustness to environmental changes and viewpoint changes. However, the merits of existing discriminative methods are quite limited in the case of DTs with complex motions, as they are not capable of reliably capturing inherent stochastic stationary properties of such video sequences.

1.2. Motivation and contribution

Reliable characterization on DT motion behaviors is crucial to the development of an effective DT descriptor. We notice that although the motion patterns of many DT sequences could be highly irregular, they are quite consistent when viewed from different spatial and temporal scales. In other words, similar mechanisms are operating at various spatial and temporal scales in the underlying physical dynamics. Such multi-scale self-similarities are referred as to *power law* or *fractal structure* [21]. In fact, the existence of fractal structures in a large spectrum of dynamic nature images has been observed by many researchers, e.g., the amplitude of temporal frequency spectra of many video sequences, including camera movements, weather and biological movements by one or more humans, indeed fits power-law models [22,23,21,24,25].

In this paper, motivated by the existence of stochastic self-similarities in a wide range of DTs, we propose to model DTs by using non-linear stochastic dynamic systems with certain inherent multi-scale self-similarities, i.e., dynamic textures are likely to be generated by some mechanism with similar stochastic behaviors operating at various spatial and temporal scales. A novel method called dynamic fractal analysis is proposed for DT description, which measures such self-similarities of the underlying system based on fractal geometry. The proposed method can be viewed as a discriminative method with generative motivation, as we assume that DT sequences are generated by some dynamic systems with self-similarities. The resulting DFS (dynamic fractal spectrum) descriptor allows us to bypass the quantitative estimation of the underlying physical model, which is challenging in practice. Meanwhile, the DFS descriptor has the merits of both categories of methods: the discriminative power of generative methods for modeling stochastic behaviors of DT and the robustness of discriminative methods to environmental changes.

A preliminary conference version of this work appeared in [26]. The main extensions of this paper include the development of an additional spatio-temporal measure of DT pixels that brings extra discriminability, the evaluation on an additional test dataset, and more detailed analysis on the proposed method. It is noted that fractal analysis has been exploited in recent literature for DT recognition; see e.g. [14,15]. These methods mainly focus on static texture classification and are applied to DT classification by either simply averaging the original features on each DT frame (e.g. [15]), or directly extending the descriptors to 3D case (e.g. [14]). Compared with these fractal-based methods, our method captures both the global self-similar behaviors on an entire DT sequence and the statistical self-similarities of the repetitive patterns on each DT slice. Thus, our method enjoys higher discriminative power in DT classification.

2. Basics on fractal analysis

Before presenting the details of the proposed method, we first briefly introduce the theory and numerical implementation of fractal analysis. Interested readers are referred to [27–29] for more details. Fractal analysis is built on the concept of *fractal dimension* which was first proposed by Mandelbrot [28] as a description for power laws. The power laws exist in numerous natural phenomena, e.g., the amplitude of temporal frequency spectra $A(f)$ of many video sequences fits $1/f^{\beta}$ power-law models [22,23,21]

$$A(f) \propto f^{-\beta}, \tag{1}$$

where f denotes the frequency.

The fractal dimension is about self-similarity defined as the power law which the measurements of objects obey at various scales. One widely used fractal dimension in Geophysics and Physics is the so-called *box-counting* fractal dimension. Let the n-dimensional Euclidean space $\mathbb{R}^n$ be covered by a mesh of n-dim hypercubes with diameter $\frac{1}{m}$. Given a point set $E \subset \mathbb{R}^n$, the *box-counting* fractal dimension $\beta(E)$ of E is defined as the following [27]:

$$\beta(E) = \lim_{m \to \infty} \frac{\log \# \left(E, \dfrac{1}{m} \right)}{-\log \dfrac{1}{m}}, \tag{2}$$

where $\#(E, \frac{1}{m})$ is the number of mesh hypercubes that intersect E for $m = 1, 2, \ldots$. In numerical implementation, it can be done by using least squares fitting in the log–log coordinate system with a finite sequence of ordered integers.

For the physical phenomena with mixtures of multiple fractal structures, the so-called multi-fractal analysis extends the fractal dimension to describe and distinguish more complex self-similarity behavior of the physical dynamic systems. The extension is done as follows. Instead of assuming all points in the set generated by the same mechanism, a measure μ is first defined such that μ obeys the local power law in terms of scale

$$\mu(B_r(x)) \propto r^{\alpha(x)}, \tag{3}$$