

Dynamic Texture Synthesis in the YUV Color-Space

Leilei XU¹ Hanqiu SUN² Jiaya JIA³ and Chenjun TAO⁴

Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, Hong Kong
{llxu,hanqiu,leojia,cjtao}@cse.cuhk.edu.hk

Abstract. Dynamic textures are representations of such textured surfaces that exhibit certain stationarity properties in time. These include sea-waves, smoke, foliage, whirlwind, dense crowds, and traffic scenes. We address the problem of *dynamic color texture* synthesis which is a natural extending of the state-of-art *Linear Dynamic System* to the YUV color space by analyzing the intrinsic relationship between the color channels and intensity channel. Our experimental results have shown good performance for the testing examples, which remodel short dynamic texture videos into infinite time domain with similar dynamic behaviors.

1 Introduction

With the fast developing of computer hardware, especially on graphics cards, realistic rendering of dynamic textures are more commonly used in computer games, special effects in TV/film producing, and rich-media digital entertainment to create virtual environments which photo-realistically simulate the natural phenomenon.

1.1 Related work

Dynamic Textures can be defined as repetitive patterns exhibiting both spatial and temporal coherence intrinsic in the process. The methods for dynamic texture synthesis in vision can be mainly categorized into two classes: *nonparametric* and *parametric*. The nonparametric methods directly sample the original information from the input image or sequence of images. Wexler's method [1] copies the patches from the input sequence and solves the synthesis process as a global optimization process. In [2], the author also copies the appropriate 2D patches or 3D voxels to synthesize by finding the optimal seam for blending using the state-of-art Graph Cut technique. Unlike the prior two methods, Video Textures [3] chooses a whole image from the sampling sequence and copy it to the suitable location to preserve temporal continuity. These nonparametric methods usually provide high quality visual effect but yield a limited amount of information for the intrinsic property of textures.

On the other hand, parametric models provide better generalization and understanding of the perceptual process, thus these models are natural choices

for analysis and controlling of textures. In [4], images and videos are treated as a superposition of image or video basis which are called *texton* and *movton*. Szummer and Picard’s work [5] models the interaction of pixels within a local neighborhood over both space and time using the spatio-temporal auto-regressive (STAR) model. Dorrto’s model [6] provides an auto-regressive random process (specifically, a *linear dynamic system*, LDS, also known as Kalman filter model) which forms the basic model for the following papers in this area.

The most related work to ours is Filip’s [7] which addresses the problem of synthesizing dynamic color textures by analyzing the eigen-system of dynamic texture images and subsequent preprocessing and modelling of temporal interpolation eigen-coefficients using a causal auto-regressive model in RGB color space. But the results turn out to be not so satisfying for some of the dynamic textures. Blurring effects can be easily observed. Doretto [6] suggests a way to synthesize color videos by applying LDS model in the combined three unfolded RGB channels ordered one after the other. The synthesized result is satisfying, but when the size of the input image sequence is large, the method will become time consuming and even cannot solve the problem because of complex matrix operations on a very large matrix.

1.2 Overview

The main goal of this paper is to provide a novel approach for dynamic color texture synthesis in the YUV color space using the LDS model by analyzing the intrinsic connection between color channels and intensity channel. In Section 2, the LDS model is briefly outlined, and based on it our proposed approach in the YUV space for dynamic color texture synthesis (or *DCTS* for short) is described in details in Section 3. The experimented results are presented in Section 4, and conclusion goes to Section 5.

2 Dynamic Texture Model

The basic model we are working with is a second-order stationary process represented by a discrete time LDS with white, zero-mean gaussian noise [8] [10]. The observation or input of the system is a sequence of τ images represented by matrix $Y = [y(1) \dots y(\tau)] \in \mathcal{R}^{m \times \tau}$ with each image represented by column vector $y(t) \in \mathcal{R}^m$; $X = [x(1) \dots x(\tau)] \in \mathcal{R}^{n \times \tau}$ with $x(t) \in \mathcal{R}^n$ stands for the hidden state vector at time t encoding the evolution of the image sequence and the initial condition $x(0)$ known. Typically, $m \gg n$ and with values of n in the range of 10 to 50. Both the observation and state variables are corrupted by additive gaussian noise, which is also hidden. The basic generative model can be written as:

$$\begin{cases} x(t+1) = Ax(t) + v(t) & v(t) \sim \mathcal{N}(0, Q) \\ y(t) = Cx(t) + w(t) & w(t) \sim \mathcal{N}(0, R) \end{cases} \quad (1)$$

where $A \in \mathcal{R}^{n \times n}$ is the state transition matrix and $C \in \mathcal{R}^{m \times n}$ is the observation matrix which encodes the appearance information. $v(t)$ and $w(t)$ are

random variables representing the state evolution noise and observation noise, respectively, which are independent of each other and the values of x and y .

Learning of the LDS model is to learn the model parameters $\Theta = \{A, C, Q, R, x(0)\}$ given the sequence of images in matrix form Y . Assumptions are adopted to solve the degeneracy in the model: $m \gg n$; $\text{rank}(C) = n$ and choose a realization that makes C orthonormal: $C^T C = I_n$ to maximum likelihood $\arg \max_{\Theta} \log p(y(1) \cdots y(\tau))$. A suboptimal *closed-form* solution for the model parameters can be learned by using principal component analysis (PCA).

Once the model parameters Θ are learned from the original sequence of images, we can synthesize new dynamic textures with infinite length by starting from a given initial state $x(0)$ and the driven noise covariance Q to calculate $x(t)$ step by step, and finally obtain the synthesized sequence $y(t)$.

3 Dynamic Color Texture Synthesis

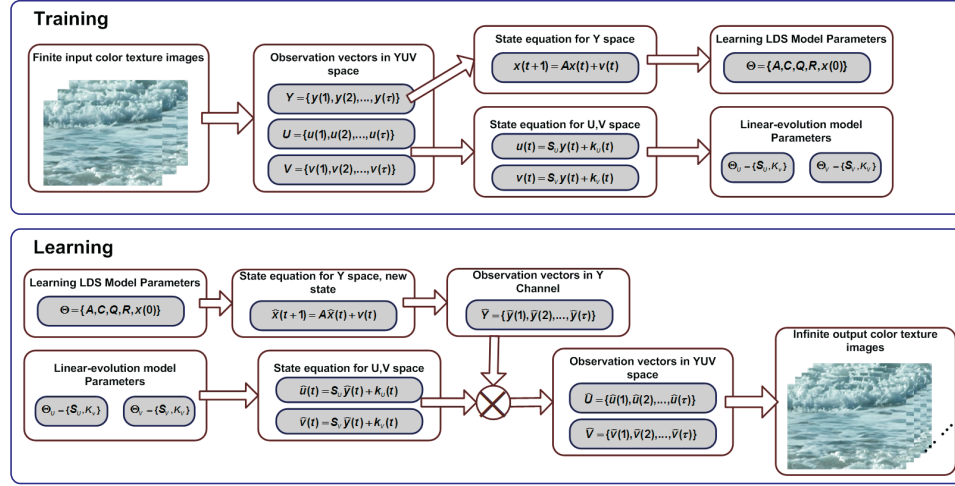


Fig. 1. The framework of our DCTS model

We work in the YUV color space, commonly used in image and video processing. $Y \in \mathcal{R}^{m \times \tau}$ (Here we use the same Y as in LDS model in most cases, it stands for intensity) is the monochromatic luminance channel, which we refer to simply as intensity, while $U \in \mathcal{R}^{m \times \tau}$ and $V \in \mathcal{R}^{m \times \tau}$ are the chrominance channels, encoding the color [11]. The framework of our DCTS model in Figure 1 outlines the processes of LDS model and the linear-evolution model between intensity channel (the intrinsic factor) and color channels (the extrinsic appearance).

In human perception, the most important clue is the intensity which gives most informative signals, and color signals can be regarded as evolving with intensity value. Because Y, U, V values are all gaussian distributed [6], we can

assume that there is an *linear-evolution* process with gaussian noise driven between intensity and color: color changes according to the underlying change of intensity. (We only concern U channel from now on, as V channel behaves the same as U).

$$u(t) = Sy(t) + k(t) \quad (2)$$

where $u(t) \in \mathcal{R}^m$ is the column vector of U channel, $y(t) \in \mathcal{R}^m$ the column vector of Y channel. $S \in \mathcal{R}^{m \times m}$. The m-vector $k(t) \sim \mathcal{N}(\mu, \Sigma)$ is random variables representing the driven noise. We call this model the *linear-evolution* model between intensity and color channels. The parameters of the model are $\Theta_u = \{S, k(t)\}$ for U channel.

There are two problems in the solution of equation 2: First, the size of matrix S is too large to be computed and stored even for a small sized video; Second, the equation is under-constrained, so there are many solutions which satisfy the requirements. Thus, we need to make reasonable assumption to uniquely define the solution which has S small in size.

3.1 Learning of Dynamic Color Texture

In order to solve the problems shown in above, we employ the theory in [13] which justifies that there is a linear relation between color and intensity. Formally, it assumes that color u_r and v_r at a pixel r is a linear function of intensity y_r : $u_r = s_r y_r + k_r$. The assumption intuitively tells that when the intensity is constant the color should be constant, and when the intensity is an edge the color should also be an edge (although the values on the two sides of the edge can be any two numbers). Rewritten the linear relation in matrix form we get:

$$u(t) = S(t)y(t) + k(t) \quad (3)$$

where $S(t) \in \mathcal{R}^{m \times m}$ is the diagonal matrix of corresponding s_r for each pixel, and $k(t) \in \mathcal{R}^m$ is the column vector for parameter k_r . We observe that if we suppose for all t , the matrix S is the same, $S(1) = \dots = S(\tau)$. Then Equation 3 will become a special case of Equation 2 where S is diagonal matrix which is small in storage and easy for matrix computation. Then model 2 becomes:

$$u(t) = Sy(t) + k(t) \quad k(t) \sim \mathcal{N}(\mu, \Sigma) \quad (4)$$

with S the diagonal matrix, and others the same as Equation 2.

Closed-form solution for Linear-evolution model The first observation in Equation 4 is that the choice of S, Σ . We want to find a unique closed-form solution which is easy for computation.

In order to make sure that the resulted $k(t)$ is gaussian distributed, in our tests, we generate a random m-vector distributed as $\mathcal{N}(0, 1)$ to be the diagonal elements in S which is distributed independently with U and Y . This ensures that $k(t)$ is also gaussian distributed. Then μ, Σ can be easily calculated given U, Y , and S .

3.2 Synthesizing of Dynamic Color Texture

Given the LDS model parameters Θ and linear-evolution model parameter Θ_u and Θ_v , we can easily generate new image sequences in the infinite time domain, by first applying LDS model to generate the new intensity images, and then given the new intensity value and Θ_u and Θ_v , linear-evolution model can calculate the new corresponding U, V channels.

4 Experimental results

Using the model we developed, we have tested a set of dynamic color textures from [9] which provides sample dynamic textures, for learning and synthesizing DC textures in the infinite time domain.

In Figure 3, some blurring effects can be seen in the generated video frames from Filip’s method, while ours has shown the better visual quality in comparison with Filip’s results and similar dynamic behavior as the original input video. Our synthesized result is comparatively the same with the LDS model in [6] applied to color video sequence. But the size of the processed matrix size is much smaller than that: for a commonly $200 \times 150 \times 100$ input color video, the largest size of matrix to be dealt with is 3 times larger than the video size. If video size grows bigger, there will be not enough memory to store all the data which will make the task impossible or really time consuming to be finished. For our DCTS model, the largest matrix size to be dealt with is only the size of the video which takes a much smaller memory. Figure 2 compares the synthesized result of our DCTS model with the ones by Filip’s method ([7]) and the input short video. It is clear that there is blurring effects in the image sequence generated by Filip’s method, while ours is fine with better visual quality also comparable dynamic behavior as to the input video.



Fig. 2. The grass video sequence. The original video size is $m = 200 \times 150$, $\tau = 100$. From top down, row 1 shows the selected frames of input video; row 2 shows the synthesized result of Filip’s method; row 3 shows the synthesized result of our DCTS model in corresponding frames as row 2.

5 Summary

In this paper, we have proposed and developed the novel approach to synthesis dynamic color textures by taking the advantage of Linear Dynamic Systems and its sub-optimal solution. We analyzed the intrinsic connection between color channels and intensity channel, and establish the relation between color information and state transition vector for the intensity channel. Our proposed DCTS approach is fast and can compress the data significantly.

Further efforts can be directed in the shifting of the control probabilities [14] of gray-level LDS model into our DCTS model to manipulate the speed, size and other characters of the dynamics. Also, feedback controls on pole distributions, presented in [15] can be employed into our model to further optimize the synthesized DCTS video results.

References

- [1] Wexler, Y., Shechtman, E., and Irani, M.: Space-Time Video Completion. CVPR (2004) 120–127
- [2] Kwatra, V., Schödl, A., Essa, I., and Bobick, A.: Graphcut textures: image and video synthesis using graph cuts. ACM SIGGRAPH 2003, ACM Press **22**(2003) 277–286
- [3] Schödl, A., Szeliski, R., Salesin, D.H., and Essa, I.: Video textures. in In Proceedings of Siggraph00 (2000)
- [4] Wang, Y. and Zhu, S.: Analysis and synthesis of textured motion: Particle, wave and cartoon sketch. IEEE Trans. on Pattern Analysis and Machine Intelligence **26**(2004) 1348–1363
- [5] Szummer, M., and Picard, R.: Temporal texture modeling. In Proceedings of IEEE International Conference on Image Processing (ICIP 96) (1996)
- [6] Doretto, G., Chiuso, A., Wu, Y.N. and Soatto, S.: Dynamic textures. International Journal of Computer Vision **2**(2003) 91–109
- [7] Filip, J., Haindl, M., and Chetverikov, D: Fast Synthesis of Dynamic Colour Textures. Proceedings of the 18th IAPR Int. Conf. on Pattern Recognition (2006) 25–28
- [8] Ljung, L.: System Identification -Theory for the User. Prentice Hall, Englewood Cliffs, NJ (1987)
- [9] Péteri, R. and Huiskes, M.: A comprehensive database of dynamic textures. URL:<http://www.cwi.nl/projects/dyntex/>
- [10] Shumway, R. and Stoffer, D.: Time Series Analysis and Its Applications. Springer (2000)
- [11] Jack, K.: Video Demystified, 3rd edition. Elsevier Science and Technology (2001)
- [12] Roweis, S., and Ghahramani, Z.: A Unifying Review of Linear Gaussian Models Neural Computation **11**(1999) 305–345
- [13] Zomet, A., and Peleg, S.: Multi-sensor super resolution. In Proceedings of the IEEE Workshop on Applications of Computer Vision (2002)
- [14] Doretto, G., and Soatto, S.: Editable dynamic textures. In ACM SIGGRAPH Sketches and Applications (2002)
- [15] Lu Yuan, Fang Wen, Ce Liu, and Heung-Yeung Shum: Synthesizing Dynamic Texture with Closed-Loop Linear Dynamic System. ECCV (2004)