Locally Linear Embedding Based Dynamic Texture Synthesis

Weigang Guo, Xinge You(✉), Ziqi Zhu, Yi Mou, and Dachuan Zheng

School of Electronics and Information Engineering,
Huazhong University of Science and Technology, Wuhan, China
youxg@mail.hust.edu.cn

Abstract. Dynamic textures are often modeled as a low-dimensional dynamic process. The process usually comprises an appearance model of dimension reduction, a Markovian dynamic model in latent space to synthesize consecutive new latent variables and a observation model to map new latent variables onto the observation space. Linear dynamic system(LDS) is effective in modeling simple dynamic scenes while is hard to capture the nonlinearities of video sequences, which often results in poor visual quality of the synthesized videos. In this paper,we propose a new framework for generating dynamic textures by using a new appearance model and a new observation model to preserves the non-linear correlation of video sequences. We use locally linear embedding(LLE) to create an manifold embedding of the input sequence, apply a Markovian dynamics to maintain the temporal coherence in the latent space and synthesize new manifold, and develop a novel neighbor embedding based method to reconstruct the new manifold into the image space to constitute new texture videos. Experiments show that our method is efficient in capturing complex appearance variation while maintaining the temporal coherence of the new synthesized texture videos.

Keywords: Dynamic texture synthesis · Dynamic system · Locally linear embedding

1 Introduction

Dimensionality reduction algorithms has been successfully applied to video analysis [1–4] for decades. A central difficulty in modeling time series data is in determining whether the model can capture the nonlinearities of the data without overfitting.

Sotto et al [5] models dynamic textures using a linear dynamic system(LDS), which was shown to be a promising technique for synthesis and analysis of dynamic textures. Texture video frames are unfolded into column vectors and constitute points set in the image space. The analysis consists in finding an appropriate space to describe the trajectory of the video frames and in modeling the trajectory basing on the dynamical system theory. The model allows for great power of editing and could create new images that were never a part of

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 287–295, 2015.
DOI: 10.1007/978-3-662-48558-3_29

the original sequence. However the visual quality of the output is usually not satisfactory because of the over-simplified linear model.

Yuan et al [6] extends this work by bringing in a feedback control to turn the system into a closed loop LDS. The feedback loop corrected the problem of signal decay, but the problem of blurry synthesized frames is still yet to be solved.

Che-Bin Liu [7] models dynamic textures using subspace mixtures, which propose to use Locally Linear coordination(LLC) [8] characterizing image manifolds and generates a much improved result.

Roberto [9] extends the SVD in basic LDS to a tensor decomposition technique(HOSVD) without unfolding the video frames into column vectors, which results in models requiring on average five times less coefficients, while still ensuring the same visual quality.

However, by using either [7] or [9], the visually quality of synthesized video sequences are still not satisfactory because of the over-simplified appearance models and observation models.

Ishan Awasthi [10] use nonlinear dimensionality reduction as the appearance model of the input texture video sequence, use a spline to move along the input manifold and learn the nonlinear mapping from the input to map the new manifold into the image space. This technique can create realistic new images, however they are often not visually consecutive when the input sequence comprises of relatively random and fast motion.

Here a similar but more formal alternative is suggested. Though we use the same appearance model, our dynamic model and observation model are totally different and more natural. Our work also chooses to use manifold based dimension reduction technique locally linear embedding (LLE) [11] to find the low-dimensional space. while instead of learning the nonlinear mapping from the input, we reconstruct the points from the low-dimensional manifold space to image space through neighbor embedding which was inspired by LLE itself.

LLE is a promising manifold learning method that has aroused a great deal of extension in machine learning and image processing. The algorithm is based on a geometric intuitions that points should preserve similar local geometry structure in both high- and low-dimensional space.

In LLE, the local properties of the data properties are constructed by writing the high-dimensional data points as a linear combination of their nearest neighbors. In the low-dimensional representation of the data, LLE attempts to retain the reconstruction weights in the linear combinations as good as possible. After the nonlinear embedding offers low-dimensional manifold representation of the texture video sequences, a Markovian dynamic model is used to maintain the temporal coherence of the low-dimensional series and create a similar manifold by driving the system with random noise. Then our approach is driven by a key question: how to map the new manifold back into the image space and constitute a new texture video. Our technique is inspired by LLE itself and still based on the assumption that data-points in the low- and high-dimensional space form manifolds with similar local geometry. Given a low-dimensional point of the new

manifold, it can be reconstructed basing on a linear mapping that weights its neighbors from the original manifold. This local geometry is preserved in the high dimensional space, and since the corresponding high-dimensional representations of the neighbors are given from the original high-dimensional data points, the high-dimensional embedding of the point can be estimated.

Our main contribution is to propose a new framework for dynamic texture synthesis. And what's more, we develop a novel technique to estimate the high-dimensional representation of variables in the low-dimensional locally linear embedding space. Our framework is not available for other existing methods and is a new strategy for dynamic texture synthesis. The rest of this paper is organized as follows: section 2, section 3 and section 4 describe the appearance model, dynamic model and observation model of our framework for dynamic texture synthesis. The experiment results are presented in section 5. And finally we conclude the paper and present some future works in section 6.

2 Appearance Model

Suppose there are N points $Y = \{y_t\}_{t=1,\ldots,N} \in R^D$ representing the input video frames which are unfolded into column vectors in a high-dimensional data. The N points are assumed to lie on a nonlinear manifold of intrinsic dimension d (typically $d \ll D$). Provided that sufficient data points are sampled from the manifold, each data point y_i and its k neighbors are expected to lie a locally linear patch of the manifold. The local geometry of each patch can be characterized by the reconstruction weights $W = \{w_{tj}\}_{t,j=1,\ldots,N}$, with which a data point is reconstructed from its neighbors. The weights can be obtained by minimizing the following cost function.

$$\varepsilon(W) = \sum_{t=1}^{N} \|y_t - \sum_{j=1}^{N} w_{tj} y_{tj}\|^2 \tag{1}$$

W is a sparse $N \times N$ matrix whose entries are set to 0, if t and j are not connected in the neighborhood graph, and equal to the corresponding reconstruction weight otherwise. The minimization is subjected to $\sum_{j=1}^{N} w_{tj} = 1$. The weights can be solved through constrained least squares, and they reflect the local geometries relating the data points to their neighbors. Then we need to compute the low-dimensional embedding representations of the data points so that they can best preserve the local geometry. Let the low-dimensional embedding coordinates be $X = \{x_t\}_{t=1,\ldots,N} \subseteq R^d$, $d \ll D$. The objective can be obtained by minimizing the following cost function.

$$\varepsilon(X) = \sum_{i=1}^{N} \|x_t - \sum_{j=1}^{N} w_{tj} x_j\|^2 \tag{2}$$

The minimization is subject to $\frac{1}{N}\sum_{t=1}^{N} x_t = 0$ and $\frac{1}{N}\sum_{i=1}^{N} x_t x_t^T = \frac{1}{N} X^T X = I_d$, I is the identity matrix. The low-dimensional embedding can be solved by computing

the eigenvectors corresponding to the smallest d nonzero eigenvalues of the inner product $(I-W)^T(I-W)$.

3 Dynamic Model

Similar to the Linear Dynamic System [5], automatic regression(AR) model is used as the dynamic model in our framework to generate the new state variables while maintaining the temporal coherence. The AR model is based on the assumption that each state variable in the time series depends on m previous state variables, here, $m = 1$.

Given the latent variables $X = \{x_t\}_{t=1,\ldots,N} \subseteq R^d$, $d \ll D$ obtained from the locally linear embedding(LLE), the sampling noise $v_t \sim N(0, Q)$ can drive the system matrix A to generate L new state variables $X^* = \{x_t^*\}_{t=1,\ldots,L} \subseteq R^d$ according to (3). L is usually two times larger than N.

$$x_t = Ax_{t-1} + v_t, v_t \sim N(0, Q) \tag{3}$$

System matrix A can be estimated based on the least squares approximation in (4).

$$A = X_{2:N} Pinv(X_{1:N-1}) \tag{4}$$

Also the noise variance Q can be obtained by (5).

$$\begin{gathered} Q = \frac{1}{T-1} \sum_{t=1}^{N-1} x_t^{'} (x_t^{'})^T \\ x_t^{'} = x_{t+1} - Ax_t \end{gathered} \tag{5}$$

4 Observation Model

Once the new state vectors X^* are generated, a novel technique is proposed to reconstruct the new state variables from the low dimensional embedding space into the high dimensional image output space. As in LLE, local geometry is characterized by how a variable corresponding a texture frame can be reconstructed by its neighbors and this property should be preserved in the low-dimensional space. The property is symmetric, and the reconstruction weights between a variable and its neighbor in the low-dimensional space should also be preserved in the high-dimensional space.

For each newly synthesized variable x_t^* in the low-dimension space, like in LLE, we first need to find its K_O nearest neighbors $X_t = \{x_{tj}\}_{j=1,\ldots,K_O}$ in X and compute the reconstruction weights of neighbors by minimizing the reconstruction error.

$$\varepsilon(W) = \|x_t^* - \sum_{j=1}^{K_O} w_{tj} x_j\|^2 \tag{6}$$

The solution to the constrained least squares problem can be find in [12].

$$\begin{aligned} G_t &= (x_t^* \Gamma^T - X_n)^T (x_t^* \Gamma^T - X_n) \\ W_t &= \frac{G_t^{-1}\Gamma}{\Gamma^T G_t^{-1}\Gamma} \end{aligned} \tag{7}$$

G_t is local gram matrix regarding x_t^*, $\Gamma = [1, \ldots, 1]^T$ is the vector of ones with K_O entries, X_n is a $d \times K$ matrix whose columns are the K_O nearest neighbors of x_t^*.

Then the optimum high-dimensional representation of x_t^* can be achieved by the linear combination of the corresponding high-dimensional neighbor $Y_t = \{y_{tj}\}_{j=1,\ldots,K_O}$ in Y basing on the reconstruction weights W_t.

$$y_t^* = \sum_{j=1}^{K_O} w_{tj} y_{tj} \tag{8}$$

The complete framework of our algorithm is summarized as Algorithm 1.

Algorithm 1. Locally Linear Embedding based Dynamic Texture Synthesis

Input: texture video sequence Y;

1: FOR each frame $\{y_t\}_{i=1,\ldots,N}$ in the input texture video sequence Y;
2: Unfold it into column vector and find the set of K_A nearest neighbors y_{tj} through KNN in Y;
3: Compute the reconstruction weights of the neighbors that minimize the reconstruction error: $\varepsilon(W) = \|y_t - \sum_{j=1}^{K_A} w_{tj} y_{tj}\|^2$;
4: END;
5: Compute the low-dimensional embedding state variables $X = \{x_t\}_{t=1,\ldots,N}$ in the $d-$dimensional space so that they can best preserve the local geometry by minimizing the cost function defined in (2): $\varepsilon(X) = \sum_{t=1}^{N} \|x_t - \sum_{t=1}^{N} w_{tj} x_{tj}\|^2$;
6: Generate L new state variables $X^* = \{x_t\}_{t=1,\ldots,L}$ according to (3);
7: FOR each newly synthesized variable x_t^*;
8: Find the set of K_O nearest neighbors $\{x_{tj}\}_{j=1,\ldots,K_O}$ through KNN in X and their corresponding high-dimensional representation $Y_t = \{y_{tj}\}_{j=1,\ldots,K_o}$;
9: Compute the reconstruction weights W of neighbors by minimizing the reconstruction error in 6: $\varepsilon(W) = \|x_t^* - \sum_{j=1}^{K_O} w_{tj} x_j\|^2$;
10: Construct the high-dimensional representation of x_t^* basing on the reconstruction weights W_t: $y_t^* = \sum_{j=1}^{K_O} w_{tj} y_{tj}$;
11: END;

Output: longer texture video $Y^* = \{y_t^*\}_{t=1,\ldots,L}$;

5 Experiments

We have synthesized many dynamic textures using out proposed method. The inputs texture sequences used in our experiments are from[1] and DynTex

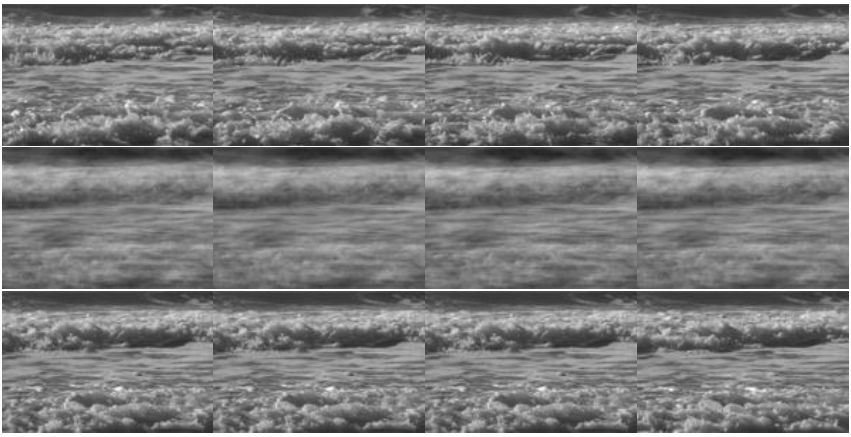

Fig. 1. The images on the top row are from the original wave sequence. The second row is synthesized by the basic LDS. The bottom row is synthesized by our method.

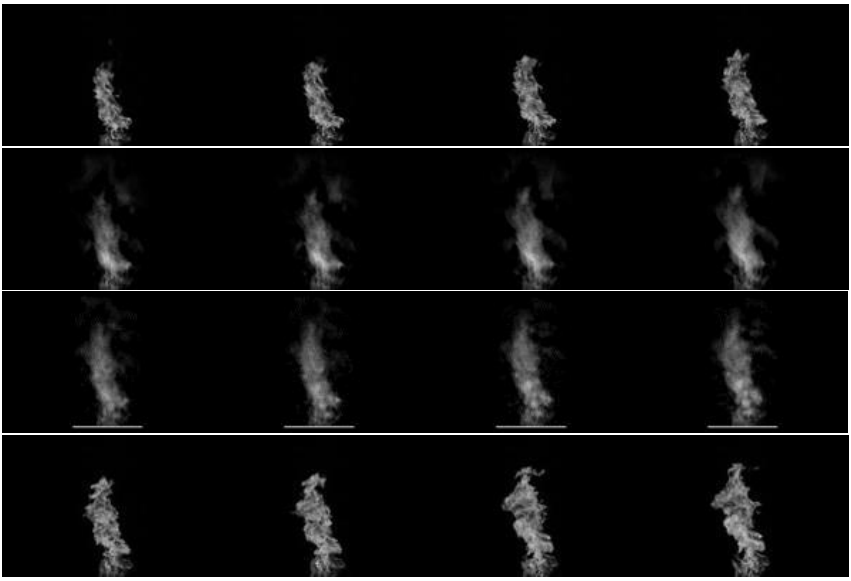

Fig. 2. The images on the top row are from the original flame sequence. The second row is synthesized by the basic LDS. The third row is synthesized by HOSVD. The bottom row is synthesized by our method.

[1] http://www.cc.gatech.edu/cpl/projects/graphcuttextures/

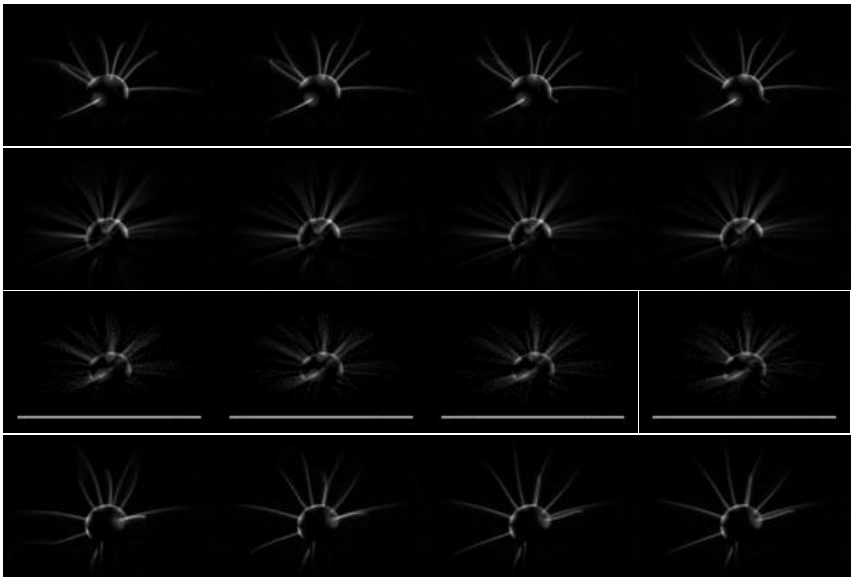

Fig. 3. The images on the top row are from the original sparkle sequence. The second row is synthesized by the basic LDS. The third row is synthesized by HOSVD. The bottom row is synthesized by our method.

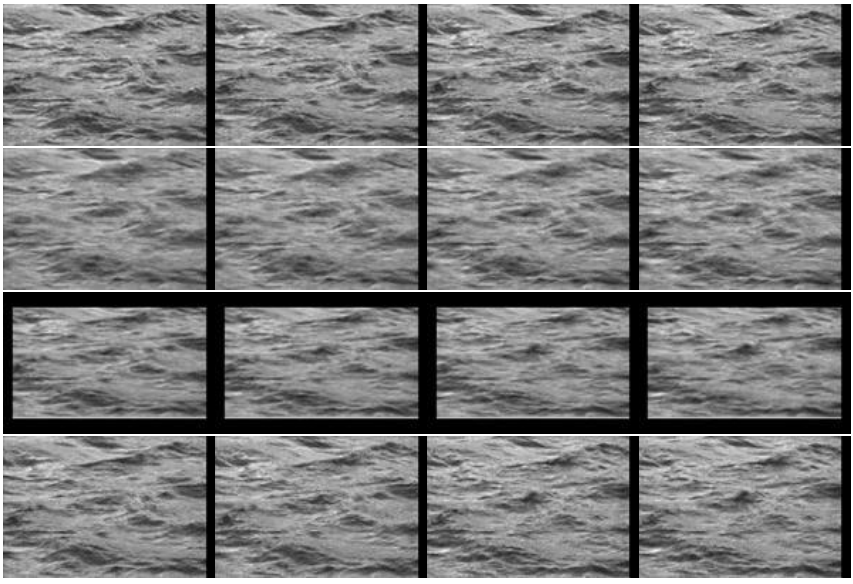

Fig. 4. The images on the top row are from the original river sequence. The second row is synthesized by the basic LDS. The third row is synthesized by HOSVD. The bottom row is synthesized by our method.

Database [13]. Most inputs videos have resolutions of 150 by 100 and lengths of 60 to 150 frames while the lengths of the outputs sequences were two times or greater than the original sequences. In our experiments, we set the low dimension d varying from 10 to 20. The neighborhood size K is between 10 and 20 in the appearance model and is between 2 and 10 in the observation model.

Fig.1, Fig.2, Fig.3 and Fig.4 shows the synthesis results for four sequences, each depicting different dynamics, by basic LDS, HOSVD [9], and our method. The HOSVD based synthesis results are downloaded from[2]. As can been seen, the basic LDS method yields blurred images and a decreasing quality as the synthesized sequence becomes longer. The HOSVD generates improved results while the visual quality is still not satisfactory. And our method can synthesize images most similar to the original frames, and like the basic LDS, the temporal coherence of the synthesized videos is also well guaranteed. What's more, the whole synthesis process is simple and in real-time.

6 Conclusions

Inspired by the promising manifold learning method LLE [11], we propose a new framework for dynamic texture synthesis. We use the LLE to capture the low-dimensional embedding of the input texture sequences. Automatic regression(AR) model is then adopted to model the texture dynamics in the low-dimensional space. And lastly we are inspired by the LLE again and develop a new technique to reconstruct the high-dimensional embedding of the newly synthesized latent variables. Both the appearance model and the observation model in our framework are nonlinear and the experiment results demonstrate that our framework can generate longer and higher-quality dynamic textures than relevant works.

Some future work will be pursued to make the proposed method more applicable. Though our proposed observation model can ensure higher visual quality of the synthesized images, the temporal coherence can be hardly maintained as the synthesized sequence becomes longer. That is because the dynamic model in our framework is too simple to capture complex temporal variation in the low-dimensional space. In the future, some other dynamic models may be created or applied in this framework to achieve more stable synthesis process.

Acknowledgement. This work is supported partially by the National Natural Science Foundation (NSFC) of China (Grant no. 61272203), the Ph.D. Programs Foundation of Ministry of Education of China (Grant no. 20110142110060).

References

1. Soatto, S., Doretto, G., Wu, Y.N.: Dynamic textures. In: 2001 Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001, vol. 2, pp. 439–446. IEEE (2001)

[2] http://lcav.epfl.ch/reproducible_research/CostantiniIP07_1

2. Wang, J., Hertzmann, A., Blei, D.M.: Gaussian process dynamical models. In: Advances in Neural Information Processing Systems, pp. 1441–1448 (2005)
3. Rahimi, A., Darrell, T., Recht, B.: Learning appearance manifolds from video. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 868–875. IEEE (2005)
4. Lin, R.-S., Liu, C.-B., Yang, M.-H., Ahuja, N., Levinson, S.E.: Learning nonlinear manifolds from time series. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 245–256. Springer, Heidelberg (2006)
5. Doretto, G., Chiuso, A., Wu, Y.N., Soatto, S.: Dynamic textures. International Journal of Computer Vision **51**(2), 91–109 (2003)
6. Yuan, L., Wen, F., Liu, C., Shum, H.-Y.: Synthesizing dynamic texture with closed-loop linear dynamic system. In: Pajdla, T., Matas, J. (George) (eds.) ECCV 2004. LNCS, vol. 3022, pp. 603–616. Springer, Heidelberg (2004)
7. Liu, C.B., Lin, R., Ahuja, N.: Modeling dynamic textures using subspace mixtures. In: 2005 IEEE International Conference on Multimedia and Expo, ICME 2005, pp. 1378–1381. IEEE (2005)
8. Teh, Y.W., Roweis, S.T.: Automatic alignment of local representations. In: Advances in Neural Information Processing Systems, pp. 841–848 (2002)
9. Costantini, R., Sbaiz, L., Susstrunk, S.: Higher order svd analysis for dynamic texture synthesis. IEEE Transactions on Image Processing **17**(1), 42–52 (2008)
10. Awasthi, I., Elgammal, A.: Learning nonlinear manifolds of dynamic textures. In: Advances in Computer Graphics and Computer Vision, pp. 395–405. Springer (2007)
11. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
12. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: 2004 Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 1, pp. I-I. IEEE (2004)
13. Péteri, R., Fazekas, S., Huiskes, M.J.: Dyntex: A comprehensive database of dynamic textures. Pattern Recognition Letters **31**(12), 1627–1632 (2010)