ELectronics EXpress

IEICE Electronics Express, Vol.7, No.18, 1329–1334

Improving dynamic texture recognition with constraint subspace learning

Tiesheng Wang[a)] **and Pengfei Shi**[b)]

Institute of Image Processing and Pattern Recognition,
Shanghai Jiao Tong University,
800 Dongchuan Road, Shanghai, 200240, China
a) *tieshengw@sjtu.edu.cn*
b) *pfshi@sjtu.edu.cn*

Abstract: We present a new framework for recognizing dynamic textures modeled by linear dynamic systems (LDS). The new framework improves previous methods by considering between-class subspace differences. Specifically, common part among different class subspaces is removed by projecting model subspaces onto a constraint subspace before recognition. This operation increases the separability among different classes of dynamic textures. The effectiveness of the new framework is shown by experimental results.

Keywords: dynamic texture recognition, constraint subspace, subspace distance, linear dynamic systems

Classification: Science and engineering for electronics

References

[1] A. B. Chan and N. Vasconceloso, "Classification and retrieval of traffic video using auto-regressive stochastic processes," *Proc. IEEE Intelligent Vehicles Symposium*, pp. 771–776, 2005.

[2] S. Fazekas and D. Chetverikov, "Dynamic texture recognition using optical flow features and temporal periodicity," *Proc. International Workshop on Content-Based Multimedia Indexing 2007*, pp. 25–32, 2007.

[3] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," *Proc. IEEE CVPR'2003*, pp. 313–320, 2003.

[4] P. K. Turaga, A. Veeraraghavan, and R. Chellappa, "Unsupervised view and rate invariant clustering of video sequences," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 353–371, 2009.

[5] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, pp. 915–928, 2007.

[6] Z. Lu, W. Xie, J. Pei, and J. Huang, "Dynamic texture recognition by spatio-temporal multiresolution histograms," *Proc. IEEE Workshop on Motion and Video Computing 2005 (WACV/MOTIONS'05)*, vol. 2, pp. 241–246, 2005.

[7] P. Saisan, G. Doretto, Y. Wu, and S. Soatto, "Dynamic texture recognition," *Proc. IEEE CVPR'2001*, vol. 2, pp. 58–63, 2001.

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010

[8] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto, "Dynamic textures," *Int. J. Comput. Vision*, vol. 51, no. 2, pp. 91–109, 2003.
[9] K. Fukui and O. Yamaguchi, "Face recognition using multi-viewpoint patterns for robot vision," *Robtics Research*, eds., pp. 192–201, 2005.
[10] R. J. Martin, "A metric for ARMA processes," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1164–1170, 2000.
[11] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
[12] A. Bjorck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 27, pp. 579–594, 1973.
[13] R. P. M. Huskies and S. Fazekas, "DynTex: a comprehensive database of dynamic textures," 2005. [Online] www.cwi.nl/projects/dyntex/

1 Introduction

Object recognition is a fundamental task in computer vision. In recent years, recognizing dynamic textures has drawn increasing interests [1, 2, 3, 4, 5, 6]. Dynamic texture recognition aims at finding dynamic visual patterns with repetitive motion, such as waving trees, flame, water waves, and traffic flow, in video. There are many practical applications of recognizing dynamic scenes. For example, in video surveillance of traffic scene, the traffic status, such as light, medium, or heavy flow, can be automatically classified based on the interpretation of dynamic textures [1]. Other examples include recognizing action [4] and facial expression [3].

A framework for dynamic texture recognition usually contains two modules. The first one is the effective representation of dynamic textures with certain features or models. Based on such representation, the second module classifies dynamic textures based on certain distance or similarity measures.

Previous methods on dynamic texture recognition roughly fall into two categories. The first category focuses on extracting local features, such as optical flow [2], local binary pattern (LBP) [5], and histogram [6], for recognition. Other information, such as temporal periodicity, can be jointly used for recognition. The second category uses linear dynamic system (LDS) models to capture the underlying dynamic processes of the appearance of dynamic textures [7, 8, 1]. In these methods, LDS model parameters, which span a subspace, are used for recognition based on some distance or similarity measures. Then, the distances or similarities among model subspaces were directly compared.

The framework proposed in this paper belongs to the second category. Though previous methods based on LDS representation may capture the underlying dynamic processes [7, 8], considering between-class differences may provide extra benefits. If the overlap or intersection among different class subspaces can be removed, the separability between different subspaces will be improved and the recognition performance will be increased accordingly. We

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010

IEICE Electronics Express, Vol.7, No.18, 1329–1334

use a transformation similar to constraint mutual subspace methods (CMSM) in [9] to improve previous methods for dynamic texture recognition. The effectiveness of using constraint subspace learning is show in our experiments.

The rest of this paper is arranged as follows. In Section 2, dynamic texture recognition using LDS representation is introduced. Constraint subspace learning is described in Section 3. Section 4 presents the proposed framework for dynamic texture recognition. The performance of the new framework is evaluated in Section 5.

2 Dynamic texture recognition based on LDS representation

The basic framework for dynamic texture recognition based on LDS representation is briefly described in this section.

The underlying dynamic processes of dynamic textures can be captured by LDS [7, 8]. Let the state at time t be a vector $x_t \in \mathbb{R}^d$ and the observed texture appearance at t be a vector $y_t \in \mathbb{R}^n$. Dynamic textures can be described by the following state space models:

$$\begin{cases} x_t = Ax_{t-1} + v_t \\ y_t = Cx_t + w_t \end{cases} \tag{1}$$

where $A \in \mathbb{R}^{d\times d}$ is the state transition matrix and $C \in \mathbb{R}^{n\times d}$ is the observation matrix. v_t and w_t are the zero-mean Gaussian noises, i.e., $v_t \sim \mathcal{N}(0, Q)$ and $w_t \sim \mathcal{N}(0, R)$, where Q and R are the corresponding covariance matrices.

When the model parameters in (1) is learned as in [8], the observation matrix C is formed by the principal eigenvectors through principal component analysis (PCA) of the observed dynamic textures. And the estimated state vector, denoted as $\hat{x}_t$, is obtained from projecting the observed vector y_t onto C, i.e., $\hat{x}_t = C^T y_t$. Accordingly, A is estimated as $A = [\hat{x}_2 \ \cdots \ \hat{x}_n][\hat{x}_1 \ \cdots \ \hat{x}_{n-1}]^\dagger$, where $\dagger$ denotes generalized inverse.

Columns of the observability matrix $\mathcal{O}_n = [C^T \ A^TC^T \ \cdots \ (A^T)^kC^T]^T$ of the LDS model span a subspace, which is used for recognition. Therefore, the closeness between dynamic textures can be measured by the distance or similarity between subspaces.

Subspaces are known as points on Grassmann manifold [11]. Principal angles between two subspaces are closely related to subspace distances. Let $U_1 \in \mathbb{R}^{n\times d}$ and $U_2 \in \mathbb{R}^{n\times d}$ be basis matrices of two subspaces. The numerically stable way to compute principal angles is through the SVD of $U_1^TU_2$ [12]:

$$U_1^TU_2 = V\Lambda W^T, \qquad \Lambda = diag(\lambda_1, \cdots, \lambda_d) \tag{2}$$

where $\lambda_i = \cos\theta_i$ is the cosine value of the ith principal angle.

Based on principal angles, Martin distance between two subspaces is defined as [10]:

$$d_M^2(U_1, U_2) = \ln \prod_{i=1}^{d} \frac{1}{\cos^2\theta_i} = -2\sum_{i=1}^{d} \ln\lambda_i, \tag{3}$$

Although there are other subspace distances [11], methods based on Martin distance have shown promising results for dynamic texture recognition [7].

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010

Therefore, we also use Martin distance in our framework. Note that we use (2) to compute principal angles while generalized SVD (GSVD) was used in [7].

Based on LDS representation and subspace distance discussed above, dynamic texture recognition can be achieved by comparing subspace distances. The simplest classifier is nearest neighbor (NN), which chooses the dynamic texture with the smallest distance as the recognized one.

3 Learning constraint subspace

Different class subspaces of dynamic textures can be first projected onto a constraint subspace to improve their separability. The aim of this operation is to remove the common part among different subspaces. Then, the projected subspaces are used for recognition. Although there are several ways to construct constraint subspace, we use generalized difference subspace suggested in [9].

Denote basis matrices for N class subspaces by $U_i \in \mathbb{R}^{n\times d}, i = 1, \ldots, N$. The sum of the N-class projection matrices is $G = \sum_{i=1}^{N} U_i U_i^T$. Since G is positive definite, based on the SVD of G, i.e.

$$G = D\Sigma D^T, \tag{4}$$

the generalized difference subspace $H \in \mathbb{R}^{n\times N_d}$ is formed by the N_d eigenvectors corresponding to the N_d smallest eigenvalues of G. Or equivalently, H is formed by discarding the first N_p eigenvectors of G corresponding to the N_p largest eigenvalues. The optimal dimension N_d of the constraint subspace is chosen experimentally.

Projecting the subspace matrix U_i onto H is $\tilde{U}_i = H^T U_i$, $\tilde{U}_i \in \mathbb{R}^{N_d\times d}$. Note that $\tilde{U}_i$ is not necessarily an orthonormal matrix as $\tilde{U}_i^T \tilde{U}_i = U_i^T H H^T U_i$. Therefore, Gram-Schimit orgthogonalization and normalization should be applied to $\tilde{U}_i$ to obtain the orthonormal basis matrix $\hat{U}_i$ of the projected subspace corresponding to the ith class.

The procedure of learning a constraint subspace is as follows. First, obtain subspace basis matrices U_i and the constraint subspace H. Second, project U_i onto H and perform Gram-Schmidt orthonormalization to obtain the new subspace bases $\hat{U}_i$.

4 Proposed framework for dynamic texture recognition

While LDS representation is useful for dynamic textures as shown in [7, 8], it may not be optimal for recognition task. As described in Section 3, the transformation with constraint subspace considers between-class subspace difference to remove the intersection among different class subspaces. Therefore, LDS based dynamic texture recognition can be augmented with constraint subspace learning to improve the separability among different models.

First, model parameters $\{C_i, A_i\}$, $i = 1, \cdots, N$ can be estimated for N classes of dynamic textures, and the observability matrices $\mathcal{O}_{n,i}$ are formed according to Section 2.

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010

IEICE Electronics Express, Vol.7, No.18, 1329–1334

Then, the basis matrix U_i can be computed from the observability matrix $\mathcal{O}_{n,i}$ via Gram-Schmidt orthogonalization. The generalized difference matrix H can be computed based on the SVD of G, the sum of projection matrices, according to the description in Section 3.

Based on the idea of constraint subspace learning, we can project the subspace basis matrix U_i onto the constraint subspace H. In this way, we obtain the projected subspaces $\tilde{U}_i = H^T U_i$. Since the projected subspace $\tilde{U}_i$ is not necessarily orthogonal, it is orthogonalized with Gram-Schmidt method and normalized to obtain the new subspace basis matrix $\hat{U}_i$.

Finally, as the standard framework in Section 2, Martin distance and NN classifier are adopted in our framework for dynamic texture recognition. The subspace basis matrices $\hat{U}_i$ are input into the standard recognition framework. The probe dynamic texture with $\hat{U}_p$ as the transformed subspace basis matrix is recognized as the dynamic texture with class label i^* according to:

$$i^* = \arg\min_i d_M^2(\hat{U}_i, \hat{U}_p). \tag{5}$$

5 Experimental results

To test the performance of the proposed method, the old DynTex database [13] was used in our experiments. There are altogether 35 color videos in this database. Each video represents one class of dynamic textures, which include tree leaves, flag, sea waves, and so on. Each video lasts for 10 seconds with 25 frames per second and the size for each frame is 400×300 pixels.

To simplify computation, we resize each frame into size 52×39 pixels and use gray scale image in experiments. We extract two video sequences from each video. Several examples are shown in Fig. 1. The first 100 frames of each video was used for training and the last 100 frames was used for test. We observed that there are obvious appearance changes in these two clipped sub-videos. In this way, there are 35 sequences for training and 35 sequences for test.

We compared two methods for dynamic texture recognition. The first one is the basic framework based on LDS and Martin distance, which is without constraint subspace learning, as described in Section 2. The second one is the presented framework based on LDS and Martin distance with constrained subspace learning, as described in Section 4. We denote the former as 'LDS' and the latter as 'CLDS'.

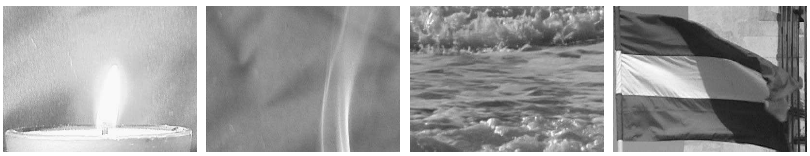

Fig. 1. Examples of dynamic textures from DynTex database [13]. From left to right: 'flame', 'smoke', 'sea_waves', and 'flag'.

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010

IEICE Electronics Express, Vol.7, No.18, 1329–1334

Table I. Performance comparison of two frameworks, LDS and CLDS, for dynamic texture recognition. d is the dimension of subspace.

	Recognition rate (%)					
Method	d=5	10	15	20	25	30
LDS	62.86	65.71	71.43	74.29	77.14	74.29
CLDS	74.29	71.43	82.86	80.0	85.71	82.86

For both methods, the first 60 principal components with PCA of each dynamic texture are chosen to estimate C in (1). Such a choice was to keep around 95% of data energy for each dynamic texture in estimating PCA subspace. For computational simplicity, $k = 1$ is chosen to form the observability matrix. And the dimension of constraint subspace was experimentally set as $N_p = 40$. Note that if there is no constraint subspace used, i.e. $N_p = 0$, CLDS reduces to LDS.

Based on the above parameter selection, the comparison of recognition performance for the two frameworks, LDS and CLDS, are presented in Table I, where the recognition rate corresponding to different choice of subspace dimension d are shown. As can be seen from the table, for the same d, CLDS has obtained improved recognition performance over LDS. When $d = 25$, the highest recognition rates are obtained for both methods, i.e., 77.14% for LDS and 85.71% for CLDS. The experimental results indicate that using a constraint subspace for dynamic texture recognition has obtained improved recognition performance since the difference between the two frameworks is whether the constraint subspace is used.

6 Conclusions

We present a new framework for recognizing dynamic textures based on LDS representation. The new framework improves the separability among different classes of dynamic textures with a transformation, which is achieved by projecting class subspaces onto a constraint subspace. As indicated by experimental results, using constrained subspace learning has obtained improved performance for dynamic texture recognition. Our future work is to use a discriminant learning framework for dynamic texture recognition by considering not only the between-class but also the within-class subspace differences.

Acknowledgements

This work was supported by the Natural Science Foundation of China (NO. 60775009).

© IEICE 2010
DOI: 10.1587/elex.7.1329
Received February 27, 2010
Accepted August 31, 2010
Published September 25, 2010