Video Feature Extraction Based on Modified LLE Using Adaptive Nearest Neighbor Approach

Preksha Bohra[1], Sunil Joshi[2]

[1]M.Tech Scholar, Department of ECE, College of Technology and Engineering, Udaipur, INDIA
[2]Professor, Department of ECE, College of Technology and Engineering, Udaipur, INDIA

Abstract—*Locally linear embedding (LLE) is an unsupervised learning algorithm which computes the low dimensional, neighborhood preserving embeddings of high dimensional data. LLE attempts to discover non-linear structure in high dimensional data by exploiting the local symmetries of linear reconstructions. In this paper, video feature extraction is done using modified LLE alongwith adaptive nearest neighbor approach to find the nearest neighbor and the connected components. The proposed feature extraction method is applied to a video. The video feature description gives a new tool for analysis of video.*

Keywords— Feature extraction, Frames, Images, LLE, ISOMAP, Manifold, PCA, Video.

I. INTRODUCTION

Digital video is an innovation which is most generally developed amid the couple of decades. The utilization of digital techniques in video engendered digital video, that allowed higher quality and, in the long run, much lower expense than prior analog technology. In straightforward terms video can be outlined as an electronic medium for the recording, copying, playback, broadcasting, and showcase of moving visual and audio media. Video is one of the medium of communication that delivers a lot of information than any other medium of multimedia system. However there are numerous complexities of video information and current video process strategies have restricted uses in fields of video information modeling, categorisation and retrieval. One among the essential step of video analysis is feature extraction[1].

Feature extraction (or dimension reduction) is a very important analysis topic in computer vision, pattern recognition and also in machine learning fields. The curse of high dimension is typically a significant reason behind limitations of the many sensible technologies, whereas the massive quantities of features could even degrade the performances of the classifiers when the size of the training set is minuscule compared with the number of features[2]. A number of feature extraction strategies are developed in past many years within which one among the essential is principle component analysis (PCA)[3], linear discriminant analysis (LDA)[3], and multidimensional scaling (MDS)[4]. These strategies are linear dimension reduction strategies and are simple and facile to implement. In recent years studies shown that several biometric systems[5] and multimedia system videos[6] are utilizing elegant nonlinear dimension reduction strategies for feature extraction. If the number of attributes is large, then the space of distinctive possible rows is exponentially large. Humans usually have problem comprehending information in several dimensions, so reducing information to a little number of dimensions is beneficial for visual image functions. Among the strategies, the foremost accepted strategies are isometric feature mapping (ISOMAP)[7], local linear embedding (LLE)[8] and Laplacian eigenmap, locality preserving projections (LPP)[9]. These strategies are quite helpful for facial or digit pictures and different real-world information sets. However, all of these strategies, either linear strategies or nonlinear ones, endeavor to find the low dimensionality features of single sample. The relationship between the samples has not been thought-about. In different words, once these collections are video sequences, these algorithms ignore the temporal coherence between frames, even if this cues give helpful information regarding the neighborhood structure and also the native geometry of the manifold.

The contribution of this paper is to define a video feature extraction technique using manifold learning and to extract the feature by applying LLE through adding virtual frames and adaptive nearest neighbor algorithm. The proposed manifold feature extraction method is applied to video trajectories. So the video manifold feature description can be a new tool for video analysis.

The rest of the paper is arranged as follows:

The video manifold feature is defined in section 2.1; In section 2.2 conditions of adding virtual frames is described; Video manifold feature extraction using modified LLE is provided in detail in section 2.3. Experimental results on video frames are presented in section 3; Finally, the conclusion is given in section 4.

II. VIDEO MANIFOLD FEATURE EXTRACTION BASED ON LLE USING ADAPTIVE NEAREST NEIGHBOR ALGORITHM

Video is depicted by a heirarchical data structure consisting of four differnet levels (video, scene, shot, frame) from top to bottom increasing in graininess, whereas a shot is that the basic unit of a video[9]. In general a video features a multiple shot, that is an ordered set of pictures. The video manifold feature is a low dimension description of the video sequences, that is extracted by modified LLE embedding using adaptive nearest neighbor algorithm and adding virtual frames.

2.1 Video Manifold Feature

Provided the given video clip is consisting of frame sequence $(f_1, f_2, \dots, f_k)$. All m×n pixel images frames f_i exist in the m×n dimensional space.

Definition : Given a dimensionality reduction method M, M is used to reduce the dimension of the original features to d,

$$M: (f_1, f_2, \dots, f_k) \rightarrow (v_1, v_2, \dots, v_k)$$

Each v_i is a d dimensional vector, and $(v_1, v_2, \dots, v_k)$ is d dimensional vector sequence, $(v_1, v_2, \dots, v_k)$ is called video manifold feature.

In general, $d \leq 3$. When d=1, $(v_1, v_2, \dots, v_k)$ is transformed into 1D vectors, so it is called video manifold feature vector.

2.2 Adding Virtual Frames

Locally linear embedding(LLE)[10][11] is an elegant nonlinear method for dimensionality reduction and manifold learning, which attempts to discover the structure of high dimensional data which lies in a nonlinear high-dimensional manifold and find their embedding in an low dimensional Euclidean space.

LLE requires that the input high dimensional data lie on a smooth and well-sampled single manifold. However, it fails to find the embedding when it is applied to multi-manifold, the original LLE algorithm failed to give a reasonable embedding, even the embedding collapse to some points [1].

The utilization of virtual samples is an effective scheme to solving the problem. In machine learning, virtual samples have been used in comprehensible learning. For example, virtual samples were generated to help extract symbolic rules from complicated learning systems such as neural network ensembles [2]. Virtual Samples are also useful in learning with imbalanced data sets. For example, in the Smote algorithm [12], virtual samples of the minority class are generated so that the number of minority training samples is increased.

The given F_i, F_{i+1} represents the last frame of the i^{th}-shot and the first frame of the j^{th}-shot respectively. $\| F_i - F_{i-1} \|^2, \| F_{i+1} - F_i \|^2$ represents Squared Euclidean distance between frame $F_{i-1,}$ and F_i, F_{i+1} respectively. Here we add m virtual examples between .

$$m = \left\lceil \| F_i - F_{i+1} \|^2 \Big/ (\| F_i - F_{i-1} \|^2 + \| F_{i+1} - F_i \|^2)/2 \right\rceil$$

frame F_i and frame F_{i+1}

(1)

$$vframe_j = rF_i + (1 - r)F_{i+1} \qquad (2)$$

Where $r = \frac{j}{m+1}, j = 1, 2, \dots, m$

2.3 Video feature extraction using LLE

Video sequence can be considered as the high-dimensional Euclidean space Rf^{N+1}, where N is the pixel number of each frame. Each frame of the video can be seen as a point in Rf^{N+1} space. A manifold learning method is used to extract the local structure embedded in space Rf^{N+1} to describe the video. The given video consists of frame sequence $(f_1, f_2, \dots, f_n)$.

The algorithm is explained in the folowing table:

TABLE I: ALGORITHM FOR MODIFIED LLE

Input: Video : Bicycle trick riding, no.2 .
Output: Mapped output in 2D and 3D scattter plot.
Procedure: Step 1: Transforming frame sequence $f_i (i = 1, \dots. k)$ into a one-dimensional vector F_i with size M*N (M,N for the number of frame rows and columns respectively). Step 2: Add virtual frames if needed. Step3 : Learn it by dimension reduction technique modified LLE. i. Find nearest neighbors and pairwise distances. ii. Constructing nearest neighbor graph either using a) k-nn (k nearest neighbor algorithm), or b) adaptive neighborhood selection. iii. Identify largest connected component of the neighborhood graph iv. Perform Dulmage-Mendelsohn permutation resulting in connected frames, set is reduced and flat region is discarded.

ISSN : 2454-1311

v. Solving MSE for reconstruction weights for all points or Computing Reconstruction Weights
vi. Define the sparse cost matrix M.
vii. Compute the embedding from the bottom eigen vectors of this cost matrix.
viii. Do Eigen-decomposition.

Step 4: Mapped output in 2D and 3D format.
Step 5: End

III. EXPERIMENTS AND RESULTS

In order to test the validity of our implementation, the testing database was collected from the Open Video Project – a shared digital video collection 'Bicycle trick riding, no. 2.mpeg' [13]. The clip opens with a man riding a bicycle in a forwards circle, pausing and balancing for a moment, then continuing in a forwards circle. Some frames extracted from the video are shown in fig 1.

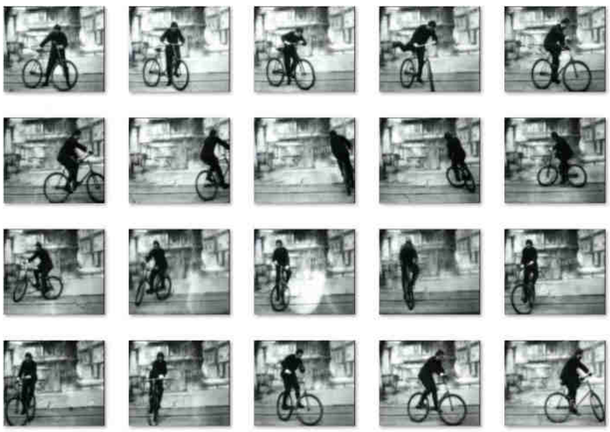
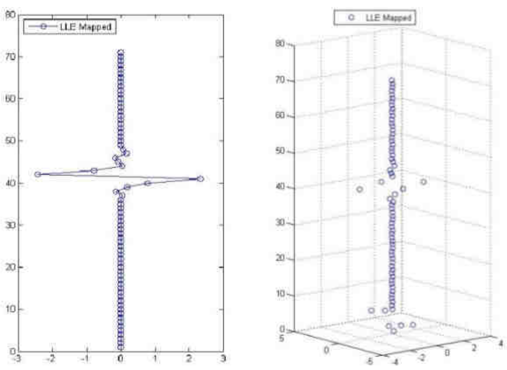


Fig.1: Some frames extracted from the video database

The following fig 2 shows the output of video feature extraction using modified LLE



(a) (b)

Fig.2 : Mapped output of modified LLE in
(a) 2D using adaptive-NN algorithm
(b) 2D using adaptive-NN algorithm

The following figure shows the mapped output using three different manifold learning techniques, LLE, ISOMAP and PCA respectively.

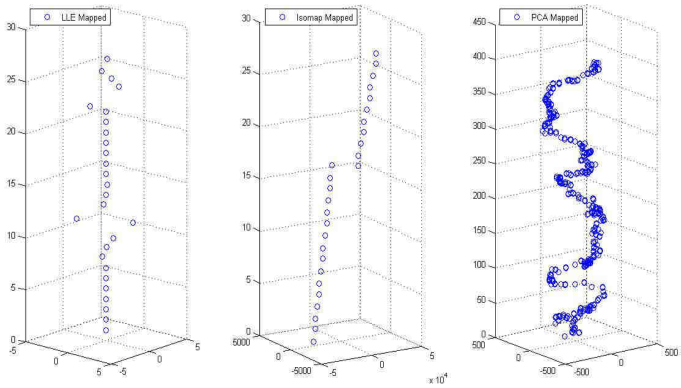


Fig. 3 : Comparison of mapped output of video feature extraction using LLE, ISOMAP and PCA

IV. CONCLUSION

This paper presents an improved technique of video feature extraction. The computation is done by using adaptive NN method. In adaptive nearest neighbour the machine itself learns and does the remaining computaion. Though it is time consuming but it overcomes the limitation of the traditional k-nearest neighbor algorithm (k-NN) which usually identifies the same number of nearest neighbors for each test example. The algorithm finds out the optimal k itself. After comparing all the three techniques, if PCA and ISOMAP is used to reduce the same data set, the resulting values are not so well organized. LLE with adaptive nearest neighbor approach is better as compared to ISOMAP and PCA.

REFERENCES

[1] [1]. Maosheng Fu , Qin Xu,Min Kong and Bin Luo,Manifold "*Feature Extraction for Video Based on Local Linear Embedding" , The 5th International Conference on Computer Science & Education,Hefei,China.*August 24-27,2010.

[2] [2] A.K Jain, B.Chandrasekaran, *Dimensionality and smaple size considerations in pattern recognition practice,in :Handbook of Statistics, (*North Holland,Amesterdam,1982)

[3] I.T.Jolliffe,*Principal Component Analysis,* (1989,Springer-Verlag,New York).

[4] [4]. T.Cox and M.Cox. *Multidimensional Scaling* (1994,Chapman & Hall, London).

[5] Ryazard Schoras," *Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems",International Journal of Biology and Biomedical Engineering.*

[6] Sen-ching S. Cheung and Avideh Zakhor," *Fast Similarity Search and Clustering of Video Sequences*

on the World-WideWeb",,IEEE Tranactions on Multimedia,Vol 7.No.3,June 2005
[7] J.B.Tenenbaum, V.de Silva, J.C.Langford. "*A global geometric framework for nonlinear dimensionality reduction", Science 290(2000) 2319-2323.*
[8] S.T. Roweis, L.K. Saul, "*Nonlinear dimension reduction by locally linear embedding", Science 290 (2000) 2323–2326*.
[9] M. Belkin, P. Niyogi," *Laplacian eigenmaps for dimensionality reduction and data representation", Neural Computation 15 (6)(2003) 1373–1396.*
[10] X. He, S. Yan, Y. Hu, H. Zhang," *Learning a locality preserving subspace for visual recognition', in: Proceedings of the Ninth International Conference on Computer Vision, France, October 2003, pp. 385–392*.
[11] Y.Rui,T.S.Huang, S.Mehrotra, "*Constructing Table-of-Content for Videos", ACM Multimedia Systems Journal, Special Issue Multimedia Systems on Video Libraries, Vol.7, No.5, pp. 359-368,* Sept. 1999
[12] R.O.Duda,P.E.Hart,*Pattern Classification and Scene Analysis(,*Wiley,New York,1973).
[13] http://www.open-video.org/details.php?videoid=4750
[14] Saul, L., Roweis, S.: "*Think Globally, Fit Locally: Unsupervised Learning of Nonlinear Manifolds". Technical Report MS CIS-02-18, University of Pennsylvania (2002).*
[15] *Robert Pless," Image Spaces and Video Trajectories: Using Isomap to Explore Video Sequences", Department of Computer Science and Engineering, Washington University in St..Louis.*