Video Classification using Slow Feature Analysis and Neural Network

Bilkis .A. Inamdar[1], Prof. Ujwal. Harode[2]

[1]*Student*, [2]*Professor*, *Pillai's Institute of Information Technology, University of Mumbai, Mumbai*

Abstract

Videos concepts based feature classification is a process, where the classification of videos will be classified on the basis of video classes. For classification initially frames will be extracted from the dataset then these frames will be processes for feature extraction. In our context we made an approach towards the slowness principle which has a fundamental ability to perform as close to the function of human brain to visualize the change in the surrounding and has the tendency to create slow motion output from the fast varying input signal due to the slowness it provides with a stable output as well. Using the frames from the input videos from the datasets features are extracted. Initially V1-like features is obtained as a result these features are then processed to slow feature analysis method which determines different category of video sections composed of natural scenes from the dataset input. Hence the system will be trained and motion components are computed thus using the knowledge based classification algorithm this videos will be classified according to their classes using a classifier i.e. feed forward back propagation neural network.

Key terms - *unsupervised learning, motion analysis, neural network.*

I INTRODUCTION

Due to explosive growth of video usage within past few years is a direct result of the huge increase in internet bandwidth speeds and increase in the availability and usage of cameras. Furthermore as the number of online video portals has drastically increased which arise a need for video classification. Our topic involves with the working principle of video classification, its advantages as well as applications of slow features analysis. Video classification problem is fundamentally different from the classical document classification. In document classification it defines a text file as one dimensional file which contains only text. On the other hand a video file is defined as 3 Dimensional which contains all 3D contents text, audio and visual. Actually motion features arise from relative motion between the different objects in the scene and the camera, designing efficient motion descriptors is a key ingredient of video analysis system. But since here we are tackling the problem of categorizing dynamic natural scenes there is a Lack of tools to classify and retrieve video content. So there is a need to develop a system which classifies videos according to its features. The proposed system will classify the videos from the datasets and based on the content the features are extracted from the video. For instance if a video comprises of several dynamic natural scenes out of which few scenes are from beaches, waterfall, ocean, windmill etc from the dataset and few are from different dynamic natural scenes from the dataset itself then while playing the video, the system will display the type of the scene motion being played. This will be done by first tagging set of video frames into V1 features and then same videos will be given to slow feature analysis and classified using classifier. In our context the motion is often correlated with effects that may be considered as interferences such as shadows, lightning, variations, etc.

II RELATED WORK

In this portion we move towards in depth of video classification and majorly focus on two main feature of proposed system: **1**.Desired motion feature and **2**.Their utilization for categorizing video.

As in previous work the use of optical flow which has been applied to natural scenes classification based on HOF, as optical flow approaches are restrained by the optical flow constraints for this the implementation in practice is not obvious and the performance of this type of motion features is subject to collapse under the content of natural video scenes. A set up of a specific application in order to classify global human action to be viewed from a distance using low resolution windows where optical flow measurements are used. The proposed texture classification which is based upon Linear Dynamical System(LDS). LDS is used in order to explicitly model texture dynamics. This has been successfully applied to various inputs ranging from dynamic texture to motion segmentation. As LDS is intrinsically limited by first order markov property and linearity assumption; therefore it is too restrictive to properly solve complex task of unconstrained dynamic scene classification. It has been recently introduce Slow feature analysis an unsupervised method where the learning principle is based on neuroscience approach, this is used to minimize temporal variation created by motion in order to learn stable representation of undergoing motion. The idea behind SFA is that perception vary on slower time scale compared to the input signal from the environment. For a given temporal input sequence (i.e. motion), the

SFA model learns to generate a "slower" and thus more invariant outputs signal. Recently, SFA has been investigated in to represent local motion for human action recognition. Interestingly, this work, closely related to ours, consolidates the relevance of using SFA to extract meaningful motion pattern for video classification.

III PROPOSED METHOD

Our method is being new as it includes unsupervised motion feature learning and it is more challenging than that of the motion descriptors used in literature. Slow Feature Analysis Principle provides a good transformation from videos into histograms of slow feature temporal averages and thus due to this the temporal dimension of the input signal is reduced to a scalar value. The exact flow of our method is depicted in figure 1.

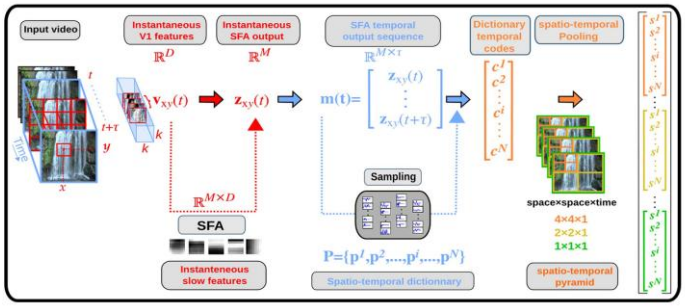

Fig.1. Block Diagram To Determine Flow of Proposed method

A. *Videos and Frame Extraction*

Initialize using datasets of yupenn dynamic natural scenes which include natural scenes of beach, elevator, forest fire, fountain, railway etc. There are 14 different dynamic natural scenes videos which comprises 30 videos in each category. The input video is processed and the number of frames is calculated along with the gray scale value for each motion components are computed which is obtained by means of gradient method.

B. *Motion features*

Motion features play an important role in video repossess. In this paper, we propose newly motion features that depicts the local motion of arbitrarily shaped video objects. Controvert to previous methods that rely on handcrafted descriptors, we propose here to represent videos using unsupervised learning of motion features. Our method is based on slow feature analysis principle to learn local motion descriptors which represent the slow and more stable motion components of training videos.

C. *Feature Extraction*

V1-like features are extracted these features is notable for detecting visual feature dimension including color, orientation and simple motion. Mathematical model function of V1 has been compared to Gabor transforms. When each frame is processed to extract V1-like features as a result each local region is represented with a vector in R^D.

D is the dimension of V1-space (space, scale, orientation) each region determined is then further processed by mapping V1-features on a set of slow features, hence generating a local low dimension representation of size R^M. An algorithm for unsupervised learning of invariant representation from data with temporal correlation is introduced i.e. slow feature analysis These invariant features of temporally signals are useful for analysis and classification here SFA is not used as a classifier but used as a features.

Slow features analysis is a new method for learning invariant or slowly varying features from a vectorial input signal, the slowness principle is proposed because it works analogous to human brain and an attempt has been made to virtualized that effect to the video classification problem, slowness principle can be explained from the figure 2.

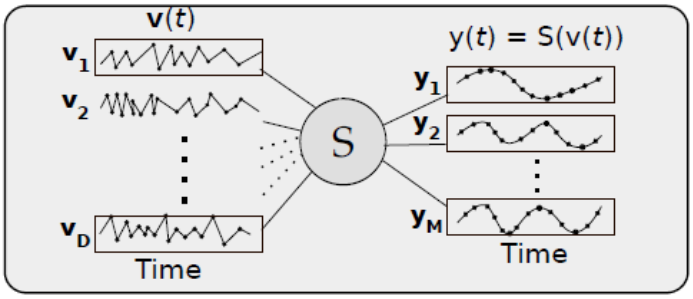

Fig. 2 Slow Feature Analysis

Given a high dimensional input signal $v(t)$ find function $S(v(t))$ such that the output signal is new representation $y(t) = [y_1(t), \dots \dots \dots, y_m(t)]^T$

$$y_j(t) = S_j(v(t)) \qquad (1)$$

Equation (1) Varies as slow as possible and still retains the relevant information.

New space representation is learned by minimizing the average square of the signal temporal derivation obtained in equation (2).

$$\min_{S_j} \langle \dot{y}_j 2 \rangle \; t \qquad (2)$$

Under the constraints

$$(y_j)t = 0 \; (zero\ mean) \qquad (3)$$

$$(y_j 2)t = 1 \; (unit\ variance) \qquad (4)$$

$$\forall j < j' : \langle y_j , y_j' \rangle t = 0 \; (decorrelation) \qquad (5)$$

where $\langle y \rangle t$ is the temporal average of y,the angular brackets indicates averaging over time.The Δ value defined by equation (2) is the objective of the optimization problem and measures the slowness of an input signal as the time average of its squared derivative. A slow value indicates small variation over time and therefore slowly varying signal. The Δ value is optimized under 3 constraints.

Equation (3) and (4) normalize all outputs to a common scale which makes their temporal derivative directly comparable. In equation (5) the output signal are decorrelated from one another and guarantees that output signal components code for different information.

Now the motion features m(t) at position (x, y) and across time (t)=(t…..t $+\tau$) is defined by threading together short term temporal sequences of SFA outputs.

With M as a slow features we have S $\in R^{M \times D}$

Motion features m(t) $\in R^{M \times \tau}$ obtained act as spatio temporal atoms corresponding to stable motion components inside a small window and is learned using a simple unsupervised sampling procedure in which motion features are sampled on training videos at random position and times. Resulting features are passed to create neural network functions with their corresponding targets (means class numbers from 1 to 14) and this can be used to feed a classifier (i.e. BPNN).

D. Classification

As a result features for a given set of input output pair obtained by slow feature analysis is further proceeding for classification. Back propagation learning algorithm is one of the most important developments in neural networks, learning algorithm is applied to multilayer feed forward networks consisting of processing elements with continuous differentiable activation functions. This algorithm provides a procedure for changing weights in a Back Propagation network to classify the given input pattern correctly in this classification it has three stages,

1. The feed forward of the input training pattern.
2. Calculation of back propagation of errors.
3. Updating weights.

Back Propagation Neural Network is a multilayer feed forward neural network consisting of input layer, hidden layer, output layer.

IV EXPERIMENTAL RESULTS

Our proposed method consists of two important modules as depicted in sequential flow in figure 3.

Module 1: Training

- Get the train videos from the dataset.
- Computing motion components
- Compute V1 features.
- Performing on train V1 features.
- Compute slow features.
- Train slow features.
- Save both trained data for V1 and slow features.

Module 2: Testing

- Get the test video.
- Compute motion components.
- Compute to extract V1 features.
- Load V1 trained data
- Perform classification on the video.
- Compute slow features to extract slow varying motion.
- Load slow features trained data
- Classify videos to categorized appropriate classes from 1 to 14

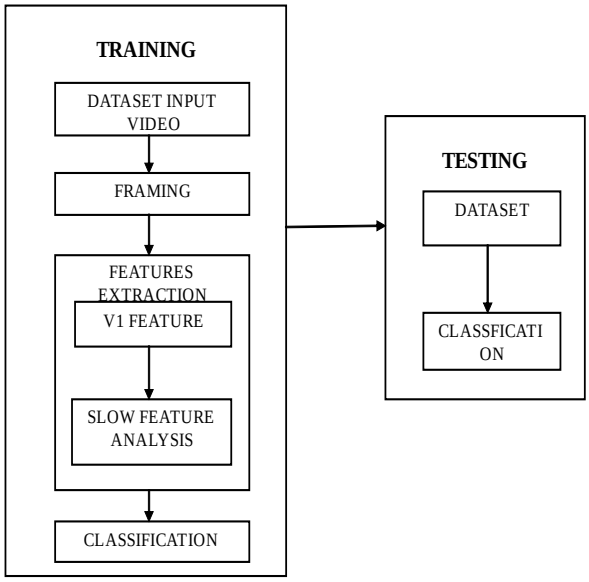

Fig.3. Data Flow Sequences for Proposed method

Fig.4. Motion components computation

Fig .5 V1 features

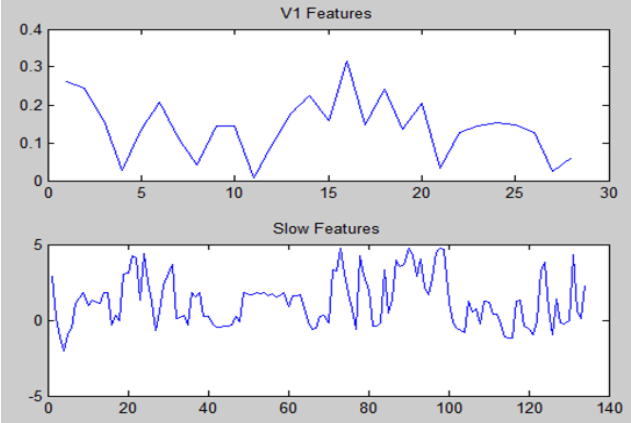

Fig .6 Graphical representation Of V1 and Slow Features

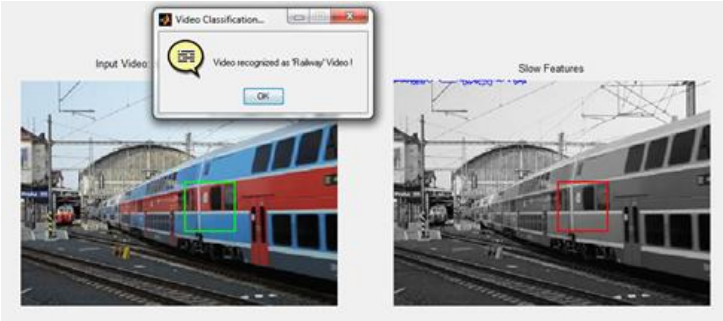

Fig .7 Classifications output for recognized video class

V CONCLUSION

The proposed slow features analysis for video classification is based on foundations of neurosciences. These features represent stable descriptions of video which is used to obtain state-of-the-art classification. The use of feed forward back propagation neural network is to classify the videos which has been successfully investigated and found to give improve classification accuracy. Thus the process of recognition by the human brain has more closely been investigated one possible condition not evaluated is that instead of using instantaneous features. Features varying across a time interval (most significant interval) could be investigated in future.

REFERENCES

[1] Christian Thériault, Nicolas Thome, Matthieu Cord, and Patrick Pérez, "Perceptual Principles for Video Classification With Slow Feature Analysis," IEEE journal of selected topics in signal processing, VOL. 8, NO. 3, JUNE 2014

[2] K. G. Derpanis, M. Lecce, K. Daniilidis, and R. P. Wildes, "Dynamic scene understanding: the role of orientation features in space and time in scene classification," in Proc. CVPR, 2012

[3] S. Klampfl andW.Maass, "A theoretical basis for emergent pattern discrimination in neural systems through slow feature extraction," Neural Comput., vol. 22, no. 12, pp. 2979–3035, Dec. 2010.

[4] A.A. Efros,A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in Proc. ICCV, 2003.

[5] S. Soatto, G. Doretto, and Y. N. Wu, "Dynamic textures," in ICCV, 2001.

[6] S. S. Beauchemin and J. L. Barron, The Computation of Optical Flow. New York, NY, USA: ACM, 1995.

[7] L. Wiskott and T Sejnowski, "Slow feature analysis: unsupervised learning of invariances," Neural Comput., vol. 14, pp. 715–770, 2002.