Abstract--- Slow Feature Analysis (SFA) has been established as a robust and versatile technique from the neurosciences to learn slowly varying functions from quickly changing signals. SFA framework is introduced to the problem of recognizing prisoner’s actions by incorporating the supervised information with the original unsupervised SFA learning. Firstly, large amount of cuboids are collected in the motion boundaries, and local feature is described with SFA method. Each action sequence is represented by the Accumulated Squared Derivative (ASD), which is a statistical distribution of the slow features in an action sequence [1]. The descriptive statistical features are extracted inorder to reduce the dimension of the ASD feature is proposed. Finally, one against all support vector machine (SVM) is trained to classify action represented by statistical features.

Index Terms--- Human Action Recognition, Pattern Recognition, Slow Feature Analysis

I. INTRODUCTION
Artificial intelligence, a subfield of computer science, addresses learning and intelligent behavior in machines. Video surveillance systems are essential in providing effective security in prisons and correctional facilities. Incidents involving inmate violence, suicide, and officer misconduct are just a few examples of the unfortunate acts that can occur within prison walls. This project employing advanced video surveillance technology enable these facilities to upgrade their systems in order to provide more comprehensive monitoring and a heightened level of safety for inmates and workers. The automatic analysis of the videos recorded and the timely warning information will help the higher officials to take action on the erring prisoners immediately.

Slowness is an important signature because meaningful variables are more persistent than the raw pixel values. Since the temporal slowness principle extracts the useful motion patterns for human motion analysis the slow feature analysis is adopted here for the problem of human action recognition. The SFA is based on the slowness principle. How does the brain process the sensory information? The work hypothesis of this project assumes that the cortex adapts itself in order to make the response of the neurons vary slowly in time. This is motivated by a simple observation: while the environment varies on a relatively slow timescale, the sensory input, e.g. in our case the gray pixel values of a camera, consists of raw direct measurements that are very sensitive even to small transformations of the environment or the state of the observer. The sensory signal vary thus on a faster timescale than the environment. The work hypothesis implies that the cortex is actively extracting slow signals out of its fast input in order to recover the information about the environment and to build up a consistent internal representation. This principle is called the slowness principle.

L. Berwin Rubia, PG Scholar, CSE, Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, India, E-mail: berwinrubia@gmail.com
K. Manimala, Assistant Professor, CSE, Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, India.
The flowchart of action recognition system is shown in figure 1. SFA takes the pixels from an input video and, by a linear transformation, extracts sets of variables which vary as slowly as possible, whilst begin decorrelated from one another. To represent action sequences, the squared first order temporal derivatives are accumulated over all transformed cuboids into one feature vector, which is termed the Accumulated Squared Derivative (ASD) feature.

The ASD feature encodes the statistical distribution of slow features in an action sequence. Then the descriptive statistical features such as geometric mean, harmonic mean, variance, standard deviation, median, range, interquartile range, mean absolute deviation, percentile, quantile, moment, kurtosis, skewness and minimum index are extracted. Finally, a one against all support vector machine (SVM) is trained to classify actions represented by descriptive statistical features.

The remainder of the paper is organized as follows. In Section II Related work is described. In Section III, the Slow Feature Analysis used is described in detail. In Section IV, the proposed recognition method is analyzed, and in Section V, final conclusion is drawn.

II. RELATED WORK

Automatic recognition of human actions is a challenging problem in machine learning. Also a long standing problem in computer vision is to develop representations which are invariant to transformations of the input. Much of the recognition progress has been through incorporating ideas from single-frame object recognition and adapting them for temporal-based action recognition. Past research in this domain can be roughly classified into two approaches: one that extracts a global feature set from a video ([2], [3], [4]), and aims to assign a single label to the entire video, using these features. This paradigm obviously requires that the observed action does not change during the duration of the video. The other approach extracts a feature set locally for a frame (or a small set of frames), and assigns an individual action label to each frame ([5], [6], [7]). If required, a global label for the sequence is usually obtained by simple voting mechanisms. The features are obtained by analysing a temporal window centred at the current frame, therefore the classification lags behind the observation, because a frame can only be classified after all frames in the temporal window have been observed. Motivated by the recent success of biologically inspired approaches for the recognition of objects in real-world applications [8], it is decided to use the SFA-based method which is motivated by the studies of the biological vision based on different theories on visual neuron modelling.
III. SLOW FEATURE ANALYSIS

Mathematically, SFA is defined as follows[10]: Given an I-dimensional input signal \( x(t) = [x_1(t), \ldots, x_I(t)]^T \) with \( t \in [t_0, t_1] \) indicating time, SFA finds out a set of input-output functions \( g(t) = [g_1(t), \ldots, g_J(t)]^T \) so that the J-dimensional output signal \( y(t) = [y_1(t), \ldots, y_J(t)]^T \) with \( y_j(t) = g_j(x(t)) \) varies as slowly as possible, i.e., for each \( j \in \{1, \ldots, J\} \),

\[
\Delta_j = \Delta(y_j) = \langle y_j^2 \rangle, \text{ is minimal,} \quad (1)
\]

Subject to

\[
\langle y_j \rangle = 0, \text{ zero mean;} \quad (2)
\]

\[
\langle y_j^2 \rangle = 1, \text{ unit variance;} \quad (3)
\]

and \( \forall j' < j : \langle y_j'y_{j'} \rangle \), decorrelation, \( (4) \)

where \( \cdot \) denotes the operator of computing the first order derivative of \( y \) and \( \langle y \rangle \), is the mean of the signal \( y \) over time. Equation (1) is the primary objective of minimizing the temporal variation of the output signal, where the temporal variation is measured by the mean of the squared first order derivative. Equation (3) means that the transformed signal should carry some information and avoid the trivial solution \( y_j(t) = \text{const} \). Equation (4) ensures that different output components carry different types of information and it also induces an order, the first output signal being the slowest one, the second being the second slowest, etc.

Slow feature functions can be obtained by the following two steps:

1. Nonlinear Expansion

The nonlinear transformation can be deemed as the linear transformation in a nonlinear expansion space. The nonlinear expansion function \( h(x) \) is defined by

\[
h(x) := [h_1(x), \ldots, h_m(x)] \quad (5)
\]

Apply a nonlinear function \( h(x) \) to expand the original signal and centralize \( h(x) \)

\[
z := h(x) - h_0 \quad (6)
\]

where \( h_0 = \langle h(x) \rangle \).

2. Solve the Generalized Eigenvalue Problem

\[
AW = BW\Lambda \quad (7)
\]

where \( A = \langle ZZ^T \rangle \), is the expectation of the covariance matrix of the temporal first order derivative of the input vector, \( B = \langle ZZ^T \rangle \), is the expectation of the covariance matrix of the input vector, \( \Lambda \) is a diagonal matrix of generalized eigenvalues, and \( W \) is the corresponding generalized eigen vectors. Assume the dimensionalities of
Matrices $A$ and $B$ are $M$, the first $K$ eigenvectors $w_1, \ldots, w_K$ ($K \ll M$) associated with the smallest eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_K$ are the nonlinear slow feature functions $g_1(x), \ldots, g_K(x)$:

$$g_j(x) = w_j^T (h(x) - h_0)$$ (8)

Here, the input-output function computes the output slowly varying signal instantaneously. Therefore, slow variation of the output signal cannot be achieved by using the temporal low-pass filter, but must be obtained by extracting aspects of the input signal that are inherently slow and useful for a higher level representation.

IV. SFA BASED ACTION RECOGNITION

SFA based human action recognition includes five major steps they are Collection of training cuboids, Slow feature function learning, Action feature representation, feature selection, and Classification.

A. Collection of Training Cuboids

The spatial and temporal information of motion is mandatory for action recognition. For that, the spatial and temporal information in the video sequence is gathered. The region in which the frame posses the motion represents the spatial information and the variation of the motion over time is temporal information. The local volume which has the spatial and temporal information is called cuboid. The cuboids are collected by randomly sampling in motion boundaries. Before the cuboid collection, perform normalization for each action sequence, so that the input signals are of zero mean with unit variance. Cuboids are collected from $d$ successive frames and the statistical feature is computed to represent the action sequence. The steps involved in collecting training cuboids are find the motion boundaries, cuboid collection, reformatting, and dimensionality reduction.

1. Find the Motion Boundaries

The motion boundary of the frame is obtained from the frame difference of the original frames. The motion boundaries of the frame differences in a bounding box will be returned if its gradient magnitude is larger than a predefined threshold $\delta$. To ensure that most informative regions are selected, the predefined threshold $\delta$ is set as a small value. This setting introduces noise regions in background and shadows. However, the effects of the selected noises can be balanced out by using the statistics over a large number of cuboids. The spatial position $(x, y)$ locates on the motion boundaries detected by sobel operator.

2. Cuboid Collection

After Sobel filtering, the pixels where the responses are ones are considered as informative region. From these response 25% of pixels are sampled uniformly and for each sampled pixel the cuboid is collected from $d$ successive frames. The patch size of the cuboid is $h \times w$, hence the cuboid size is $h \times w \times d$. In experiments the cuboid size is $16 \times 16 \times 7$.

3. Reformatting

To include temporal information, the input vectors to SFA were formed by the pixel intensities of $d$ consecutive frames. Each input vector is reformatted by $\Delta t$ successive frames, so SFA counts the temporal information in the
neighbor frames. Figure 2 shows the reformatting process. After the cuboid is reformatted the input vector at each time includes $\Delta t$ successive patches. The pixel intensities of the 3 patches are concatenated along the row order to form a long vector.

![Figure 2: Reformatting Process](image)

4. **Dimensionality Reduction**

After the nonlinear expansion, which is performed before SFA learning, the dimensionality of input vector increased greatly. For example, the quadratic expansion increases the dimensionality from $n$ to $n + n \times (n + 1) / 2$. Thus, before the nonlinear expansion perform Principal Component Analysis (PCA) to reduce the dimensionality of the input vector to 50, which is sufficient for the subsequent experiments. The first principal component, which accounts much of the variability in the data, is obtained by choosing the eigenvector that corresponds to the largest Eigen value of the covariance matrix.

**B. Supervised Slow Feature Function Learning**

The collected local cuboids are labelled by action categories. Then, the SFA learning is performed to extract slow feature functions for each action category independently. Finally, the statistical feature is computed with all slow feature functions.

**C. Action Feature Representation**

SFA minimizes the average squared derivative, so the fitting degree of a cuboid to a certain slow feature function can be measured by the squared derivative of the transformed cuboid. If the value is small, the cuboid fits the slow feature function. Otherwise, the cuboid does not fit the function. For cuboid $C_i$ and slow function $F_j$, the squared derivative $v_{ij}$ is

$$
\sqrt{\sum_{i=1}^{d-\Delta t} [C_i(t+1) \otimes F_j - C_i(t) \otimes F_j]^2}
$$

(9)

where $\otimes$ is the transformation operation.

Then accumulate the square derivatives over all cuboids to form ASD feature:

$$
f_{ASD} = \sum_{i=1}^{N} V_i
$$

(10)

where $N$ is the total number of cuboids in the current snippet and $V_i = <v_{i,1}, v_{i,2}, \ldots, v_{i,k}>^T$. Since the number of cuboids detected in a snippet may differ from that in another snippet, it is necessary to normalize the feature vector.
D. Feature Selection

Since the ASD feature is of dimension 200, training the SVM is time consuming. Hence the descriptive statistical features such as geometric mean, harmonic mean, variance, standard deviation, median, range of values, interquartile range, mean absolute deviation, percentiles, quantiles, central moments, kurtosis, skewness and minimum index are extracted from the ASD feature.

E. Classification

After the computation of the statistical features, a multiclass one against all SVM is trained by the statistical features for each action category. The statistical features are computed from d successive frames. Thus, for a sequence with N frames, at most N-d+1 features can be obtained. Accordingly, N-d+1 labels can be assigned by classifier. Finally, the majority voting rule is used to determine the label of the sequence.

V. Conclusion

The video surveillance in prison cannot gather any information to recognize human actions from the video itself or from the frames also video footages are analyzed manually only after abnormal events occurred in jail. Hence online analysis of video sequences inside prison with the advanced information technology methods make chance to deal with the abnormal cases at the very beginning. Here, slow feature analysis is employed in this project to extract the invariant features. To evaluate the capability of the SFA-based method to human action recognition the Weizmann human action dataset and UT Interaction dataset is adopted in the experiments. The interesting actions considered are single person actions such as bend, jump, run, walk, wave, wave2 and multiperson actions such as kick, push, handshake, punch, and hug. The learned slow feature functions encode discriminative information for the subsequent recognition. Further, the descriptive statistical features are extracted from the Accumulated Squared, which is a statistical distribution of the slow features in an action sequence. Finally, the one against all support vector machine (SVM) is applied to recognizing prisoner’s actions in given videos.

References