**Experiment No.11 (CNN)**

**Aim-** Pattern recognition application and Understanding of Convolutional Neural Network Algorithm.

**Pattern Recognition -** Pattern recognition is a process that looks at the available data and tries to see whether there are any regularities within it. There are two main parts:

1. Explorative part, where the algorithms are looking for patterns in general
2. Descriptive part, where the algorithms start to categorize the found patterns

Unlike with computer vision, the pattern recognition can be anything:

1. Texts or words
2. Images
3. Sentiments (emotions)
4. Sounds
5. Other elements and information

Pattern recognition is a process that takes in raw data and makes an action based on the category of the pattern. It optimally extracts patterns based on certain conditions and separates one class from another. Pattern recognition was often achieved using linear and quadratic discriminants, the k-nearest neighbour classifier or the Parzen density estimator, template matching and Neural Networks.

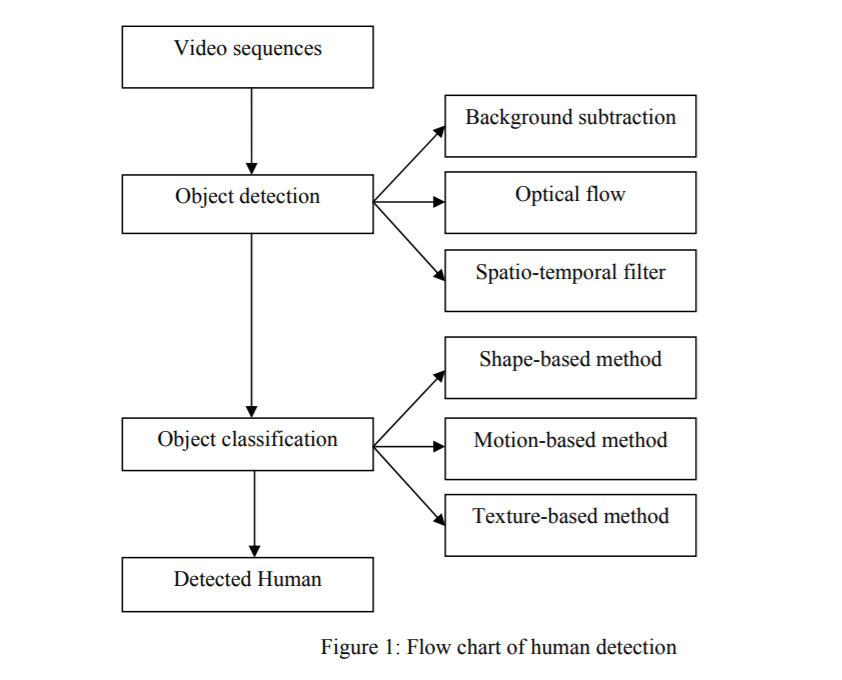
**Application of pattern recognition to be discussed-** Human detection in surveillance videos

**Introduction -** Detecting human beings in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in a dense crowd, person identification, gender classification, fall detection for elderly people, etc. The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field. Most existing digital video-surveillance systems rely on human observers for detecting specific activities in a real time video scene. But there are limitations in the human capability to monitor simultaneous events in surveillance displays.

Hence human motion analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition. An intelligent system detects and captures motion information of moving targets for accurate object classification. The classified object is being tracked for high-level analysis.

**Techniques-** The detection process generally occurs in two steps: Object detection and Object classification. Object detection could be performed by background subtraction, optical flow and spatio-temporal filtering. The object classification methods could be divided into three categories: shape-based, motion-based and texture-based.

**Block Diagram –**



**Dataset used –**

1. KTH dataset is the largest available and most standard dataset widely used for benchmarking results for human action classification. The dataset contains six activities (boxing, hand waving, handclapping, running, jogging, and walking) performed by 25 subjects in four different scenarios.
2. Weizmann human action dataset contains a total of ten actions performed by nine people, to provide a total of 90 videos. Sample sequences are shown in Figure 5. The dataset contains videos with a static camera unlike KTH, where some of the videos had zooming and also the videos have simple background. As this dataset contains ten activities, which is more compared to six activities of KTH dataset

**Applications -** For an intelligent video surveillance system the detection of a human being is important for abnormal event detection, human gait characterization, people counting, person identification and tracking, pedestrian detection, gender classification, fall detection of elderly people, etc

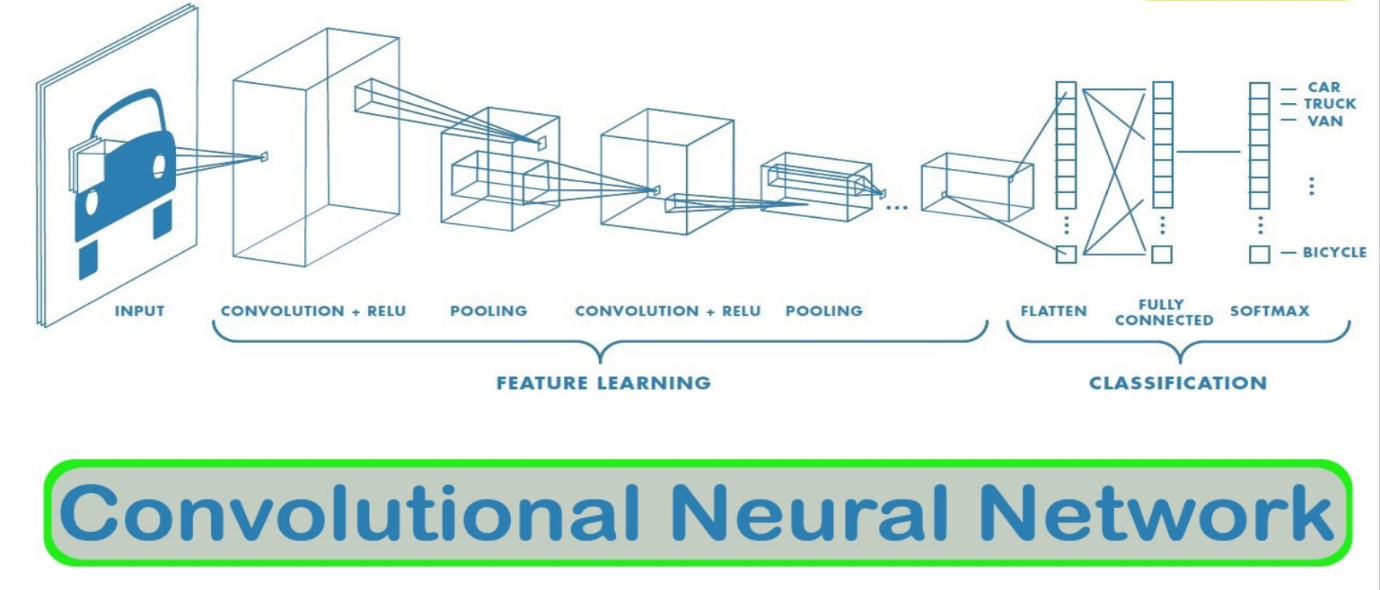
**Result -** A significant amount of work has been done with a view to detect human beings in a surveillance video. However the low-resolution images from the surveillance cameras always make this work challenging. Most of the object detection methods rely on known operation environment. The model adaptation speed based on observed scene statistics could be improved in future for faster adaptation of changed background and better persistency. But occlusion is a major problem for background segmentation technique. Optical flow and Spatio-Temporal filter techniques address this issue to some extent where the object of interest is occluded by a fixed object. But it is always difficult to detect an object in motion which is occluded by objects with similar shape and motion. One solution could be constructing 3D image for a 3D system by using volume information obtained from multiple cameras. From the machine vision perspective it is hard to distinguish an object as a human due to its large number of possible appearances. Moreover the motion of human is not always periodic.

**Conclusion -** Detecting human beings accurately in a surveillance video is one of the major topics of vision research due to its wide range of applications. It is challenging to process the image obtained from a surveillance video as it has low resolution. A review of the available detection techniques is presented. The detection process occurs in two steps: object detection and object classification. In this paper, all the available object detection techniques are categorised into background subtraction, optical flow and spatio-temporal filter method. The object classification techniques are categorised into shape-based, motion-based and texture-based method. The characteristics of the benchmark datasets are presented and major applications of human detection in surveillance video are reviewed.

**Introduction to CNN** - In [deep learning](https://en.wikipedia.org/wiki/Deep_learning), a convolutional neural network (CNN, or ConvNet) is a class of [deep neural networks](https://en.wikipedia.org/wiki/Deep_neural_network), most commonly applied to analyzing visual imagery. They have applications in [image and video recognition](https://en.wikipedia.org/wiki/Computer_vision), [recommender systems](https://en.wikipedia.org/wiki/Recommender_system), [image classification](https://en.wikipedia.org/wiki/Image_classification), [medical image analysis](https://en.wikipedia.org/wiki/Medical_image_computing), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [brain-computer interfaces](https://en.wikipedia.org/wiki/Brain%E2%80%93computer_interface), and financial [time series](https://en.wikipedia.org/wiki/Time_series). CNNs are [regularized](https://en.wikipedia.org/wiki/Regularization_(mathematics)) versions of [multilayer perceptrons](https://en.wikipedia.org/wiki/Multilayer_perceptron). A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

#### **Purpose of Convolutional Neural Network** - Mainly to process and analyse digital images, with some success cases involving processing voice and natural language.

**Structure (with an example of vehicle classification) –**



**The CNN is a combination of two basic building blocks -**

1. **The Convolution Block**— Consists of the Convolution Layer and the Pooling Layer. This layer forms the essential component of Feature-Extraction
2. **The Fully Connected Block**— Consists of a fully connected simple neural network architecture. This layer performs the task of Classification based on the input from the convolutional block.

**There are 4 major operations in CNN image detection/classification -**

1. Convolution
2. Activation map
3. Max pooling
4. Flattening
5. Fully connected layer

**Difference between ANN and CNN-** The major difference between a traditional Artificial Neural Network (ANN) and CNN is that only the last layer of a CNN is fully connected whereas in ANN, each neuron is connected to every other neurons.

#### **Advantage of using CNN -**

Little dependence on pre processing, decreasing the needs of human effort developing its functionalities.

**Conclusion :** After doing this task on CNN, I understood how Convolutional Neural Network works, it’s block diagram, working & I also performed CNN application based task on Human detection in surveillance videos regarding the same.