CHAPTER-1

INTRODUCTION

1.1 Overview

Machine Vision is a study that aims at practical implementation of systems that allows machines to recognize objects from acquired image data and perform useful tasks from that recognition. Machine vision has now been established as a respected field of research and application with strains of sound theoretical foundation. There are various limitations of human mind i.e. fatigue, less accuracy, inability to work in life risking situation etc. These limitations results in reduced efficiency if the human mind is used in activities like machine vision. Thus, there occurred the need to make use of a machine that could imitate a human mind as closely as possible. Addressing a Machine Vision problem requires expert's knowledge in all the aspects of Image processing. Machine Vision is also known as computer or robot vision and image understanding but these terms vary only in the area of research. For instance, image understanding is studied by image and signal processing researchers. Computer vision hails from computer science and artificial intelligence studies, and robot vision comes from the field of robotics. The availability of high-powered computers, cheaper and more capable cameras, affordable processing power along with the realistic images have proved to be the catalyst for making Machine vision a reality. Computer vision as a field is an intellectual frontier. Like any frontier, it is exciting and disorganized. It is a rapidly growing field, which works by extracting descriptions of the world from pictures or sequences of pictures. It is being used in a variety of applications ranging from mobile robot applications to realistic rendering of synthetic scenes in computer graphics.

Since early 1970s object tracking has attracted extensive attention among researchers. Any vision system that monitors the real world process has to perform foreground segmentation. This is considered to be the foremost fundamental step in most of the applications like video conferencing. Once the foreground and the background images are separated, the background can be replaced by another image, which then beautifies the video and protects the user privacy. Object tracking is an important task within the field of computer vision whose goal is to consistently infer the motion of the desired targets, e.g. feature points, contours, regions of interest, and articulated objects etc., from image sequences captured by single or multiple cameras. In other words, Object tracking bridges the low level image processing techniques with the high level video content analysis. More precisely, it is called Visual Object tracking in the cases where the intended objects bears some semantic knowledge. Video Analysis is the major step to be performed to accomplish Object tracking. There are three key steps in video analysis:

- Detection of interesting moving objects:

The aim is to analyze the objects of interest, so that these objects are first separated from the overall image. It is done by a detection phenomenon and then foreground background separation is performed. This step generates a background and a separate foreground image.

- Tracking of such objects from frame to frame:

As soon as the objects of interests have been retrieved, their motion is then continuously tracked depending on the area of application like vehicles on road, planes in air, people walking on road etc. To perform tracking, objects are represented by their shapes and appearances. Shape could be chosen as points, primitive geometric shapes, Object silhouette and contour, articulated shape models or skeletal models. Some of commonly used object appearance representations could be probability densities of object appearance, templates, active appearance models and multi-view appearance models.

- Analysis of object tracks to recognize their behavior:

After detection and tracking has been done the behavior of the objects are then analyzed in order to perform desired operations when specific conditions are met.

1.2 Image Processing

Image Processing is one of the potent branches of Signal Processing. It is used in diverse application areas such as biomedical imaging, biometrics, multimedia computing, remote sensing, texture understanding, and content based image retrieval, compression and many more. In an image processing system, the input is an image, such as photographs or frames of video and the output can be either an image itself or a set of characteristics or parameters defined for the image. In an image formation the two most important components are illumination and the different reflectance models. The basic principle behind the formation of any digital image is sampling and quantization.

The main component of Image Processing is a Digital Camera that captures the images of a three dimensional object. The camera, video camera, and camcorder are used interchangeably. The sensor primarily used in this camera is either charge coupled device (CCD) or CMOS sensor depending on the applications. In a CCD camera there is very large number of small photo diodes called photosites. Each image cell has some accumulated electric charge which is transported and recorded with appropriate analog to digital conversion. In CMOS sensors there are a number of transistors that amplify the signal at each pixel location which is then read individually. It has a less light sensitivity as number of transistors used is more. CMOS sensors are less expensive and have lower power consumption however; they are comparatively noisier than CCD sensors. There are various parameters of a digital camera that determines their choice in a particular application:

i. Shutter Speed

When a camera is placed in a bright sunlight then the aperture which is located at the back of the camera lens shrinks in order to allow only a small amount light to enter whereas in cloudy days it expands to allow large amount of light. The aperture functions just like the human eye. The shutter speed decides the amount of time during which the light passes through the aperture. The shutter opening and closing takes place for the time period that is dependent on the requirement of the light.

ii. Focal Length

For a particular camera, distance between the focal plane of the lens and the surface of the sensor array is the measure of focal length. It is one of the critical parameter in selecting the amount of required magnification which is desired from the camera.

iii. Scanning order

To render a video there are two different techniques namely progressive scan and interlaced scan. Progressive scan differs from interlaced scan in that the image is displayed on a screen by scanning each line in a sequential order unlike the interlaced scan where scanning is done alternately. In other words, in progressive scan, the image lines are scanned in numerical order i.e. 1, 2, 3 down the screen from top to bottom, instead of in an alternate order i.e. lines or rows 1, 3, 5, etc. Followed by lines or rows 2, 4, 6.

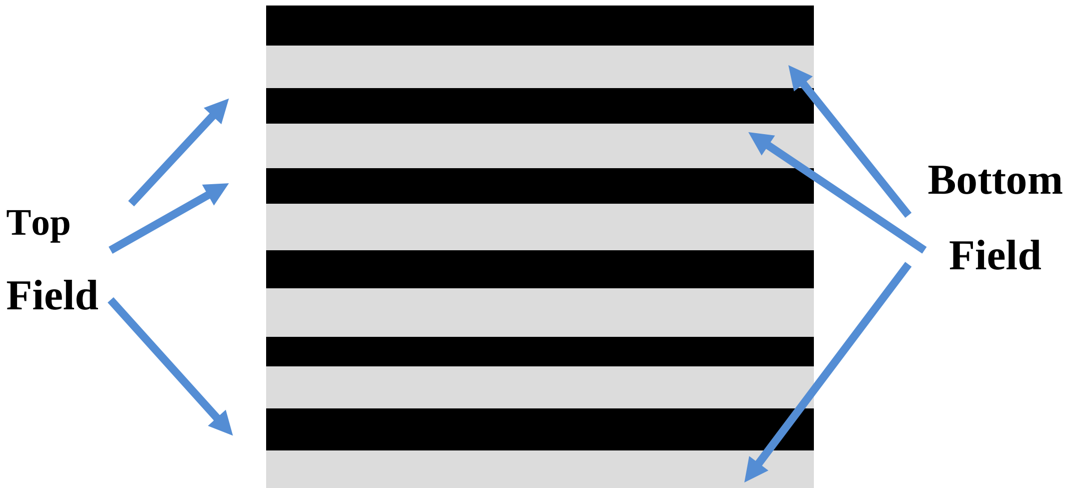


Fig1.1. Interlaced scan

By using progressive scan the image is displayed on the screen every 60^{th} of a second whereas it is displayed every 30^{th} second in interlaced scanning. Progressive scan produces a smoother, more detailed image screen that is perfectly suited for viewing fine details, such as text, and is also less susceptible to interlace flicker. The primary intent of progressive scan is to refresh the screen more often.

iv. Resolution

Spatial resolution of any image is the image size in pixels that correspond to the size of the CCD array in the camera. There are various resolutions with which any digital camera captures images varying from low (320x240, 352x288), to medium (640x480), to high resolution (1216x912, 1600x1200). The normal cameras we use have the capability of producing 16 million colors at each pixel. Sensor resolution is the smallest feature size of the objects present in a scene that our imaging system could easily distinguish. In other words, it is the measure of the number of rows and columns of the CCD array.

v. Field of View (FOV)

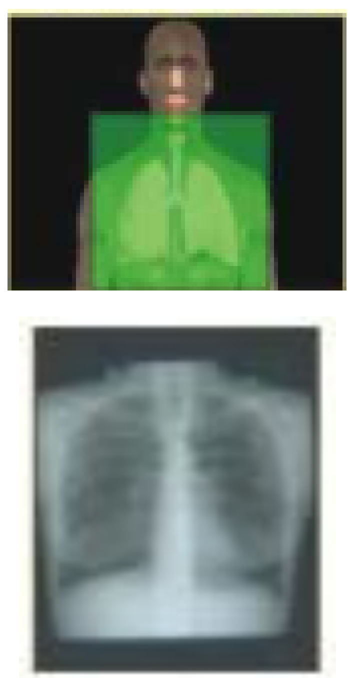
It can be defined as the area that the camera can capture. It is chosen as the horizontal dimension of the inspection region containing the objects of interest. The angle of view of a camera is the angle over which the sensor can view through the lens. The field of view is another way of representing the angle of view and is expressed as a measurement of the object area rather than an angle. The captured images could be zoomed which is performed by interpolating the pixels. It doesn't increase the information content of the image.

There are some other supplementary features supported in cameras like shutter control, focus control, exposure time control along with various triggering features.

1.2.1 Analog Image Processing

Generally, Image processing refers to Digital Image Processing but a continuous analogue to it is the optical and analog image processing which is its one sub-category. There are two types of images, analog and digital which are used in their specific applications like medical imaging, photography, displays etc. Analog images are for human perception. They may include things such as photographs, paintings, TV images, and all of our medical images recorded on film or displayed on various display devices like computer monitors CROs etc. as shown in Fig.1.2. In analog image we see various levels of brightness which in other word is the film density and colors. It is in a continuous form that means not broken into many small individual pieces. The optical image of an object is obtained by the light distribution coming from every point of the object at the image plane of an optical system. When all rays from an object point unite to a single image point then an ideal image is said to be formed. And, the process of acquisition of the image before beginning with its processing is also known as imaging.

Analog Image



For Human veiwing

Fig 1.2 An example of analog image

In advanced information technology Optical techniques are being increasingly applied due to its inherent speed and potentials for high-density information storage. The few of the active components that are used are optical fibers, photonic chips, flat panel displays, and optical fiber based active components optical variable devices etc. There are new developments that are going on to redefine the present standards of data processing and storage in areas such as optical computing, optical storage and integrated optics.

1.2.2 Digital Image Processing

Digital Image Processing commonly referred to as image processing is defined as a systematic process of manipulation of images using the computer algorithms to enhance, restore and understand the information contained in them. Digital Images can be produced by a variety of physical devices like still and video cameras, X-ray devices, electron microscopes, radar, and ultrasound. Image is a two-dimensional signal that consists of pixels. The pixels are derived by

Sensing the scene data using a camera or a scanner. Each pixel in image can be indexed as values of a two-dimensional function f(x ,y), where x and y are spatial coordinates, and the amplitude of any coordinates (x, y) is called the intensity or grey level of the image at that point. When x, y and the amplitude values of f ( ) are all finite, discrete quantities i.e. 0 or 1 for a black and white image then it becomes a digital image.

By convention, we take upper left corner as the origin for analysis. Sampling gives the total number of pixels in the picture and quantization determines the range of values that a pixel can attain. As already mentioned, the resolution of an image is the number of pixels in a given area. Unlike signal processing where sampling means samples per unit time in image processing it is defined as samples per unit distance. As a subcategory of the vast field of digital signal processing, Digital image processing (DIP) has many advantages over Analog image processing. DIP enables a much wider range of algorithms to be applied to the input data and can avoid the problems such as the build-up of noise and signal distortion during processing. Since images are defined over two or more dimensions digital image processing may be modeled in the form of multidimensional systems. A digital image processing allows the use of much more complex algorithms for image processing, and hence, can offer both more sophisticated performance at simple tasks, and the implementation of methods which would otherwise be impossible in implementation by analog means. In particular, digital image processing can be applied in the areas of:

i. **Classification**: In classification different predefined data classes are present and a large data set having similar and distinct elements are to be put distinctly in these distinct classes. Classification algorithms typically employ two phases of processing: *training* and *testing*. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, *i.e. training class*, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features.

The description of training classes is an extremely important component of the classification process. In supervised classification, *statistical* processes (*i.e.* based on an *a priori* knowledge of probability distribution functions) or *distribution-free* processes can be used to extract class descriptors. Unsupervised classification relies on *clustering* algorithms to automatically segment the training data into prototype classes. In either case, the motivating criteria for constructing training classes are that they are:

- *independent*, *i.e.* a change in the description of one training class should not change the value of another,
- *discriminatory*, *i.e.* different image features should have significantly different descriptions, and
- *reliable*, all image features within a training group should share the common definitive descriptions of that group.

ii. **Feature extraction**: It is a method of capturing visual content of images. To analyze any object for some specified applications, from the given image the features of the object is to be obtained. If the whole image is used for analysis the whole problem becomes more complex. The features retrieved should be such that :

- They should carry enough information about the image.
- They should be easy to compute.
- They should relate well with the human perceptual characteristics.

iii. **Pattern recognition**: Pattern recognition observes the patterns of features that repeat across different samples. For instance, an image of a ploughed field may have a stripe pattern whose feature can be determined by Fourier analysis. Pattern recognition pertains not to the single stripe pattern, but to the pattern of a number of different image areas having that same stripe pattern, by which they are classified together [1] .Pattern recognition is employed for various applications like automatic fruit ripening system, human activity detection, and gesture recognition by analyzing the patterns in the subsequent frames of the video.

iv. **Projection** : Subspace projection techniques are often employed as it offers the advantages of dimensionality reduction, less complexity and efficiency. They include:

- Principal Component analysis (PCA),
- Linear Discriminant analysis (LDA),
- Independent Component analysis (ICA) ,and
- Non-negative matrix factorization (NMF).

v. **Multi-scale signal analysis** : In several applications it is required to analyze an image at various resolutions in order to extract its very fine details. Wavelet Analysis is the basis of such applications. Wavelets Transforms are an important sub class, and have a variety of applications like face recognition, image compression, denoising images etc.Thus, Multi-scale signal analysis is widely applied in the fields of Computer vision, Face detection, Feature detection, Lane departure warning system, Non-photorealistic rendering, Medical image processing, Microscope image processing Morphological image processing, Remote sensing, etc.

1.2.3 Operations performed in DIP

Numerous operations varying from the very basic to high level which are prominently used in digital image processing few of which are mentioned as under:

(i) Pixelization

The word *pixel* is based on a contraction of *pix* ("pictures") and *el* (for "element"). Pixelation is a term sometimes used to describe the act of turning a printed image into a digitized image file (such as the GIF file that is used to display an image on a Web page). As the image is captured, it is processed into a vectorized or rasterized file that can be used to illuminate color units called pixels on a display surface

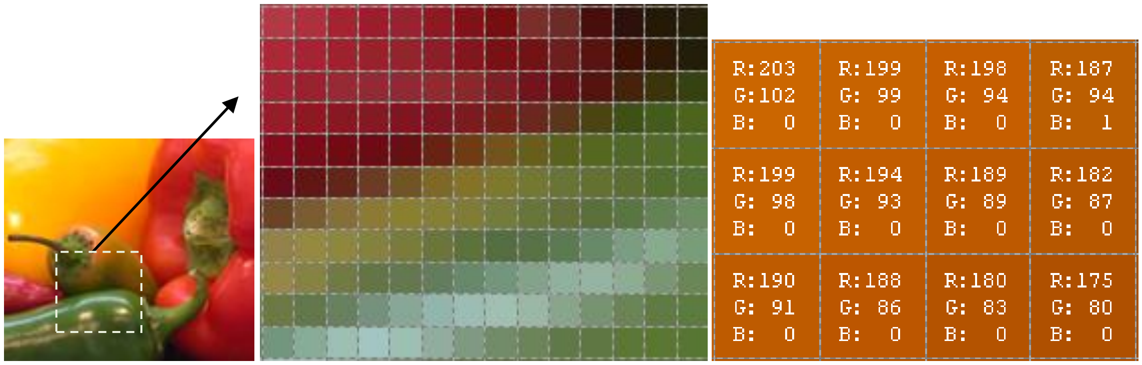


Fig 1.3 Illustration of an RGB image using 'imtool' of matlab

(ii) Linear Filtering

Linear filtering involves methods that modify the pixels based on its neighborhood. Such methods based on the local neighborhood are the simplest ones. In Linear filtering based methods the linear combination of the neighbors is being used for modification. It is used for:

- Integrating information over constant regions.
- Scaling.
- Detecting changes

(iii) Principal components analysis

Principal component analysis (PCA) is a mathematical procedure that transforms a large number of correlated variables into a comparatively smaller number of uncorrelated variables called principal components [2]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Majorly PCA is applied to:

- Discover or to reduce the dimensionality of the available data set.
- Identify new meaningful underlying variables.
- Avoid the problem of curse of dimensionality.

(iv) Independent component analysis

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed Non- Gaussian and mutually independent and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA.

(v) Hidden Markov models

The Hidden Markov Model (HMM) is a powerful statistical tool for modeling generative sequences that can be characterized by an underlying process generating an observable sequence [3]. HMMs have found application in many areas of interest in signal processing, and in particular speech processing, but have also been applied with success to low level NLP (Natural Language Processing) tasks such as part-of-speech tagging, phrase chunking, and extracting target information from documents

(vi) Partial differential equations

Partial differential equations (PDEs) have led to an entire new field in image processing and computer vision. PDE-based methods are one of the mathematically best-founded techniques in image processing. PDE-based image processing techniques are mainly used for smoothing and restoration purposes. Many evolution equations for restoring images can be derived as gradient descent methods for minimizing a suitable energy functional, and the restored image is given by the steady-state of this process.

(vii) Self-organizing maps

Self-organizing maps (SOMs) are a data visualization technique invented by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing neural networks. The Self-Organizing Map is one of the most popular neural network models. It belongs to the category of competitive learning networks. The Self-Organizing Map is based on unsupervised learning, which means that no human intervention is needed during the learning and those little needs to be known about the characteristics of the input data. We could, for instance use the SOM for clustering data without knowing the class memberships of the input data.

The SOM can be used to detect features inherent to the problem and thus has also been called SOFM, the Self-Organizing Feature Map. The problem that data visualization attempts to solve is that since humans simply cannot visualize high dimensional data, so innovative techniques need to be developed to help us understand this high dimensional data. The way SOMs go about reducing dimensions is by producing a map of usually 1 or 2 dimensions which plot the similarities of the data by grouping similar data items together. So SOMs accomplish two things, they reduce dimensions and display similarities.

(viii) Neural networks

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The problem is that there is no way of knowing if the system is faulty or not, unless an error occurs.

(ix) Wavelets

Wavelets are used in a wide range of applications such as signal analysis, signal compression, finite element methods, differential equations, and integral equations. Wavelets can be categorized into discrete (DWT) and continuous (CWT) wavelet transforms.

1.3 Video Processing

Video is a data file that contains the sequence of images with respect to time and is obtained from a recording device using a universal standard bus i.e USB cable attachment. The files are then loaded into a computer software program or peripheral device. The frame rate is required to be at least 15 frames per second (fps). For the smoother appearance 30 fps is necessary. Video Processing is one of the applications of Image Processing. In video processing discrete images are subsequently treated and analyzed. Video processing uses hardware, software, and combinations of the two for editing the images and sound recorded in video files. Extensive algorithms in the processing software and the peripheral equipment allow the user to perform editing functions using various filters. The desired effects can be produced by editing frame by frame or in larger batches.

These days video processing has become a very important phenomenon and are being widely used either in television systems, video post production or even in common life. Despite the fact that there is availability of professional hardware video processing solutions but software video processing is still very popular mainly because of the great flexibility it offers. Most modern personal computers come with software that allows the user to compile images and videos, edit images, and create videos on a limited level. Storyboards allow the addition of audio files and the adjustment of visual images, transitions, and audio files, which, together, determine the overall length of the video. Videographers, Electrical Engineers, and Computer Science professionals use programs that are capable of a wider range of functions. Signal processing usually involves applying a combination of pre-filters, intra-filters, and post filters. Video processing is practically used in many applications like lane tracking and departure warning systems, surveillance systems for security, keeping track of the targets and many more.

1.4 Research Aims and Objectives

Appearance modeling is very important for background modeling and object tracking. Subspace learning based algorithms have been used to model the appearances of objects or scenes. Current vector subspace-based algorithms cannot effectively represent spatial correlations between pixel values. Current tensor subspace-based algorithms construct an offline representation of image ensembles, and current online tensor subspace learning algorithms cannot be applied to background modeling and object tracking.

The goal of our algorithm i.e. 3D-ITSL (3-Dimensional Incremental Tensor Subspace Learning) is efficient foreground segmentation and robust tracking in the complex scenes containing dynamic changes like noisy images, illumination variations , occlusions, random background motion etc. It models the appearances of the targets of the scenes which are a prerequisite for any background modeling and object tracking task. The appearance models are constructed by incrementally learning a tensor subspace representation. A three dimensional tensor is chosen for the analysis as it efficiently captures the intrinsic spatio-color-temporal characteristics (SCT) of the scenes. The method is online and the appearance model keeps on updating or modifying as the new information in the scenes comes in like a new object enters, previous object leaves, trees waving, sun rise, sun-set. The proposed algorithm provides us the correct and the most recent background models which is very important to achieve foreground segmentation and object tracking.

Singular Value Decomposition (SVD) of the whole Tensor data which is in the form of Unfolding Matrices is performed to obtain the basis vectors that create the tensor subspace and the incremental learning of the tensor begins by adaptively updating the sample mean and the eigen basis for each unfolding matrix. As soon as the foreground segmentation has been performed the next task lies in tracking of the objects. The tracking method should be such that it include elements of nonlinearity and non-Gaussianity in order to model accurately the underlying dynamics of a physical system. Moreover, it is typically crucial to process data on-line as it arrives, both from the point of view of storage costs as well as for rapid adaptation to changing signal characteristics Thus, keeping these requirements under consideration Particle filter is used for tracking which is performed by maintaining the object's state continuously.

Particle filters are sequential Monte Carlo methods based on point mass (or "particle") representations of probability densities, which can be applied to any state-space model and which generalize the traditional Kalman filtering methods. Several variants of the particle filter such as SIR, ASIR, and RPF are introduced within a generic framework of the sequential importance sampling (SIS) algorithm. The new tracking algorithm captures the appearance characteristics of an object during tracking and uses a particle filter to estimate the optimal object state.

1.4.1 Strengths

In this work, we have developed an incremental tensor subspace learning algorithm based on subspace analysis within a multi-linear framework. The appearance of an object or a scene and the changes in appearance over time are modeled by incrementally learning a low dimensional tensor subspace representation which is updated incrementally as new images arrive. We have applied the proposed incremental tensor subspace learning algorithm to foreground segmentation and object tracking. Our foreground segmentation algorithms for grayscale image sequences or color image sequences capture the intrinsic spatiotemporal characteristics of scenes based on a likelihood function which is constructed on the basis of the learned tensor subspace model. Our tracking algorithm captures the appearance characteristics of an object during tracking, and uses particle filtering to propagate the sample distributions over time.

Experimental results show that our proposed algorithms for foreground segmentation and object tracking are robust to noise or low quality images, occlusions, lighting changes, scene blurring, objects with small apparent sizes, and object pose variations. Consequently, 3D-ITSL performs effectively in modeling appearance changes of objects or scenes in complex scenarios.

- It efficiently captures the SCT characteristics of the scenes and robustly handles the dynamic variations.
- The method operates in an online fashion and undergoes continuous updation in an incremental manner, and hence real time.
- It gives a better quality outputs as compared to the other Tensor based algorithm [6]
- The algorithm can work on the image scenes containing multiple objects and can handle even occlusions. Thus, multiple object detection, segmentation as well as multiple object tracking could also be achieved.
- The tracking system is robust.

1.4.2 Weaknesses

- The proposed method is not capable of handling Shadows. Thus, shadow effects in the example videos are not considered in the proposed work.
- Some optimization techniques could be applied to tracker, to generate better results. These optimization techniques could be PSO (Particle Swarm Optimization),BBO(*Biogeography-based optimization)* ,ACO (Ant colony optimization) could further improve the efficiency of the algorithm.

1.5 Distinct Nature of the proposed method

Our Algorithm possesses a number of distinct characteristics as compared to previously developed vector and tensor based algorithms like Vector Subspace leraning [4], weighted incremental tensor subspace learning [5] , Tensor subspace learning, Incremental Rank (R_1,R_2,R_3) tensor subspace learning [6]. Till now the existing methods for foreground segmentation and tracking methods based on higher order subspace are of atmost third order that contains spatial information of the scenes in the form of their corresponding gray values and other is the temporal information. The higher order tensor analysis was used in order to capture most of the information as accurately as possible. Our proposed algorithm enhances the distinct characteristics by employing a higher-order tensor representation, which is a novel effort in its own way. A three tensor is being used that deals with the changes in the color appearances. The image is considered to be 3-D tensor, with color component as the third axis and the fourth order decomposition is performed to deal with the changes in the color appearances with time.. The incremental subspace learning for 4-order tensors is also developed for the applications of background modeling and object tracking. However, the complexity of the proposed method is somewhat more than that of the previously developed methods.

1.6 Organization of the Thesis

In this study we have firstly introduced domain of the work.

Chapter 1 deals with an overview of the area of research and its other variants. The distinct nature of the proposed algorithm is discussed with its corresponding strengths and weaknesses.

Chapter 2 discusses the related work, firstly in the area of computer vision and then in the field of computer vision. The comparison of our proposed algorithm with the latest techniques is also discussed in this chapter.

Chapter 3 familiarizes reader with the area of foreground segmentation from the very basic definition towards its various approaches. It touches upon the tensor based foreground segmentation, which is our main area of work.

Chapter 4 throws light on the field of object tracking, again starting from the very basic definition and moving towards its various widely used methods. Occlusion concept is briefly discussed at the end.

Chapter 5 deals with the notion of Multilinear Algebra i.e. algebra of tensors. The software tool used for the implementing the algorithms is also briefly discussed.

Chapter 6 concludes the thesis with the results and discussions of our work. Future scope is also highlighted.