

**Complex Systems**

**(6CS014)**

**Title: Object Recognition and Tracking**

|  |  |
| --- | --- |
| Student Id | : 2049822 |
| Student Name | : Sashank Dulal |
| Supervisor | : Mr. Rupak Koirala |
|  |  |
|  |  |
| Submitted on | : Wednesday, September 16,2021 |

**Acknowledgements**

**Table of Contents**

[Introduction 4](#_Toc89096103)

[Aims and Objectives 5](#_Toc89096104)

[Aims 5](#_Toc89096105)

[Objectives 6](#_Toc89096106)

[Literature Review 7](#_Toc89096107)

[1. Detection, Recognition and Tracking of Moving Objects from Real-time Video via SP Theory of Intelligence and Species Inspired PSO 7](#_Toc89096108)

[2. Object Recognition and Tracking for Remote Video Surveillance 7](#_Toc89096109)

[3. Multiple Object Tracking Using SIFT Features and Location Matching 10](#_Toc89096110)

[4. Multi-Object Recognition and Tracking with Automated Image Annotation for Big Data Based Video Surveillance 11](#_Toc89096111)

[5. Real-time Moving Object Recognition and Tracking Using Computation Offloading 12](#_Toc89096112)

[6. A probabilistic integrated object recognition and tracking framework 13](#_Toc89096113)

[Analysis of literature review 14](#_Toc89096114)

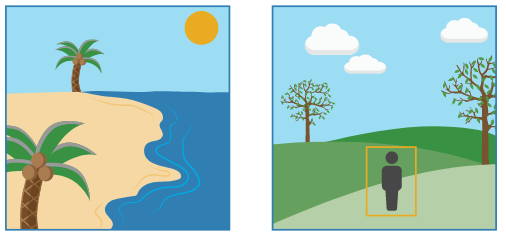
[Conclusion 14](#_Toc89096115)

**Table of Figures**

[Figure 1: Architecture of the proposed system 9](#_Toc89096232)

# Introduction

Object detection is a technique used in computer vision to identify the objects within an image or video. The main objective of this method is to be able to replicate the recognition abilities of humans with the use of various processing techniques and methods (mathworks, Inc, 2021). The technologies “object detection” and “object recognition” are similar in terms of their objectives but the major difference is that the object is identified as well as located in the image using the object detection technique (mathworks, Inc, 2021).



Similarly, object tracking is another image processing technique that involves detection of object, creation of unique identifier for each object in the video and tracking of the object as they are moving around different parts of a frame in a video (viso.ai, 2021).

# Aims and Objectives

## Aims

* Gain an in-depth understanding of object detection and tracking.
* Research on various techniques available for detection and tracking of objects.

## Objectives

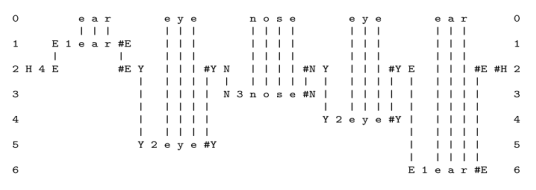
* Research on similar systems and published literatures.
* Analyze their qualities and drawbacks.
* Research on the complexity of the techniques available.

# Literature Review

## Detection, Recognition and Tracking of Moving Objects from Real-time Video via SP Theory of Intelligence and Species Inspired PSO

This paper mainly focuses on analyzing the SP Theory of Intelligence and address the challenges in tracking moving objects. In this paper, the authors have extracted multiple alignments, parsed raw data to gather noise-free alignment and reach best-possible solution with the user of family resemblance concept in order to efficiently recognize the object. Then the recognized object is tracked using Particle Swarm Optimization (PSO) inspired by species.

In this paper, the authors have proposed using SP theory of intelligence to detect objects which is mainly efficient in NLP due to its ability to process noisy and erroneous data and after the detection of objects have been done, species-inspired Particle Swarm Optimization (PSO) was to be used to track the objects. The use of SP system to detect the object can be demonstrated by the image below in which depicts how a person’s face with all the features are parsed.



Then the knowledge of the structure of features (Nose, Ears etc) are aligned with every feature on the object we are tracking. Then the achieved alignment is checked with the pattern in row 2 (i.e. comparing for head). Because of this process, much of the errors due to noisy data is eliminated.

Now for the tracking of objects, a PSO framework is used where

(Ray, et al., 2017)

## Object Recognition and Tracking for Remote Video Surveillance

This paper propose an object recognition and tracking system for surveillance in real-time. To meet the real-time requirement, the authors have come up with the idea of using statistical morphological skeleton to achieve accuracy in localization, low complexity in terms of computation and more strength against noise. Recognition of object is achieved by the comparison of the approximation of the skeleton function obtained by the analysis of image against the one got from the model stored in the database and the tracking task is achieved by using “Kalman filter” to the set of quantities observable gathered from the skeleton and with the help of other characteristics of the object that is moving in the scene.

In the proposed system, a camera is used to gather video of the surveillance scene. Then an algorithm is applied to every frame of the video which gives out a binary image where each possible object is represented with a blob moving inside the scene. The absolute difference between the background image and current image is calculated at a fixed time interval. The formula to calculate the difference is given below:



Where D(x,y) is the difference, I(x,y) is the current image and BCK(x,y) is the background image. The resulting image is N \* M in size. Then a hysteresis function having 2 thresholds, THRin and THRout with 2 states (i.e. object and background) is applied to the difference image to decide whether a point is either moving object or a background.

With this method, with every calculation, a binary image with changing pixels denoted as 1 and background pixels as 0 is generated. Then the Kalman filter gets applied to every pixel of the image to get the predicted estimated background. The filter is denoted by the equations below:



Where S(i,p) is the level of grayness of the background, S(i+1,p) is the same quantity of grayness in the next frame and µ(i) is an estimated system error. Similarly, l(i,p) is the level of gray in current image point p and η(i,p) represents the estimate of noise in the input image.

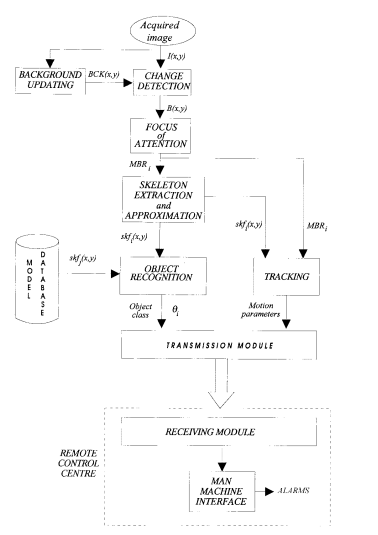


Figure : Architecture of the proposed system

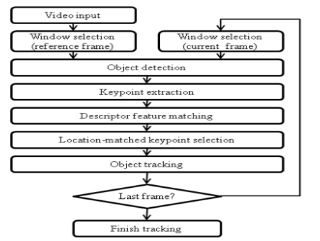
To be continued

(Foresti, 1999)

## Multiple Object Tracking Using SIFT Features and Location Matching

This paper presents a system that is able to rack objects with stable results by removing location-mismatched points. It uses Shift Invariant Transform (SIFT) algorithm which generates local features that are immune to changes in the scale of image, noise, lightings and distortion. In D.G. Lowe’s (who is the author of the SIFT algorithm) proposed paper, the matches having distance ratio greater than 0.8 were rejected so the authors have proposed a method to enhance the efficiency of the algorithm by recognizing location-mismatched points and preventing their participation in the matching of keypoints. Then tracking is applied to the video which is used to determine the position and size of the window.

The proposed algorithm in this paper uses location-matched keypoints from the keypoints computed from SIFT algorithm, the tracking is achieved with a rectangle window surrounding the object on the reference frame and the current frame and location-mismatched keypoints are avoided. Firstly, rectangle window surrounds object in reference frame, keypoints are calculated through SIFT algorithm and the features get stored. Then, keypoints are genereated in the next frame using same methods, matching keypoints are selected based on the distance ratio and the matched keypoints are selected as the candidate keypoints for tracking. Then the location-matching keypoints are determined which is used for stable tracking of the objects. The overall flow of the algorithm is shown in the image below.



The reference frame and the current frame are used to generate keypoints through SIFT algorithm which is as follows:



With this technique, the resulting data are “im” which consists of pixel values of the test image, “des” is a matrix of descriptor vectors, and “loc” consists of location, scale and orientation values for every keypoints. The matching of distance in SIFT is mostly done with the help of “des” data. Since all of the keypoints are classified into two parts, the object and the background and the background could change while the object is moving, the keypoint extraction must consider the fact. The decision is taken with the help of formula below:



Remaining

(Ha & Moon, 2011)

## Multi-Object Recognition and Tracking with Automated Image Annotation for Big Data Based Video Surveillance

In this paper, an improved region based scalable convolution neural network (IRS-CNN) based multiple object tracking model has been discussed. It uses Automatic Image Annotation (AIA) in order to improve the detection capacity and reduce computation time.

In the IRS-CNN method, the anomaly detection takes place in three stages namely AIA model to annotate images, RPN(Region Proposal Networks) and R-CNN. During the first stage, the image annotations are generated. Then a CNN creates regions in the image and finally at the last stage, the regions created in previous stage is used to detect the anomalies in the image. The IRS-CNN model is significant in size and consists of smaller sized sub-networks to detect anomalies. Similarly, the scaling awareness with AIA enables the IRS-CNN model to detect anomalies efficiently no matter the size and distribution.

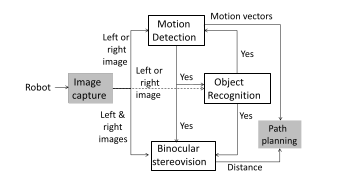
Furthermore, CNN with WARP (Weighted Approximated Ranking) method was used to improve the time complexity of the annotation process. Five convolutional layers with 3 connected layers was used from CNN model. Similarly, the Region Proposal Networks (RPN) was made in such a way that it inputs images with varying dimensions and outputs groups of rectangular objects as proposals with their objectless values. In order to facilitate faster resource sharing with R-CNN, each network was made to share similar convolution layers and for region proposal, a smaller network gets an input of n\*n spatial window where the sliding windows get mapped to lesser dimensional feature set. To train the RPN, a binary label (normal or abnormal) for each of the anchors (rank of every region box) was assigned.

Hence the paper concluded that the proposed IRS-CNN based MOT system was better than the existing RS-CNN method with the use of AIA to increase efficiency in detection and decrease computation complexity.

(Vijiyakumar, et al., 2021)

## Real-time Moving Object Recognition and Tracking Using Computation Offloading

This paper introduces computation offloading technique to overcome the limitations of hardware required for real-time movement tracking and object recognition on robots. Through this technique, the computation is migrate to the servers such that the computation time decreases significantly. But the speed of the process mainly depends on the size of the data being transmitted and network bandwidth available so the decision on whether to perform such complex calculations on the locally available processor or to process it through a server is based on the computational requirement and based on the requirement, the selection of resource is done by the robot.



The proposed system mainly contains five different modules which is depicted in the figure above. The gray colored modules are done by the robot itself and the other white colored modules may either be executed on the robot or on the server depending on the scenario discussed above. First of all, the image is captured by the robot in regular time intervals. As many changes may occur during capture of images such as change in background, color and objects. If the robot finds it complex to track and recognize objects at any instance, then the whole recognition and tracking is offloaded to the server and if the calculated complexity is low, then the robot performs those tasks on its own processors.

The system built by the authors comprised of a mobile robot (pioneer 3DX ) with Intel 2GHz processor as an onboard processor, was equipped with two fixed cameras whereas for a remote server, an Intel Xeon eight core 2.33 GHz processor running Linux server was used. The robot’s camera captured at the rate of 20FPS and the images were 640\*480 resolution. The visual cortex model was used for object recognition which involves extraction of features and comparing with the database. Then the amount of computation required is based on the number of compared features and the database size.

(Nimmagadda, et al., 2010)

## A probabilistic integrated object recognition and tracking framework

In this paper, a probabilistic integrated object recognition and tracking system PIORT has been introduced. The system composes of a static module for recognition that outputs probabilities for every pixel of the image from a group of local features and the probabilities update dynamically which are used by a module for tracking decision that is capable of handling occlusions (full and partial). Similarly, two methods are derived from the PIORT framework which uses color features such as RGB properties and one of them use Bayesian technique depending on the maximum possibility and another one used neural networks.

The proposed framework comprises of three different modules in which the first one is responsible for the recognition of the current frame (static recognition) which outputs probability images which indicate probabilities of each pixel belonging either to the object of interest or the background. Then in a second module, dynamic recognition is done where the result of first step are utilized to update a second probability images set. After the dynamic recognition has been done, the tracking decision module determines binary images for each of the objects gathered from the dynamic recognition probabilities in the previous stage which provides the predicted motion of the object in terms of changes in scale and translation of the object.

The experimental data have shown that the neural network based approach performs similar to the Bayesian approach and sometimes outperforms it. Similarly, the proposed PIORT methods have performed far much better compared to other tracking methods published that work on video taken by a moving camera and contains occlusions.

Remaining

(Serratosa, et al., 2012)

# Analysis of literature review

With the in-depth study of above literatures, it can be said that most of the studies are mainly focused towards algorithmic approach to solve the object recognition and tracking but one paper presented with a unique aspect of computational offloading. This technique can be useful in many scenarios involving limited resources for computation such as self-driving vehicles, delivery drones and mobile robots. Similarly, only two papers focused on the real-time aspect which means they could be used in situations that require instant recognition and tracking such as surveillance and so on. On the other hand, the sixth paper proposed a completely different probabilistic approach for recognition and tracking of objects along with promising results on moving videos also. Furthermore, two of the papers introduced approaches for tracking multiple objects with the help of feature extraction, location matching, CNN and AIA which can help tackle many problems such as mass surveillance, traffic management, contact tracing and so on.

# Conclusion