



INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student : Mathew Francis
Roll Number : 146102001
Programme of Study : Ph.D.

Thesis Title: Visual Object Tracking in Dynamic Scenes

Name of Thesis Supervisor(s) : Dr. Prithwijit Guha

Thesis Submitted to the Department/ Center : Electronics & Electrical Engineering

Date of completion of Thesis Viva-Voce Exam : 30-September-2023

Key words for description of Thesis Work : Tracking, Localization, Object Model, Siamese, Motion

SHORT ABSTRACT

Visual object tracking is one of the popular and legacy problems in computer vision. Tracking has applications in a wide spectrum of domains with applications in agriculture, automobile, surveillance, defence, and entertainment. It remains a challenging research problem on account of various factors like occlusions, shape deformations, illumination changes and background clutter. The object tracking performance depends on the joint efficiencies of a target object modeling scheme and a localization strategy. Accordingly, this thesis proposes three major contributions. First, a colour feature based representation that models the foreground while considering the background context. The target is further localized by using a meta-heuristic search strategy based on breeding fireflies. The breeding fireflies technique is realized through a combination of Real Coded Genetic Algorithm (RGA) and the Firefly Algorithm (FA). The second contribution involves target representation in the sparse representation framework. Here, the object is modeled using a (weighted) distribution of sparse codes. The weights are derived from a foreground-background classifier. Here, the target is localized by using the Firefly-RGA approach. The third contribution uses target representation using deep network embeddings. Here, a Siamese network is used to derive the object location predictions. Additionally, a multi-part object model is explored for handling occlusions. All three proposals use the appearance features to predict the object positions. Thus, complementary information is derived by inter-frame dense motion estimation. The motion based predictions are used to enhance the accuracy of the three proposed appearance based trackers. Thus, all the proposed trackers are realized in an ensemble framework of appearance and motion based predictions. The colour feature based tracker has an object model based on handcrafted features derived from colour distribution. The sparse representation-based tracker learns the model by using features from only the first frame of the sequence. In the Siamese tracker, a pre-trained deep network is used for extracting image features. The performances of the classical methods are limited by the feature used. For example, colour distribution-based features get disturbed by illumination changes and structure-based features face challenges from geometric transformations. In contrast, the deep network used in the Siamese tracker is trained on a very large set of general images (for image classification task) and is thus capable of generating rich feature representations. It is also observed that the Siamese tracker has significantly outperformed the other two trackers. The proposed trackers are benchmarked on the VOT2018, OTB2015 and UAV123 datasets and are compared against a number of baseline methods formulated in traditional and deep learning framework.