Computer Vision and Image Understanding 114 (2010) 297–310

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Online adaptive radial basis function networks for robust object tracking

R. Venkatesh Babu [a,*], S. Suresh [b], Anamitra Makur [c]

[a] Exawind, Bangalore, India
[b] Department of Electrical Engineering, Indian Institute of Technology, Delhi, India
[c] School of Electrical and Electronics Engineering, NTU, Singapore

ARTICLE INFO

Article history:
Received 31 July 2007
Accepted 23 October 2009
Available online 29 October 2009

Keywords:
Extreme learning machine
Face tracking
Non-rigid object tracking
Online adaptive object modeling
RBF networks

ABSTRACT

Visual tracking has been a challenging problem in computer vision over the decades. The applications of visual tracking are far-reaching, ranging from surveillance and monitoring to smart rooms. In this paper, we present a novel online adaptive object tracker based on fast learning radial basis function (RBF) networks. Pixel based color features are used for developing the target/object model. Here, two separate RBF networks are used, one of which is trained to maximize the classification accuracy of object pixels, while the other is trained for non-object pixels. The target is modeled using the posterior probability of object and non-object classes. Object localization is achieved by iteratively seeking the mode of the posterior probability of the pixels in each of the subsequent frames. An adaptive learning procedure is presented to update the object model in order to tackle object appearance and illumination changes. The superior performance of the proposed tracker is illustrated with many complex video sequences, as compared against the popular color-based mean-shift tracker. The proposed tracker is suitable for real-time object tracking due to its low computational complexity.

1. Introduction

Visual tracking of object in complex environments is currently one of the most challenging and intensely studied tasks in machine vision field. The objective of object tracking is to faithfully locate the targets all through the successive video frames. In recent years, considerable effort has been made towards real-time visual tracking, especially in adverse conditions such as occlusion, background clutter, appearance, and illumination changes of the object of interest [1]. Most of the existing tracking algorithms can be broadly classified into the following four categories.

(1) *Gradient-based methods* locate target objects in the subsequent frames by minimizing a cost function [2,3].
(2) *Feature-based approaches* use features extracted from image attributes such as intensity, color, edges, and contours for tracking target objects [4–6].
(3) *Knowledge-based tracking algorithms* use *a priori* knowledge of target objects such as shape, object skeleton, skin color models, and silhouette [7–10].
(4) *Learning-based approaches* use pattern recognition algorithms to learn the target objects in order to search them in an image sequence [11–14].

Recently, Mean-Shift Tracking (MST) [6], a feature-based approach that primarily uses color-based object representation, has attracted much attention due to its low computational complexity and robustness to appearance change. All the extensions of MST algorithms assume that the histogram of the tracked object does not change much during the course of tracking. However, they all suffer from fundamental problems that arise due to complexity of object dynamics, change in camera viewpoint and lighting conditions. These adverse situations call for online adaptation of the target model. MST uses global color histogram as object model, for which there exists no principled way of updating the model, to tackle the object dynamics. Various solutions to this problem have been proposed. Online feature selection algorithms with weighted color-based features have been presented in [15,16]. Recently, in [17], an appearance generative mixture model based MS tracker has been presented. Here, static histogram is updated online using expectation maximization technique. However, the limitation of the algorithm is that it assumes that the key appearances of the object can be acquired before tracking, though the fact is that in real-time tracking, collecting key appearances is a difficult task. Hence, it is essential to consider an object model which evolves over time in order to effectively capture the object dynamics.

Learning-based approaches allow highly complex, non-linear modeling, with scope for dynamic updation. This framework has been successfully exploited in a number of applications such as pattern recognition [18], remote sensing [19], dynamic modeling

* Corresponding author.
E-mail addresses: venkatesh.babu@gmail.com (R. Venkatesh Babu), suresh99@gmail.cpm (S. Suresh), eamakur@ntu.edu.sg (A. Makur).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.10.004

and control and medicine [20]. The increasing popularity of neural networks in many fields is mainly due to their ability to learn the complex non-linear mapping between the input-output data and generalize them. Complex decision boundaries are realizable through this framework [21]. Also, neural networks carry the advantage of needing no prior assumptions about the input data.

In neural network learning algorithms [22,23], iterative search methodology has been widely used for network parameters update. Hence, the learning process is computationally intensive and may require several hours to train the network. Also, one has to select proper learning parameters to avoid sub-optimal solutions due to local minima. Extensive training duration and issues in selection of learning parameters lead to the development of an alternative algorithm which can be implemented in real-time. Recently a fast learning neural algorithm called 'extreme learning machine' (ELM) was presented for Single hidden Layer Feed-forward Network (SLFN) [24]. SLFN with sigmoidal or radial basis activation functions are found to be effective for solving a number of real world problems. In fast learning algorithms [24], it is shown that for SLFN with radial basis activation function, random parameter selection (mean and variance) and analytically calculated output weights can approximate any continuous function to desired accuracy [25]. Here, the output weights are analytically calculated using Moore–Penrose generalized pseudo-inverse [26]. This algorithm overcomes many issues in traditional gradient algorithms such as stopping criterion, learning rate, number of epochs and local minima. Due to its shorter training time and generalization ability, it is suitable for real-time applications. The performance of the fast learning algorithm has been found to be better on various real world problems, as against the other neural network approaches [24].

Learning-based tracking algorithms were rarely used for general purpose object tracking. Typical learning-based object trackers are designed to track specific objects, which require off-line learning phase [27–29]. This is due to the difficulty in adapting the neural networks for tracking purpose. Adapting a tracking problem into a classification problem gives a wider scope for modeling the objects using neural networks [30]. In some existing approaches, the context of application needs to be fixed beforehand [12,13]. The context enables the user to train the classifier with as many images, in order to distinguish the relevant object from the others. In [13], a model that learns the relationship between the local motion of an object and its corresponding appearance in the image, is developed for the purpose of tracking. A severe disadvantage of these approaches is that the context is clamped and cannot be flexed. Besides, the performance of the technique critically depends on the number of training instances utilized. The requirement of enormous labeled training examples in the process of modeling leads to the disadvantage of the technique being laborious and context-specific. Recently in [14], ensemble of weak classifiers are trained to distinguish between the object and background. Here, the tracking problem is handled as a binary classification and ensemble of N weak classifiers are used to develop the confidence map for tracking. In the subsequent frames, weights of K best weak classifiers are updated and N–K new weak classifiers are added. The weights of all N weak classifiers are updated such that the new set of weak classifiers form strong classifier. The process of adding/deleting the weak classifiers and updating their ensemble weights increase the computational complexity.

In this paper, we present a robust online adaptive tracker using object/non-object classifiers [30]. The basic building block of the classifiers are radial basis function network. Here, the center and width of the radial basis function network are selected randomly and output weights are calculated analytically using least square algorithm [31]. Posterior probability of the object pixels [32] is used as confidence map and their weighted average is used to find the object location in the subsequent frames. In the subsequent frames, the classifiers are adapted to handle the change in object dynamics. For online adaptation, only fewer pixels are used to update the output weights using recursive least square algorithm. The proposed scheme is computationally less intensive and effective under varying object dynamics.

The paper is organized as follows: Section 2 describes the overview of the proposed object tracker. Section 3 presents the details of main modules of RBF networks based object tracker. Experimental results and discussions are presented in Section 5. Finally, Section 6 concludes the paper.

2. Overview of online adaptive neural tracking system

In this paper, we present an online adaptive object tracking algorithm using radial basis function networks. The major components of object tracking algorithm are object model development, object localization and online model adaptation. The schematic diagram for the proposed online adaptive neural tracking system is shown in Fig. 1. In object tracking, first one needs to develop a model for the object of interest (target) from the given initial video frame. Next, the object localizer estimates the target location in the subsequent frames using the object model. Also, the object model is adapted online to accommodate the changes in the target model due to the object dynamics.

Fig. 1a illustrates the object model development using two RBF networks. Initially the object of interest is localized by the user input by drawing a rectangle around the object of interest. The object–background separation module separates the object from the surrounding background pixels by estimating the likelihood map. Using the estimated likelihood values, pixels are classified as either object or non-object pixels. The feature extraction module, extracts features like color and location information of the labeled object and non-object pixels. The functional relation between the extracted features (U) and the class labels (C) is estimated using the real-time learning radial basis function networks. The objective of the object model development phase is to accurately identify the object from the background. Hence, two RBF classifiers are used for this purpose. The object (N_o) and non-object (N_b) classifiers are tuned to maximize classification accuracy for object and background pixels correspondingly. Only the reliable object pixels, which are classified as object in both classifiers, are used as the target/object model. Here, the posterior probability of the pixels that belong to the object class represents the object model.

The object tracking phase of the proposed algorithm is shown in Fig. 1b. The object localization starts at the center of the object window in the frame where it was previously tracked. In order to find the object pixels, the features are extracted from this location and are tested with both object and non-object classifiers. The displacement of the object (D_1) is given by the shift in centroid of the object pixels. The object location is iteratively shifted and tested until convergence. The cumulative displacement indicates the shift in object location for the current frame.

The third component in the proposed tracking algorithm is the model adaptation phase. In this phase, the parameters of object and non-object classifiers are adapted based on the features extracted from the most recent frame. This phase compensates for the changes in non-object and object pixels as they evolve with time. The proposed tracking algorithm uses online learning algorithm to update the object/non-object classifiers. Similar to model development component, in this phase, spirally sampled pixels