Optik 126 (2015) 1449–1456

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Robust visual tracking based on joint multi-feature histogram by integrating particle filter and mean shift

Jianfang Dou*, Jianxun Li

Department of Automation, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 3 March 2014
Accepted 14 April 2015

Keywords:
Visual tracking
Particle filter
Mean shift
Scale invariant feature transform
Template matching

ABSTRACT

In this paper, we proposed a robust visual tracking and simultaneous recognition method in a particle filter frame work. First, an appearance model based on Multi-Part Joint Color Texture and Edge of Orientation Histogram with corrected background weighted histogram is Build, which is adopted to calculate the likelihood of particles; then, perform scale invariant feature transform (SIFT) keypoint detection and matching to predict the coarse position of the target; thirdly, fine tune the target position by enlarged searching region to improve the proposal distribution of particle filter, mean shift is embedded into particle filters, in which a smaller number of samples is used to estimate the posterior distribution than conventional particle filters by shifting samples to their neighboring modes of the observation so that samples are moved to have large weights. Finally, estimating the target position. Experimental results demonstrate that this algorithm can track the object accurately in conditions of scale modifications, rotation, abrupt shifts, as well as clutter and partial occlusions occurring to the tracking object with good robustness.

1. Introduction

Object tracking in a video sequence is an important problem in computer vision with applications in areas like video surveillance, vehicle navigation, perceptual user interface and augmented reality [1]. Object surveillance may provide crucial information about the behavior, interaction, and relationship between objects of interest. Since automated video surveillance systems comprise modules from low-level to high level processing, e.g., object detection, tracking, event analysis and classification, robustness and efficiency in each module are particularly important. However, the task of robust tracking is very challenging regarding illumination variation, background clutter, fast motion, occlusion, structural deformation, real-time restriction, etc.

In order to solve the above problems, a variety of algorithms have been proposed. These tracking algorithms can be divided into two categories. This method finds the local maximum of probability distribution in the direction of gradient. Mean shift tracking [2], derived by using the Bhattacharyya coefficient as the similarity metric, is another technique that is relatively robust. Other similarity metrics, such as the Kullback–Leibler divergence. It performs relatively well in the presence of short-term occlusions, however, when the colors of object are similar to the background, tracking drift or tracking failure may occur. The second category is a probabilistic method that involves target estimation. The preventative method is a particle filter, which is a multi-hypothesis tracking algorithm under the Bayesian framework. Reference [3] used particle filers to estimate the pdf of state vector for non-linear and non-Gaussian distributions. Although particle filters are shown to be robust in visual tracking through occlusions, cluttered backgrounds, the use of particle filters is hampered due to the computational cost from using a large number of particles.

On the other hand, point correspondences, like scale-invariant point features, were found effective in relating salient features in two image frames. Lowe [4] proposed SIFT (scale-invariant feature transform) for finding local point features. In SIFT, each point feature is described by a vector of size 128, formed from the gradient directions and magnitudes around the point. One disadvantage of using point features is that it is prune to noise and image variations. When using point feature correspondences for object tracking, one may consider a set of correspondence features on an object under a given motion model. Random sample consensus of point feature correspondences, like RANSAC [5], is a useful tool to achieve this. Using the RANSAC, one may choose a subset of correspondence point features that satisfy the model.

* Corresponding author.
E-mail address: specialdays_2010@163.com (J. Dou).

http://dx.doi.org/10.1016/j.ijleo.2015.04.031

Template matching [6] based schemes have been reported in literature which use templates of objects created by a separate system for matching in consecutive frames in order to track them.

In this paper, we proposed a robust visual tracking and simultaneous recognition method in a particle filter frame work. First, an appearance model based on Joint Color Texture and Edge of Orientation Histogram is Build, which is adopted to calculate the likelihood of particles; then, perform scale invariant feature transform (SIFT) keypoint detection and matching to predict the coarse position of the target; thirdly, fine tune the target position by enlarged searching region to improve the proposal distribution of particle filter, mean shift is embedded into particle filters, in which a smaller number of samples is used to estimate the posterior distribution than conventional particle filters by shifting samples to their neighboring modes of the observation so that samples are moved to have large weights. Finally, estimating the target position.

The remainder of this paper is organized as follows: Section 2 briefly reviews the background of this paper. The proposed robust tracking method JH_MSEPF (Joint Histogram Mean Shift Embedding Particle Filter) is described in Section 3, respectively. Experiments and results are provided and analyzed in Section 4. Finally, our work is summarized and conclusions are drawn in Section 5.

2. Related work

Mean shift for visual object tracking has drawn much interest. Early work on mean shift tracking was proposed by Comaniciu et al. [2], where tracking is formulated by maximizing the Bhattacharyya coefficient between the reference and the target object. Collins [7] extended the mean shift by introducing kernel bandwidth normalization using the method in [8]. It performs an extensive search within a range of ellipse scales and is computationally intensive. Yilmaz [9] proposed a level-set to use an asymmetric kernel for the mean shift, while Sumin and Xianwu [10] proposed anisotropic mean shift, where the center, size and orientation of the bounding box are simultaneously estimated during the tracking. Parameswaran et al. [11] proposed mean shift with a tunable bandwidth according to the object deformation. To make the tracker more robust, Xu et al. [12] proposed to compute the Bhattacharyya coefficient from joint spatial and color histograms of non-overlapping multi-scale regions. All these methods use the anisotropic kernel for the mean shift. Despite the efforts, mean shift-based tracking may drift away or lose tracking through frames, especially in the presence of object occlusions, or when the foreground object appears in similar color distributions to the background.

Particle filters have drawn much interest due to their robust tracking performance, especially in the complex video scenarios. A major barrier of using particle filters for visual tracking is its heavy computation, where a large number of particles is required as the result of a large size state vector used to characterize a visual object. Following the early work of CONDENSATION [13], a variety of particle filters have been proposed for visual tracking [14,15]. Wang et al. [14] separated the state vector into shape and appearance-based sub-vectors by using Rao–Blackwellized particle filters, where the appearance is treated by linear models, thereby reducing the number of particles required in the particle filter. Michael and Shephard [15] introduced an auxiliary particle filter (APF) that generates particles from the importance distribution depending on the most recent observation, resulting in an improved performance as compared to [13]. Further improvement is made by combining mean shift and particle filters, referred to as the hybrid method, into an integrated single scheme. Shan et al. [16] proposed to embed the mean shift in particle filters to track human hands, where particles with large weights from the mean shift [17] are combined in the observation model, thereby reducing the degeneracy and requiring fewer particles than the conventional particle filter. Zhong and Hao [18] proposed to weight particles by using the observation model, followed by applying the mean shift on particles with large weights, called elite particles.

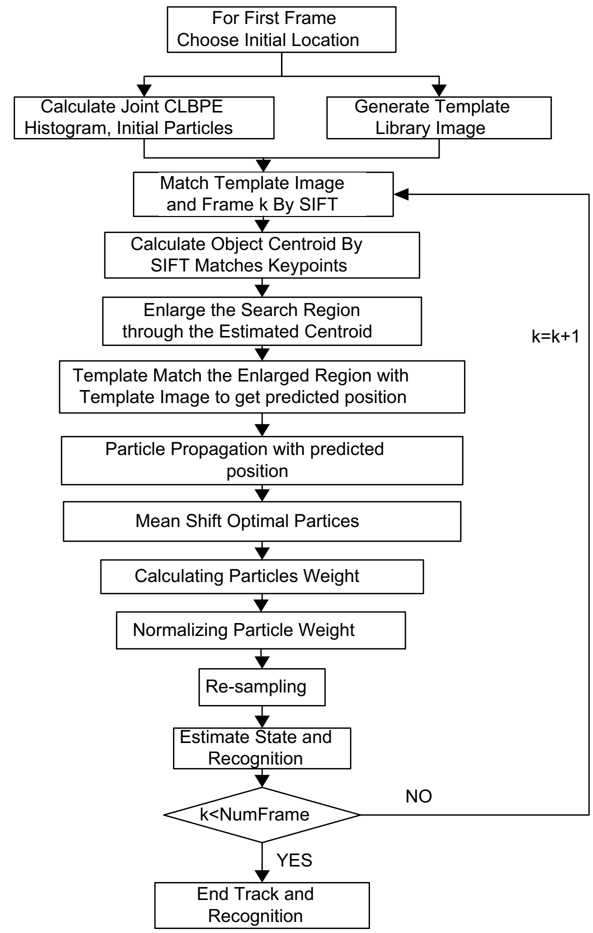

Fig. 1. Flow chart of the proposed algorithm.

Recently, Zhou et al. [19] proposed an expectation–maximization algorithm that integrates SIFT features and MS with good performance reported. However, this method does not handle scale and orientation. In [20], affine structure was estimated by SIFT feature point correspondences to achieve box size robustness, while object box center was estimated from MS. The method, however, does not treat the box orientation as a parameter. Further more, no refined selection of point features from SIFT was done.

3. Proposed method

The process of the proposed method is depicted in Fig. 1.

3.1. SIFT matching

The SIFT descriptor was proposed by Lowe [4]. The SIFT algorithm is a local feature extraction algorithm, which finds extrema points in scale space, and extracts a position, scale, rotation invariant feature vector for each extrema point. It includes two stages: SIFT features extraction and similarity measure.

3.1.1. SIFT features extraction

SIFT features extraction includes 4 steps as follows: