Computer Vision and Image Understanding 125 (2014) 97–114

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Online domain-shift learning and object tracking based on nonlinear dynamic models and particle filters on Riemannian manifolds ☆

Zulfiqar Hasan Khan, Irene Yu-Hua Gu *

Dept. of Signals and Systems, Chalmers Univ. of Technology, 41296 Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 16 December 2012
Accepted 26 March 2014
Available online 26 April 2014

Keywords:
Domain-shift object learning
Manifold appearance learning
Visual object tracking
Infrared object tracking
Riemannian manifolds
Covariance tracking
Object partition
Particle filters
Occlusion handling
Nonlinear dynamic model

ABSTRACT

This paper proposes a novel online domain-shift appearance learning and object tracking scheme on a Riemannian manifold for visual and infrared videos, especially for video scenarios containing large deformable objects with fast out-of-plane pose changes that could be accompanied by partial occlusions. Although Riemannian manifolds and covariance descriptors are promising for visual object tracking, the use of Riemannian mean from a window of observations, spatially insensitive covariance descriptors, fast significant out-of-plane (non-planar) pose changes, and long-term partial occlusions of large-size deformable objects in video limits the performance of such trackers. The proposed method tackles these issues with the following main contributions: (a) Proposing a Bayesian formulation on Riemannian manifolds by using particle filters on the manifold and using appearance particles in each time instant for computing the Riemannian mean, rather than using a window of observations. (b) Proposing a nonlinear dynamic model for online domain-shift learning on the manifold, where the model includes both manifold object appearance and its velocity. (c) Introducing a criterion-based partial occlusion handling approach in online learning. (d) Tracking object bounding box by using affine parametric shape modeling with manifold appearance embedded. (e) Incorporating spatial, frequency and orientation information in the covariance descriptor by extracting Gabor features in a partitioned bounding box. (f) Effectively applying to both visual-band videos and thermal-infrared videos. To realize the proposed tracker, two particle filters are employed: one is applied on the Riemannian manifold for generating candidate appearance particles and another is on vector space for generating candidate box particles. Further, tracking and online learning are performed in alternation to mitigate the tracking drift. Experiments on both visual and infrared videos have shown robust tracking performance of the proposed scheme. Comparisons and evaluations with ten existing state-of-art trackers provide further support to the proposed scheme.

1. Introduction

Tracking visual object from videos has drawn increased interest in recent years [1,2]. Many robust trackers have been successfully developed however challenges remain especially when dynamic objects contain large out-of-plane changes and significant appearance changes, experience long-term partial occlusions or intersections with other objects in dynamic background and crowded scenes. Online learning of object appearances is an essential issue for mitigating the tracking drift, where off-line learning is often unrealistic for dealing with real world problems. One of the main tasks for online learning is to estimate current statistics, parameters or states of non-stationary objects from new observations. It enables a tracker to adaptively utilize a timely reference object model for tracking. A main challenge for online learning of visual object is the ambiguity on image changes that can be caused either by object itself (e.g. deformation, pose, self-occlusion) or by other occluding object/background. It is desirable that an online learning method adapts to changes in object intrinsic parameters (e.g. pose, appearance and shape) and is insensitive to extrinsic variations (e.g. illumination, occlusion, background, camera motion and viewpoint). Several online learning methods have been proposed. For example, Refs. [3,4] uses vector space learning that incrementally updates the sample mean of object appearance. Refs. [5,6] proposes manifold learning where dynamic object appearance is described as a point moving on a smoothed manifold surface. Since images of dynamic objects, especially out-of-plane objects, reside in nonlinear spaces, or manifolds [7], using a set

☆ This paper has been recommended for acceptance by Nikos Paragios.

* Corresponding author.
E-mail address: irenegu@chalmers.se (I.Y.-H. Gu).

http://dx.doi.org/10.1016/j.cviu.2014.03.013
1077-3142/

of low-dimensional subspaces on manifolds has led to more robust online domain-shift learning as comparing with those using a single vector space.

Using covariance matrices of image features as object descriptors has drawn increasing interest lately for visual tracking [8]. It enables efficient description of object features, and shows to be robust and versatile for variations in illuminations, views and poses at modest computational cost. The space of non-singular covariance matrices of image features can be formulated as connected points (corresponding to smoothly shifting in different subspaces) on a Riemannian manifold. This paper addresses online domain-shift learning and tracking on Riemannian manifolds by using a set of weighted candidate covariance matrices as object descriptors.

1.1. Related work

Many object tracking methods have been proposed recently. One category of tracking methods models the images of objects on a vector space. For example, a mean shift-based tracker [9] seeks local modes through the kernel-based pdf estimate. It maximizes the similarity between a reference and a candidate object. However, mean shift tracker may fail if a target object has similar colors to occluding objects/background or too large changes in appearance due to pose changes. To tackle the problem, Ref. [10] proposes to use combination of anisotropic mean shift for describing global object appearance and SIFT-RANSAC for describing local object features for effective object tracking. CONDENSATION tracker [11] uses particle filters (PFs) and learned dynamic model, together with visual observations, to estimate the posterior probability of state vector. Ref. [3] incrementally learns the appearance subspace and updates the mean image by using a PF and extended sequential Karhunen–Loeve algorithm, that is performed by linear operations on a single subspace. However, for 2D planar images, dynamic object appearances especially objects containing large out-of-plane pose changes lie in different subspaces, or, on a differentiable manifold. Although these trackers perform rather well for some scenarios, challenges remain when objects contain significant out-of-plane pose changes, especially when the object size in image is large. Another category of tracking methods model the image of an object on a manifold that is locally Euclidean. For example, Ref. [5] uses a conjugate gradient and Newton's method for subspace tracking on Grassmann and Stiefel manifolds with applications to orthogonal procrustes. Ref. [12] proposes piecewise geodesics on complex Grassmann manifolds using projection matrices for subspace tracking, where simulations are performed on synthetic signals from an array of sensors. Ref. [13] proposes a visual tracking approach by applying a Kalman filter to velocity vectors in the tangent planes of Grassmann manifolds. Ref. [14] applies visual tracking on Grassmann manifolds by using sliding-window observations. Much manifold tracking work is focused on covariance tracking, probably due to their robust performance. The first significant work on covariance tracking is proposed by Porikli et al. [6]. A variety of methods are then proposed that further improve and enhance the method. Among them, Ref. [15] uses an exhaustive search and a distance measure to find the best matching where model updating is performed using the Lie group structure on the manifold. The method is able to track objects with moderate out-of-plane pose changes however significant ones remain a challenge. Ref. [16] uses PFs [17] and affine invariant metrics [18] on Riemannian manifolds for seeking objects with similar appearances and tracking object bounding boxes with central location, width and height as the parameters. Ref. [19] incrementally learns covariance matrices of object appearances by applying the log-Euclidean metric [20] on a time window of manifold points, and uses PFs to track the object bounding box parameterized by its scale and central location. However, none of these methods simultaneously estimates affine bounding box parameters and covariance matrix of object appearance. Ref. [21] employs the log-Euclidean metric on a time window of manifold points for tracking affine box parameters of a moving object, and incrementally learns the eigen object representation in the tangent planes of manifold points. However, using a time window in log-Euclidean metric could affect the tracker in adapting abrupt object motions. Ref. [22] proposes a head pose estimation by using covariance matrices for object appearances and a nearest centroid classifier. The method is mainly related to an image descriptor for static images. Ref. [23] utilizes PFs on the manifold to estimate the target position and time-varying noise covariance with synthetic 2D position points for tracking target trajectories. The method is only simulated on tracking point data rather than images/video objects. Ref. [24] proposes nonlinear mean shift on a manifold for image segmentation and filtering. The basic formulation of Riemannian manifolds for imaging application is described in a very concise and thorough way, however, it has not addressed the issue of video object tracking. Ref. [25] proposes a Kalman filter on the manifold for visual object tracking. The weakness is that it cannot well handle nonlinear and non-Gaussian visual objects. Ref. [26] applies offline manifold training from a face dataset of different poses, and subsequent online learning of local linearity of appearance manifolds by PFs with a coarse-to-fine factorized sampling. It searches object locations during tacking, however, does not handle affine object transformations. The two categories of trackers above both belong to subspace tracking and modeling techniques. A third category of methods is based on tracking-by-detection, where an online discriminative classifier is trained to separate a target object from the background/clutter [27–31]. One can see that using learned classifiers for tracking is a rather different methodology: classifiers are usually formed by online boosted learning (e.g. AdaBoost) using a set of positive and negative visual samples corresponding to objects and background/clutter. It is also worth noting that the first category of methods uses features in a vector space while the second category uses features embedded on a manifold. The difference on whether basis matrices of feature vectors are defined in a vector space or on a manifold leads to key difference of these methods, where the latter category offers domain-shift possibility. Further, an important difference of these two categories to the third category of tracking-by-detection methods is on how the reference model is built, where the last category uses a set of training data to obtain the model. Although the final tracking performance is examined in a same way, parameters that impact the performance are rather different in the tracking-by detection category. The performance of tracking-by-detection trackers are largely influenced by how discriminative the online learned classifier with respect to the selected sets of positive and negative samples (in terms of the size of training samples and their statistical coverage).

Despite relatively good tracking results in these covariance trackers, several limitations are posed on their performance. One main limitation is tracking objects with fast or variable speed out-of-plane (or, non-planar) pose changes. This may be due to the use of *a time window of manifold observations* in the Riemannian mean to generate a new manifold point estimate [6]. The second limitation may occur when two objects have similar statistics however different spatial distributions, as the conventional covariance descriptor lacks of spatial information. The covariance matrices in these trackers are computed by first describing each pixel by a feature vector (e.g. pixel position, color and intensity values, first and 2nd derivatives), followed by computing the covariance over an entire set of feature vectors within a candidate object box. The third limitation could be caused by inaccuracies in updating new manifold points using predicted velocity vectors.