Object tracking using learned feature manifolds ☆

Yanwen Guo [a,*], Ye Chen [a], Feng Tang [b], Ang Li [c], Weitao Luo [a], Mingming Liu [a]

[a] *National Key Lab for Novel Software Technology, Nanjing University, Nanjing 210023, PR China*
[b] *Hewlett-Packard Laboratories, Palo Alto, CA 94304, USA*
[c] *University of Maryland, College Park, MD 20740, USA*

A R T I C L E   I N F O

A B S T R A C T

Local feature based object tracking approaches have been promising in solving the tracking problems such as occlusions and illumination variations. However, existing approaches typically model feature variations using prototypes, and this discrete representation cannot capture the gradual changing property of local appearance. In this paper, we propose to model each local feature as a feature manifold to characterize the smooth changing behavior of the feature descriptor. The manifold is constructed from a series of transformed images simulating possible variations of the feature being tracked. We propose to build a collection of linear subspaces which approximate the original manifold as a low dimensional representation. This representation is used for object tracking. Object location is located by a feature-to-manifold matching process. Our tracking method can update the manifold status, add new feature manifolds and remove expiring ones adaptively according to object appearance. We show both qualitatively and quantitatively this representation significantly improves the tracking performance under occlusions and appearance variations using standard tracking dataset.

1. Introduction

Object tracking is a central problem in computer vision with many applications, such as activity analysis, automated surveillance, traffic monitoring, and human–computer interaction. It is essentially the problem of finding the most likely estimate of the object state given a sequence of observations. Object tracking is challenging because of:

- *Complex object appearance*. The object may have complicated appearance which is hard to model. Furthermore, it may undergo significant changes due to the pose and scale variations as well as non-rigid object motions.
- *Occlusions*. The object may be occluded by the background or other moving objects, making it difficult to be localized.
- *Complex object motion*. This is caused by either the moving pattern of the object or by camera motion accompanied by object motion.

There are two key components in an object tracking algorithm: object representation and dynamics. Object representation tries to model the object as accurately as possible so that the tracking algorithm can accurately describe the complex object appearance.

Object dynamics model how the object appearance evolves over time to be able to handle appearance variations. The two problems are usually coupled together: the object representation should be designed to be easily updated to model appearance variations, while the object dynamics should be able to take advantage of the characteristics of object representation for model update.

Traditional methods for representing the object, such as global histogram based approach in meanshift tracking [1] and PCA subspace based approach in EigenTracking [2], are global approaches which describe the object to be tracked as a whole. Such methods work well in many practical applications, but have several intrinsic limitations. First, it is usually very difficult for a global representation to capture local details and as a result unable to model complex appearances. Second, global representations are not robust to partial occlusion. Once the object is occluded, the whole feature vector of object representation is affected. Third, global representations are hard to update.

Recently, local representations have opened a promising direction to solve these problems by representing an object as a set of local parts or sparse local features. Part-based trackers generally use sets of connected or global visual properties incorporated local parts or components [3–6]. The parts used for object representation are updated during tracking by removing old parts that exhibit signs of drifting and adding new ones for easy accommodation of appearance changes. Feature-based trackers often represent the target by a set of sparse local features such as SIFT [7] and affine invariant point detectors [8] which are often invariant to changes

in rotation, scale, illumination and viewpoint. These approaches first localize the features at a sparse set of distinctive image points by feature detectors. Then the feature vectors, usually named as descriptors, are computed based on the local image statistics centered at these locations. Two major advantages of sparse local features are the invariance to image changes and robustness to occlusions. Existing local feature based approaches typically model how the local features vary using prototypes. However, this discrete representation cannot capture the gradual changing property of local appearance.

In this paper, we propose a local feature based manifold representation for object tracking. The object is represented by a set of sparse local feature manifolds. Each local feature manifold is computed from a series of SIFT feature descriptors [7] that correspond to different appearances of the same object feature under simulated variations of practical situations. To build it, we first detect a set of interest points on the object by the state-of-the-art feature detectors. For each feature point, we transform the image regions surrounding it for simulating real object changes. A feature manifold is thus obtained by exploring the ensemble of descriptors extracted on the transformed image regions. Such a manifold is an informative yet robust representation in that it captures the local appearance variations of a part of the object over time, making the local representation more robust against object changes. The local feature variation is complicated and nonlinear in practice as an example illustrated in Fig. 1 which shows a feature on a walking man. As can be observed, the feature appearance changes dramatically during the move. As a result, the feature manifold is a highly nonlinear appearance manifold. For computational efficiency, we apply incremental principal component analysis to it and yield a collection of linear subspace approximation.

To model geometric relations among local features, the feature manifolds are organized as a feature manifold graph which is used to represent the target object to be tracked. Each local feature manifold describes object appearance details and relationships among them encode object structure. Such geometric relationships are elastic and have the flexibility to handle objects with coherent motion and a certain amount of variations caused by viewpoint changes and articulated motions. An advantage of the feature manifold graph is that locally the manifold graph reinforces the power of feature description and characterizes variations of object appearance by learning a series of descriptors, while globally it encodes object structure with the geometric relations among those manifolds. Such characteristics make it suitable for many vision tasks.

We apply the feature manifold graph to object tracking as an application. With the feature manifold graph representation, the target object is tracked based on graph-based feature-to-manifold tracking. During tracking, features are extracted in a candidate region of the current frame and then matched with the manifold. Object position is located by integrating all matching in the manifold graph. Since features may appear and disappear due to

viewpoint changes and occlusions, our dynamic model is designed to be able to add new feature manifolds and remove expiring ones adaptively and dynamically. To the best of our knowledge, this is the first paper that applies manifold learning to local features for object tracking.

The rest of this paper is organized as follows. Section 2 describes the related work on object tracking. We present our feature manifold model in Section 3. Section 4 shows our main tracking paradigm. Experiments and analysis are given in Section 5, and the whole paper is concluded finally.

2. Related work

Object tracking using local features has been explored by previous researchers. In [9], Shi and Tomasi proposed a method to select corner-based features that are most reliable for tracking. Collins et al. developed an algorithm for unsupervised learning of object models as constellations of features, and proposed a discriminative feature tracker [10]. A simultaneous modeling and tracking method is proposed in [11] to learn the object model during tracking. The object features are selected manually and tracked individually. The posterior distribution of appearance and shape is built up incrementally using an exemplar-based approach. In [12], the object is represented as view-dependent object appearance models corresponding to different viewing angles. This collection of acquired models is indexed with respect to the view sphere. These models are matched to each frame to estimate object motion. In [13], the authors proposed a "feature harvesting" approach that has a training phase to learn the object geometry and appearance using a randomized tree classifier. The online tracking then becomes a detection problem using the learned classifier. Liu et al. proposed to jointly track different types of features by representing the objects of interest with the hybrid templates [14]. A generative model is developed to learn the template and to estimate object location and scale.

It is noted that the state-of-the-art local features such as SIFT [7] and SURF [15] have been used for object tracking recently. In [16], an attributed relational feature graph which represents the object using SIFT features with geometric relations is proposed for object tracking. Zhou et al. presented a SIFT based mean shift tracking algorithm [17]. The similarity between two neighboring frames is measured in terms of color and SIFT correspondence by using an expectation–maximization algorithm. In [18], He et al. proposed to represent the object by a set of SURF features of interest. Object motion is estimated in terms of maximum likelihood feature motion observations. In [19], Sun and Liu proposed an object tracking method which is based on the combination of local SIFT description and global PCA representation. The method is constructed in the framework of particle filter. In fact, the changing of feature appearance is smooth and highly nonlinear in nature, which is hard to be modeled using discrete prototypes. In this
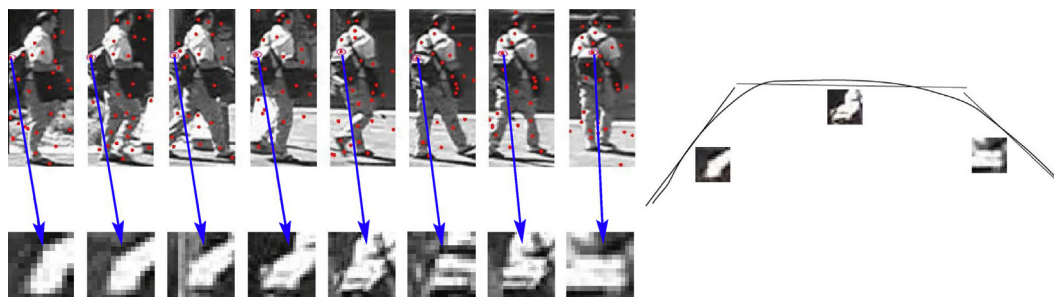

Fig. 1. Appearance variations of a feature during tracking. Feature patches of different frames are shown on the left and a feature manifold is visualized on the right.