Image and Vision Computing 30 (2012) 332–344

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

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

Shape based appearance model for kernel tracking☆

Zhijie Wang *, Mohamed Ben Salah, Hong Zhang, Nilanjan Ray

Department of Computing Science, University of Alberta, 221 Athabasca hall, University of Alberta, Edmonton, Alberta, Canada T6G 2E8

ARTICLE INFO

Article history:
Received 25 May 2011
Received in revised form 12 March 2012
Accepted 25 March 2012

Keywords:
Object tracking
Appearance model
Shape cue

ABSTRACT

This paper investigates kernel based tracking using shape information. A kernel based tracker typically models an object with a primitive geometric shape, and then estimates the object state by fitting the kernel such that the appearance model is optimized. Most of the appearance models in kernel based tracking utilize the textural information within the kernel, although a few of them also make use of the gradient information along the kernel boundary. Interestingly, shape information of a general form has never been fully exploited in kernel tracking, despite the fact that shape has been widely used in silhouette tracking at the cost of intensive computation. In this paper, we propose an original way to incorporate shape knowledge into the appearance model of kernel based trackers while preserving their computational advantage versus silhouette based trackers. Experimental results demonstrate that kernel tracking is strongly improved by exploiting the proposed shape cue through comparisons to both kernel and silhouette trackers.

1. Introduction

Object tracking consists of following moving objects through an image sequence. It is a challenging problem in computer vision which serves various applications such as vehicle navigation, medical image processing, and video analysis [45].

To locate a given object from one image to the next in a sequence, the current trackers are generally based on the photometric characteristics or shape of the objects. The photometric appearance is used assuming that the object has a profile which is distinctive against the background [14,33]. Differently, shape based trackers follow the object having a given geometric description modulo some deformations [11]. Other visual artifacts have also been used to assist tracking such as 3D object views, but photometric characteristics and shape constraints remain the most useful and they determine the efficiency of the tracker considerably. Besides the efficiency concern, there are also other practical needs which determine considerably the framework adopted for designing the tracker. In some applications, tracking consists of following the object grossly placed about a mean position [9,45] and only some bounding figures, such as rectangles or ellipses, are used to circumscribe the moving object. However this is insufficient for other applications where accurate object boundary is rather needed to identify the object, classify it, or interpret its behavior [11,45]. As a result, two major groups of tracking methods, silhouette based and kernel based, have been proposed depending on their specific advantages and on the targeted applications. Silhouette based trackers employ rich but inefficient contour models to estimate accurate object shape as shown in Fig. 1a, while on the contrast kernel based trackers employ rough but efficient primitive geometric kernel models to estimate the object's basic information, e.g., location and scale as shown in Fig. 1b.

In silhouette based tracking, the object contour is generally described by a closed curve which encloses the region corresponding to the photometric and/or shape constraints. The variational framework has been the most convenient for this kind of methods [11,28,33], where the problem constraints are embedded in an objective functional whose minimum corresponds to positioning the curve by the region of interest boundaries. The shape constraint has been intensively used to assist tracking by active contours. In [8], a Fourrier-based shape prior model is used to constrain the deformation of a classic active contour (snake). In [11], it has been demonstrated that object tracking can be improved considerably when a model of the moving object boundary shape is available. A principal component analysis (PCA) is used to learn a geometric shape description of the moving object from a training set of various object outlines. The principal components are systematically embedded in a shape prior of an active curve/level set formulation. The tracking process updates continuously the shape prior based on the previously observed contours. Moreover, shape priors have been also used in [46] for multi-object tracking but it has been activated only when an occlusion is detected. The tracker is mainly based on the appearance term which acts when there is no occlusion between moving objects. Meanwhile, the shape prior model is updated using the observed contours in the tracking trail. All the above algorithms pose the silhouette tracking as a minimization problem, and the minimization of such functionals relies on gradient descent. As a result, the algorithms converge to local minima, can be affected by the initialization, and, more importantly, are notoriously slow in spite of the various

☆ This paper has been recommended for acceptance by Stefanos Zafeiriou.

* Corresponding author. Tel.: +1 780 492 7188.
E-mail address: zhijie@ualberta.ca (Z. Wang).

0262-8856/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2012.03.003

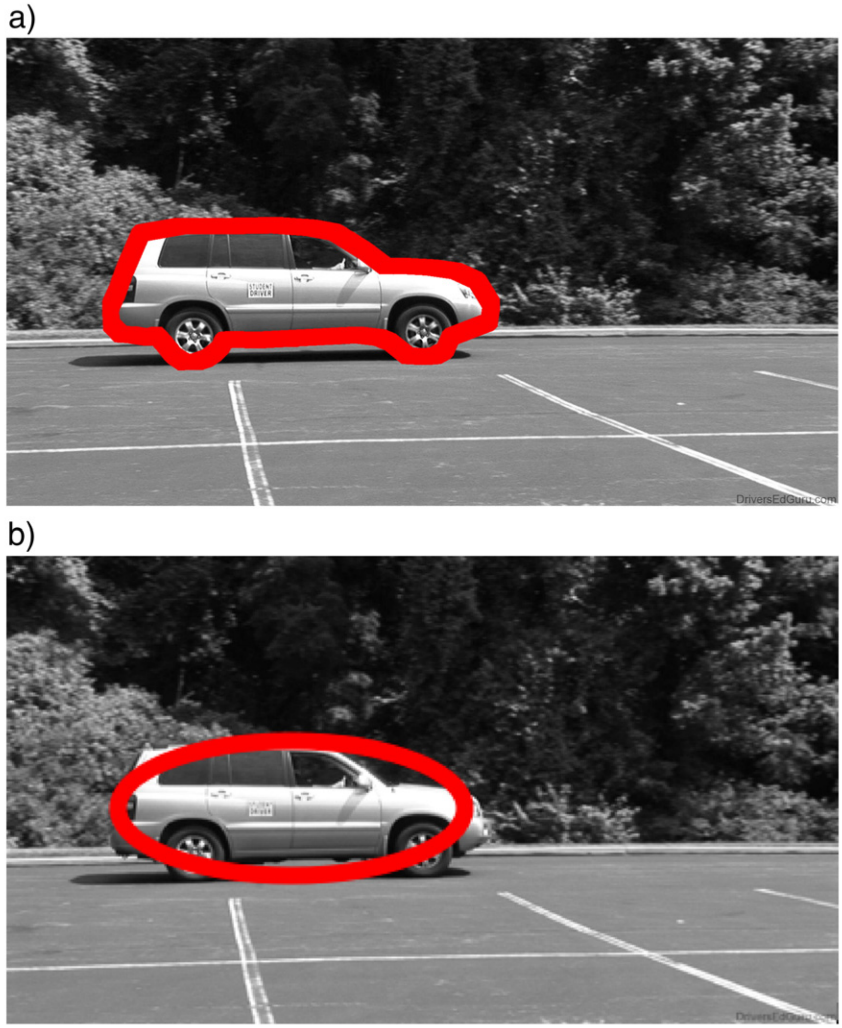

Fig. 1. (a) An example of a silhouette based tracker estimating the object boundary. (b) An example of a kernel (ellipse in this case) based tracker estimating the object state. Silhouette based trackers are mainly used when the exact object boundary is required, which implies a high computational load. Kernel based trackers are rather used when only basic object information (location and scale, etc.) needs to be estimated, which generally requires less computations.

computational artifacts which can speed their execution [40]. To overcome these problems, another group of algorithms pose the silhouette tracking as an inference problem, and the goal is to estimate the posterior density of a contour given image observations. In this context, the shape information constrains the object model to transform only in ways characteristic of the object of interest. For instance, in [17,35], an object is represented with an active shape model that undergoes constrained transformation defined by the shape prior information. While all the aforementioned shape models are shown to drastically improve the tracking performances, they still suffer from the inefficiency problem resulting from the high dimensional solution space, and they are of no consequence where the object need not be identified nor its behavior interpreted based on its shape. Practically, the inefficiency in terms of execution load of these silhouette based tracking methods is the major impediment in many applications, particularly those which require real time processing.

In contrast, kernel based trackers have been widely adopted because of their relative simplicity and low computational cost. The purpose is to estimate the object state efficiently by computing the motion of the object and modeling it with a primitive geometric shape. In this framework, many methods have been proposed and resulted in efficient and accurate algorithms. These methods differ mainly in terms of their appearance representations which did not fully exploit shape knowledge due to the lack of the object boundaries' information. Most of the cues that have been used in this framework are based on regional information, and some of them on boundary information. Among the regional based cues, color histogram might be the most popular. It is scale and rotation invariant and it can handle occlusion in a certain degree [15,20,34]. As global statistic information, color histogram does not provide a discriminative localization ability. In addition to that, there are also models based on histograms of edge orientation [5,20] and motion [34,38]. These two cues are quite sensitive to clutter. Similarly, texture has also been used as a cue in the appearance models of [4] and [5], which are based on wavelet and steerable pyramid, respectively. Another approach to modeling the object appearance is using PCA to build a subspace representation from a set of templates [21,37,38]. SIFT (Scale Invariant Feature Transform) descriptors have also been used as features in appearance models in [7,24] to increase the discriminative ability. All the regional based cues mentioned above are computed from some manually designed statistics. There are also a few papers which use machine learning methods to learn appearance models directly from training sets. For example in [1], the authors use a support vector machine to learn the appearance model from a set of templates, and in [2] the authors learn the appearance model via Adaboost from a set of Haar-like features. All the existing regional cues used in the above works are based on the information within the kernel (not the object) boundary that contains absolutely no shape information.