Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 25 (2007) 1205–1216
www.elsevier.com/locate/imavis

Robust tracking with motion estimation and local Kernel-based color modeling ☆

R. Venkatesh Babu [a,*], Patrick Pérez [b], Patrick Bouthemy [b]

[a] *Department of Electrical Engineering, Indian Institute of Science, Bangalore, India*
[b] *IRISA/INRIA-Rennes, France*

Received 9 February 2006; received in revised form 15 June 2006; accepted 12 July 2006

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. Mean-shift tracker, which gained attention recently, is known for tracking objects in a cluttered environment. In this work, we propose a new method to track objects by combining two well-known trackers, sum-of-squared differences (SSD) and color-based mean-shift (MS) tracker. In the proposed combination, the two trackers complement each other by overcoming their respective disadvantages. The rapid model change in SSD tracker is overcome by the MS tracker module, while the inability of MS tracker to handle large displacements is circumvented by the SSD module. The performance of the combined tracker is illustrated to be better than those of the individual trackers, for tracking fast-moving objects. Since the MS tracker relies on global object parameters such as color, the performance of the tracker degrades when the object undergoes partial occlusion. To avoid adverse effects of the global model, we use MS tracker to track local object properties instead of the global ones. Further, likelihood ratio weighting is used for the SSD tracker to avoid drift during partial occlusion and to update the MS tracking modules. The proposed tracker outperforms the traditional MS tracker as illustrated.

Keywords: Visual tracking; Mean-shift; Object tracking; Kernel tracking

1. Introduction

The objective of object tracking is to faithfully locate the targets in successive video frames. The major challenges encountered in visual tracking are cluttered background, noise, change in illumination, occlusion and scale/appearance change of the objects. Considerable work has already been done in visual tracking to address the aforementioned challenges. Most of the tracking algorithms can be broadly classified into the following four categories.

(1) *Gradient-based methods* locate target objects in the subsequent frame by minimizing a cost function [1,2].
(2) *Feature-based approaches* use features extracted from image attributes such as intensity, color, edges and contours for tracking target objects [3–5].
(3) *Knowledge-based tracking algorithms* use *a priori* knowledge of target objects such as shape, object skeleton, skin color models and silhouette [6–9].
(4) *Learning-based approaches* use pattern recognition algorithms to learn the target objects in order to search them in an image sequence [10–12].

Visual tracking in a cluttered environment remains one of the challenging problems in computer vision for the past few decades. Various applications like surveillance and monitoring, video indexing and retrieval require the ability to faithfully track objects in a complex scene involving appearance and scale change. Though there exist many

☆ This work was supported by an ERCIM post-doctoral fellowship at IRISA/INRIA, Rennes, France.
* Corresponding author. Tel.: +91 80 23722879.
E-mail addresses: venkatesh.babu@gmail.com (R.V. Babu), perez@irisa.fr (P. Pérez), Patrick.Bouthemy@irisa.fr (P. Bouthemy).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.07.016

techniques for tracking objects, color-based tracking with kernel density estimation, introduced in [13,8], has recently gained more attention among research community due to its low computational complexity and robustness to appearance change. The reported work in [13] is due to the use of a deterministic gradient ascent (the "mean shift" iteration) starting at the location, corresponding to the object location in previous frame. A similar work in [8] relies on the use of a global appearance model, e.g., in terms of colors, as opposed to very precise appearance models such as pixel-wise intensity templates [14,15].

The mean-shift algorithm was originally proposed by Fukunaga and Hostetler [16] for clustering data. It was introduced to image processing community by Cheng [17] a decade ago. This theory became popular among vision community after its successful application to image segmentation and tracking by Comaniciu and Meer [18,5]. Later, many variants of the mean-shift algorithm were proposed for various applications [19–24].

Though mean-shift tracker performs well on sequences with relatively small object displacement, its performance is not guaranteed when the objects move fast as well as when they undergo partial occlusion. Here, we attempt to improve the performance of mean-shift tracker when the object undergoes large displacements (when the object regions do not overlap between the consecutive frames) and in the event of partial/full occlusion. The problem of large displacements is tackled by cascading an SSD tracker with the mean-shift tracker. An SSD tracker based on frame-to-frame appearance matching, is useful in finding the object location in successive frames. However, the problem with SSD tracker is its short-term memory which can cause drifting problems or even complete loss in worse cases. On the other hand, MS trackers which rely on persistent global object properties such as color, can be much more robust to detailed appearance changes due to shape and pose changes. However, MS tracker has problems with large displacements. It thus seems interesting to combine the advantages of the two aforementioned trackers.

In order to improve the performance of MS tracker, in the event of the object undergoing partial occlusion, we propose to rely on a number of elementary MS modules (tracking points) embedded within the object, rather than on a single global MS tracker representing the whole object. We also address the issue of large scale changes due to camera operations.

For each of the above-mentioned challenges, solutions proposed so far have been within the realm of pure MS trackers: incorporation of a dynamic model (e.g., using Kalman filter in [13,25] or particle filter in [26,27]) to cope with large displacements, occlusions and, to some extent, with scale changes; simple linear histogram updates with fixed forgetting factor [27] for on-line adaptation of reference model; rather complex procedures [28,29] for addressing the generic problem of scale changes (independent of their origin).

The novelty of the proposed approach lies in a one-step approach which exploits the fact that the reference color model and instantaneous motion estimation based on pixel-wise intensity conservation, complement one another. The latter is provided by greedy minimization of the intensity sum-of-squared differences (SSD), which is classic in point tracking and motion field estimation, by block matching. Scale changes of the object that are due to camera zoom effect or ego-motion are estimated by approximating the dominant apparent image motion by an affine model. By using local object color models (tracking points embedded on the object) instead of a global one, the performance of the tracker is greatly improved when the object undergoes partial occlusion.

The paper is organized as follows. Section 2 explains the proposed SSD/MS combined tracker to track fast moving objects. The problem of occlusion handling is discussed in Section 3. The results illustrating the performance of proposed tracker are given in Section 4. Concluding remarks are given in Section 5.

2. Proposed combined tracker

In this work, tracking is done in Kalman filter framework. The object to be tracked is specified by the location of its center and scale (for a fixed aspect ratio) in the image plane. The objective of the tracking algorithm is to find the object location in successive frames. In this work, we cascade SSD tracker with MS tracker to obtain better tracking performance. The measurements obtained by the combined tracker module are used for estimating the states of the Kalman filter. The overview of the proposed system is illustrated in Fig. 1.

The state-space representation of the tracker used in Kalman filter framework is given below:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ x_t \\ y_t \\ s_{t+1} \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \\ s_t \end{bmatrix} + \mathbf{w}_t, \qquad (1)$$

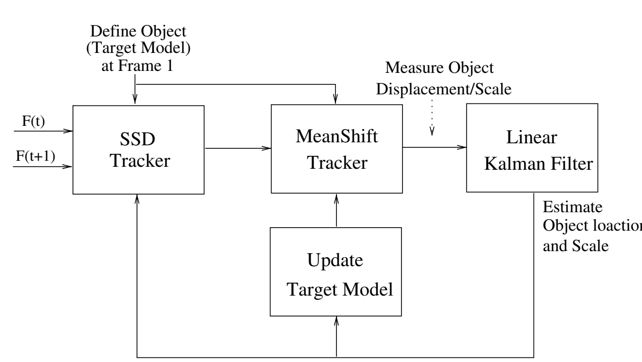

Fig. 1. Overview of the proposed tracking system.