Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 985–1001
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Real-time object tracking using bounded irregular pyramids ☆

R. Marfil *, L. Molina-Tanco, J.A. Rodríguez, F. Sandoval

Departamento de Tecnología Electrónica, E.T.S.I. Telecomunicación, Universidad de Málaga, Campus de Teatinos, 29071-Málaga, Spain

Received 18 July 2005; received in revised form 7 June 2006
Available online 8 February 2007

Communicated by S. Dickinson

Abstract

Target representation and localization is a central component in visual object tracking. In this paper a new approach for target representation and localization is presented. This approach tackles two of the most important causes of failure in object tracking: changes of object appearance and occlusions. We propose a modified template matching approach which does not require a priori learning of object views. This method allows to track non-rigid objects in real-time by employing a weighted template which is dynamically updated, and a hierarchical framework that integrates all the components of the tracker. Our hierarchical tracker allows tracking of multiple objects with low increase of computational time. The capability of the tracking system to handle occlusions and target distortions is demonstrated for several video sequences.

Keywords: Non-rigid object tracking; Multiple object tracking; Target representation and localization; Hierarchical template matching; Template-based tracking

1. Introduction

Object tracking has been one of the main fields of study in Computer Vision for the last 20 years. Real-time object tracking is a critical task in many applications such as surveillance, object-video compression and driver assistance. Typically, a visual tracking system can be divided into two major components: (i) target characterization and localization, and (ii) filtering and data association (Comaniciu et al., 2003). The first component is mostly a bottom-up process, which must be capable of dealing with changes in appearance and partial occlusions of the target, while the second component is usually a top-down process dealing with the dynamics of the objects and the evaluation of different assumptions. The way the two components are weighted and integrated in the same framework depends on the final application, having a great influence in the robustness and efficiency of the tracking process.

In this paper, we propose a system to track one or multiple objects in cluttered scenes. Typically, when the goal is to track objects in cluttered scenes, the application relies more on target representation and localization (Comaniciu et al., 2003). Therefore, this work puts the emphasis in the first component, presenting a new approach to target representation and localization which allows handling of occlusions and appearance changes of the objects and performs the tracking process in real time.

Five main object tracking approaches have been developed depending on the target representation (Cavallaro et al., 2005):

- Model-based methods (Koller et al., 1993), which employ a priori knowledge about the geometry of objects in a given scene. Therefore, they present two major drawbacks: the need for object models with

☆ This work has been partially granted by the Spanish Ministerio de Educación y Ciencia and FEDER funds project No. TIN2004-05961 and by EURON project VISOR.
* Corresponding author. Fax: +34 952131447.
E-mail address: rebeca@uma.es (R. Marfil).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.11.013

detailed geometry for all objects that could be found in the scene, and the lack of generality. Besides, they are usually computationally expensive.

- Appearance-based methods (Jepson et al., 2003), which track connected regions that roughly correspond to the 2D shapes of the objects based on their dynamic model. The tracking strategy relies on information provided by the entire region. Examples of such information are motion, colour and texture. These methods cannot usually cope with complex deformations of the tracked object.
- Contour-based methods (Blake et al., 1995), which track only the contour of the object. Usually they use active contour models like snakes, B-splines or geodesic active contours.
- Feature-based methods (Shi and Tomasi, 1994), which use features of an object to track parts of it. Although these approaches are very stable even in case of partial occlusions, the problem of grouping the features to determine which of them belong to the same object is its current major drawback.
- Hybrid methods (Cavallaro et al., 2005), which are designed as a hybrid between a region-based and a feature-based technique. They exploit the advantages of the two by considering first the object as an entity and then by tracking its parts. The main drawback of these approaches is their high computational complexity.

This paper is concerned with tracking objects in image sequences using a template-based appearance model. The aim is robust real-time tracking under severe changes of viewpoint in the absence of an a priori model. Appearance models can be divided in (Jepson et al., 2003):

- Template-based models (Nguyen and Smeulders, 2004) which use an image sample or template of the target to track.
- View-based models (Ho et al., 2004), which are usually learned with Principal Component Analysis. They have the advantage of modelling variations in pose and illumination. However they also have the disadvantages of being object specific and requiring training prior to tracking in order to learn the subspace basis.
- Motion-based models, which usually have problems when motions of the target and background are similar. They are usually improved by accumulating an appearance model through time (Irani et al., 1994) or estimating both motion and appearance simultaneously (Jepson et al., 2003). These methods are computationally expensive.
- Global statistic based methods (Comaniciu et al., 2003), which use image statistics to represent the tracked object.

The use of local and global image statistics, such as color histograms, have been popular for tracking. Colour distribution can provide an efficient feature for tracking as it is robust to partial occlusion, scaling and object deformation. It is also relatively stable under rotation in depth in certain cases (Nummiaro et al., 2003). Therefore, colour distributions have been used to track non-rigid objects like heads (Raja et al., 1999) or hands (Martin et al., 1998). A variety of statistical techniques have been used to model the colour distribution (Elgammal et al., 2002). Thus, Raja et al. (1999) modelled the colour distribution of an object using a mixture of Gaussians fitted using the EM (Expectation Maximization) algorithm. The major drawback of this parametric technique is to choose the right number of Gaussians for the assumed model. To avoid this problem, non-parametric techniques using histograms can be used. Although colour histograms is not the best non-parametric density estimate (Scott, 1992), it has been successfully used to track hands (Martin et al., 1998) or other non-rigid objects against cluttered backgrounds (Comaniciu et al., 2003). Besides, colour histograms can be easily quantized into a small number of bins to satisfy the low-computational cost imposed by real-time processing. One of the main drawbacks with colour histograms is that, if only spectral information is used to characterize the target, the similarity function can have large variations for adjacent locations on the image lattice and the spatial information is lost. To find the maxima of such functions, an expensive exhaustive search must be applied (Comaniciu et al., 2003). In order to avoid it, the similarity function can be regularized by masking the objects with an isotropic kernel in the spatial domain (Elgammal et al., 2002).

Template-based models can be seen as a way to combine colour information with spatial information. The classical idea behind template tracking is that an object is tracked through a video sequence by extracting an example image of the object in the first frame—a template—and then finding the region which matches the template as closely as possible in the remaining frames. The underlying assumption behind this classical idea is that the appearance of the object remains the same throughout the entire video. This assumption is generally reasonable for rigid objects during a certain period of time, but breaks in the case of non-rigid objects which modify their appearance with time. A naive solution to this problem is to update the template every frame (or every n frames) with a new template extracted from the current image at the current location of the template. The problem with this approach is occlusions. What happens if the template is updated in a frame where the object is occluded? This work will address the two main drawbacks of classical template matching approaches to tracking, namely:

- mismatches between template and object appearance,
- partial and total occlusion of the object.

To do that, the tracker should: (i) update the template to accommodate the change of object appearance and, (ii) detect the occlusion and recapture the object when the occlusion ends.