Available online at www.sciencedirect.com
SCIENCE DIRECT®

Image and Vision Computing 24 (2006) 680–692

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Integrating region and boundary information for spatially coherent object tracking

Desmond Chung [a], W. James MacLean [a,*], Sven Dickinson [b]

[a] Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada M5S 3G4
[b] Department of Computer Science University of Toronto, Toronto, Canada M5S 3G4

Received 23 July 2004; received in revised form 12 August 2005; accepted 9 September 2005

Abstract

The problem of segmenting image sequences based on 2D motion has been under study for many years now. Most early approaches were either region-based, doing some sort of robust motion estimation, or boundary-based, preferring instead to track the bounding contours of the moving image region. In this paper, we explore an approach based on a synergy between these two previous approaches. For example, while motion constraints are often in violation of their underlying assumptions at region boundaries, image edges are a rich source of information. The approach we propose uses feed-forward to use region-based information to propagate boundary estimates, feedback to use boundaries to improve motion estimation, and finally uses motion-based warping to compare image appearance between frames in order to provide additional information for the boundary estimation process.

We show results from an implementation in which a hierarchical, layered-motion estimation using parametric models is coupled with a distance-transform based active contour. The system is shown to provide stable and accurate segmentation in sequences with background motion, and multiple moving objects. Quantitative measures are proposed and reported for these sequences. Finally, a modification is detailed which allows the system to incorporate a Condensation algorithm tracker, but without requiring off-line learning in advance.

Keywords: Segmentation; Motion estimation; Boundary recovery; Parametric models; Motion layers

1. Introduction

The detection and measurement of object motion in image sequences is a central problem in computer vision and video processing. Accurate and reliable estimates of object motion and spatial extent are required for tasks such as video coding, object recognition, object avoidance during navigation, and accurate determination of the observer's motion in an environment where objects may have their own motion independent of that of the observer.

Attempts at motion estimation in image sequences have typically focused on the problems of optical flow computation and motion layer segmentation, yet has paid relatively little attention to recovering accurate boundaries of moving objects. When determining the motion of an image pixel, layered motion approaches generally utilize no information about the motion of neighbouring pixels and, as such, often yield support maps that are highly sparse. On the other hand, the object tracking community has typically focused on tracking the shape of a moving object, often assuming manual initialization of the tracking region, active contour, or model pose (in the case of model-based tracking). Trackers that do not assume an a priori model typically focus on object boundaries while ignoring the rich motion information encoded within the object boundaries.

Each of these paradigms assumes a model of spatial coherence. The motion community seeks to label the pixels defining the region of the moving object, while the boundary-based tracking community seeks to label the pixels defining the boundary of the moving object. Each approach is not without its limitations. Motion constraints can be weak in areas of limited texture, while boundary constraints can be weak in areas of limited contrast. We attempt to bring together these two components in a novel manner to detect, track, and recover the shape of a moving object, effectively drawing on the strength of each component to overcome the weakness of the other. The approach, which is described in detail in Section 3,

* Corresponding author. Tel.: +1 416 946 7285; fax: +1 416 946 8734.
E-mail addresses: desmond.chunglincheung@utoronto.ca (D. Chung), maclean@eecg.toronto.edu (W.J. MacLean), sven@cs.toronto.edu (S. Dickinson).

0262-8856/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2005.09.020

is general, and makes no assumptions about a static background, a static camera, or the number of moving objects. In the following sections, we review related work, provide an overview of the approach, describe the components in detail, and demonstrate the approach on image sequences in which both the object and the background/camera are moving. We conclude with a discussion of the limitations of the approach, along with directions for future research.

2. Related work

The notion of spatial coherence in visual motion analysis has been in the literature for quite some time. Yuille et al. ([1–3]) explore the notions of both spatial and temporal coherence, and present a mathematical theory which claims to provide explanations for a variety of motion perception phenomena, including the aperture problem and motion capture. The theory proposes the notion of a dense velocity field defined even where there is no local image evidence for motion, and which is estimated from image measurements.

Previous work on object segmentation and tracking can be divided into region based approaches ([4–10]) and boundary-based approaches ([11,12]; [13,14], [15]). Among the region-based approaches, some ([6–9]) can be classified as layered approaches, with the latter two using models to describe image regions. In [8], layered flow is computed using octagonal-shaped regions to limit the region of support for a particular motion. In addition to the parameters for each motion, parameters for the size, shape and pose of each region are also computed, as well as a visibility ordering. This method does not attempt to fit an accurate boundary to any region. In [9], an elliptical appearance model is learned and tracked. Both short-term (2-frame) and longer-term structure is represented and tracked. A model-less, layered approach is taken in [10], where the authors perform motion-based segmentation by computing motion parameters for fixed regions, then merging based on adjacency and similarity of motion parameters. Another model-less approach is found in [4,5], where a method for nonparametric flow estimation is given. A Markov random field model is used to provide a prior encoding the notion that neighbouring image points are likely to be related. A mean-field approximation is used to make the method computationally feasible, but the method gives no explicit estimate of the region boundaries.

Boundary-based approaches fall into two categories: probabilistic contour tracking ([11,12]) and active contour approaches ([16]; [13,14], [15]; [17,18]). A probabilistic formulation of curve tracking is presented in [11] that propagates a set of sampled states to approximate a posterior distribution on possible states given the observed data. The method requires an initial curve template and a learning phase to acquire a motion model, after which it can operate uninitialized on new image sequences.

It appears to require a stationary background as this is how the motion model is learned, and no results to the contrary are shown. This approach is furthered in [12] where the motion-learning phase is replaced with an explicit boundary motion model augmented with rough estimates of boundary motion derived from a layered flow model. The model does not directly address the issue of grouping motion edges to identify the boundary of a particular object. The approach in [12] can also be thought of as region-based in that an explicit flow model is provided for non-boundary regions.

The first attempt at integrating region information with an active contour boundary model is found in [16]. Here the authors use a constant, affine or homomorphic warp computed through a correlation approach to compute the displacement of the entire region, and to update the active contour between frames. The active contour is then allowed to settle on image edges. Motion information is not used between frames to update the contour after the initial warp, and as such is not expected to discover evolving object structure except through image edges. All sequences presented by the authors have a single moving object against a static background or a single motion over the entire image.

Perhaps the closest work to that described in this paper is the geodesic active contour formulation, which automatically handles contour splitting and merging based on a single energy function. It is proposed in [13–15], which assumes a static background, so that an image differencing approach can be used to detect motion. Difference images, local intensity statistics and intensity warping within the active contour region are all used to control the active contour. Whereas the geodesic active contour framework focuses on a more elegant active contour formulation (while assuming a simpler motion model), we opt for a more elegant motion formulation while assuming a standard active contour model. As a result, while our active contour implementation is not currently topologically adaptive (although, in principal, we could also employ geodesic active contours), our approach does not assume a static background or a static camera. More recently, work by [17,18] employs level set contours for object boundary tracking, using colour and texture information to drive the active contour, in a manner similar to our proposed intensity constraints.

3. Integrating region and boundary constraints

An overview of the proposed system is shown in Fig. 1. The basic approach is to use a feed-forward, feed-back approach to combine region-based information, in our case motion constraints and intensity-consistency constraints, and boundary-based information, in our case the object boundary as determined by an active contour. Instead of relying solely on either type of information, we use the two types together to improve the results.

In our region-based module, gradient-based motion constraints are used to compute a parametric, layered flow model to estimate local image motion. Since motion constraints are often too sparse to perform proper boundary estimation due to lack of texture, we use the estimated motions to perform an intensity-based consistency check of each pixel against each motion layer. This is achieved by warping each pixel location according to the recovered motions, and comparing image intensities in the two frames accordingly. This yields a denser