Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Computer Vision and Image Understanding 102 (2006) 260–316

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

Erratum

A multi-Kalman filtering approach for video tracking of human-delineated objects in cluttered environments

Jean Gao*, Akio Kosaka, Avinash C. Kak

Robot Vision Lab, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

Received 12 November 2002; accepted 27 October 2004
Available online 4 April 2006

Abstract

In this paper, we propose a new approach that uses a motion–estimation based framework for video tracking of objects in cluttered environments. Our approach is semi-automatic, in the sense that a human is called upon to delineate the boundary of the object to be tracked in the first frame of the image sequence. The approach presented requires no camera calibration; therefore it is not necessary that the camera be stationary. The heart of the approach lies in extracting features and estimating motion through multiple applications of Kalman filtering. The estimated motion is used to place constraints on where to seek feature correspondences; successful correspondences are subsequently used for Kalman-based recursive updating of the motion parameters. Associated with each feature is the frame number in which the feature makes its first appearance in an image sequence. All features that make first-time appearances in the same frame are grouped together for Kalman-based updating of motion parameters. Finally, in order to make the tracked object look visually familiar to the human observer, the system also makes its best attempt at extracting the boundary contour of the object—a difficult problem in its own right since self-occlusion created by any rotational motion of the tracked object would cause large sections of the boundary contour in the previous frame to disappear in the current frame. Boundary contour is

DOI of Publisher's note: 10.1016/j.cviu.2006.02.003.

* Corresponding author. Fax: +1 817 272 3784.
E-mail addresses: jgao@ecn.purdue.edu (J. Gao), kosaka@ecn.purdue.edu (A. Kosaka), kak@ecn.purdue.edu (A.C. Kak).

doi:10.1016/j.cviu.2004.10.012

estimated by projecting the previous-frame contour into the current frame for the purpose of creating neighborhoods in which to search for the true boundary in the current frame. Our approach has been tested on a wide variety of video sequences, some of which are shown in this paper.

Keywords: Tracking; Kalman filtering; Object tracking; Normalized cross-correlation; Perspective; Segmentation; Motion estimation; Recursive motion estimation; Feature extraction; Correspondence problem; Extended Kalman filtering; Boundary extraction; Region growing; Semi-automatic segmentation; Human-in-the-loop segmentation; Video surveillance; Video tracking

1. Introduction

Object tracking has received considerable attention during the past several years [5,15,16,21,24,26,29,31,33]. Applications of object tracking can be found in areas as diverse as video editing for publishing and entertainment, video surveillance, object-based coding for MPEG-4, query formation for MPEG-7, etc.

The approaches suggested so far for object tracking can be classified into automatic and semi-automatic categories. The fully automatic approaches, such as those proposed by [4,14,18,20,23], work mostly for simple objects executing simple motions against clutter-free backgrounds. An example would be a bright light source moving against a uniform dark background. Tracking under such conditions is relatively easy for the obvious reason that the object can be trivially segmented from the background. Automatic methods are not the focus of this paper, since we are specifically interested in complex objects executing complex motions against cluttered backgrounds.

The semi-automatic methods are all based on the rationale that if the human could help out with the initial segmentation of the object to be tracked, the computer could then be relied upon to track the extracted form in subsequent frames. This rationale underlies the many contributions in the semi-automatic category. The published literature on such semi-automatic methods uses two different approaches for motion estimation. While some researchers, such as [5,21,24], perform motion estimation by establishing feature correspondences between the frames of a video sequence, others do motion estimation by first calculating optical flows.

The optical-flow based and the feature-correspondence based methods for motion estimation have their own advantages and disadvantages. Optical flow based methods theoretically treat tracking as a segmentation of the flow field and group together the optical flow vectors that exhibit the same motion [1,27,34]. The tracking performance of these methods depends on the accuracy of the estimated motion field which is error-prone in the vicinity of intensity discontinuities in images [2,10,17,34]. Additionally, tracking can often require dense optical flow fields, which may result in burdensome computations.

With regard to the feature-correspondence based approaches for motion estimation, it is possible to use color, texture, contour, edge, illuminance, etc., for establishing correspondences between successive frames and to then determine the motion model parameters that best fit the entire set of observations [5,21,24,33].