Real-Time Tracking for Visual Interface Applications in Cluttered and Occluding Situations

David Bullock, John Zelek
School of Engineering, University of Guelph
Guelph, Ontario, Canada, N1G 2W1
dbullock/jzelek@uoguelph.ca

Abstract

Visual interface systems require object tracking techniques with real-time performance for ubiquitous interaction. A probabilistic framework for a visual tracking system which robustly tracks targets in real-time using color and motion cues is presented. The algorithm is based on particle filtering techniques of the I-Condensation filter. An innovation of the paper is the use of motion cues to guide the propagation of particle samples which are being evaluated using color cues. This results in a probabilistic blob tracking method which is shown to greatly outperform conventional blob trackers when in the presence of occlusion and clutter. A second innovation presented is the use of motion-based temporal signatures for the visual recognition of an initialization cue. This allows for passive initialization of the tracking system. The application presented here is the task of digital video annotation using a hand-held marking device.

1 Introduction

Computer vision technologies offer a natural medium for human-computer interfaces. People naturally tend to express themselves using gestures, expressions, and actions. There have been extensive research initiatives addressing the problem of real-time tracking for interface purposes. However, most solutions have claimed success by employing one of several simplification techniques: using a constrained and uncluttered environment [1], relaxing the real-time constraint [2], or assuming that minimal occlusion will occur [3, 4]. Unless highly constrained, all visual interfaces will inevitably be faced with ambiguous data. This ambiguity can result from the tracked object being temporarily occluded, agile target movement, or the target becoming camouflaged by clutter that is similar in appearance. A visual interface system must have the ability to identify ambiguous scenarios, and be able to re-acquire the true target once it becomes visible again. Our system is able to maintain robustness even in the presence of ambiguity by employing two powerful tracking techniques. First, particle filtering is used which represents the target posterior distribution as a collection of weighted samples in state space. This is a natural mechanism for maintaining multiple modes (hypotheses) and propagating the uncertainty over time. This is a strength of the Condensation algorithm [5] compared to conventional predictive filters such as the Kalman filter [6] which are limited to uni-modal Gaussian posterior representations. Secondly, our system employs a motion detection algorithm and uses it to redistribute a portion of the samples into high-probability regions of state space using importance sampling [7]. The sample weight represents the likelihood of a particular sample being the true target location and is calculated by comparing the image data to the a priori target color model. The motion information is not used to gage the likelihood of a sample, but rather it only identifies areas in the state-space that should be investigated by samples. This is an effective technique as targets are most commonly lost during periods of movement, and this method allows the system to investigate these regions of movement in a principled Bayesian manner.

A second innovation of this paper is the implementation of a passive initialization scheme for activating the interface device. Vision-based interface devices are typically employed in office or classroom environments in which there may exist extended periods of time between device usages. During these periods of inactivity it is inappropriate to keep the system in the continuous tracking mode, as this can result in temporary misclassification of an individual simply passing through the field-of-view as being a target user. Therefore, a method of initialization and activating the system is required. Ideally this method should not require the explicit intervention of the user by having them hit an activation key. Rather, a passive initialization strategy is presented in this paper in which target individuals are recognized by having them perform a simple pre-defined visual initialization cue which the system is tuned to detect via template