Chapter IX
Subsequence-Wise Approach for Online Tracking

Bin Shen
Tsinghua University, Beijing, China

Yu-Jin Zhang
Tsinghua University, Beijing, China

ABSTRACT

This chapter is concerned with online object tracking, which aims to locate a given object in each of the consecutive frames. Many algorithms have been proposed to deal with this problem. Most make a decision in each frame, failing to consider the inner relationship among these decisions. However, the relation among these decisions is important if they can be fused together. Intuitively, human beings do not make a decision in a single frame and always explore the temporal information contained in neighboring frames, and then make decisions for all these neighboring frames. This chapter proposes a novel framework, which views the tracking as a sequence of decisions, with each subsequence of decisions corresponding to a subsequence of the video. Such an approach is described as a subsequence-wise approach. This framework considers the relation among the states in different frames, and it is ready to be incorporated into many related tracking techniques.

1. INTRODUCTION

With the development of computer vision, online object tracking is becoming a more and more active research area. It plays an important role in a lot of applications, such as area surveillance, navigation, video compression, and human computer interfaces. Besides, tracking paves the way for further process of videos, such as object classification or recognition.

1.1 Existing Approaches

Lots of algorithms have been proposed to deal with this task, from the simple feature point matching method to non-rigid object tracking.

The general idea inside these approaches can be simply described as two steps: (1) make use of the information available to model the target object or both the target and background; (2) decide where the target is in the current frame. For example, particle filter based tracking algorithm (Arulampalam, 2002, Isard, 1998) adopts the information available in the past frames to get the priori probability of the target's state for the current frame, and then the measurements are used to get the posterior probability distribution function via Bayesian Theorem. Based on the posterior the state in the current frame of the target is estimated. Mean shift based tracking algorithm (Cheng, 1995, Comaniciu, 2000) treats tracking as a mode seeking process. The model of the target for tracking is constructed based on the passed frames, and then it deploys the mean shift method to search the optimal mode the current frame. These algorithms make a decision based on the feature extracted from the current frame, while they fail to take into account the constraint among the decisions in consecutive frames, which we call a subsequence. Some other algorithms (Grabner, 2006, Nguyen, 2006) are aimed to explore not only the spatial context of the object but also the temporal spatial context. However, they still make one decision at a time, which is for the current frame, and do not consider the innate relation among the decisions in neighboring frames.

Now, to express our idea clearly, we introduce the term *tracking unit*. Here, we call a repeated component of a video sequence a tracking unit, if and only if the tracking algorithm treats each such component equivalently. For example, these algorithms listed previously treat every frame equivalently, and make a decision in an individual frame, no matter whether they make use of the temporal context. Therefore, the tracking unit for them is a single frame. While experiences tell us that when we humans track a target using our eyes, we do not make decisions about the states in several continuous frames separately. Or rather, we will explore some inner relation among these decisions and fuse them before releasing the decision result. For instance, we are able to estimate the occluded target states if we know how the target enters occlusion and how it gets out. Therefore, it is intuitive that the constraint among decisions in consecutive frames will be helpful for our final better decisions if we can make use of it. From this example, we can further see that choosing a frame as the tracking unit may be not a good choice, for it probably fails to estimate the target's state when it is occluded. Before seeking a better choice for tracking unit, we denote these aforementioned algorithms as frame-wise approach, since they take a single frame as the tracking unit.

1.2 Proposed Approach

As discussed previously, we shall select a new tracking unit rather than a single frame in order to reasonably take into account the relation among decisions in consecutive frames. Naturally, we divide a video sequence into several subsequences, and a subsequence is adopted as the tracking unit in our framework. We denote our tracking framework as subsequence-wise approach. In this framework, we will not only be able to explore both spatial and temporal context in video, but also refine the decisions in the subsequence, and then output a reasonable sequence of decisions for all the frames in a subsequence. This approach can ensure that the set of decisions for a video subsequence is optimal according to some criteria, while the frame-wise approach can only ensure that each of the decisions is optimal to some criteria. Note that, the quality of a tracking algorithm on a video sequence should be measured based on the set of decisions in all frames, neither in a single frame nor in a subsequence, because when tracking a given object, our goal is to find the optimal path through the entire sequence. Apparently, subsequence-wise optimum is a better approximation of global optimum for a sequence than frame-wise