Chapter IX
Occlusion Sequence Mining for Activity Discovery from Surveillance Videos

Prithwijit Guha
Indian Institute of Technology - Kanpur, India

Amitabha Mukerjee
Indian Institute of Technology - Kanpur, India

K.S. Venkatesh
Indian Institute of Technology - Kanpur, India

ABSTRACT

Complex multiobject interactions result in occlusion sequences, which are a visual signature for the event. In this work, multiobject interactions are tracked using a set of qualitative occlusion primitives derived on the basis of the persistence hypothesis: objects continue to exist even when hidden from view. Variable length temporal sequences of occlusion primitives are shown to be well correlated with many classes of semantically significant events. In surveillance applications, determining occlusion primitives is based on foreground blob tracking and requires no prior knowledge of the domain or camera calibration. New foreground blobs are identified as putative objects that may undergo occlusions, split into multiple objects, merge back again, and so forth. Significant activities are identified through temporal sequence mining; these bear high correlation with semantic categories (e.g., disembarking from a vehicle involves a series of splits). Thus, semantically significant event categories can be recognized without assuming camera calibration or any environment/object/action model priors.
INTRODUCTION

Unsupervised activity discovery remains one of the more challenging areas in computer vision. Key problems involve the identification of object and action features, and temporal data mining for abstracting the activity from the data. This chapter presents a set of occlusion features that are tracked over time to yield a surprisingly rich inventory of actions involving interactions among multiple objects.

Object interactions in 3D space often leave their imprint in image space in terms of occlusions. Instead of treating occlusions as a problem, we show that temporal sequences of occlusion phenomena constitute a qualitative signature for large classes of events. In particular, events involving actual contact (e.g., push, embark, hit) necessarily involve overlap in image space for part of the event history. However, many classes of noncontact situations also result in occlusions, and the sequence of such occlusions can (depending on viewpoint) lead to characterization of specific events (e.g., overtaking, crossing). We claim that occlusions among objects constitute an inexpensive and cognitively important cue to reasoning about interactions in space.

Here we explore the limits of what can be learned based on occlusion phenomena. The primary advantage of such an approach is that unlike quantitative approaches using supervised priors for object behavior recognition (Gavrilla, 2005; Haritaoglu, Harwood & Davis, 2000; Zhao & Nevatia, 2004), occlusion signatures do not require any priors for either objects or events. From a cognitive perspective, categorizing events by combining occlusion with other low-level features such as trajectory and segmentation may constitute a key part of the process leading to formation of image schema (Mandler, 1992). Working together with preattentive cues such as image flow and motion, temporal learning in sequences of occlusion phenomena may constitute a prelinguistic model for concept formation for both activities and objects. These links to cognitive processes also reflect computational efficiencies to be gained by focusing attention and avoiding more expensive 3D computations as called for in Granlund (2003).

In order to discover events from sequences of occlusion states, we mine variable-length temporal sequences of occlusion primitives and learn signatures for a wide class of actions. For example, a group of people hugging each other; a person coming on a bicycle, getting off, and going into a building; a crowd of people embarking a tempo; a person boarding and riding a cycle (Figure 1) are all events that have stable signatures in terms of occlusion primitives or O-primitives. The feature

Figure 1. Person boarding and riding a cycle: (a) Cycle static and person moving—both are isolated from each other; (b) person and cycle are in contact—both are static; (c) person and cycle are in contact—both are moving.
Related Content

Both Hands’ Fingers’ Angle Calculation from Live Video
[www.igi-global.com/article/both-hands-fingers-angle-calculation/72312?camid=4v1a](www.igi-global.com/article/both-hands-fingers-angle-calculation/72312?camid=4v1a)

Technology of Music Score Watermarking
[www.igi-global.com/chapter/technology-music-score-watermarking/27054?camid=4v1a](www.igi-global.com/chapter/technology-music-score-watermarking/27054?camid=4v1a)

Inexpensive, Simple and Quick Photorealistic 3DCG Modeling
[www.igi-global.com/chapter/inexpensive-simple-quick-photorealistic-3dcg/77562?camid=4v1a](www.igi-global.com/chapter/inexpensive-simple-quick-photorealistic-3dcg/77562?camid=4v1a)

A Feature Selection Approach for Network Intrusion Classification: The Bi-Layer Behavioral-Based
[www.igi-global.com/article/a-feature-selection-approach-for-network-intrusion-classification/103958?camid=4v1a](www.igi-global.com/article/a-feature-selection-approach-for-network-intrusion-classification/103958?camid=4v1a)