Visual Tracking Using Multimodal Particle Filter

Tony Tung, Kyoto University, Kyoto, Japan
Takashi Matsuyama, Kyoto University, Kyoto, Japan

ABSTRACT

Visual tracking of humans or objects in motion is a challenging problem when observed data undergo appearance changes (e.g., due to illumination variations, occlusion, cluttered background, etc.). Moreover, tracking systems are usually initialized with predefined target templates, or trained beforehand using known datasets. Hence, they are not always efficient to detect and track objects whose appearance changes over time. In this paper, we propose a multimodal framework based on particle filtering for visual tracking of objects under challenging conditions (e.g., tracking various human body parts from multiple views). Particularly, the authors integrate various cues such as color, motion and depth in a global formulation. The Earth Mover distance is used to compare color models in a global fashion, and constraints on motion flow features prevent common drifting effects due to error propagation. In addition, the model features an online mechanism that adaptively updates a subspace of multimodal templates to cope with appearance changes. Furthermore, the proposed model is integrated in a practical detection and tracking process, and multiple instances can run in real-time. Experimental results are obtained on challenging real-world videos with poorly textured models and arbitrary non-linear motions.

Keywords: Depth, Earth Mover, Human, Multimodal, Particle Filter, Tracking, Video, Vision

1. INTRODUCTION

Visual tracking of human body parts is widely used in many real-world applications, such as video surveillance, games, cultural and medical applications (e.g., for motion and behavior study). The literature has provided successful algorithms to detect and track objects of a predefined class in image streams or videos (Yilmaz, Javed, & Shah, 2006; Wu, Lim, & Yang, 2013). Simple objects can be detected and tracked using various image features such as color regions, edges, contours, or texture. On the other hand, complex objects such as human faces require more sophisticated features to handle the multiple possible instances of the object class. For this purpose, statistical methods are a good alternative. First, a statistical model (or classifier) learns different patterns related to the object of interest (e.g., different views of human faces), including good and bad samples. And then the system is able to estimate whether a region contains an object of interest or not. This kind of approach

DOI: 10.4018/ijncr.2014070104
has become very popular. For example, the face detector of (Viola, & Jones, 2001) is well known for its efficiency. The main drawback is the dependence to prior knowledge on the object class. As the system is trained on a finite dataset, the detection is somehow constrained to it. As a matter of fact, most of the tracking methods were not designed to keep the track of an object whose appearance could strongly change. If there is no a priori knowledge on its multiple possible appearances, then the detection fails and the track is lost. Hence, tracking a head which turns completely, or tracking a hand in action remain challenging problems, as appearance changes occur quite frequently for human body parts in motion.

In order to leverage visual tracking under challenging conditions, we introduce a multimodal framework based on the well-known particle filter model (Isard, & Blake, 1998). Our global model integrates various cues such as color, motion, and also depth to perform robust tracking. In addition, Earth Mover distance (Rubner, Tomasi, & Guibas, 1998) has been chosen to compare color models due to its robustness to small color variations, and drift effects inherent to adaptive tracking methods are handled using extracted motion features (e.g., optical flows). As well, an online adaptive process updates a subspace of multimodal templates so that the tracking system remains robust to occlusions and appearance changes.

The tracking system is integrated in a practical workflow containing two modes, switching between detection and tracking. The detection steps involve trained classifiers to update estimated positions of the tracking windows. In our experiments, we use the cascade of boosted classifiers of Haar-like features by (Viola, & Jones, 2001) to perform head detection. Other body parts can be either detected using this technique with ad-hoc training samples, or chosen by users at the initialization step (i.e., pick and track method), or as well can be deduced based on prior knowledge on human shape features and constraints. Our experimental results show accuracy and robustness of the proposed method on challenging video sequences of humans in motion. For example, we use videos of yoga performances (stretching exercises at various speeds) with poorly textured regions, and arbitrary non-linear motions were used for testing (See Figure 1), and also multiple view videos of multiple people interacting during a group discussion in various environments (e.g., meeting room, conference hall) as can be seen in Sect. 5.

The rest of the paper is organized as follows. The next section gives a recap of work related to the techniques presented in this work. Section 3 presents an overview of the algorithm (initialization step and workflow). Section 4 describes the proposed multimodal particle filter framework. Section 5 presents

Figure 1. Body part tracking with multimodal particle filter (using color and motion). Here, body parts located by the tracker are highlighted in green, while regions located by the detector (e.g., face) are highlighted in red. The proposed model is robust to strong appearance changes.
Related Content

A New Differential Evolution Based Metaheuristic for Discrete Optimization
[www.igi-global.com/article/new-differential-evolution-based-metaheuristic/45884?camid=4v1a](www.igi-global.com/article/new-differential-evolution-based-metaheuristic/45884?camid=4v1a)

Rough Set Based Clustering Using Active Learning Approach
[www.igi-global.com/article/rough-set-based-clustering-using/62069?camid=4v1a](www.igi-global.com/article/rough-set-based-clustering-using/62069?camid=4v1a)

Corticofugal Modulation of Tactile Responses of Neurons in the Spinal Trigeminal Nucleus: A Wavelet Coherence Study
[www.igi-global.com/chapter/corticofugal-modulation-tactile-responses-neurons/4969?camid=4v1a](www.igi-global.com/chapter/corticofugal-modulation-tactile-responses-neurons/4969?camid=4v1a)
Simulating Spiking Neural P Systems Without Delays Using GPUs
www.igi-global.com/article/simulating-spiking-neural-systems-without/57968?camid=4v1a