

Multi-face tracking to help AI follow the action

July 11 2018, by Chung-Ching Lin

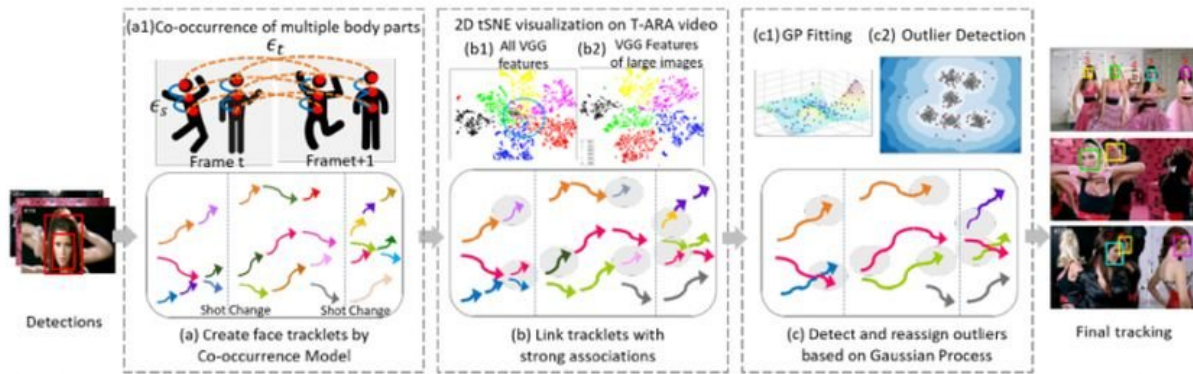


Figure 1. Three core algorithmic components of our method for multi-face tracking in a video sequence. Credit: IBM

At the recent 2018 Conference on Computer Vision and Pattern Recognition, I presented a new algorithm for multi-face tracking, an essential component in understanding video. To understand visual sequences involving people, AI systems must be able to track multiple individuals across scenes, despite changing camera angles, lighting, and appearances. The new algorithm enables AI systems to accomplish this task.

Previous work in this area has largely focused on tracking a single person or several persons within a shot. The next step is to track multiple people throughout a whole video consisting of many different shots. This task is

challenging because people may leave and re-enter the video repeatedly. Their appearances can change drastically thanks to wardrobe, hairstyle, and makeup. Their poses change, and their faces may be partially occluded by viewing angle, lighting, or other objects in the scene. The camera angle and zoom changes too, and characteristics like poor image quality, bad lighting, and motion blur can increase the difficulty of the task. Existing face recognition technologies may work in more constrained cases, where the images are of good quality and show a person's full face, but fail in unconstrained video, where people's faces may be in profile, occluded, cropped, or blurry.

A method for multi-face tracking

Collaborating with Professor Ying Hung, of Department of Statistics and Biostatistics in Rutgers University, we developed a method to spot different individuals in a video sequence and to recognize them if they leave then re-enter the video, even if they look very different. To do this, we first create tracklets for the people present in the video. The tracklets are based on co-occurrence of multiple body parts (face, head and shoulders, upper body, and whole body) so that people can be tracked even when they are not fully in view of the camera (e.g., their faces are turned away or occluded by other objects). We formulate the multi-person tracking problem as a graph structure $G = (v, \epsilon)$ with two types of edges: ϵ_s and ϵ_t . Spatial edges ϵ_s denote the connections of different body parts of a candidate within a frame and are used to generate the hypothesized state of a candidate. Temporal edges ϵ_t denote the connections of the same [body parts](#) over adjacent frames and are used to estimate the state of each individual person in different frames. We generate face tracklets using face-bounding boxes from each individual person's tracklets and extract facial feature for clustering.

The second part of the method connects tracklets that belong to the same person. Figure 1(b) shows 2-D tSNE visualization of extracted VGG-

face feature on a music video. It shows that compared to all features (b1), feature of large images (b) are more discriminative. We build unambiguous connections between tracklets by analyzing the objects' face image resolution and the relative distances of extracted deep features. This step generates an initial clustering result. Empirical studies show CNN-based models are sensitive to image blur and noise because the networks are generally trained on high-quality images. We generate robust final clustering results by using a Gaussian Process (GP) model to compensate for the deep feature limitations and to capture the richness of data. Different from CNN-based approaches, GP models provide a flexible parametric approach to capture the nonlinearity and spatial-temporal correlation of the underlying system. Therefore, it is an attractive tool to be combined with the CNN-based approach to further reduce the dimension without losing complex and import spatial-temporal information. We apply the GP model to detect outliers, remove the connections among outliers and other tracklets, and then reassign the outliers to refined clusters formed after the outliers are disconnected, thus yielding high-quality clusters.

Multi-face tracking in music videos

To evaluate the performance of our approach, we compared it against state-of-the-art methods in analyzing challenging datasets of unconstrained videos. In one series of experiments, we used music videos, which feature high image quality but significant, rapid changes in scene, camera setting, camera movement, makeup, and accessories (such as eyeglasses). Our algorithm outperformed other methods with respect to both clustering accuracy and tracking. Clustering purity was substantially better with our algorithm compared with the other methods (0.86 for our algorithm versus 0.56 for closest competitor using one of the music videos). In addition, our method automatically determined the number of people, or clusters, to be tracked without the need for manual video analysis.

Tracking performance of our algorithm was also superior to state-of-the-art methods for most metrics, including Recall and Precision. Our [method](#) noticeably increased most tracked (MT) and reduced instances of identity switching (IDS) and track fragments (Frag). The video below shows sample tracking results in several [music videos](#). Our algorithm tracks multiple individuals reliably across different shots in the entire unconstrained videos, even though some individuals have very similar facial appearance, multiple main singers appear in a cluttered background filled with audiences, or some faces are heavily occluded. This framework for multi-face tracking in unconstrained video is an important step in improving [video](#) understanding. The algorithm and its performance are described in more detail in our CVPR paper, "A Prior-Less Method for Multi-Face Tracking in Unconstrained Videos."

More information: A Prior-Less Method for Multi-Face Tracking in Unconstrained Videos: openaccess.thecvf.com/content_CVPR_2018_paper.pdf

This story is republished courtesy of IBM Research. Read the original story [here](#).

Provided by IBM

Citation: Multi-face tracking to help AI follow the action (2018, July 11) retrieved 30 April 2024 from <https://phys.org/news/2018-07-multi-face-tracking-ai-action.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
