Persistent People Tracking and Face Capture Over a Wide Area

Gérard Medioni and Yinghao Cai
Institute for Robotics and Intelligent Systems
University of Southern California, Los Angeles, California 90089
{medioni, yinghaoc}@usc.edu

Abstract

Persistent people tracking over a wide area is important for many applications and has received much attention in recent years. For real-time and forensic applications, it is very helpful to capture the facial images of people tracked. In this paper, we address the problem of people tracking and face capture over a wide area. First, we propose a people tracking and face capture system using a single PTZ camera. The availability of high resolution facial images opens up new opportunities for further biometric analysis. Then, we present a new solution to persistent people tracking across multiple cameras with non-overlapping fields of view. The development of these methods should provide new insights into long-term people tracking over a wide area.

1. Introduction

The ever increasing number of cameras installed raises the need for automatic methods for wide area surveillance. One problem associated with wide area surveillance is persistent people tracking, i.e. to maintain identities of people when they move from one camera to another. In forensic applications, it is very helpful to capture the facial images of people tracked as well. However, regular CCTV cameras only capture a small number of pixels on the face region, resulting in poor performance of face detection and recognition. PTZ cameras which can pan, tilt and zoom are powerful tools in far-field scenarios since zooming offers the option of a close view at a high resolution on demand.

We address here the problem of persistent people tracking and face capture over a wide area. First, we propose a people tracking and face capture system using a single PTZ camera. The availability of high resolution facial images opens up new opportunities for further biometric analysis. Then, we present a new solution to persistent people tracking across multiple cameras with non-overlapping fields of view.

2. People Tracking and Face Capture using a PTZ Camera

There are two modes in our proposed system, the zoomed-out mode and the zoomed-in mode. We receive image sequences from the PTZ camera and switch between these two modes to capture faces from multiple people in far-field scenarios.

The system first detects and tracks pedestrians in zoomed-out mode, then selects, using a scheduler, a person to zoom in. After zoom in, we come back to the wide area mode, and solve the person-to-person, face-to-person and face-to-face data association problems.

A. Person Detection

We employ the person detector proposed by Huang and Nevatia [5]. To reduce false alarms from the detector, detected regions which do not contain foreground pixels are filtered out. Given the camera calibration with respect to the world, we further remove false alarms of detections using a prior on human height.

B. Multiple Target Tracking

We combine tracking-by-detection and visual tracking approaches in a unified framework. Reliable trajectories which are obtained by incremental linking of detection responses frame by frame are used to build the appearance model for visual tracking. If detector fails, e.g, there is no detection response at a frame, a separate visual tracker [3] for each person is then triggered to resume tracking. Integrating category information with person-specific information together allows for more accurate tracking results than using either tracking-by-detection or visual tracking approach alone.

C. Camera Scheduling Module

After obtaining the state information of each person in the view, camera scheduling module is then triggered to determine the schedule and allocate resources to observe each person in the view. We only zoom in to people who are moving towards the camera. The camera observes each person in turn and spends the rest of time transitioning between the zoomed-in mode and the zoomed-out mode.

Figure 1. Snapshots of our PTZ people tracking and face capture system on outdoor scenarios. (a) Zoomed-out mode. (b) Zoomed-in mode. (c) Associating faces to people and trajectories.

D. Person-to-Person Association Module

When the camera finishes observing the first person on the schedule list, the camera zooms back out to the wide angle position. The person-to-person association module is then triggered to acquire the next person on the list. The person-to-person association module aims to link tracklets obtained before zooming in with those tracklets obtained right after zooming in. The affinity between tracklets is defined in terms of both location and appearance similarity.

E. Face Detection Module

We employ a real-time face detection proposed by Froba and Ernst [4]. This face detector runs at 20 fps on 640×480 image sequences and can detect faces with less than 45^o out-of-plane rotation.

F. Face-to-Face Association Module

In face-to-face association, face detection responses which belong to the same person are linked together. In order to increase the discriminative power of the association, color and LBP histograms are computed both from the face region and from a region below the face.

G. Face-to-Person Association Module

After zooming in, the system tries to center the image on the face region. Since it might be possible to detect multiple faces in a single zooming in, face-to-person association module is applied to indicate which face is associated with which person. Finally, we annotate the high resolution faces to people and trajectories obtained in the tracking stage illustrated in Figure 1 (c). Please see [2] for more details.

3. People Tracking Across Multiple Cameras

Instead of matching snapshots of people across cameras, we explore what kind of context information from videos can be exploited to track people across multiple cameras. We introduce two kinds of context information here, spatio-temporal context and relative appearance context.

Since one person can not appear at different locations at one time, this spatio-temporal constraint indicates a natural way of collecting positive/negative samples for discriminative appearance learning. Target-specific discriminative appearance models are learned by Adaboost algorithm to effectively distinguish different people from each other.

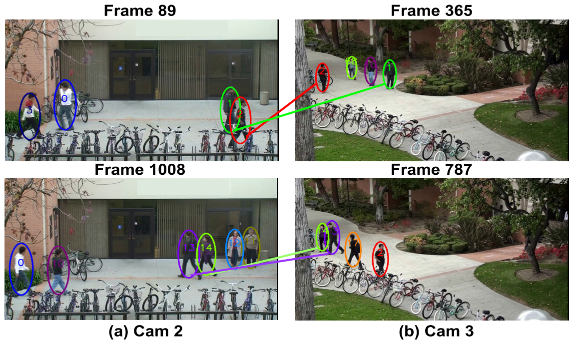

Figure 2. Persistent people tracking across multiple cameras. People moving between cameras are linked with solid lines

On the other hand, people often walk in groups in crowded scenes. If cameras are not geographically far apart, the same sets of people tend to re-appear in neighboring cameras. This group information provides important visual context for individual appearance matching. Relative appearance context models the inter-object appearance similarities for people walking in proximity. Since the imaging conditions within cameras in a short period of time are similar, the appearance differences between pairs of people in proximity under two views are similar. Relative appearance context helps disambiguate individual appearance matching across cameras. We show improved performance with context information for inter-camera multiple target tracking. Please see [1] for more details.

4. Conclusions

We have presented approaches to persistent people tracking and face capture over a wide area which could prove most helpful in forensic applications. Our future work will be focused on persistent people tracking over a wide area with a face ID.

Acknowledgement This research was supported by Award No. 2011-IJ-CX-K054, awarded by the National Institute of Justice, Office of Justice Programs, U. S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

[1] Y. Cai and G. Medioni. Exploring context information for inter-camera multiple target tracking. *WACV*, 2014. 2
[2] Y. Cai, G. Medioni, and T. B. Dinh. Towards a practical ptz face detection and tracking system. *WACV*, 2013. 2
[3] T. B. Dinh, N. Vo, and G. Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments. In *CVPR*, 2011. 1
[4] B. Froba and A. Ernst. Face detection with the modified census transform. In *FG*, 2004. 2
[5] C. Huang and R. Nevatia. High performance object detection by collaborative learning of joint ranking of granules features. In *CVPR*, 2010. 1