

Event Detection in Pedestrian Detection and Tracking Applications

Philip Kelly, Noel E. O'Connor, and Alan F. Smeaton

Centre for Digital Video Processing, Adaptive Information Cluster,
Dublin City University, Ireland

Abstract. In this paper, we present a system framework for event detection in pedestrian and tracking applications. The system is built upon a robust computer vision approach to detecting and tracking pedestrians in unconstrained crowded scenes. Upon this framework we propose a pedestrian indexing scheme and suite of tools for detecting events or retrieving data from a given scenario.

Key words: Pedestrian Detection, Tracking, Stereo, Event Detection

1 Introduction

Robust detection and tracking of humans is a key enabling technology for many applications. It is key to knowing who is where in a scene *and* what their actions have been. It potentially allows other layers in an application's framework to infer beliefs about those people. However, depending upon the end-user application, a variety of event detectors may need to be defined. In this paper, we propose a framework for detecting user-defined events at run-time which may then be used in a variety of surveillance applications.

2 Pedestrian Detection and Tracking

Many of the person detection and tracking techniques described in the literature make assumptions about the environmental conditions, pedestrian and background colour intensity information [1], occlusions [2], the pedestrian flow, pedestrian movements, or that a person will exist within the scene for a significant number of frames un-occluded[3]. We have developed a robust pedestrian detection and tracking system [5] for a single stereo camera that makes *none* of the above assumptions and in addition requires *no* external training. In addition, it is camera viewpoint invariant and is able to handle a large variety of pedestrian appearances, making a single constraining assumption that a person is standing vertically with respect to the groundplane. During tracking, the system can obtain statistics about each pedestrian, such as colour appearance models, the duration of time spent in the camera's field of view and the path they have traversed through the scene. Since stereo information is used, both the 3D position, velocity and the height of a pedestrian can be accurately obtained.

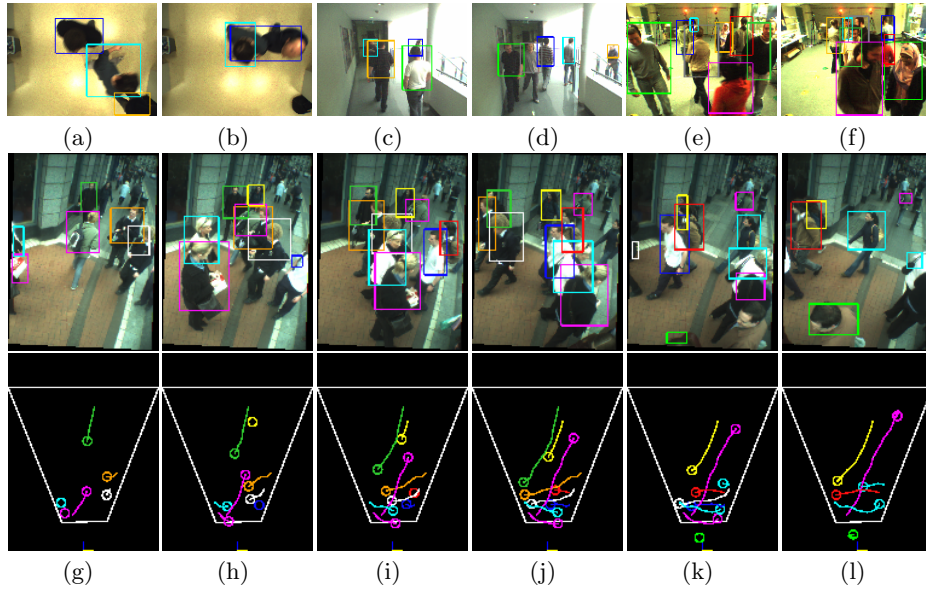


Fig. 1. Examples; (a)-(f) Pedestrian Detection; (g)-(l) Pedestrian Tracking.

Evaluation of this system on 3,500 manually annotated ground truth pedestrians from sequences with varying camera height, camera orientation and environmental conditions reveal extremely accurate performance of our proposed approach. Figure 1(a)-(f) presents some illustrative results where each detected pedestrian is enclosed by a bounding box of a certain colour. An example of a tracking sequence can be seen in figure 1(g)-(l), where the second row of images depict the scene from a *plan-view* or *birds-eye view* orientation. In these plan-view images, the white lines indicate the bounds of the scene, the position of detected pedestrians in that frame are illustrated by a circle of the same colour as their bounding box, and tracks are depicted as “tails” from the centre of the circle to previous positions in the scene.

3 Event Detection

Beyond detection and tracking, many applications require a further level of processing whereby the *actions* of pedestrians must be interpreted. For some applications these event detectors can be hard-coded into the application framework, for example surveillance applications which determine pedestrian flow densities during specific time periods. For many applications, however, the exact event detector *cannot* be hard-coded into the system as; (1) the event definition is dependent on an undefined scene; or (2) the event is itself undefined. The first scenario is typical of many Ambient Intelligence (AmI) applications [6], as although the event required to be detected is known (e.g. detect pedestrians in

a designated area waiting to cross the road), information about the scene (e.g. the exact designated area) is unknown. The second scenario is typical of general purpose surveillance applications where events are determined by a user for detection either at run-time or *after* the event has occurred. Take for example a typical CCTV surveillance system, if the surveillance video is augmented with robust pedestrian tracking and statistical information it becomes possible to quickly search the video for specific events via the augmented meta-data. For example, if a lost child’s appearance is known (e.g. the colour of their clothes and their height) then all possible detections of the child can be subsequently flagged from every camera in the system. Other examples include adding “watched” areas in a scene (e.g. to determine the number of pedestrians who entered a specific shop) or flagging “unusual” events such as lingering pedestrians (e.g. to review all footage of persons who were in the vicinity of an area that was vandalised for a significant amount of time).

4 Proposed Approach

Building on our prior work, we have developed a pedestrian indexing scheme and suite of tools for detecting events and retrieving data from a given scenario. One such tool is the creation of 3D *hotspot* regions from 2D plan-view images - see the 2D yellow coloured area in figure 2(a). In this figure, it should be noted that a background colour model has been projected onto the image-plane to allow the gauging of distance and orientation within the plan-view image. Using this plan-view image, a hotspot is simply created in the proposed system by circling a region of interest within the plan-view image. The resultant hotspot can be seen as a 3D area of interest which can be incorporated into the definition of events for some application scenarios. For example, in the framework of an automated pedestrian traffic light system it can be used to define an area where pedestrians tend to wait before crossing the road – see figures 2(b)-(f). A detected event, which in turn leads to a changing of the traffic lights, could then be defined using this hotspot region and an event (or query) syntax. An example of the syntax is, $Event(cross) = hpt(1), p(g5), t(10s)$, which declares an event called *cross* to be detected if on hotspot number 1 ($hpt(1)$) there are greater than 5 pedestrians ($p(g5)$) who have been waiting for more than 10 seconds ($t(10s)$). The proposed system can be easily used across scenes with differing underlying structure whilst the event syntax allows a number of user-defined events to be created and searched for during run-time in content-based retrieval applications.

5 Future Work

These techniques will underpin a user-defined event detection system which can be simply tailored for a suite of pedestrian detection and tracking applications. Currently the event syntax consists of a number of features including; hotspots, timings (such as the length of time an object is in the scene), and pedestrian numbers. However, the syntax should be extended to incorporate other features,

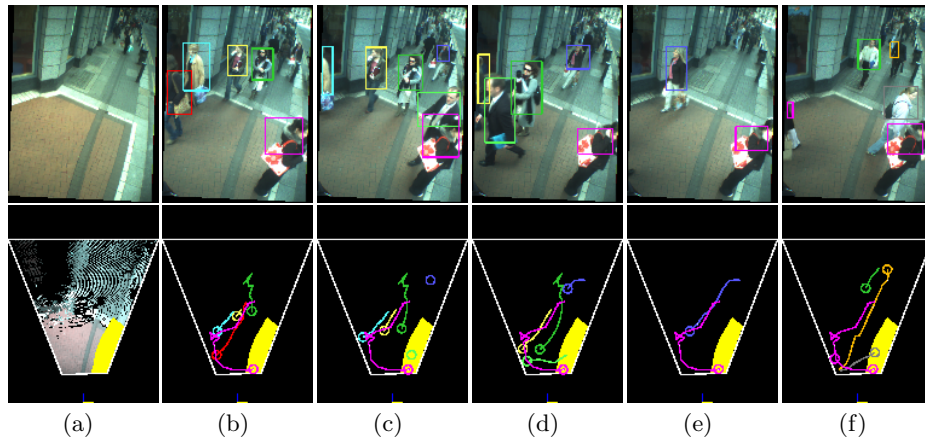


Fig. 2. Pedestrian crossing application; (a) Hotspot; (b)-(f) Waiting to cross the road.

including; time of day, pedestrian interactions (such as pedestrians walking in a group or on their own) and pedestrian statistics such as colour, height, velocity and position).

Acknowledgements

This material is based on works supported by Science Foundation Ireland under Grant No. 03/IN.3/I361. The authors wish to acknowledge the support of the European Commission under the FP6-027026-K-SPACE contract. This paper was awarded a best poster prize at the 2007 K-Space Jamboree Workshop for PhD students.

References

1. A.W. Senior, "Tracking with probabilistic appearance models," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance Systems*, June 2002, pp. 48–55.
2. M. Harville, "Stereo person tracking with adaptive plan-view templates of height and occupancy statistics," *International Journal of Computer Vision*, vol. 22, pp. 127–142, 2004.
3. A.E Elgammal and L.S. Davis, "Probabilistic framework for segmenting people under occlusion," in *IEEE International Conference on Computer Vision*, 2001, vol. 2, pp. 145–152.
4. P. Kelly, E. Cooke, N. E. O'Connor, and A. F. Smeaton, "Pedestrian detection using stereo and biometric information," in *International Conference on Image Analysis and Recognition*, September 2006, pp. 802–813.
5. P. Remagnino and G.L. Foresti, "Ambient intelligence: A new multidisciplinary paradigm," in *IEEE Transactions on Systems, Man and Cybernetics*, January 2005, vol. 35, pp. 1–6.