Detection of people in military and security context images

Thomas Shannon, Ben Wiltshire, and Emmet Spier

An autonomous person detection solution could help alert surveillance operators to potential issues, reducing the cognitive burden and achieving more with less manpower.

A major challenge to the use of aerial platforms for gathering intelligence is the restricted line of sight in cluttered and congested urban environments. It is likely that ground-based acquisition systems will become more prevalent in future operations, exploiting new and emerging electro-optic surveillance technologies. Recent operations in Northern Ireland, Iraq, and Afghanistan by British forces and in Somalia by US troops have shown that for ground forces to fight effectively in built-up areas, or to act as aids to the civil power, they must have access to current and pertinent intelligence about potential threats.

Previous research detecting objects in images has employed computer vision feature descriptors, such as histograms of oriented gradients (HOG) using the support vector machine (SVM) approach (learning models with algorithms for data and pattern analysis). Although this method is robust in detecting persons in images of limited quality, it fails in cases where the individual is partially occluded or overlaps with another subject.

To address this challenge, we considered how a person could be partly hidden by physical structures, by handling personal infantry weapons, or by the tactical pose they had adopted. We applied current computer vision techniques to achieve reliable detections within 2D images by investigating an approach described by Felzenszwalb and coworkers. This technique is based on the construction of cascaded, non-linear classifiers from part-based deformable models. In contrast to the HOG-SVM method based on a hard decision algorithm, this new approach uses a probabilistic framework for object class detection. The published improvements in detection performance compared to existing methods (reduced false positives and false negatives) in the presence of partial occlusion show that this approach holds promise. Figure 1 depicts an example problem and solution as described by a bounding box and classification.

Figure 1. An example of occluded person detection.

We first evaluated the baseline performance of this approach using 170 images of 345 labeled and upright pedestrians at different scales, orientations, actions, and degrees of occlusion within the Penn-Fudan image data set. Then, in 431 images, we established how well the method detected experienced volunteers undertaking low-level infantry tactics or innocent civilian activities in the open, and when obscured by structures.

We identified detections in each image by a bounding box and a confidence score corresponding to the classifier’s assessment of a person’s presence within the box. If a box registered the ground truth presence of a person, then it was defined as a true positive detection, otherwise it was regarded as a false positive. Based on the confidence score we were able to filter the results to obtain a desirable performance and a trade-off between the rate of false and true positives. We defined false negative detections when the overlap with the ground truth annotations

Continued on next page
Figure 2. True positive and false positive detections. (Blue box = ground truth. Red Box = detection.)

Figure 3. False negative outcomes.

was less than 50%. Figure 2 depicts examples of positive results. Figure 3 shows two false negatives. (Blue box = ground truth. Red box = detection.)

Our findings concluded that the method of Felzenszwalb and coworkers could potentially provide a useful person detection tool, yielding a precision of approximately 70% for a recall rate of around 85% when applied to our military and security context imagery. We are continuing to work on improving detection speed per image, which is currently between 5 and 10 seconds, depending on processor performance and image pixel resolution.

Bounding the location of people in a scene could enable us to apply emerging 2D pose classification algorithms to identify both the likely activity and the possible intent. Our work is now focused on further improving detection by tracking, and through the use of background removal for those applications using fixed cameras.

This research was supported by the Defence Science and Technology Laboratory, UK.

Author Information

Thomas Shannon, Ben Wiltshire, and Emmet Spier
2d3 Sensing
Oxford, UK

Tom Shannon is an ex-infantry field officer with more than 30 years’ experience as a practicing chartered professional engineer, medical physicist, and computer vision scientist. His focus is on clinical and machine vision applications applied to the analysis of human motion and shape.

Ben Wiltshire has worked in the defense industry for 10 years developing and testing algorithms for a wide range of applications, including command and control software, missile guidance systems, and image analysis, while also completing a PhD in remote sensing and computer vision.

Emmet Spier has 15 years of experience in industry and academia developing computer vision and machine learning systems. Recently he was a lead engineer for a new motion imagery exploitation tool, and technical lead for several defense enterprise projects.

References