Carried Object Detection and Tracking using Geometric Shape Models and Spatio-Temporal Consistency

Aryana Tavanai, Muralikrishna Sridhar, Feng Gu, Anthony G. Cohn, and David C. Hogg

School of Computing, University of Leeds
Leeds, LS2 9JT, United Kingdom
{fy06at,scms,f.gu,a.g.cohn,d.c.hogg}@leeds.ac.uk*

Abstract. This paper proposes a novel approach that detects and tracks carried objects by modelling the person-carried object relationship that is characteristic of the *carry event*. In order to detect a generic class of carried objects, we propose the use of geometric shape models, instead of using pre-trained object class models or solely relying on protrusions. In order to track the carried objects, we propose a novel optimization procedure that combines spatio-temporal consistency characteristic of the carry event, with conventional properties such as appearance and motion smoothness respectively. The proposed approach substantially outperforms a state-of-the-art approach on two challenging datasets PETS2006 and MINDSEYE2012.

1 Introduction

Detection and tracking of carried objects is an important component of vision systems whether these are surveillance systems that aim to detect events such as leaving, picking up or handing over a luggage, or robots that learn to perform better in indoor environments by analysing events where humans manipulate carried objects. Despite significant progress in object detection and tracking, the task of detecting and tracking carried objects well enough to be able to use them for activity analysis is still a challenging problem. This task is elusive due to the wide range of objects that can be carried by a person and the different ways in which carried objects relate to the person(s) carrying it e.g. carrying, dropping, swinging, picking it up, occluding etc.

An early approach [2] demonstrated that pre-trained object-class models for specific types of objects may be useful in domains where the variety of carried objects is relatively small and is known in advance, the objects are of sufficient size and there is limited clutter in the background. To generalise to a more realistic setting, researchers have focused on *indirect* ways of characterising carried objects, which first aim to identify the person region and background and then attempt to explain the remaining regions in terms of carried objects. The first of these approaches looked for carried objects in protrusions which are regarded

* The financial support of the EU Framework 7 project Co-RACE (FP7-ICT- 287752), and the DARPA Mind's Eye program (project VIGIL, W911NF-10-C-0083) is gratefully acknowledged.

as the part of foreground that is different from the person region. This approach evolved starting from an early work - *Backpack* [6] - that proposed temporal templates as a way of characterising the person region. Subsequent researchers have extended this approach by introducing refinements - such as modelling variances from the temporal templates [1] and 3-D exemplar temporal templates corresponding to different viewpoints of a walking person together with spatial priors in a very recent work [4]. Other indirect approaches have built a pre-trained appearance model of *persons without carried objects* and they detect *person carrying objects* as anomalies [9].

We propose a novel approach for carried object detection and tracking with the following contributions. (1) we perform object detection by using geometric shape models to characterise carried objects. In this way, we avoid using specific pre-trained object class models as in [2]. (2) our approach integrates detection and tracking by incorporating normal motion properties that apply generically to most carried objects such as spatio-temporal smoothness that have been widely used in the tracking literature, but have not been exploited for the carried object task. (3), and most importantly, we propose a novel approach for carried object detection and tracking by characterising carried objects given that only the *carry* event occurs i.e. that these objects follow a person's trajectory with a temporally continuous and characteristically consistent spatial relationship with respect to the person. Accordingly, we introduce an optimisation strategy that starts with a small set of detections with possibly false positives and increasingly incorporates a learned person-object spatial relationship that characterises the carry event. This procedure starts building longer tracks that tend to approximate the true carried object trajectory, while also rejecting the false positives. The learned spatial relationship leads to significant improvement compared to using a static spatial prior [4]. §5 shows that the proposed approach significantly improves the performance over a state-of-the-art carried object detector [4] on the PETS2006 and MINDSEYE2012 (www.visint.org) datasets. Dataset and code can be found at: http://www.engineering.leeds.ac.uk/computing/research/vision/CODT.

2 Proposed Formulation

We consider a video $\mathcal{I}$ which is a time series of images $\{I^1, ..., I^t, ..., I^N\}$. For this video, we obtain a corresponding sequence of foreground regions $F = \{f^1, ..., f^t, ...f^N\}$ and a set of person tracks $P = \{p_1, ..., p_M\}$. Here a person track $p_i \in P$ is a time series of *segmented person regions* $\{..., p_i^t, ...\}$. In addition, we define $\mathcal{R}$ as a set of candidate object regions, from which a set $\mathcal{O}$ of all possible candidate object tracks may be sampled. We describe the procedure for obtaining the foreground, person tracks, object detections in §4.

In this work, we make the simplifying assumption that *carry* is the only event that governs the relationship between a person p_i and an associated set of carried object tracks $O \subseteq \mathcal{O}$ i.e. the carried objects are not picked up, dropped or given to another person. That is, if a carried object track $o_j \in O$ is associated with a person track p_i, then there exists a bijective relationship between the corresponding regions $o_j^t \in o_j$ and $p_i^t \in p_i$. We also assume that the carried object tracks are independent of each other.

Under these assumptions, our task is to find a set of carried object tracks O associated with each person track p_i. Accordingly, for each person track p_i we formulate our task as finding an optimal set of carried object tracks $\hat{O}$ that maximises the following objective.

$$\hat{O} = \arg\max_{O \subseteq \mathcal{O}} \prod_{o_j \in O} \mathcal{P}(o_j|\Theta_{\mathsf{O}})\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})\mathcal{P}(o_j|\Theta_{\mathsf{S}}) \quad (1)$$

In the above equation, the probability distribution $\mathcal{P}(o_j|\Theta_{\mathsf{O}})$ prefers tracks that consists of regions which correspond to certain geometric shapes, as detailed in §2.1. The probability distribution $\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})$ models the person-object relationship that is characteristic of the carry event (§2.2). The probability distribution $\mathcal{P}(o_j|\Theta_{\mathsf{S}})$ parametrised by the smoothness model Θ_{S} in the above equation regards a track o_j being more likely, if the sequence of carried object regions constituting this track are smooth with respect to motion and appearance and if it has other desirable properties such as minimum overlap with other tracks, minimum gap and maximum possible length. These measures are computed similarly to [15].

2.1 Geometric Object Shape Models $\mathcal{P}(o_j|\Theta_{\mathsf{O}})$

We regard a candidate object track $o_j \in O$ as more likely to be a carried object if the shape of the region is likely to be any of the pre-defined generic geometric shapes. The distribution $\mathcal{P}(o_j|\Theta_{\mathsf{O}})$ in equation 1 measures this likelihood with respect to a set of geometric shape models Θ_{O}. Assuming independence between an object region o_j^t and the rest of the object regions in an object track o_j, we factorise the likelihood $\mathcal{P}(o_j|\Theta_{\mathsf{O}})$ as $\mathcal{P}(o_j|\Theta_{\mathsf{O}}) = \prod_{o_j^t \in o_j} \mathcal{P}(o_j^t|\Theta_{\mathsf{O}})$. We marginalise across each of the object shape models $\theta \in \Theta_{\mathsf{O}}$ and assume a uniform prior distribution $\mathcal{P}(\theta)$ across these models to obtain $\mathcal{P}(o_j^t|\Theta_{\mathsf{O}}) = 1/|\Theta_{\mathsf{O}}|\Big(\sum_{\theta \in \Theta_{\mathsf{O}}} \mathcal{P}(o_j^t|\theta)\Big)$.

We consider a convex shape model with parameter $\theta_c \in \Theta_{\mathsf{O}}$ and an elongated shape model with parameter $\theta_e \in \Theta_{\mathsf{O}}$ since many carried objects have a shape that is approximately convex (e.g. briefcases, suitcases, petrol cans) or elongated (e.g. objects with an elongated part - shovels, guns, brooms). We evaluate the probabilities $\mathcal{P}(o_j^t|\theta_c)$ and $\mathcal{P}(o_j^t|\theta_e)$ for the convex and elongated model as an exponential distribution $1/z_0 \exp(\theta_c \mathcal{C}(E(o_j^t)))$ and $1/z_1 \exp(\theta_e \mathcal{E}(E(o_j^t)))$ over a convexity measure $\mathcal{C}(E(o_j^t))$ and a parallel measure $\mathcal{E}(E(o_j^t))$ respectively. Here, $E(o_j^t)$ refers to the set of edges that form the boundary of the object region o_j^t. In §4, we describe our novel level-wise mining approach for extracting the set $\mathcal{R}$ of candidate object regions, where each such region is formed by a set of edges. We compute the degree of convexity $\mathcal{C}(E(o_j^t))$ for a region o_j^t, using the method in [16]. In order to compute the degree of parallelism, $\mathcal{E}(E(o_j^t))$, we only consider those candidate sets of contour segments $E(o_j^t)$ which can be partitioned into two non-overlapping proximal groups of nearly co-linear contour segments, that are roughly parallel to each other. We combine a measure of co-linearity [13] within each group with the degree of parallelism across the two groups.

2.2 Person-Carried Object Relationship $\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})$

We regard a candidate object track $o_j \in O$ as more likely to be a carried object associated with a person p_i if: (i) the track o_j follows p_i's trajectory with spatio-

temporal consistency characterised by the *carry event*; (ii) the object regions $o_j^t \in o_j$ overlaps with protrusions corresponding to the person region $p_i^t \in p_i$. Both these person-carried-object relationships are modelled by the probability distribution $\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})$ with carriedness parameter set Θ_C. Given model parameters θ_r for protrusions, θ_s for spatio-temporal consistency and the foreground regions F, we factorise this distribution whose two terms that capture person-object spatial relation and protrusions respectively, as explained below.

$$\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}}) = \prod_{o_j^t \in o_j} \mathcal{P}(o_j^t|p_i^t, \theta_s)\mathcal{P}(o_j^t|p_i^t, f^t, \theta_r)$$

Person-Object Spatial Relation. A novel way of characterising carried objects given that only the *carry* event occurs is that they follow a person's trajectory with a temporally continuous and characteristically consistent spatial relationship with respect to the person. To quantify this, we propose a voting measure that counts the number of times the relative position of a pixel with respect to the centroid of a person's region falls within a detection.

Let $dx_{p_i^t}, dy_{p_i^t}$ be the offset of a pixel relative to the centroid $(x_{p_i^t}, y_{p_i^t})$ of the i'th person's bounding box p_i^t at time t i.e. $(x_{p_i^t} + dx_{p_i^t}, y_{p_i^t} + dy_{p_i^t})$ is the absolute position of the pixel relative to the image frame I^t. We define a function $\delta(dx_{p_i^t}, dy_{p_i^t}, o_j^t)$ as follows.

$$\delta(dx_{p_i^t}, dy_{p_i^t}, o_j^t, i) = \begin{cases} 1, & \text{if } (x_{p_i^t} + dx_{p_i^t}, y_{p_i^t} + dy_{p_i^t}) \in o_j^t \\ 0, & \text{if } (x_{p_i^t} + dx_{p_i^t}, y_{p_i^t} + dy_{p_i^t}) \notin o_j^t \end{cases}$$

Using the above definition we define the heatmap H of a relative offset $(dx_{p_i^t}, dy_{p_i^t})$ position as the following.

$$H(dx_{p_i^t}, dy_{p_i^t}) = \sum_{o_j \in O} \sum_{o_j^t \in o_j} \delta(x_{p_i^t} + dx_{p_i^t}, y_{p_i^t} + dy_{p_i^t}, o_j^t, i)$$

Given a set of tracks O associated with a person p_i, the intensity values in the heatmap measure the number of votes for each relative offset pixel $(dx_{p_i^t}, dy_{p_i^t})$ given by the tracks in O. Since we expect carried objects to have a consistent relative location with respect to the person and noise to be more randomly distributed, the heatmap captures the locations relative to the person where carried objects is most likely to exist. This is as a result of these locations receiving higher votes in the heatmap due to the repeated presence of potential carried objects even though they may be sparsely detected in the video.

We regard a detection o_j^t as more likely to be a carried object if it covers pixels with high intensity values in the heatmap. We model the *relative positional probability* $\mathcal{P}(o_j^t|p_i^t, \theta_s)$ as follows.

$$\mathcal{P}(o_j^t|p_i^t, \theta_s) = \frac{1}{z_3} \exp\Big(\theta_s \sum_{(x,y) \in o_j^t} H(x - x_{p_i^t}, y - y_{p_i^t})\Big) \tag{2}$$

This distribution tends to get closer to the true distribution of the carried objects' relative location with respect to a person (Fig. 1) with the increasing number of true detections over the false detections, as further described in §3.

Protrusions. Areas corresponding to protrusions have been shown to be likely carried object regions with respect to the region of the person carrying it. For each person region p_i^t, we obtain a protrusion region α_i^t by subtracting the person region p_i^t from the foreground region f^t in frame I^t and considering only

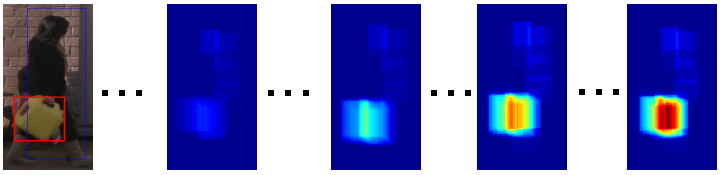


Fig. 1: An illustration of the learned spatial distribution of the object relative to the person approximates the true relative position in the leftmost figure.

a subregion of α_i^t in the vicinity of the person (defined by the detected person bounding box). We regard a region o_j^t as more likely to be a carried object if it overlaps significantly with α_i^t. Accordingly we compute the degree of overlap $\mathcal{V}(\alpha_i^t, o_j^t) = (\alpha_i^t \cap o_j^t)/(\alpha_i^t \cup o_j^t)$ and then evaluate $\mathcal{P}(o_j^t|p_i^t, f^t, \theta_r)$ using an exponential model $1/z_2 \exp(\theta_r(\mathcal{V}(\alpha_i^t, o_j^t))$.

3 Event Driven Optimisation.

We now describe the main novelty of the paper which is an event driven optimisation. According to this scheme, the optimal solution of the objective function in equation 1 emerges as a result of iterations which involve cyclic interactions between the two components of the objective function. We define the first component, $\mathcal{P}(o_j|\Theta_{\mathsf{O}})\mathcal{P}(o_j|\Theta_{\mathsf{S}})$, as a product of the probability distributions corresponding to the detection strengths and spatio-temporal continuity respectively. The second novel component $\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})$ is the relative positional probability distribution that models the person-object spatial relationship which is characteristic of the *carry event*.

We first describe the basic search procedure in the optimisation process before discussing the role of these two components. For each person track p_i, the optimisation involves starting with an initial set of tracklets O^0 and then applying a sequence of moves to iteratively obtain a sequence of hypothesised tracklets $(O^1, ..., O^k, ...)$. The objective function given in equation 1 is used at each step k in the iteration to decided whether to accept the new hypothesis O^k or to persist the previous hypothesis O^{k-1}. We adopt two simple moves, (i) form larger tracklets from smaller ones by randomly choose a tracklet and then linking this tracklet to a neighbouring tracklet, which is chosen uniformly at random (u.a.r) from the set of neighbouring tracklets. (ii) split larger tracklets into smaller ones by choosing a tracklet u.a.r from the set of tracklets and then selecting a location along this chosen tracklet u.a.r and finally breaking it into two smaller tracklets at this location. After a relatively large number of iterations, we terminate the optimisation process and regard the final set of tracklets of length more than one as the optimal set of carried objects $\hat{\mathcal{O}}$. In the following we first introduce the basic tracking system to which we add the contribution of the heatmap and an attention-like mechanism leading to three variants of the optimisation process. We evaluate each of these variations in the experimental section.

Basic Tracking System (BTS). When this procedure is used *only* with the first component, it tends to result in carried object tracks that have higher detection probabilities $\mathcal{P}(o_j|\Theta_{\mathsf{O}})$ and are smooth with respect to the properties captured in $\mathcal{P}(o_j|\Theta_{\mathsf{S}})$. We call such a system as the basic tracking system, that we refer to in our experimental section.

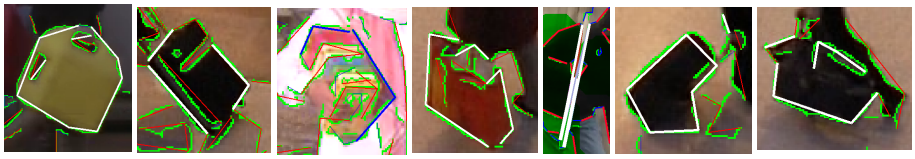


Fig. 2: Examples of using Geometric Shape Models for carried object detection. Green segments represent edges from the Canny edge detector and the solid convex/elongated objects mined in a level-wise fashion.

Heatmap Driven System (HDS). The introduction of the 2nd component i.e. the relative positional probability distribution $\mathcal{P}(o_j|p_i, F, \Theta_{\mathsf{C}})$ tends to favour the formation of object tracks whose objects firstly overlap with protrusions, and secondly (more importantly) those tracks that overlap with the heatmap given in equation 2. That is these tracks tend to accumulate higher values of the positional probability distribution and therefore have the characteristics of a carried object, as described in §2.2.

Attention Driven System (ADS). To further capitalise on the potential of this relative positional probability distribution, we introduce an *attention-like mechanism* into the optimisation process, where we start by considering only those object detections that have high detection likelihoods and we call these initial tracklets of length one as initial seed tracklets. At each iteration, the link move forms larger seed tracklets by *focussing* on connecting *only* seed tracklets to other seed tracklets or non-seed detections (tracklet of length one). Similarly the split move operates only on the seed tracklets.

At each iteration, only the seed tracklets contribute to the computation of the heatmap. As the heatmap becomes more well defined with further iterations, some of those non-seed tracklets with higher positional probability distributions (although they may have relatively lower detection likelihoods) tend to be included as seed tracklets. These updated seed tracklets are used for applying moves in the next iteration. In this manner, an attention-like mechanism begins to evolve with a tendency to *select* object tracklets that correspond to the true carried objects, *against* other false positive candidate tracks. Due to the cyclic interactions between the two components of the objective function, the optimisation often starts with a sparse set of detections with possibly several false positives and starts building longer tracks that tend to approximate the true carried object trajectory, while rejecting the false positives.

4 Object Detection

We now describe the procedure for obtaining foreground, person regions and carried object detections (Fig. 2) respectively. We start by computing a sequence of foreground regions for a video using an off the shelf foreground extraction technique [11]. We then obtain person tracks by detecting a set of person regions in each frame and then we track all these detections using a dynamic programming based tracker [10]. The person regions in each frame are obtained in three steps. First we detect bounding boxes corresponding to the person detections obtained using a standard object detector with a trained person model [5]. Second, we obtain bounding boxes that are body part estimates inside each of the person

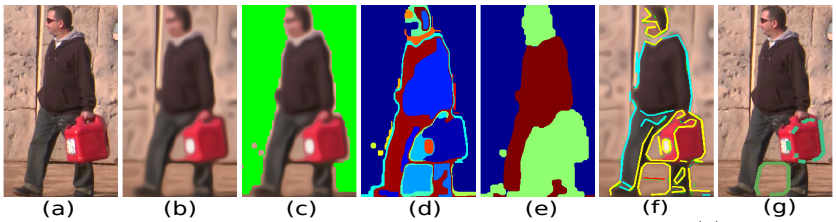


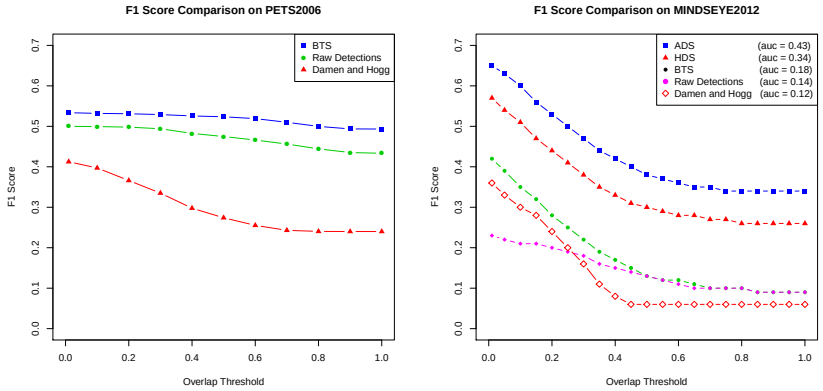
Fig. 3: The process of obtaining candidate carried object detection. (a) We first obtain the image corresponding to the person detection; (b) We then apply the method in [8] to enhance edges corresponding to natural boundaries; (c) We apply foreground extraction on b (background shown in green); (d) We apply colour based segmentation to c; (e) We identify the two largest segments (given in red) in d, which tend to correspond to regions on the person. The carried object is more likely to be present in the non-person regions (shown in green); (f) Using the regions identified in e, many of the line segments belonging to the person are removed (coloured with cyan); (g) The result of applying level-wise mining to the remaining edges (coloured yellow in f) to obtain candidate carried object regions (coloured in green), as an input to the event driven optimisation.

bounding boxes using articulated pose estimation code [14]. Finally we take the union of the regions circumscribed by each part to be a segmentation of the person.

In order to find likely candidates for carried object detections, we first remove a majority of line-segments that form the boundaries of persons but not of the objects using a procedure illustrated in Fig. 3 (a-f). This approach drastically reduces the set of line segments enabling us to generate a smaller set of candidate object regions from the remaining set of line segments. We then search this set for candidate object detections o_j^t, where each detection is just a subset of line segments forming a fully or partially connected chain (Fig. 3.g), that are likely to belong to any of the geometric shapes under consideration. In order to search efficiently, we use a level-wise mining procedure, where two candidate $k-1$ subsets are merged if they share $k-2$ segments and accepted as a k candidate set o_j^t, if the likelihood score $\mathcal{P}(o_j^t|\theta)$ of o_j^t, with respect to a geometric shape model $\theta \in \Theta_{\mathsf{O}}$ is above a minimum conservative threshold.

5 Experimental Setup

The experimentation consists of two aspects, first of which is a comparison between the proposed approach and the state-of-the-art protrusion based Damen and Hogg's carried object detector [4], henceforth DHD. Secondly, we would like to further explore the true potential of our approach, by alternating certain key components and identifying their effects in terms of detection performance. As a result, a benchmark dataset, namely the PETS2006 dataset is used for the first aspect of baseline comparison. On the other hand, a much more complex dataset, the MINDSEYE2012 dataset, is used in a set of more extensive experiments, which are aimed at the exploration of key components of the proposed approach. The corresponding evaluation is concentrated on the detection performance of the compared approaches and thus it is done with respect to spatio-temporal localisation of each carried object per frame by computing the



(a) PETS2006 (b) MINDSEYE2012

Fig. 4: Result plots of compared methods in terms of F1 scores as the threshold of overlap increases on both PETS2006 and MINDSEYE2012.

standard overlap ratio also used in [4], except that we also vary the overlap threshold and report results for each value.

Datasets. All seven videos of the third camera were chosen, due to its view angle, for PETS2006, similar to [4]. Overall 70 video clips were created by a third party from the MINDSEYE project year 2 dataset, with an average length of 200 frames. The complexity of this dataset results from variations in camera settings, environmental factors, e.g. changes in light conditions (e.g. brightness due to weather), moving trees and grasses in the background, as well as a greater variety of carried object types. The ground-plane homography estimation of PETS2006 was provided as part of the sample set, while that of MINDSEYE2012 is done for each camera setting. Human tracks of both datasets are generated through first applying basic background subtraction to obtain foreground segmentation and then using an off-the-shelf tracker [10].

Parameter Settings. In our experiments, we tune the parameter set Θ_{O} (corresponding to the geometric shape models), Θ_{S} (modelling smooth trajectories), and $\theta_r \in \Theta_{\mathsf{C}}$ (concerning the overlap between the protrusion and the object mask respectively), on a separate subset of the Mindseye project. Values of these parameters are independent from any particular selection of subset, containing a reasonable number of videos. This is because general geometric properties Θ_{O} (e.g. convexity) are invariant across samples from any dataset. As focus of this work is to prove a concept, only the convex shape model is investigated. Similarly Θ_{S} are generic due to similar motion patterns in the datasets (e.g. people walking). Finally, for $\theta_r \in \Theta_{\mathsf{C}}$, irrespective of the dataset and the perspective, it is reasonable to assume that the protrusion mask corresponds to a part or whole of the object. This is due to the assumption that the person and the carried object together constitute the foreground mask. In addition, we set the parameter $\theta_s \in \Theta_{\mathsf{C}}$ equal to 1 over the length of the person track in consideration, acting

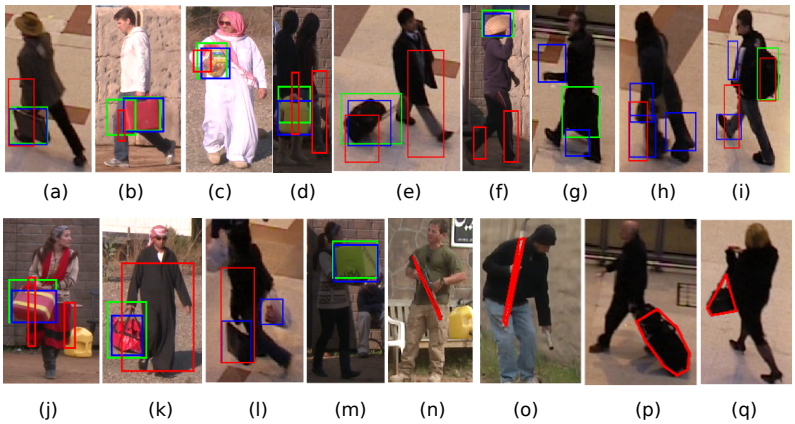


Fig. 5: Illustration of the successes and the failures of our approach and also a comparison with [4]. For images (a)-(m), boxes coloured in green correspond to ground truth, red to [4] and blue to those obtained using the proposed approach. Note that the ground truth is sometimes imperfect eg. (l). Images (n)-(q) illustrate the obtained contour of the detected object using the proposed method.

as a normalisation factor. Default parameter settings of the detector [4] are used for both datasets, as it is often considered most suitable for general uses.

6 Results and Analysis

Results. Fig. 4a gives F1 curves for DHD, raw detections (RD) and the BTS, with Fig. 4b additionally showing HDS and ADS, with each being better than the previous. Even though BTS out-performs DHD on both datasets, ADS significantly outperforms DHD and all other variations of the system.

Qualitative Analysis. In this section we also present a qualitative analysis of the results on PETS2006 and MINDSEYE2012 by summarising successes and failure cases Fig 5. (a)-(f) illustrate how our approach is able to detect different types of objects such as boxes, bags, plastic bags and suitcases. This highlights the merits of performing generic object detection without specific object models. (g)-(i) show a few cases where our approach performs poorly, as the edges do not sufficiently demarcate the object from the person. The (c,d,n,o) images illustrate that our approach is also able to detect objects that are not protrusions. (a,b,c,f,j,k,l) highlight some typical cases where the protrusion based approach [4] fails whilst ours succeeds. (d) illustrates a situation when multiple persons are close by, or when the person's bounding box is displaced. (f) illustrates a case where the influence of a relatively strong prior on the position of the object in relation to the person can hinder the detection of an object (e.g. basket) above a person's head. (n,o,p,q) also illustrate that our approach can localise an object accurately with a contour around it.

7 Summary and Future Work

We have introduced a vision system that performs carried object detection and tracking. Our approach characterises carried objects in terms of generic shape properties such as convexity, whilst taking account of the fact that they are

often, but not always, protrusions on a person silhouette, and exploiting the property that they have continuous and spatially consistent trajectories relative to the person carrying them. In addition, an iterative event driven optimisation process, which uses a heatmap and attention like mechanism, is introduced to obtain an optimal set of object detections. Experimental results show that our approach significantly outperforms a state-of-the-art technique [4], especially the ADS system where both a heatmap and attention-like mechanism are employed, on two challenging datasets. A future extension of this work would be to include other geometric shapes and events such as drop, pick-up, give etc.

References

1. C. Benabdelkader and L. S Davis. Detection of people carrying objects: A motion-based recognition approach. *Proc. Intl Conf. Automatic Face and Gesture Recognition*, pages 378–384, 2002.
2. A. Branca, M. Leo, G. Attolico, and A. Distante. Detection of objects carried by people. *Proc. Intl Conf. Image Processing*, 3:317–320, 2002.
3. R. Cutler and L. Davis. Robust real-time periodic motion detection, analysis, and applications. *PAMI*, 22(8), 2000.
4. D. Damen and D. Hogg. Detecting carried objects from sequences of walking pedestrians. *PAMI*, 34(6):1056–1067, 2012.
5. P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 32(9):1627–1645, 2010.
6. I. Haritaoglu, R. Cutler, D. Harwood, and L.S. Davis. Backpack: Detection of people carrying objects using silhouettes. *CVPR*, 1:102–107, 1999.
7. D. Harwood I. Haritaoglu and L.S. Davis. W4: Real-time surveillance of people and their activities. *PAMI*, 22(8), 2000.
8. D. Kroon and C. H. Slump. Coherence filtering to enhance the mandibular canal in cone-beam ct data. In *Proceedings of the 4th Annual Symposium of the IEEE-EMBS Benelux Chapter*, pages 41–44, 2009.
9. H. Nanda, C. Benabdelkedar, and L. S. Davis. Modelling pedestrian shapes for outlier detection: A neural net based approach. *Proc. Intelligent Vehicles Symp*, pages 428–433, 2003.
10. H. Pirsiavash, D.Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *CVPR*, pages 1201–1208, 2011.
11. C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *PAMI*, 22:747–757, 2000.
12. D. Tao, X. Li, S. J. Maybank, and W. Xindong. Human carrying status in visual surveillance. *CVPR*, 2006.
13. K. Tsuda, M. Minoh, and K. Ikeda. Extracting straight lines by sequential fuzzy clustering. *Pattern Recognition Letters*, 17(6):643–649, 1996.
14. Y. Yang and D. Ramanan. Articulated pose estimation using flexible mixtures of parts. *CVPR*, 2011.
15. Qian Yu and Gerard Medioni. Multiple-target tracking by spatiotemporal monte carlo markov chain data association. *PAMI*, 31, 2009.
16. J. Zunic and P. L. Rosin. A convexity measurement for polygons. *PAMI*, 26:173–182, 2002.