Optimized Pedestrian Detection for Multiple and Occluded People

Sitapa Rujikietgumjorn and Robert T. Collins

Computer Science and Engineering, The Pennsylvania State University

sur167@cse.psu.edu; rcollins@cse.psu.edu

Abstract

We present a quadratic unconstrained binary optimization (QUBO) framework for reasoning about multiple object detections with spatial overlaps. The method maximizes an objective function composed of unary detection confidence scores and pairwise overlap constraints to determine which overlapping detections should be suppressed, and which should be kept. The framework is flexible enough to handle the problem of detecting objects as a shape covering of a foreground mask, and to handle the problem of filtering confidence weighted detections produced by a traditional sliding window object detector. In our experiments, we show that our method outperforms two existing state-of-the-art pedestrian detectors.

1. Introduction

Detecting people is a difficult problem due to body pose articulation and variation in human shapes and appearances. Notwithstanding these difficulties, good progress has been made on the problem of detecting individual walking pedestrians through the use of statistical machine learning methods for training pedestrian object detectors [4]. However, although the current results are promising, the problem becomes more difficult when there are many people at different locations and scales in a scene. There is still need for improvement, and one outstanding problem is how to handle the detection of multiple occluded people.

Conventional approaches find an object based on running a single-object detector on sliding windows throughout the image spatially and across scale levels, followed by non maximum suppression to remove multiple responses of the detector on the same individual at slightly shifted spatial locations and neighboring scales. However, when multiple people are close together or even partially overlapped in the image, non-maximum suppression can remove a correct cluster of responses, resulting in a missed detection. As shown in Figure 1 using the PLS detector [20] with its default non maximum suppression method, correct candidate detections were present that were suppressed by stronger

Figure 1. (a) PLS detector using its default non maximum suppression method misses one person. (b) Examining candidate object center locations from the PLS detector without non maximum suppression shows that candidate detections were generated for that person, but they were subsequently suppressed.

neighboring detections. The aim of this paper is to show that quadratic optimization for reasoning about overlapping detections can improve the performance of a pedestrian detection system, especially when there are multiple overlapping people.

We propose an approach that reasons directly over the space of overlapping object detections by formulating a quadratic objective function that contains both unary scores measuring the quality of an individual detection, and pairwise scores measuring the joint compatibility of pairs of overlapping detections. Loosely speaking, the unary scores reward candidates with high detector confidence, whereas the pairwise scores impose a penalty for excessive amounts of overlap between two candidates. The problem is to find a binary vector that maximizes the quadratic objective function, which is a classic problem of quadratic unconstrained binary optimization (QUBO). Although this is an NP-hard problem, efficient approximate methods are available that yield high quality solutions on large problem sizes.

An example usage of our framework is shown in Figure 2. First, a large set of possible detection candidates (Figure 2(b) and 2(f)) is generated, based on shape covering of a foreground mask in the top row, or sampling from a confidence map produced by a standard pedestrian detector (with non-maximum suppression disabled) in the bottom row. In both cases, the object configuration that maximizes the quadratic objective function is found and shown in Figure 2(c) and 2(g). Our framework thus looks for a subset of
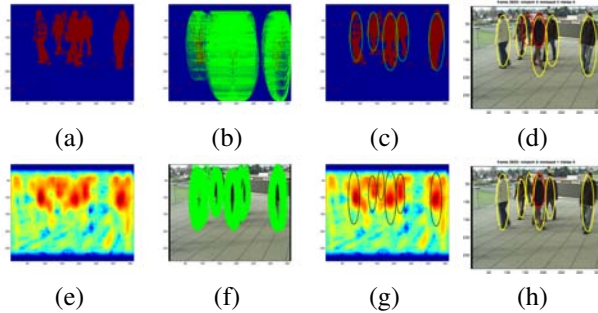
Figure 2. Example usage of our proposed algorithm. The top row uses a foreground shape covering approach to produce a large set of candidate detections. The bottom row samples from confidence maps produced by a sliding window human detector to generate candidate detections. In both cases, we perform the same binary quadratic optimization procedure to choose the solution set of candidates that maximize a quadratic objective function. (a) foreground image. (b) candidate detections generated from (a). (c)(d) our detection results. (e) sliding window detector confidence map (only one scale level shown). (f) candidate detections sampled from the confidence map. (g)(h) our detection results.

candidate detections to optimize the tradeoff between unary confidence scores and pairwise overlap penalties.

2. Related Work

Object detection has been widely studied and significant progress has been made in recent years. Of the several approaches that have been proposed for detecting pedestrians, one common method uses a pre-trained classifier within a sliding window to scan the whole image looking for people at all locations and scales. For example, [1] proposes a state-of-the-art human detection method using the Histograms of Oriented Gradients (HOG) descriptor and SVM as a linear classifier. [7] compares this HOG-based approach with other methods on large datasets including more than 20,000 images, and shows that it outperformed other methods. Many variants have been proposed, seeking improvements through novel features and/or classifiers [22, 19, 14, 3] or by leveraging feature space transformation methods, e.g. Partial Least Squares (PLS) analysis [20], for feature generation and dimensionality reduction. Since sliding window methods usually generate multiple overlapping detections on a person, the common final step is to apply non-maximum suppression as an attempt to remove false positive detections [9, 11].

In video, alternative approaches based on motion or background subtraction can be used to detect pedestrians. One line of work proposes to segment foreground blobs into human shapes using an MCMC-based optimization approach to determine the number and configuration of overlapping shapes [23, 24, 10]. These methods generate good results for detection and counting tasks, but the sampling-based MCMC search mechanism is very time-consuming.

We propose to use Quadratic Unconstrained Binary Optimization (QUBO) for pedestrian detection in order to reason more directly/thoroughly about overlapping detection candidates and their associated confidence scores and overlap penalties. Recently, several other papers have also explored the application of combinatorial optimization methods to the problem of object detection [8, 2, 18, 6, 21]. [2] introduces a unified model for recognizing multi-class object configurations by learning the spatial arrangement of those objects. A quadratic cost function is used to represent criteria relating object confidence, overlapping hypotheses, and spatial interactions between pairs of objects from different classes. [18] apply a similar framework to crowd scene analysis and integrate it with a term related to crowd density. Both of these papers use a heuristic greedy search method to seek a solution to the QUBO problem. Another heuristic method for solving binary optimization problems is Tabu search, a stochastic local search approach which has been used in various fields [12, 17], e.g. to solve the multi-assignment problem in Intelligent Visual Surveillance [5].

3. Proposed Optimized Detection Framework

Figure 3 presents a "big picture" overview of how our approach would be incorporated into a typical pedestrian detection pipeline. Our framework consists of two steps. First, an existing pedestrian detector is applied to produce a detection confidence score map, or if that is not available as output, a set of unfiltered bounding boxes with associated confidence scores. Our method then samples a large but finite set of plausible candidates. A unary confidence score is computed for each candidate, to represent the quality of that proposed detection. Furthermore, for candidates that overlap, a pairwise score is computed to specify a penalty that will be incurred if both candidates are kept in the final solution. The purpose of this penalty is to prohibit excessive amounts of overlap while still allowing some amount of reasonable overlap to occur.

In the second step, unary and pairwise scores are grouped into a cost matrix to form the objective function for a quadratic unconstrained binary optimization (QUBO) problem. In this QUBO problem, the unknown binary variables to solve for represent whether to keep or discard each pedestrian candidate from the final solution set of detections. An optimization algorithm is then applied to search for an assignment of 0's and 1's to candidates yielding a high, ideally the maximum, objective function value.

Section 3.1 discusses generation of candidates and objective function scores for two very different types of detection approach. The first is a shape covering approach, based on finding size and placement of a set of pedestrian shapes in order to cover the pixels of a foreground mask computed by, e.g., background subtraction. The second ap-
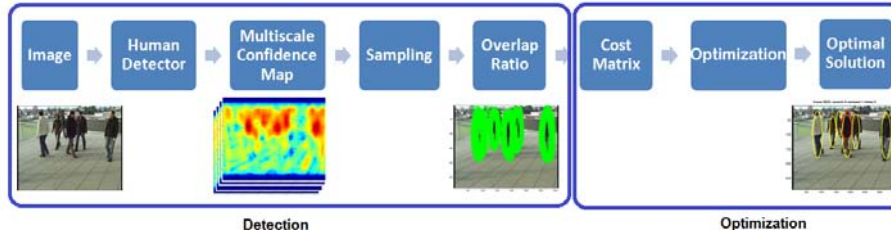
Figure 3. Overview of the Proposed Optimized Detection Framework

proach uses a multi-scale confidence map produced by an existing, publicly available appearance-based detector. Section 3.2 discusses the second step of transforming into a binary quadratic objective function and efficient methods for generating high-quality approximate solutions to the resulting QUBO problem.

3.1. Generating a Set of Detection Candidates

3.1.1 Detection Candidates by Shape Covering

Previous works have considered the problem of detecting people as a "shape covering" of foreground mask data [10, 23]. That is, given a foreground mask computed by background subtraction or motion analysis, a solution is sought as to the number, location, size and possibly articulation of a set of shapes to cover as many foreground pixels as possible while leaving as many background pixels as possible uncovered. To avoid unnecessary proliferation of overlapping shapes, these methods augment the covering quality term of the objective function with either prior terms on the number of objects present, or with data terms that penalize object overlap. Both [10, 23] used an expensive Markov Chain Monte Carlo stochastic search procedure to find a good shape covering. In this paper, we address the shape covering problem using QUBO.

To use this shape covering approach in practice, we first generate a lookup table relating location (x,y) in the image to expected height and width of a pedestrian centered at that location. In our experiments we have created this lookup data from ground truth camera calibration information, however in Section 3.1.3 we discuss how this prior size information can be learned automatically from a training sequence. Given an automatically computed foreground mask, a candidate set of shapes is generated by methodically sampling midpoint locations every 10 pixels in x and y, looking up the expected width and height at each location, and computing a unary score c_i for each candidate x_i. We use three common pedestrian poses shown in Figure 4 as the three shapes that can be proposed for shape covering.

For each candidate x_i, three unary confidence scores are computing using each of these three pedestrian shapes, scaled to the size of the detection candidate bounding box. The maximum score is selected. Each score $c_i(x_i)$ is com-

Figure 4. Pedestrian silhouette shapes used in shape covering.

puted as

$$c_i(x_i) \; = \; \mathrm{on}(x_i) \; - \; \alpha \, \mathrm{off}(x_i) \tag{1}$$

where function $\mathrm{on}(x_i)$ returns the number of foreground pixels that are "on" within the shape, and $\mathrm{off}(x_i)$ returns the number of off pixels. In all of our experiments, $\alpha = 0.5$ and candidates with $c_i \leq 0$ are discarded.

3.1.2 Detection Candidates by Bounding Box Filtering

Many object detection approaches use a sliding window based detector to generate a confidence score map, and then generate a set of final detections through a process of non maximum suppression. For our approach, we modify an existing Histogram of Oriented Gradients (HOG) based pedestrian detector [1], available in OpenCV, and apply it at multiple scales without non maximum suppression to generate a multi-scale detection confidence map. In the experiment, a set of 500 detection candidates is then randomly sampled from each detection scale, with the likelihood of a candidate being sampled at location (x,y) being proportional to the detector confidence at that location and scale. Figure2(f) shows candidate samples generated from the confidence score map in Figure2(e) (only one scale level of the map is shown).

3.1.3 Learning a Prior on Bounding Box Size

The camera viewpoint obviously affects the range of detection scales that are observed. For example, images taken by a camera near eye level will have a large allowable range of scales, whereas an elevated camera much farther away may not see any difference in pedestrian size across the image. Sliding window detectors that do not have access to such information are prone to a greater number of false positives, due to detections that are either too large or too small. At the very least, having access to a minimum and maximum
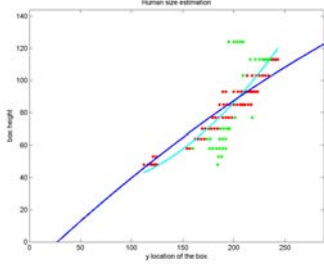
Figure 5. Expected pedestrian bounding box size learned as a regression function between y location in the image and height of a detected pedestrian at that location.

scale at which to expect detections is helpful, yet even this small amount of information is scene specific and troublesome to set by hand.

When camera calibration information is available, expected size effects can be computed automatically; however, calibration information is often not available. Some previous works have used apriori pedestrian height distributions to estimate camera calibration information from noisy pedestrian detections [13, 18]. In our work, we employ a simple online learning approach that learns a regression function on expected bounding box height versus location in the image from a set of detections of pedestrians at different locations in images taken from a stationary camera view. Figure 5 shows a sample plot comparing image y location with height of detected boxes in the image. When there are multiple heights of people observed over many different image locations, we can compute an approximate expected height model from this data. The light blue line is a quadratic regression function learned for one frame only, while the dark blue line is the height approximation found using data acquired over several frames. These approximation curves could be used to identify outlier detections, represented by green dots, that do not have an appropriate size. However, rather than apply a hard threshold to filter improperly sized detections, we reduce unary confidence scores according to the dissimilarity between a candidate bounding box scale and the expected detection height at that location. The detection confidence score will be greatly reduced when the detected box is much larger or smaller than the learned size estimate, reducing the chances that the candidate will be kept in the solution vector returned by QUBO.

3.2. Quadratic Unconstrained Binary Optimization

Binary optimization is the problem of finding a binary vector $x = [x_1, x_2, ..., x_n]$ that maximizes an objective function $f(x)$. The objective function is typically represented by a multi-linear polynomial expression of degree of 1 (linear), 2(quadratic) or possibly higher order. In this

work, we use the quadratic objective function

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j + \sum_{i=1}^{n} c_i x_i \qquad (2)$$

where $x_i \in \{0, 1\}, i = 1, ..., n$ are the binary variables to be solved for, c_i are n unary coefficients, and c_{ij} are $O(n^2)$ pairwise coefficients. The goal is to assign 0,1 values to each x_i in a way that maximizes the objective $f(x)$. Noting that $x_i = x_i^2$ for binary variables, we can combine all the coefficients c_i and $c_{i,j}$ into a single square matrix Q and solve the binary integer maximization

$$\max_{x \in \mathbb{B}^n} f(x) = \max_{x \in \mathbb{B}^n} x'Qx \qquad (3)$$

The resulting Quadratic Unconstrained Binary Optimization (QUBO) problem is known to be NP-hard [17].

3.2.1 Objective Function

The quadratic objective function in Eq. 2 is computed by combining unary and pairwise terms. Each unary score c_i is a measure of confidence that candidate x_i represents a person, while the pairwise scores $c_{i,j}$ penalize excessive overlap between pairs of candidates. These pairwise terms are based on the *overlap ratio*, computed as the intersection area (in pixels) of two overlapping shapes, divided by the area of the smaller one. If there is only a small overlap penalty, it might be reasonable to keep both candidates. On the other hand, if there is a large overlap, a higher penalty is applied, and it is probably better to keep only one of them.

Figure 6 presents an example of objective function computation. The three elliptically shaped candidates from left to right have unary values 3425, 4412 and 3658 computed from Eq. 1. For each pair of distinct overlapping candidates x_i and x_j, the overlap penalty $c_{i,j}(x_i, x_j)$ is -4594 for shapes x1 and x2, -1998 for shapes x1 and x3, and -3432 for shapes x2 and x3. From Eq. 3, we want to find a maximal value of $x'Qx$ with the constraint that $x = [x_1, x_2, x_3]$ can take only binary values. We find that the optimal solution $[1, 0, 1]$ specifies that candidates x1 and x3 should be kept, while candidate x2 should be discarded. Note that if we applied a traditional, greedy non maximum suppression approach where the candidate with highest confidence score is chosen and overlapping candidates of lesser score are suppressed, we would have chosen to keep only the middle candidate x2, while suppressing the other two.

In general, the matrix Q in the objective function of Eq. 3 can be formed as

$$Q = w_1 \, U \quad - w_2 \, P \qquad (4)$$

where U is a diagonal unary score matrix, P is the pairwise score matrix (overlap ratios), and w_1, w_2 are relative
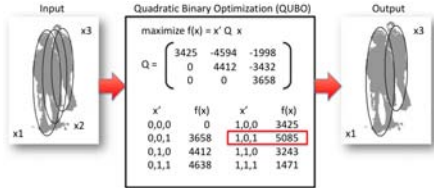
Figure 6. Using quadratic binary optimization to find the best set of detection candidates.

weights determined as described in the following section, with $w_1 + w_2 = 1$. Both U and P are normalized to be values between 0 and 1. Furthermore, we can extend Eq. 2 by adding a second unary term that represents the score from a second detector or other additional information. For example, in our experiments we have explored combining unary scores computed from foreground shapes with scores computed from a detector confidence map. In this case, the matrix Q is formed as

$$Q = w_1 \, U_1 \; - w_2 \, P \; + w_3 \, U_2 \qquad (5)$$

with $w_1 + w_2 + w_3 = 1$.

3.2.2 Weight Parameter Estimation

Even though the unary and pairwise scores are both normalized values between 0 and 1, they represent very different types of information, one being an appearance-based detection confidence and the other being an area ratio. Furthermore, the amount of "acceptable" bounding box overlap for a given situation may depend on the expected density of people in the scene as well as on the camera viewpoint. For this reason, it is better to weight the relative contributions of the unary and pairwise terms with weighting parameters learned from representative training data.

We use the Pattern Search algorithm [15] to find appropriate weight values $w_1 + w_2 = 1$ that maximize detection performance on a separate training dataset. During the search, different proposed weight values are evaluated by plugging them into Eq 4, solving the resulting QUBO problem, and comparing the results against known ground truth. The pattern search algorithm is ideal for use with an algorithmic objective function like this because it requires no knowledge of the gradient of a function, nor even that the function be differentiable.

3.2.3 Solving QUBO

Although QUBO is NP-hard, several methods are available for efficiently finding good approximate solutions. In this paper we compare three algorithms: Tabu search, Greedy forward search, and quadratic programming.

Tabu search is an efficient stochastic local search approach for handle QUBO problems of large size, and has been successfully used in many applications [5, 12, 17]. An important concept of Tabu search is the use of adaptive memory; starting with an initial solution point, the method sequentially adjusts the value of a small subset of variables in an attempt to find an improved solution in a local neighborhood, while maintaining a "tabu" list to make sure that previously searched locations are not soon revisited.

Greedy algorithms can also yield good approximate solutions in practice [2]. We test the greedy forward search method used in current state-of-the-art papers [2, 18]. The single candidate that gives the highest unary score is first selected. Then, given a set of previously selected candidates, each remaining candidate is tentatively added to the set, and the set yielding the highest new objective function value becomes the new current solution set. The process stops when adding any single candidate to the solution set will reduce the objective function value.

As a third approach, we relax the binary variable constraints into $0 \leq x_i \leq 1$ to transform QUBO into a continuous quadratic programming problem. We used Matlab's trust-region method to solve for the maximization [16]. A subsequent rounding procedure forces the continuous vector result into a binary solution vector.

4. Experimental Results

In this section we evaluate our proposed quadratic binary optimization framework to detect overlapping pedestrians. Three methods are tested for generating candidates and forming the quadratic objective function: shape covering of a foreground pixel mask, sampling from a confidence map produced by a multi-scale sliding window detector, and a hybrid approach that generates candidates from a detector confidence map but also incorporates foreground shape covering information into the objective function as a second unary confidence term. We also evaluate three methods for solving the constructed QUBO optimization problem: Tabu search, greedy algorithm, and relaxation into a continuous quadratic program. We use the ITS Multistart Tabu algorithm by Palubeckis [17][1]. We follow the Greedy approach used in [2, 18]. We use Matlab's trust-region method to solve the relaxed quadratic programming problem.

Dataset We perform quantitative evaluation of our approach using a pedestrian dataset from EPFL called the Terrace sequences[2] and on the PETS 2009 dataset. These two datasets were chosen because of the large number of occlusions, a challenging issue in pedestrian detection.

---

[1]The Iterated Tabu Search (ITS) code provided by Palubeckis is available for research use

[2]from CVLab, Ecole Polytechnique Federale de Lausanne (http://cvlab.epfl.ch/data/pom/)
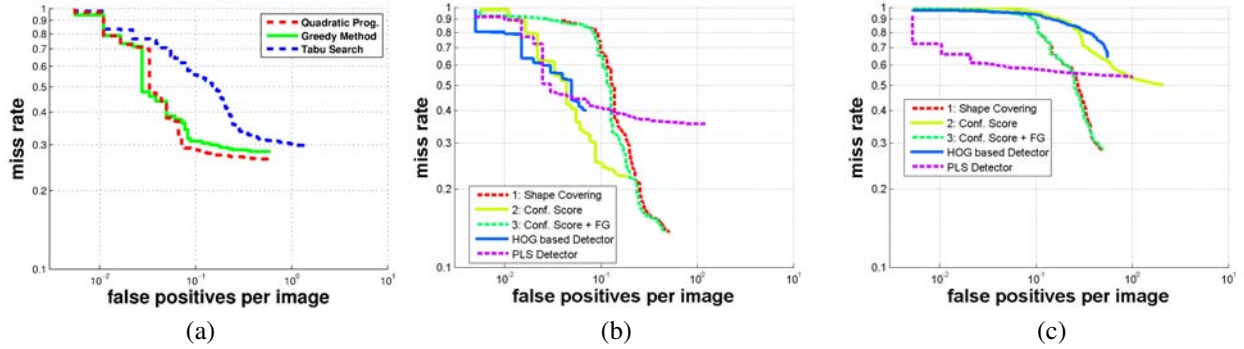
Figure 7. Plots of miss rate vs FPPI (lower curves are better). (a) Comparison between Tabu search, Greedy, and Quadratic programming methods using foreground shape covering. (b) Comparison of three QUBO variants and two baseline methods on the EPFL dataset. (c) Comparison on the PETS 2009 dataset.

Figure 7(a) shows a quantitative comparison between Tabu Search, greedy algorithm, and quadratic programming results when candidates are generated based on foreground shape covering. Both greedy algorithm and quadratic programming perform reasonably well. They have lower miss rate and fewer false positives than Tabu search. The overall accuracy of greedy method is 0.6849, slightly lower than quadratic programming's overall accuracy of 0.6925.

Based on this result, we decided to use quadratic programming to evaluate our three options for generating candidates and unary confidence scores (foreground shape covering, detector confidence, confidence+foreground shape cover). These three approaches are compared by plotting Miss Rate vs False Positives Per Image (FPPI) for two datasets, EPFL (Fig. 7(b)) and PETS 2009 (Fig. 7(c)). Lower curves are better. Two other approaches compared in those plots are OpenCV's HOG-based human detector [1] and the PLS detector of [20], both using their default non-maximum suppression methods.

Approach 1 based on finding shape covering of a foreground mask works surprisingly well given the simplicity of the approach compared to the sophisticated appearance-based detectors it is being compared against. Comparing approaches 1 and 2, using detector confidence scores yields better results on the EPFL dataset, whereas foreground shape covering gives the better result on PETS. This can be explained by the size/resolution of pedestrians in those two datasets. In EPFL, people are large, with clearly visible edge appearance information, whereas the small / low-resolution crowds in PETS is a situation where fitting body shapes to cover foreground blobs yields better results. In approach 3 we attempt to improve results using a hybrid of detector confidence combined with foreground covering as a second unary term, but with mixed results. However, it demonstrates that the QUBO framework is flexible enough to be applied to a variety of detectors or combinations of unary and pairwise information.

Table 1 shows quantitative results of the five methods tested. The first three are all variants of our proposed QUBO approach, and all of them have higher numbers of true positives and higher accuracy than the state-of-the-art HOG and PLS detectors. Here, accuracy is defined as
$$accuracy = (tp + tn)/(tp + tn + fp + fn).$$

Figure 8 shows several illustrative results from the three QUBO variants and two baseline methods HOG and PLS. Yellow ellipses in the images represent true positives, blue boxes are false positives, and red boxes are false negatives.

5. Conclusion

We have presented a framework for improving pedestrian detection performance in cases where there are multiple, overlapping objects. A QUBO framework is adopted where a quadratic objective function is formed from unary confidence scores and pairwise overlap penalties. The unary terms are not limited to a specific type of detector and can be applied to various types of detection confidence scores with some adjustment. Solving for the binary solution vector that maximizes this quadratic objective function automatically balances the trade off between encouraging multiple, high-quality detections, while discouraging excessive amounts of overlap. Since finding exact solutions for large-scale QUBO problems is not possible, we evaluate three approximate methods: heuristic Tabu search, greedy forward search, and relaxation to a continuous quadratic program. All three variants find good quality approximate solutions in the experiments, however the last two methods quantitatively outperform the first.

Our results show that the use of binary quadratic optimization to explicitly reason about pedestrian candidate confidences and overlaps yields a performance improvement over existing detection methods that use non-maximum suppression, in terms of lower miss rates and lower false positives. Furthermore, the proposed method can be used to improve the performance of any existing

Method	Dataset A					Dataset B				
	TP	FN	FP	Accuracy	Precision	TP	FN	FP	Accuracy	Precision
shape covering	602	98	112	0.7414	0.8431	3142	1240	92	0.7023	0.9716
confidence score	544	156	34	0.7411	0.9412	2183	2199	399	0.4566	0.8455
confidence score+FG	560	140	46	0.7507	0.9241	3142	1240	94	0.7020	0.9710
HOG based detector	437	263	14	0.6120	0.9690	1580	2802	106	0.3520	0.9371
PLS detector	403	297	47	0.5395	0.8956	2017	2365	193	0.4409	0.9127

Table 1. Quantitative comparison of the five methods tested. Dataset A is the EPFL Terrace dataset. Dataset B is the PETS 2009 dataset.

sliding window detector that produces either a detection confidence map or a set of unfiltered, thresholded bounding boxes with associated confidence scores. We have also demonstrated how our method can be used to detect people as shape coverings of a foreground mask. This covering approach can be generalized to use a library of realistic pedestrian shapes, such as those in [10, 23].

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:886–893, June 2005. 2, 3, 6

[2] C. Desai, D. Ramanan, and C. Fowlkes. Discriminative models for multi-class object layout. *IEEE International Conference on Computer Vision (ICCV)*, pages 229–236, Sept. 2009. 2, 5

[3] P. Dollar, Z. Tu, P. Perona, and S. Belongie. Integral channel features. pages 91.1–91.11, 2009. 2

[4] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, Apr. 2012. 1

[5] I. Dotu, P. Van Hentenryck, M. A. Patricio, A. Berlanga, J. Garca, and J. M. Molina. Real-time tabu search for video tracking association. *Principles and practice of constraint programming*, pages 21–34, 2009. 2, 5

[6] I. Endres, V. Srikumar, M.-W. Chang, and D. Hoiem. Learning shared body plans. *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 3130–3137, June 2012. 2

[7] M. Enzweiler and D. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, Dec. 2009. 2

[8] P. Felzenszwalb and R. Zabih. Dynamic programming and graph algorithms in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):721–740, 2011. 2

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2009. 2

[10] W. Ge and R. Collins. Marked point processes for crowd counting. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2913–2920, June 2009. 2, 3, 7

[11] D. Geronimo, A. Lopez, A. Sappa, and T. Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1239–1258, 2010. 2

[12] F. Glover. Tabu search: A tutorial. *Interfaces*, 20(4):74–94, July 1990. 2, 5

[13] J. Liu, R. Collins, and Y. Liu. Surveillance camera auto-calibration based on pedestrian height distributions. *British Machine Vision Conference*, pages 117.1–117.11, 2011. 4

[14] Y. Liu, S. Shan, W. Zhang, X. Chen, and W. Gao. Granularity-tunable gradients partition (GGP) descriptors for human detection. *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1255–1262, June 2009. 2

[15] N. Meinshausen, P. Bickel, and J. Rice. Efficient blind search: Optimal power of detection under computational cost constraints. *The Annals of Applied Statistics*, 3:38–60, 2009. 5

[16] J. More and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 3:553–572, 1983. 5

[17] G. Palubeckis. Multistart tabu search strategies for the unconstrained binary quadratic optimization problem. *Annals of Operations Research*, 131(1-4):259–282, Oct. 2004. 2, 4, 5

[18] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert. Density-aware person detection and tracking in crowds. *IEEE International Conference on Computer Vision (ICCV)*, pages 2423–2430, Nov. 2011. 2, 4, 5

[19] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2007. 2

[20] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human detection using partial least squares analysis. *IEEE International Conference on Computer Vision (ICCV)*, pages 24–31, Sept. 2009. 1, 2, 6

[21] H. Trinh, Q. Fan, P. Gabbur, and S. Pankanti. Hand tracking by binary quadratic programming and its application to retail activity recognition. *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1902–1909, 2012. 2

[22] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *IEEE International Conference on Computer Vision (ICCV)*, 1:90–97, 2005. 2
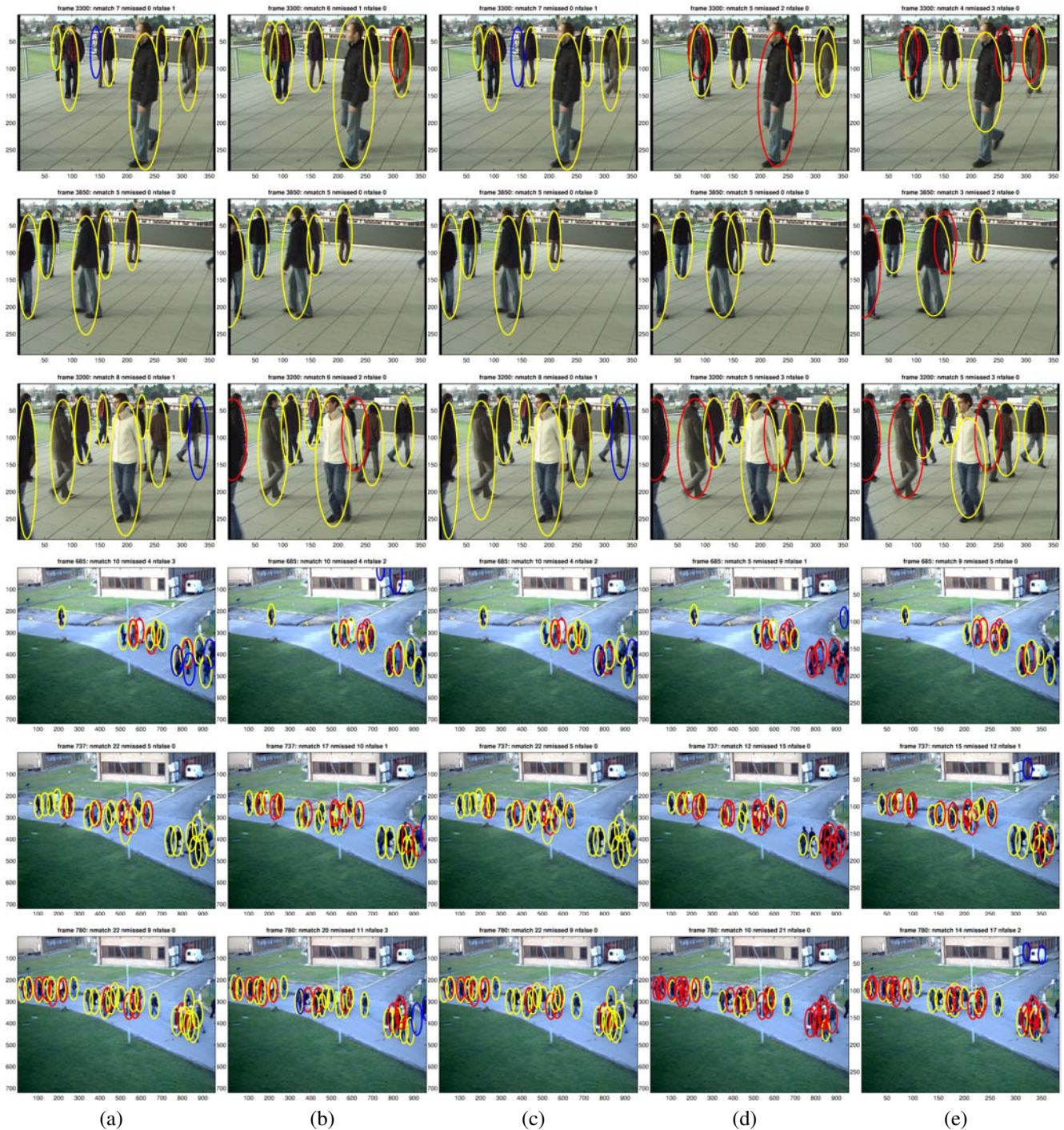
Figure 8. Comparison of three QUBO variants and two baseline methods. Yellow means correct detection, red is a missed detection, and blue is false positive. (a) Our approach 1: foreground shape covering. (b) Our approach 2: detector confidence score. (c) Our approach 3: hybrid of confidence score and shape covering. (d) OpenCV HOG detector. (e) PLS detector.

[23] T. Zhao and R. Nevatia. Bayesian human segmentation in crowded situations. *IEEE Computer Vision and Pattern Recognition (CVPR)*, page 459466, June 2003. 2, 3, 7

[24] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2:406–413, 2004. 2