A Framework for Evaluating Stereo-Based Pedestrian Detection Techniques

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

***Abstract*—Automated pedestrian detection, counting, and tracking have received significant attention in the computer vision community of late. As such, a variety of techniques have been investigated using both traditional 2-D computer vision techniques and, more recently, 3-D stereo information. However, to date, a quantitative assessment of the performance of stereo-based pedestrian detection has been problematic, mainly due to the lack of standard stereo-based test data and an agreed methodology for carrying out the evaluation. This has forced researchers into making subjective comparisons between competing approaches. In this paper, we propose a framework for the quantitative evaluation of a short-baseline stereo-based pedestrian detection system. We provide freely available synthetic and real-world test data and recommend a set of evaluation metrics. This allows researchers to benchmark systems, not only with respect to other stereo-based approaches, but also with more traditional 2-D approaches. In order to illustrate its usefulness, we demonstrate the application of this framework to evaluate our own recently proposed technique for pedestrian detection and tracking.**

***Index Terms*—Benchmarking, disparity estimation, evaluation, pedestrian detection, stereo vision.**

I. Introduction

ACCURATE detection and tracking of pedestrians are two essential components required by a variety of applications that include, amongst others, ambient intelligence (AmI), automated surveillance, image compression, and content-based multimedia storage and retrieval. Given this large number of potential applications, pedestrian detection and tracking has become an extremely active research area in computer vision. This has resulted in a significant amount of prior work proposing pedestrian segmentation techniques using a myriad of approaches. These include algorithms based on traditional monocular image processing techniques, such as blob analysis [1], template matching [2], statistical shape models [3], and classification based on low-level features [4]. Many of these monocular techniques produce good results when presented with constrained scenarios that allow assumptions to be made about the camera parameters, environmental conditions, pedestrian flow, and appearance. Unfortunately, few of these, if any, produce reliable results for long periods of time in unconstrained environments [5]. However, in recent times, techniques that use 3-D stereo information have become more prevalent as its use carries with it some distinct advantages over conventional 2-D information [1], [5], [6]

Given competing classes of approaches, an issue arises of how to perform quantitative assessment of a given system's performance for use in cross-system evaluations, whereby the system in question may be *either* monocular or stereo-based. To date, comparisons between these two classes of pedestrian detection technique have been difficult as the community's standard test sequences are monocular and provide no depth data [7] or the means to obtain it via stereo techniques. As such, the level of quantitative benchmarking between the two classes of systems, and indeed between solely stereo-based systems, has been minimal.

In this paper, we provide a framework for evaluating the accuracy of both monocular and short-baseline stereo-based pedestrian detection algorithms. This framework provides the means for quantitatively evaluating a proposed pedestrian detection technique using two different methodologies: 1) using traditional 2-D image plane comparison techniques and 2) via 3-D groundtruth information. In addition, provision is made for evaluating stereo-based approaches at both the component and system levels. A major flaw in many stereo-based techniques is that very few attempt to obtain the highest *quality* disparity map that is possible within reasonable time constraints given the scene features and temporal information. Of course, low-quality or sparse-disparity maps do not inherently mean that pedestrian detection will fail. However, for many techniques, obtaining the best possible disparity map could potentially lead to improved pedestrian detection results over a variety of scenarios. To this end, the evaluation framework incorporates the means to quantitatively evaluate a chosen disparity estimation algorithm via a number of novel synthetic image-pairs. Finally, a contribution of this study is to make all test data and associated groundtruths publicly available to download.[1]

The paper is organized as follows. Section II gives an overview of the related work in this area with respect to groundtruth evaluation. Section III presents the details of our proposed approach for disparity estimation evaluation within stereo-based pedestrian detection algorithms. In Section IV, we provide sequences and groundtruths for a number of challenging stereo test scenarios and propose two different methodologies for evaluation. We demonstrate the application and usefulness of the proposed framework in Section V by applying it to our own approach to pedestrian detection. Finally, Section VI details conclusions and future work.

II. Related Work

There already exist a number of open metrics-based evaluation campaigns that include a number of test data-sets augmented with groundtruth data. These include ETISEO,[2]

Manuscript received October 30, 2007; revised March 7, 2008. First published July 9, 2008; current version published August 29, 2008. This work was supported by Science Foundation Ireland under Grant 03/IN.3/I361. This paper was recommended by Guest Editor M.-T. Sun.

The authors are with the Centre for Digital Video Processing, Adaptive Information Cluster, Dublin City University, Dublin 9, Ireland (e-mail: kellyp@eeng.dcu.ie).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.928228

[1] [Online]. Available: http://www.cdvp.dcu.ie/datasets/

[2] [Online]. Available: http://www.etiseo.net

OTCBVS,[3] CAVIAR,[4] BEHAVE,[5] and PETS.[6] Within these evaluation campaigns, some of the test sequences provided directly target pedestrian detection. However, these sequences are monocular, meaning that researchers working on stereo-based approaches are unable to assess performance using these evaluation frameworks. This has led some researchers of wide [8]–[10] and short-baseline [5], [7], [11] stereo-based approaches to create their own data sequences for evaluation purposes. Although this allows them to evaluate their stereo-based systems against their own groundtruth, this practice yields little in terms of quantitative benchmarking between proposed systems, as, from these techniques, only [11] has made a subset of its stereo test sequences publicly available to the community. However, the released sequences are not augmented with groundtruth data. In addition, the more challenging sequences applied in [11] have not been included within the released data-sets. As such, the released sequences of [11] are not highly complex and do not reflect the high level of difficulty which arises in real-world pedestrianized scenarios.

The robust segmentation and tracking of pedestrians under unconstrained conditions introduces a multitude of complicating factors such as varying environmental conditions, complex pedestrian interactions, occlusion, and the huge amount of possible variations in pedestrian pose, size, and appearance in a given scene. These difficulties make it one of the most challenging problems in computer vision. However, many techniques simply ignore some of these issues. For example, assumptions have been made that a pedestrian is fronto-parallel to the camera image plane [4], that all foreground objects in the scene are pedestrians [12], or that people enter the scene unoccluded [13] or one at a time [14]. Clearly, these types of assumptions can limit the practical applicability of a developed technique in some real-world scenarios.

In this study, we strive to provide a set of test sequences that minimize such simplifying assumptions and incorporate the multiple levels of difficulty that are inherent in real-world pedestrianized scenarios. In addition, we also believe that a proposed algorithm should be evaluated at both the component and system level. For single-camera-based approaches, this may involve evaluating an underlying background subtraction algorithm—such frameworks have already been proposed by other works [15] and thus are not addressed here. For stereo-based approaches, an evaluation of the underlying disparity estimation technique should be made—this is a contribution of this work and, to the knowledge of the authors, an important subcomponent of stereo-based pedestrian detection algorithms that has not been evaluated before.

III. Disparity Estimation Data-Set

In order to quantitatively evaluate a disparity estimation approach, a data-set with groundtruth disparities is required. Unfortunately, standard disparity data-sets, such as the *Tsukuba*, *Venus*, or *Map* data-sets [16], [17] may not be applicable for pedestrian detection algorithms. This can be due to the lack of a groundplane region within the scene, which is a constraint required in some approaches [12], or the inability to obtain the background models required by a proposed algorithm. In addition, as pedestrian detection techniques are generally designed for real world pedestrianized scenes, it would be advantageous to determine the robustness of the disparity estimation technique within this context for a range of challenging scenarios such as varying lighting conditions, shadows, a lack of texture at depth discontinuities, homogeneous foreground and background regions, and, most importantly, pedestrians exhibiting a variety of clothing, poses, distances, and scales.

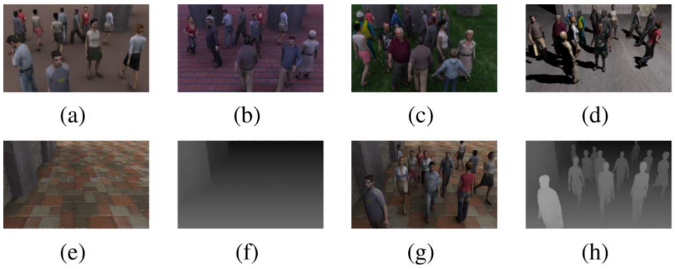

Fig. 1. Synthetic scenes. (a)–(d) Four foreground synthetic scenes. (e) Background scene. (f) Background groundtruth disparity. (g) Foreground scene. (h) Foreground groundtruth disparity.

To this end, we have developed a new synthetic data-set designed to incorporate a number of difficulties associated with typical pedestrianized scenarios. This data-set consists of eight scenes, where each 3-D scene was designed to incorporate a flat groundplane, one or more background objects, and a number of varying pedestrian models [see Fig. 1(a)–(d)]. Throughout the data-set, a variety of texture maps were chosen for the groundplane. These vary from textured or tiled surfaces to a single homogeneous colour. Finally, a variety of ambient and directional lighting sources were introduced, designed to mimic the lighting conditions of both indoor and outdoor scenarios. Depending on the lighting conditions, shadows (both cast- and self-shadows) range from subtle to strong.

For each 3-D scene, two sets of rectified stereo-pairs were created (from a virtual stereo-camera with a baseline of 100 mm); the first rendering [see Fig. 1(e)] contains no foreground objects and can be used to initialize background models; the second incorporates a number of foreground pedestrians [see Fig. 1(g)]. In addition, each rendering of the scene is accompanied by a groundtruth disparity map [see Fig. 1(f) and (h)]. Using these 16 synthetic stereo-pairs, a quantitative evaluation of a proposed disparity estimation technique can be undertaken and can be benchmarked with respect to other disparity estimation techniques. We recommend using the Middlebury open-source evaluation test bed[7] for this benchmarking. This process is outlined in Section V-A.

IV. Pedestrian Detection Data-Sets

In order to quantitatively evaluate the final system-level output of a pedestrian detection technique, we propose the application of two different methodologies. The first technique, presented in Section IV-A, evaluates the proposed

[3] [Online]. Available: http://www.cse.ohio-state.edu/otcbvs-bench/

[4] [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

[5] [Online]. Available: http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/

[6] [Online]. Available: http://www.petsmetrics.net

[7] [Online]. Available: http://vision.middlebury.edu/stereo/

TABLE I
DATA-SET SEQUENCES. NOTE: GND. REFERS TO THE NUMBER OF GROUNDTRUTHED PEDESTRIANS WITHIN A SEQUENCE

Sequence	*Gnd.*	*Frames*	*Frame Rate (Hz)*	*Time (in mins.)*
Overhead	657	418	≈ 6.5	1.10
Corridor	1027	697	≈ 5.3	2.26
DCU Corner	5126	828	≈ 4.6	3.02
2D Total	6810	1943	≈ 4.6–6.5	6.38
Vicon 1	198	198	≈ 5.4	0.86
Vicon 2	526	263	≈ 5.5	0.94
Vicon 4	1296	324	≈ 5.3	1.18
Vicon 8_A	2104	263	≈ 5.4	0.96
Vicon 8_B	2120	265	≈ 5.4	1.00
3D Total	6244	1313	≈ 5.3–5.5	4.94
Total	13054	3256	≈ 4.6–6.5	11.32

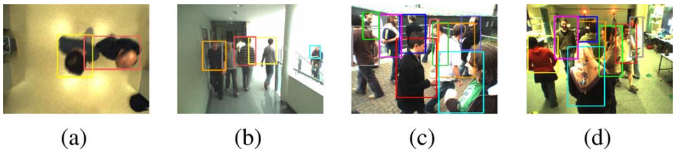

Fig. 2. Data-set sequence example results. (a) *Overhead*. (b) *Corridor*. (c) *DCU Corner*. (d) *Vicon* 8_A.

approach using traditional 2-D image-plane comparison techniques. This methodology could also be used to evaluate monocular approaches to pedestrian detection. The second technique, described in Section IV-B, evaluates the proposed stereo-based technique using 3-D groundtruth information. For each approach, a number of rectified and synchronized input stereo-video sequences were captured via a Digiclops stereo-camera[8] (which handles the camera synchronization internally and image rectification via an SDK). These experimental sequences were designed to stress-test a proposed technique in several areas, such as disparity estimation, foreground segmentation, pedestrian detection, and tracking. When capturing each sequence, no restrictions or instructions were provided to subjects as to where they could go, what they could do, or what they could wear. An overview of all sequences is provided in Table I. Example frames can also be viewed in Fig. 2.

A. 2-D Pedestrian Detection Data-Set

The first methodology proposed to evaluate a pedestrian detection technique is based on traditional 2-D image plane comparison techniques. In this approach, the synthetic data of Section III plus the first three real-world test scenarios were manually groundtruthed by positioning a separate bounding box around each person in the right stereo-camera image. In this process, a person is defined as *someone who has a section of their body above the waist, no matter how small, visible in the image*. In this process, the only constraint placed on creating groundtruth regions was that people who are further than 8 m from the camera are *not* considered valid pedestrians (however, this only effects the *Corridor* sequence).

Using these groundtruth annotations, a quantitative evaluation of a proposed monocular or stereo-based pedestrian detection technique can be undertaken via the calculation of a number of evaluation metrics. In this work, the *Localized Object Count Precision* and *Recall* metrics of [18] were chosen. For these metrics, a person within the video is declared as detected if their bounding box correctly overlaps that of a groundtruth annotation box by a predefined amount (set to 50% in our experiments). However, as both the groundtruth annotations *and* the outputs of our work (for both methodologies in this section) are available for download via the website outlined in Section I, a variety of other benchmarking metrics from the results of these sections can be easily obtained by interested readers.

B. 3-D Pedestrian Detection Data-Set and Evaluation

There are limitations associated with the proposed 2-D groundtruth evaluation process. They include the following.

- How accurate is the system for a maximum distance of 5, 6, 7, or 8 m? Does the system's performance degrade gradually or is there a threshold distance after which there is a large drop off in performance?
- How accurate are the 3-D statistics, such as height and 3-D position, obtained for each pedestrian detected?
- How does the system perform with respect to varying pedestrian numbers?
- Are missed pedestrians evenly distributed throughout a scene or does a proposed technique have difficulty segmenting pedestrians within specific areas?

To help answer some of these questions, we also propose to evaluate a system using a second methodology based on groundtruth data captured using a 3-D Vicon infrared motion analysis system.[9] The Vicon system is an automated motion capture system that tracks the 3-D position of infrared reflective markers in 3-D space with a high degree of accuracy (up to 1 mm in a 6-m space). For these experiments, five different test sequences were recorded—see Table I, where each sequence, *Vicon* n, consists of n people in the scene. For each sequence, the 3-D origin and coordinate axes of the Vicon and stereo-camera systems were coaligned. When generating these sequences, every person was tracked by the Vicon system using infrared reflective markers situated on the top of each person's head. In addition, the 3-D position of the groundplane was obtained via markers—therefore allowing the groundtruth height of each person in the scene to be obtained.

In order to evaluate a stereo-based pedestrian detection system, a comparison between the detected and groundtruth 3-D position of a pedestrian is proposed. However, in order for this comparison to be made, these two must correspond to the same real-world feature point. To expand on this point, consider that a 3-D groundtruth position is situated on the top of a person's head. However, a number of pedestrian detection techniques define the 3-D position of a detected person as the centroid of the person's body [10] or head region [19]. These vertical offsets in position features should not be incorporated into a 3-D positional evaluation metric. In order to address this, we propose to remove them by orthographically projecting both the detected and groundtruth 3-D position of a pedestrian onto the groundplane—via the technique outlined in [12]. Let these orthographically projected 3-D points for a detected d and groundtruth g pedestrian be represented by d^{3d} and g^{3d}, respectively. In addition, let the associated heights of d and g be represented by d^h and g^h, respectively.

[8] [Online]. Available: http://www.ptgrey.com

[9] [Online]. Available: http://www.vicon.com

We propose the use of four evaluation metrics: precision, recall, average 3-D positional error (APE), and average 3-D height error (AHE). The precision and recall metrics are similar to those applied in Section IV-A. However, we propose that a match between d and g be determined via their 3-D information and a biometric pedestrian model—based on the the application of the *Golden Ratio* $\Phi = \sqrt{5} * 0.5 + 0.5$ [20]. Using Φ and g^h, it is possible to determine an estimation of the groundtruth person's shoulder width g^s via $g^s = g^h/\Phi^3$. Using this value, a match is made between d and g if $g^{3d} - d^{3d} < g^s$, i.e., if d^{3d} is within shoulder distance of a groundtruth pedestrian 3-D position. The second two metrics, APE and AHE, are employed to provide an evaluation of detected pedestrians' 3-D positional and height accuracy, respectively. They are calculated as $\text{APE} = 1/n \sum_{i=1}^{n} \left| g_i^{3d} - d_i^{3d} \right|$ and $\text{AHE} = 1/n \sum_{i=1}^{n} \left| g_i^h - d_i^h \right|$, for n matched pedestrians.

It should also be noted that this data-set provides the means for further evaluation techniques which are beyond the scope of traditional 2-D evaluation techniques. For example, the 3-D positions of all *undetected* pedestrians in a sequence can be automatically obtained. Although this has minimal benchmarking value, this overview can be highly informative as the visualization of missed pedestrians can aid the interpretation of objective results, allowing the identification of areas within the scene where a proposed technique performs poorly.

Using this methodology, *some* of the outstanding issues from a 2-D evaluation process can be addressed. Unfortunately, due to limitations in the Vicon system's field of view, not all of the required data to answer all the outstanding questions could be gathered. In practice, the Vicon system provided an elliptical area of reliable tracking that measured 5.5 m in length and 3.15 m in width with respect to the stereo camera's principal axis. As pedestrians were therefore limited to this area, some of the outstanding questions relating to a proposed technique's performance at distances greater than 5.5 m could not be addressed. As part of future work, it is intended to recalibrate the Vicon system in a larger room, thereby incorporating a larger field of view to provide a means to answer some of the remaining questions.

V. Using the Proposed Evaluation Framework

In order to illustrate the usefulness of the proposed framework, here, an evaluation of a short-baseline stereo-based pedestrian detection technique [6] is carried out.

A. Evaluating Disparity Estimation

As outlined in Section III, we recommend using the Middlebury College open-source stereo algorithm evaluation test bed for this benchmarking. The Middlebury framework provides a standalone C++ implementation of many stereo algorithms. In addition, it provides a module for the quantitative evaluation of disparity results. From this module, two standard metrics are recommended [17]: 1) the rms (root-mean-squared) disparity error between a computed pixel's test and groundtruth disparity map and 2) the percentage of badly matching pixels (%BMP)—for this metric, a pixel is defined as being badly matched if the difference between its computed and groundtruth disparity is outside some error tolerance δ.

TABLE II
Disparity Results for all 16 Synthetic Scenes

Technique	*Average RMS error*	*Average %BMP*
Minimum Middlebury result	3.17	16.47
Authors' Technique	1.55	3.44

In order to benchmark our disparity estimation technique, 700 algorithms within the Middlebury evaluation framework were identified. Within these algorithms, three different types of optimization were tested—winner-takes-all, dynamic programming, and scanline optimization. For each stereo-image pair, the Middlebury algorithm with the minimum rms error was obtained—an average of 3.17 for these values over the 16 test image pairs was obtained (see Table II). In addition, a similar value for an average minimum %BMP was obtained—note that, in our experiments, δ was set to the average minimum rms error values from the 16 synthetic scenes within the test data-set (i.e., $\delta = 3.17$). This value of δ can be seen as representative of the inherent difficulty associated with the synthetic data-set. As the rms error is mathematically the spatial equivalent to the standard deviation of the disparity of pixels within a test disparity map from their respective groundtruth values, using this value of δ, the %BMP can be viewed as the percentage of pixels outside one standard deviation of the minimum average best rms error results for the 16 synthetic image-pairs.

Using the corresponding results of our proposed disparity estimation technique from these 16 stereo-image pairs, a comprehensive quantitative evaluation of its performance can be obtained. From Table II, it can be seen that a significant reduction in both average rms error and %BMP was obtained. These results highlight the advantage of applying the proposed disparity estimation technique, which has been specifically developed for applications involving pedestrian detection. Interested readers are directed to [6] for further details on the specifics of this experiment.

In this experiment, global-based optimization techniques were not employed due to the computational complexity involved in obtaining a result from the test stereo-pairs, resulting in processing times of up to 3 h for a single disparity map on a standard 2-GHz laptop (compared with < 10 s for the proposed approach). However, subsequent experiments have shown that incorporating global-based algorithms from the Middlebury framework only improves the overall rms and %BMP values by 0.16 and 0.82, respectively—these results are still outperformed by those obtained by the proposed system.

B. Evaluating Pedestrian Detection

The precision and recall values for the output of our system are presented in Table III, with a selection of results presented visually in Fig. 2. Within the context of the proposed evaluation framework, each individual sequence metric score is not highly important. They are, however, provided for benchmarking purposes, in the hope that the proposed framework will be adopted by the community. Here, a focus is made on how the information provided by the proposed framework can be adopted to help uncover the strengths and weaknesses in a proposed pedestrian detection technique.

TABLE III
PRECISION AND RECALL EVALUATION OF ALL SEQUENCES. NOTE: APE AND AHE ARE BOTH DEFINED IN CENTIMETERS

Sequence	*Detected*	*Correct*	*Precision*	*Recall*	*APE*	*AHE*
Synthetic	95	93	97.89	95.88	–	–
Overhead	641	610	95.16	92.85	–	–
Corridor	969	883	91.12	85.98	–	–
DCU Cor.	4866	4749	97.60	92.65	–	–
2D Total	6571	6335	96.41	91.72	–	–
Vicon 1	198	198	100.0	100.0	10.5	7.3
Vicon 2	533	526	98.69	100.0	11.7	8.4
Vicon 4	1301	1291	99.23	99.61	8.2	9.8
Vicon 8_A	1980	1976	99.80	93.92	10.0	10.4
Vicon 8_B	1942	1942	100.0	91.60	7.4	10.6
3D Total	5954	5933	99.65	95.02	8.9	10.0
Total	12525	12268	97.95	93.29	8.9	10.0

The performance of the proposed system with respect to varying pedestrian numbers can be viewed in the *Vicon* scenario results, where an increase in numbers results in a decrease in recall and AHE performance (mainly due to increasing occlusion). However, the precision metric remains relatively constant, revealing robustness against false-positives. In addition, the APE metric reveals a consistency in the level of accuracy within the 3-D position of detected pedestrians.

In addition, due to the differing surveillance scenarios within the proposed evaluation framework, a comparison of performances across all test sequences can help determine areas of weakness in a proposed pedestrian detection technique. For example, from Table III, the system performance of the 2-D sequences—in particular, the *Corridor* sequence—is less than those of the 3-D sequences. This drop is partly due to an increase in over-segmentation and missed pedestrians for the proposed system with respect to increased pedestrian distance from the camera. In addition, the examined technique does not robustly handle people ascending/descending stairs in the *Corridor* sequence resulting in lower performance scores.

VI. CONCLUSIONS AND FUTURE WORK

Here, we presented a framework for evaluating the accuracy of pedestrian detection algorithms, both at system and component levels, via a number of publicly available test sequences and groundtruth sets that incorporate many of the challenges that are inherent for pedestrian detection in real application scenarios. However, it is acknowledged that there are certain areas of evaluation that have *not* been incorporated to date within our framework. It is intended to capture further sequences in order to address these deficiencies. These include, amongst other scenarios: 1) pedestrians situated beside background objects of similar range, such as walls; 2) rapidly changing lighting conditions; 3) nonpedestrian objects; 4) pedestrians sitting, jumping, and running; and 5) sequences taken from different stereo (and infrared [21]) camera setups. Finally, we wish to extend the groundtruth data beyond pedestrian detection and into the area of pedestrian tracking.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of Dr. K. Moran, DCU School of Sports Science, and Dr. R. McDonnell, Graphics Vision and Visualisation Group in TCD, for their help in the creation of test sequences and groundtruth data.

REFERENCES

[1] L. Zhao and C. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Trans. Intell. Transportation Syst.*, vol. 1, no. 3, pp. 148–154, Sep. 2000.

[2] D. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, vol. 1, pp. 87–93.

[3] A. Baumberg, "Learning deformable models for tracking human motion," Ph.D. dissertation, Sch. of Computer Studies, Univ. of Leeds, Leeds, U.K., 1995.

[4] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 878–885.

[5] M. Harville, "Stereo person tracking with adaptive plan-view templates of height and occupancy statistics," *Int. J. Comput. Vis.*, vol. 22, pp. 127–142, 2004.

[6] P. Kelly, "Pedestrian detection and tracking using stereo vision techniques," Ph.D. dissertation, Sch. of Electron. Eng., Dublin City Univ., Dublin, Ireland, 2007.

[7] M. Harville, "Stereo person tracking with short and long term plan-view appearance models of shape and color," in *Proc. IEEE Conf. Adv. Video and Signal Based Surveillance*, 2005, pp. 522–527.

[8] A. Mittal and L. Davis, "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo," in *Proc. Eur. Conf. Comput. Vis.*, 2002, vol. 1, pp. 18–36.

[9] S. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 133–146.

[10] M. Keck, J. Davis, and A. Tyagi, "Tracking mean shift clustered point clouds for 3d surveillance," in *Proc. ACM Int. Workshop Video Surveillance and Sensor Networks*, 2006, pp. 187–194.

[11] Real-Time Stereo Vision Based on the Uniqueness Constraint: Experimental Results and Applications [Online]. Available: www.vision.deis.unibo.it/smatt/stereo.htm

[12] P. Kelly, N. O'Connor, and A. Smeaton, "Pedestrian detection in uncontrolled environments using stereo and biometric information," in *Proc. ACM Int. Workshop Video Surveillance and Sensor Networks*, 2006, pp. 161–170.

[13] A. Senior, "Tracking with probabilistic appearance models," in *Proc. ECCV Workshop Perform. Eval. Tracking and Surveillance Syst.*, 2002, pp. 48–55.

[14] S. Khan and M. Shah, "Tracking people in presence of occlusion," in *Proc. Asian Conf. Comput. Vis.*, 2000, pp. 1132–1137.

[15] L. Brown, A. Senior, Y. Tian, J. Connell, A. Hampapur, C. Shu, H. Merkl, and M. Lu, "Performance evaluation of surveillance systems under varying conditions," in *Proc. IEEE Int. Workshop PETS*, 2005, pp. 1–8.

[16] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta, "Occlusion detectable stereo-occlusion patterns in camera matrix," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1996, pp. 371–378.

[17] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, pp. 7–42, 2002.

[18] V. Mariano, J. Min, J.-H. Park, R. Kasturi, D. Mihalcik, H. Li, D. Doermann, and T. Drayer, "Performance evaluation of object detection algorithms," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, vol. 3, pp. 965–969.

[19] J. Batista, "Tracking pedestrians under occlusion using multiple cameras," in *Image Anal. Recognit.*, 2004, pp. 552–562.

[20] P. Kelly, E. Cooke, N. O'Connor, and A. Smeaton, "Pedestrian detection using stereo and biometric information," in *Proc. Int. Conf. Image Anal. Recognit.*, 2006, pp. 802–813.

[21] M. Bertozzi, E. Binelli, A. Broggi, and M. D. Rose, "Stereo vision-based approaches for pedestrian detection," in *Proc. IEEE Int. Workshop Object Tracking and Classification in and Beyond the Visible Spectrum*, 2005, vol. 3, pp. 16–22.