Isotropic Granularity-tunable gradients partition (IGGP) descriptors for human detection

Yazhou Liu
http://www.cse.oulu.fi/YazhouLiu
Janne Heikkila
http://www.cse.oulu.fi/JanneHeikkila

Machine Vision Group
University of Oulu
Oulu, Finland

Abstract

This paper presents a new descriptor for human detection in still images. It is referred to as isotropic granularity-tunable gradients partition (IGGP), which is extended from granularity-tunable gradients partition (GGP) descriptors. The isotropic representation is achieved by aligning the features with different orientation channels according to their principal angles. The benefits of this extension are two folds: firstly, since the partitions' sizes of all the orientation channels are equal, the noise introduce by the small partitions in the original GGP descriptors is eliminated and the performance can be essentially improved; secondly, the integral image based fast computation is applied and more than 20 times speedup has been achieved. In addition, we introduce a new human dataset HIMA. Unlike the previous available human datasets which are mainly captured on the street views for automobile safety or robotics, HIMA dataset is captured on the outdoor work fields for industry safety. The major challenges include: extreme light conditions, occlusion and strong noise. We benchmark several promising detection systems, providing an overview of state-of-the-art performance on the HIMA set. Experimental results show that the proposed method can yield very competitive results in both the detection speed and accuracy.

1 Introduction

Human detection research has received more and more attention in recent years because of increasing demands in practical applications, such as smart surveillance system, on-board driving assistance system and content based image/video management system. Even through remarkable progress has been achieved [2, 3, 4, 5, 8, 10, 20, 22, 25, 28], finding the human is still considered as one of the hardest task for object detection. The difficulties come from the articulation of human body, the inconsistency of clothes, the variation of the illumination and the unpredictability of the occlusion.

Varieties of features have been invented to overcome the difficulties mentioned above. Earlier works for human detection started from Haar-like features, which have been applied to face detection task successfully [19, 21, 24]. Because of the large variation of human clothes and background, some researchers turned to the contour based descriptors. Gavrila [9] presented a contour based hierarchical chamfer matching detector. Lin et al. [14, 15]

BMVC 2010 doi:10.5244/C.24.63

extended this work by decomposing the global shape models into parts to construct a parts template based hierarchical tree. Ferrari el al. [7] used the network of contour segments to represent the shape of the object. Wu and Nevatia [29] used edgelet to represent the local silhouette of the human.

After the invention of the SIFT descriptor[17], more researchers have used the statistical summarization of the gradients to represent human body. Such as the position-orientation histogram features proposed by Mikolajczyk et al. [18]; the histograms of oriented gradients (HOG) proposed by Dalal et al. [1, 30] and it's improvements [6]; the implicit shape model proposed by Leibe et al. [12, 13]; the covariance matrix descriptor proposed by Tuzel et al. [23]; and the HOG-LBP descriptor proposed by Wang et el. [26].

Recently, granularity-tunable gradients partition (GGP) descriptor was proposed by Liu et al. [16], in which granularity is used to define the spatial and angular uncertainty of the line segments in the Hough space. In the formulation of GGP, the feature extraction contains two steps: firstly, the image is parsed as the combination of the generalized lines by orientation space partition; secondly, the heterogeneous GGP feature vector is calculated within the generalized lines. By this means, the GGP descriptor can encode both the geometrical structure and the statistical summarization of the objects.

The rationale of GGP is reasonable but there are some difficulties in its implementation part. Take Figure.1(c) for example, the partitions are uneven for the channels whose principal orientations are not equals to 90° or 0°. Since the size of the partitions are different, on the one hand, the center partitions with bigger size become dominant and the contributions of the other partitions are suppressed; on the other hand, the minor partitions can introduce noise because its insufficient gradient points makes the feature values become statistically unstable. Moreover, the shapes of the partitions are different, which makes the fast computation intractable.

The difficulties mentioned above motivate the works of this paper. By introducing the isotropic feature representation, a substantial performance improvement is observed and the computation complexity is reduced from $O(n*w*h)$ to $O(n)$, where n is the number of orientation partition and (w,h) is the size of the feature window. Practically, 23 times speedup is achieve by IGGP over GGP.

The rest of the paper is organized as follows: Section 2 introduces the basic idea of IGGP descriptor; Section 3 provides the computational details; and Section 4 contains the experimental results on INRIA dataset and our new HIMA dataset.

2 Isotropic Granularity-tunable gradients partition (IGGP) descriptors

Granularity-tunable gradients partition (GGP) descriptor was proposed for human detection by Liu et al. [16], in which the authors defined the *generalized line* in the Hough space by extending the classic definition of lines with spatial and angular uncertainties. These uncertainties are referred to as granularity. By adjusting the granularity, GGP provides a container of descriptors from deterministic to statistic.

In their work, the feature extraction procedure was represented as:

$$f(I;\vartheta,\tau) = S(T(I;\vartheta,\tau)) \tag{1}$$

where τ is the granularity parameter and ϑ is the feature parameter. The feature extraction

function $f(\cdot)$ is represented by the composition of two functions: image parsing (IP) function $T(\cdot)$ and image description (ID) function $S(\cdot)$. Intuitively, the GGP feature extraction procedure contains two steps: firstly, the original image is parsed into the geometrical structures by $T(\cdot)$; and secondly, the descriptions of these geometrical structures are generated by $S(\cdot)$.

The granularity control is accomplished by image parsing function, which divided original images into partitions that corresponding to the generalized lines defined in the Hough space. The overall feature extraction can be summarized as follows: firstly, *orientation partition* divide the original gradient image into different channels according to their gradient orientations; secondly, *Space partition* further partition each channel image into parallel line belt regions, where the tangent angle of each partition line equals to the gradient orientation of each channel image; thirdly, the GGP descriptors are extracted for the partitions which corresponding to the generalized lines in the Hough space. As shown in Figure.1(a)∼(c), the yellow rectangle is the feature window and the red partitions are the generalized lines with the maximum strength in each orientation.

This partition method is straightforward but suffer from some difficulties:

- The selection of the partition with the maximum strength is biased for some orientation. Take the 45° channel image for example, as shown in the Figure.1(c), the partitions in the feature window are of different sizes and the partitions at the middle position have much bigger areas than the others. Therefore, when we selecting the partition with the maximum gradient strength, the middle partitions have bigger possibility to be selected due to its dominant partition size.

- The small partitions may introduce the noise into the descriptor. GGP aiming at selecting the most prominent line structure for each orientation, but if in the flat region or texture region, the trivial structures in the small partitions can be amplified by its normalization method and therefore introduce the noise into the descriptor.

- The partitions are of different shapes that make the fast calculation become inapplicable. As shown in Figure.1(c), the shapes of the partitions of different orientations are quite different. The partitions in the 90° channel image are rectangles but the partitions in the 45° channel image are triangles, pentagons and hexagons. Therefore, it is hard for integral image based fast feature calculation.

In order to overcome the difficulties mentioned above, we developed the isotropic granularity-tunable gradients partition (IGGP). *Isotropic* means for all the orientations, the partitions of a feature are of the same size and shape.

Given image I, using the filter $[-1,0,1]$, a gradient image dI is generated as shown in Figure.1(a)∼(b). Then this gradient image dI is divided into n disjoint orientation channels as:

$$\left\{ Q_{\theta_i} | i = 1,\ldots,n; \bigcup_{i=1}^{n} Q_{\theta_i} = dI; Q_{\theta_l} \bigcap Q_{\theta_o} = \emptyset, l \neq o \right\} \tag{2}$$

where n is the number of orientation partition and each orientation channel Q_{θ_i} only contain the pixels whose norm angle can be quantized as θ_i and all the other pixels' strength will be set to zero, refer to [16] for more details. The θ_i is referred to as the principal angle of each channel.

A feature $R(x_c, y_c, w/2, h/2)$ is specified by it center (x_c, y_c) and half window size $(w/2, h/2)$. IGGP create a feature window R_{θ_i} for each orientation channel Q_{θ_i} by rotate the feature rectangle around its center by angle $\theta_i - 90°$. Therefore, the height edge of the feature window

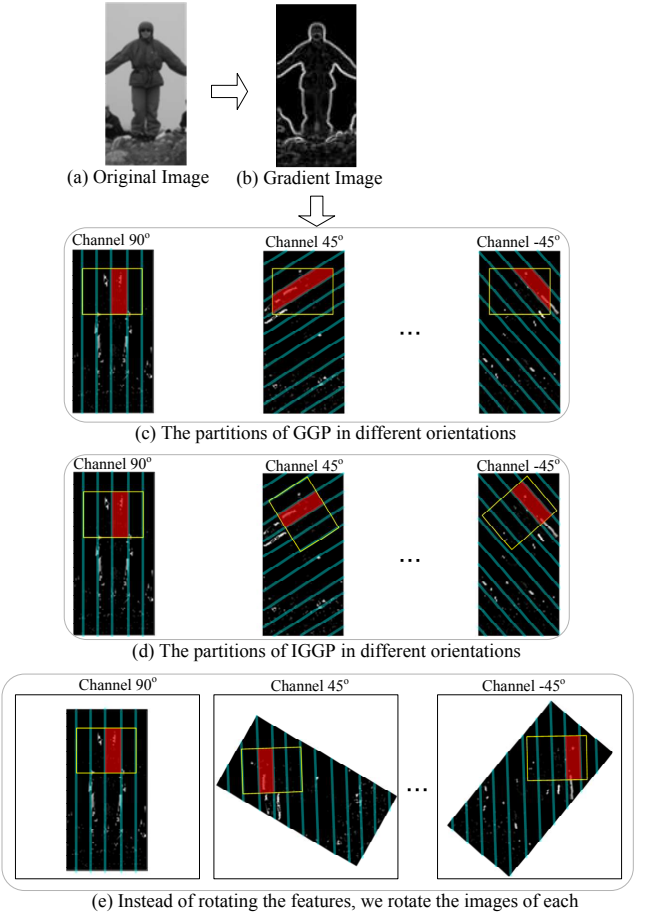

Figure 1: The difference between GGP and IGGP

will be parallel to the partitions, as shown in Figure.1(d). The partitions for all the orientation channels have the same sizes and shapes. In order to develop the integral image based fast calculation, we rotate the channel images instead of the features which will be detailed in the following section.

3 Integral Image based Fast Computation of IGGP

For a given image I, we divide it into n orientation channels as Equ.2. Then each orientation channel Q_{θ_i} is rotated by angle $90^\circ - \theta_i$ as shown in Figure.1(e). This rotated channel is referred to as Q'_{θ_i}. The position (x,y) in the Q_{θ_i} coordinate frame is mapped to (x',y') in the Q'_{θ_i} coordinate frame. We maintain six computation images for each channel:

- Q'_{θ_i}: the rotated orientation image:

$$Q'_{\theta_i}(x',y') = \begin{cases} s & \text{if the quantized gradient angle at } (x,y) \text{ is } \theta_i \\ 0 & \text{others} \end{cases} \tag{3}$$

  where s is the gradient strength at (x,y).

- C'_{θ_i}: the counter image:

$$C'_{\theta_i}(x',y') = \begin{cases} 1 & \text{if } Q'_{\theta_i}(x',y') > 0 \\ 0 & \text{others} \end{cases} \tag{4}$$

- X'_{θ_i}: x position image:

$$X'_{\theta_i}(x',y') = \begin{cases} x' & \text{if } Q'_{\theta_i}(x',y') > 0 \\ 0 & \text{others} \end{cases} \tag{5}$$

- Y'_{θ_i}: y position image:

$$Y'_{\theta_i}(x',y') = \begin{cases} y' & \text{if } Q'_{\theta_i}(x',y') > 0 \\ 0 & \text{others} \end{cases} \tag{6}$$

- $X2'_{\theta_i}$: x position square image:

$$X2'_{\theta_i}(x',y') = \begin{cases} x'^2 & \text{if } Q'_{\theta_i}(x',y') > 0 \\ 0 & \text{others} \end{cases} \tag{7}$$

- $Y2'_{\theta_i}$: y position square image:

$$Y2'_{\theta_i}(x',y') = \begin{cases} y'^2 & \text{if } Q'_{\theta_i}(x',y') > 0 \\ 0 & \text{others} \end{cases} \tag{8}$$

For simplicity, we will drop the subscript θ_i and superscript $'$ in the following descriptions. The integral images of these computation images are calculated and referred as iT, where T is the symbol of the computation image. For example, iQ represent the integral

image of the orientation image Q. For a given rectangle $r(x,y,w,h)$ and a computation image T, the summation within the rectangle can be represented as:

$$T_r = iT(x+w,y+h) - iT(x+w,y) - iT(x,y+h) + iT(x,y) \tag{9}$$

Since all the partitions in the feature window are the rectangles with upright positions, as shown in Figure.1(e), the heterogeneous features can be calculated by the integral image with constant number of computations. Here we just give an example on how to calculate the standard deviation of positions along the tangent direction of partition r in Equ.10. The computation of other elements are straightforward.

$$v_{tang} = \sqrt{Y2_r/C_r - m_y^2}/h \tag{10}$$

where:

m_y	—	the mean position in y direction, can be calculated as $m_y = Y_r/C_r$;
$Y2_r$, C_r and Y_r	—	can be calculated as Equ.9;
h	—	the height of partition r;

Given a feature window $R(x_c,y_c,w/2,h/2)$, the computation complexity of IGGP is $O(n)$, where n is the number of orientation partition; the computation complexity of GGP is $O(n*w*h)$. In practical computation, a 23 times speed up can be achieved by IGGP over GGP.

4 Experiments

We evaluate the proposed method on both the public INRIA human dataset and our new human dataset, which is referred to as HIMA dataset. Firstly, IGGP is evaluated against the-state-of-the-art detectors on the INRIA dataset based on both the FPPW and FPPI criteria; secondly, a new human dataset which captured on the outdoor work fields is introduced and detailed evaluation results on this challenging dataset are presented.

For IGGP, all the training data come from INRIA dataset, including 2416 positive training samples and 2436 background images. The size of our normalized sample is 64×128. A LogitBoost classifier with rejection cascade is build for human detection. The weighted linear regression function is used as the weak classifier. For each stage of cascade, we use the total 2416 positive training samples and 10000 negative training samples. For the first stage, the negative samples are randomly selected from the background images; and for the following n_{th} stages, the negative samples are selected by bootstrapping of previous $n-1$ stages. For each stage, the minimum detection rate is 99.7% and maximum false positive rate is 35%. The final detector contains 30 stages and 538 weak classifiers.

4.1 Detection Results on INRIA dataset

INRIA human data set is one of the most widely used dataset for human detection, and we firstly evaluate the proposed method on this dataset. We use two different criteria: 1) The miss rate vs. false positive per window (FPPW); and 2) The miss rate vs. false positive per image (FPPI). Since IGGP is an extension of GGP and we've claim that IGGP can reduce the noise effect that induced by the minor partitions in GGP, it is worth to make detailed comparisons between these two methods. Based on the FPPW criteria, the detection results is presented in Figure.2(a). The overall performance of IGGP is superior to GGP and the miss

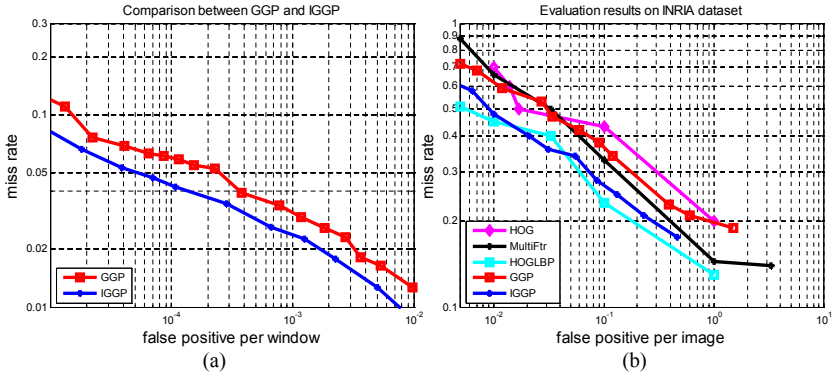

Figure 2: Evaluation results on INRIA dataset. (a) The comparison between GGP and IGGP base on FPPW criteria. (b) The comparison between IGGP and the state of the art methods base on FPPI criteria.

SeqIndex	Frames	Annot.	Visi. Annot.	Scenario	Challenges
1	300	620	452	Snow	Low light
2	300	552	470	Warehouse	Smoke
3	700	1460	1199	Mine	Low light, cluttered background
4	2500	2131	1733	Stree view	Small person size, occlusion
5	700	1382	1362	Harbor	Small person size, cluttered background
6	700	507	422	Harbor	Small person size, cluttered background
7	700	641	457	Harbor	Small person size, occlusion
ALL	5900	7293	6095	N/A	Outdoor working condition

Table 1: The summarizations of the seven subsets of HIMA dataset

rate is reduce by 2.1% at 10^{-4} FPPW. In addition, processing 105483 scan windows cost GGP 263 seconds and IGGP 11.3 seconds on a PC with Intel 2.39GHz processor. 23 times speedup has bee achieved by this simple extension. The evaluation results based on FPPI criteria are presented in Figure.2(b). We also quote some of the best published detectors: **HOG** detector from [1]; the **HOG-LBP** detector from [26]; and the **MultiFtr** detector from [27]. These results show that IGGP is slightly behind the currently best detector HOG-LBP.

4.2 Detection Results on HIMA dataset

In this section, we firstly introduce a new human dataset which is referred to as HIMA dataset. Unlike the previous available human datasets which are mainly captured on the street views for automobile safety or robotics, HIMA dataset is captured on the outdoor work fields and its target application is for outdoor industry machines. The major challenges includes: extreme light conditions, occlusions and strong noise. More specially, HIMA dataset contain 7 subsets which captured in five different outdoor working scenarios: snow, warehouse, mine, street and harbor. Totally there are 5900 frames, 7298 annotated people and 6095 of them are fully visible. The detailed summarization of each subset can be seen in Table.1. The sample images of these subsets can be seen in Figure.4(a)∼(g).

Besides the IGGP, we select five recently published promising detection systems as the baselines. These methods includes: **Haar** from [11] is a Haar filter based human detector;

Method Name	Image Size	Stride	Scale Step	Time (Second/image)
Haar	1024x768	(2, 2)	1.05	0.12
HOG	1024x768	(8, 8)	1.05	1.3
Pid	1024x768	(2, 2)	1.05	33
HOGLBP	1024x768	(4, 4)	1.3	69
PLS	1024x768	(4, 8)	1.15	324
IGGP	1024x768	(4, 4)	1.2	11.3

Table 2: Average detection time on HIMA dataset

HOG from [1] is the most widely used baseline for human detection; **Pid** from [14] is a pose invariant descriptor using combination of contour and HOG feature; **HOG-LBP** from [26] and **PLS** from [22] are the best results in year 2009. For **HOG** detector, the default minimum detection window is 64×128. In order to enable it to detect smaller persons, we resize the input images to 2 times of its original size before detection and we refer to this detector as **HOG_resize**. All of these detectors are from the authors, and their training data may come from different dataset. We use FPPI as the criteria and the evaluation results on the HIMA subsets can be seen from Figure.3(a)∼(g). **IGGP** achieves the leading performance on 5 of the subsets, especially on SET2, SET3 and SET5 it outperform the other detectors by a big margin. The performance of **HOG-LBP** is also impressive, and it achieves the best performance on 3 of the subsets.

We also evaluate the detection speed of all the detectors and the results can be seen in Table.2. Ideally, we should make sure that all the methods have the same number of scan window then evaluate their speed. But since some of the detectors are provided as the binary and it is not easy to count their actual number of scan window. Therefore, we just provide the detection parameters here. For **Haar** and **HOG**, they have been optimized for multi-thread processing in OpenCV. After **Haar** and **HOG**, **IGGP** take the third place and using 11.3 second to detect a 1024×768 image with specified parameters.

The sample of detection results on the seven subsets of HIMA dataset are shown in Figure.4(a)∼(g).

5 Conclusion

A new descriptor, isotropic granularity-tunable gradients partition (IGGP), has been presented in this paper, which extended from the GGP descriptor by introducing the isotropic feature representation. This simple extension eliminate the noises caused by uneven partition size in GGP and make the integral based fast computation possible. Therefore, both the accuracy and the speed have been improved substantially. In addition, a new challenging human dataset HIMA has been introduced, which captured in the wild working conditions. Extensive evaluations on both the public INRIA dataset and new HIMA dataset show the superior performance of the proposed descriptor.

References

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

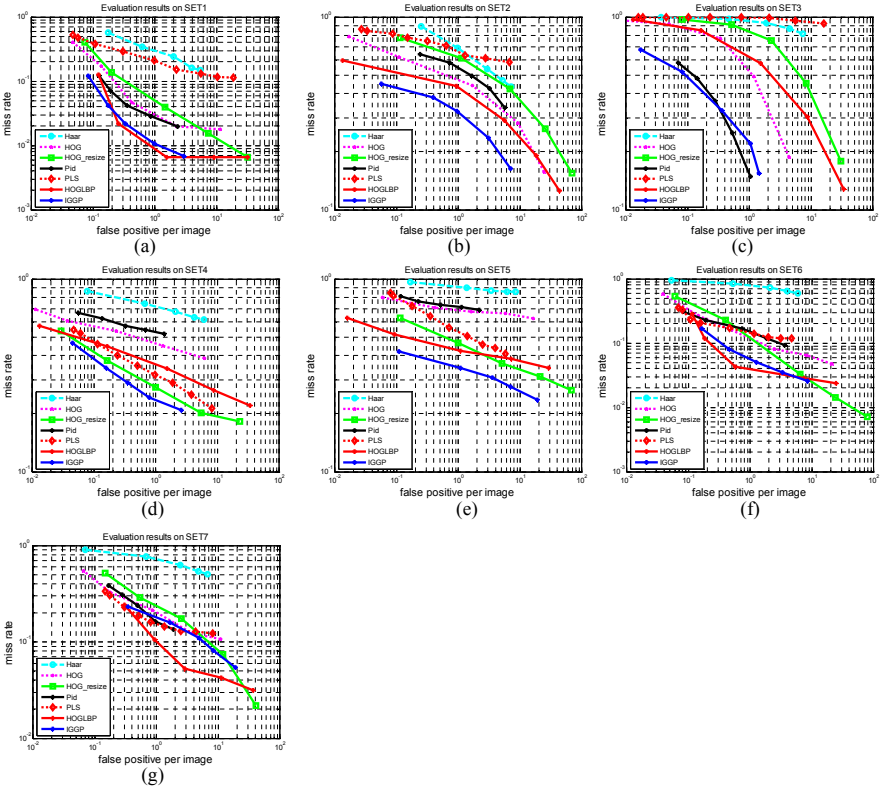

Figure 3: Evaluation results on HIMA dataset

[2] Navneet Dalal, Bill Triggs, and Cordelia Schmid. Human detection using oriented histograms of flow and appearance. In *ECCV*, pages 428–441, 2006.

[3] Piotr Dollar, Boris Babenko, Serge Belongie, Pietro Perona, and Tu Zhuowen. Multiple component learning for object detection. In *ECCV*, pages 211–224, 2008.

[4] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: A benchmark. In *CVPR*, 2009.

[5] Andreas Ess, Bastian Leibe, and Luc Van Gool. Depth and appearance for mobile scene analysis. In *ICCV*, pages 14–21, 2007.

[6] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.

[7] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool. Object detection by contour segment networks. In *ECCV*, pages 14–28, 2006.

[8] Juergen Gall and Victor Lempitsky. Class-specific hough forests for object detection. In *CVPR*, 2009.

[9] Dariu M. Gavrila. Pedestrian detection from a moving vehicle. In *ECCV*, 2000.

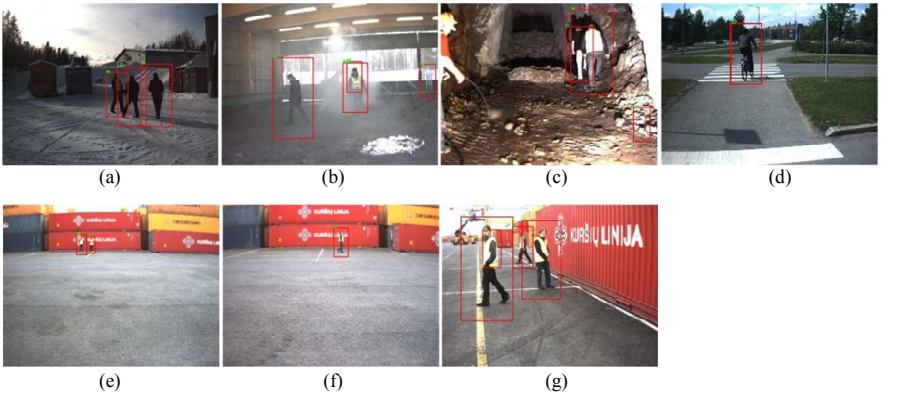

Figure 4: Sample detection results on HIMA dataset

[10] Feng Han, Ying Shan, Harpreet S. Sawhney, and Rakesh Kumar. Discovering class specific composite features through discriminative sampling with swendsen-wang cut. In *CVPR*, 2008.

[11] Hannes Kruppa, Modesto Castrillon Santana, and Bernt Schiele. Fast and robust face finding via local context. In *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 157–164, 2003.

[12] Bastian Leibe, Ales Leonardis, and Bernt Schiele. Combined object categorization and segmentation with an implicit shape model. In *ECCV Workshop on Statistical Learning in Computer Vision*, Prague, 2004.

[13] Bastian Leibe, Edgar Seemann, and Bernt Schiele. Pedestrian detection in crowded scenes. In *CVPR*, volume 1, pages 878 – 885, 2005.

[14] Zhe Lin and Larry S. Davis. A pose-invariant descriptor for human detection and segmentation. In *ECCV*, pages 423 – 436, 2008.

[15] Zhe Lin, Larry S. Davis, David Doermann, and Daniel DeMenthon. Hierarchical part-template matching for human detection and segmentation. In *ICCV*, 2007.

[16] Yazhou Liu, Shiguang Shan, Wenchao Zhang, Wen Gao, and Xilin Chen. Granularity-tunable gradients partition (ggp) descriptors for human detection. In *CVPR*, pages 1255 – 1262, 2009.

[17] David G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.

[18] Krystian Mikolajczyk, Cordelia Schmid, and Andrew Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In *ECCV*, pages 69–81, 2004.

[19] Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio. Example-based object detection in images by components. *TPAMI*, 23:349–361, 2001.

[20] Patrick Ott and Mark Everingham. Implicit color segmentation features for pedestrian and object detection. In *ICCV*, pages 724–730, 2009.

[21] Constantine Papageorgiou and Tomaso Poggio. A trainable system for object detection. *IJCV*, 38:15–33, 2000.

[22] William Robson Schwartz, Aniruddha Kembhavi, David Harwood, and Larry S. Davis. Human detection using partial least squares analysis. In *ICCV*, 2009.

[23] Oncel Tuzel, Fatih Porikli, and Peter Meer. Human detection via classification on riemannian manifolds. In *CVPR*, 2007.

[24] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, pages 511–518, 2001.

[25] Paul Viola, Michael J. Jones, and Daniel Snow. Detecting pedestrians using patterns of motion and appearance. In *ICCV*, pages 734–741, 2003.

[26] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In *ICCV*, pages 32–39, 2009.

[27] Christian Wojek and Bernt Schiele. A performance evaluation of single and multi-feature people detection. In *30th DAGM symposium on Pattern Recognition*, page 82ÍC91, 2008.

[28] Christian Wojek, Stefan Walk, and Bernt Schiele. Multi-cue onboard pedestrian detection. In *CVPR*, 2009.

[29] Bo Wu and Ram Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *ICCV*, 2005.

[30] Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng. Fast human detection using a cascade of histograms of oriented gradients. In *CVPR*, pages 1491– 1498, 2006.