# Human Detection by Quadratic Classification on Subspace of Extended Histogram of Gradients

Amit Satpathy, *Student Member, IEEE*, Xudong Jiang, *Senior Member, IEEE*, and How-Lung Eng, *Member, IEEE*

***Abstract*— This paper proposes a quadratic classification approach on the subspace of Extended Histogram of Gradients (ExHoG) for human detection. By investigating the limitations of Histogram of Gradients (HG) and Histogram of Oriented Gradients (HOG), ExHoG is proposed as a new feature for human detection. ExHoG alleviates the problem of discrimination between a dark object against a bright background and vice versa inherent in HG. It also resolves an issue of HOG whereby gradients of opposite directions in the same cell are mapped into the same histogram bin. We reduce the dimensionality of ExHoG using Asymmetric Principal Component Analysis (APCA) for improved quadratic classification. APCA also addresses the asymmetry issue in training sets of human detection where there are much fewer human samples than non-human samples. Our proposed approach is tested on three established benchmarking data sets – INRIA, Caltech, and Daimler – using a modified Minimum Mahalanobis distance classifier. Results indicate that the proposed approach outperforms current state-of-the-art human detection methods.**

***Index Terms*— Histogram of gradients, human detection, HOG, dimension reduction, asymmetric principal component analysis.**

## I. INTRODUCTION

COMPUTER vision and machine intelligence have been the foci of researchers since the invention of computers. Over recent years, researchers have been working on replacing humans with computers to take over the labour-intensive and time-consuming tasks. One of the key areas is object detection from images and videos. Particularly, human detection has been gaining popularity. Several areas of applications have spurred the interest of human detection such as human-computer interaction for video games, robotics, video surveillance, smart vehicles etc. However, human detection is a challenging problem due to the huge intra-class variation stemming from colour, clothing, pose and appearance variations. Furthermore, external conditions such as partial occlusions, illumination and background clutter further compound the problem.

The work on human detection are categorized under 2 approaches - feature development and classifier development. Humans can either be detected wholly or by parts. There have been numerous features proposed by researchers for holistic human detection such as Haar Wavelet features [35], Haar-like features and motion information [46], edge templates [12], Implicit Shape Models [21], Adaptive Contour Features [11], [27], Histogram of Oriented Gradients (HOG) [3] and the Covariance descriptor [45]. Holistic human detection methods, in general, underperform when there are heavy occlusions or extreme pose variations in the image.

In contrast to holistic human detection methods, parts-based human detection [1], [9], [10], [23], [24], [32], [33], [50], [55] is able to handle occlusions more robustly. The main concern in parts-based methods is high false positives number. Hence, most research in this area is devoted to determining robust assembly methods to combine detections and eliminate false positives. Furthermore, for good performance, high resolution images are needed to extract sufficient and robust features for each human part. In practice, these kind of images are usually not available. There are also some work that propose hybrid features [1], [5]–[7], [25], [41], [48], [50], [52], [54]. However, the improved performance comes at the expense of increasing the feature dimensionality.

Feature development methods for human detection usually use classifiers such as Support Vector Machines (SVM) [3], [8]–[10], [22], [25], [29], [30], [33]–[37], [48], [50], [59] and cascade-structured boosting-based classifiers [4]–[7], [11], [13], [14], [16], [24], [26], [27], [32], [38], [44]–[47], [49], [51], [53]–[55], [58], [60]. Among methods that use SVM classifiers, linear SVM classifiers are generally preferred for speed and to minimize the overfitting problem of non-linear SVM kernels.

Some work has been done on classifier development for human detection. The reason for such a direction stems from the problem of large intra-class variation of humans. Building features that can handle the large intra-class variation of humans is difficult. However, since classifiers use statistical methods for classification, intra-class variations can be better controlled. Some popular classifier frameworks include

Manuscript received May 13, 2012; revised December 23, 2012 and April 25, 2013; accepted May 15, 2013. Date of publication May 22, 2013; date of current version November 28, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hsueh-Ming Hang.

A. Satpathy is with the School of Electrical and Electronics Engineering, Nanyang Technological University, 639798 Singapore, and also with the Institute for Infocomm Research, Agency for Science, Technology & Research, 138632 Singapore (e-mail: amit0010@ntu.edu.sg; satpathya@i2r.a-star.edu.sg).

X. Jiang is with the School of Electrical and Electronics Engineering, Nanyang Technological University, 639798 Singapore (e-mail: exdjiang@ntu.edu.sg).

H.-L. Eng is with the Institute for Infocomm Research, Agency for Science, Technology & Research, 138632 Singapore (e-mail: hleng@i2r.a-star.edu.sg).

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

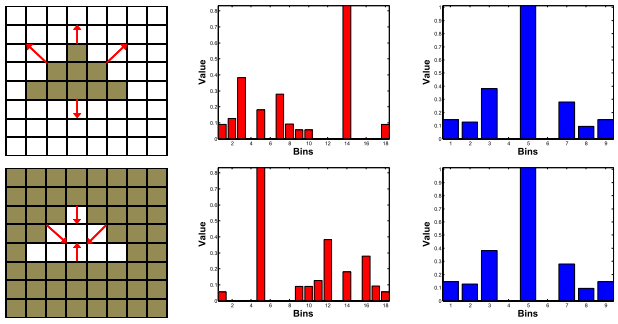
Digital Object Identifier 10.1109/TIP.2013.2264677

Wu *et al.*'s Cluster Boosted Tree [53], [55], Maji *et al.*'s [30] Intersection Kernel for SVM, Multiple Instance Learning [4], [47] and Lin *et al.*'s [24] "seed-and-grow" scheme.

HOG is the most popular feature used for human detection [1], [3], [8]–[10], [22], [25], [29], [30], [36], [37], [41], [48], [50], [52], [54], [59], [60]. It densely captures gradient information within an image window and is robust to small amounts of rotation and translation within a small area. It was introduced to solve the issue of differentiation of a bright human against a dark background and vice versa by Histogram of Gradients (HG) [28]. HOG maps all gradients of opposite directions to the same orientation. However, for the *same cell*, HOG also maps all gradients of opposite directions to the same orientation. As such, it is unable to differentiate some local structures and produces the same feature for some different local structures. Following preliminary work in [39], this paper proposes a new feature called the *Extended Histogram of Gradients* (ExHoG) by observing the inherent weaknesses in HOG and HG. The proposed feature differentiates most local structures which HOG misrepresents. It also alleviates the brightness reversal problem of human and background.

Linear SVM classifiers are popular in human detection as non-linear SVM classifiers have a high computational burden for high-dimensional features. Furthermore, the degree of non-linearity of SVM classifiers is unknown and cannot be easily controlled. As a result, non-linear SVM classifiers may suffer from overfitting during training. However, the classification capability of linear SVM is limited. Based on preliminary results in [40], we propose a classification framework that includes a modified Minimum Mahalanobis Distance classifier and Asymmetric Principal Component Analysis (APCA) [17], [18]. This paper discovers that the boundary between human and non-human can be well approximated by a hyper-quadratic surface. However, for high-dimensional features, the estimated eigenvalues in some feature dimensions deviate greatly from that of the data population which results in overfitting of the quadratic classifier. Hence, there is a need to reduce ExHoG dimensions to minimize the overfitting problem. Furthermore, training sets, usually, contain much fewer positive samples than the negative ones which results in the negative covariance matrix being more reliable than the positive covariance matrix. Using PCA is inefficient as the unreliable dimensions from the less reliable covariance matrix are not effectively removed. To tackle the problems of dimensionality reduction and the asymmetry issue of human training sets, we propose using APCA for dimension reduction. As a result, the projected features allow for a more robust classifier to be trained that less overfits the training data.

We present comprehensive experimental results to verify the validation of the proposed approach on 3 different data sets - INRIA, Caltech and Daimler - which contain humans in different contextual domains. Results are evaluated for INRIA and Caltech using the new evaluation framework of per-image methodology [8] and for Daimler using the standard per-window methodology. Results show that the proposed approach outperforms the compared holistic human detection methods across all 3 data sets.




(a) HG (b) HOG

Fig. 1. Problem of HG and its solution by HOG.

## II. Extended Histogram of Gradients

### A. Limitations of Histogram of Gradients and Histogram of Oriented Gradients

Given an image window, first order gradients are computed. The image window is divided into grids of cells of $\varepsilon \times \varepsilon$ pixels. $\xi \times \xi$ cells are grouped into a block. Histogram of Gradients (HG) is computed for each cell. The gradient magnitudes of the pixels with the corresponding directions are voted into the $L$ histogram bins of each cell. The $\xi \times \xi$ histograms are normalized using clipped L2-norm. The histograms are then concatenated to form a $(\xi \times \xi \times L)$-D feature vector for the block. All the overlapping block features are collected to form a combined feature vector for the window.

HG differentiates situations where a bright object is against a dark background and vice versa as it considers gradient directions from 0° to 360°. This makes the intra-class variation of the humans *larger*. In Fig. 1, the situations of a dark object against a bright background and vice-versa in the two different cells are illustrated. As it can be observed, the HG features for the 2 situations are different.

To solve the problem of HG, Histogram of Oriented Gradients (HOG) [3] was proposed that treats all opposite directions as same orientation. This is illustrated in Fig. 1 where the HOG representations for both the situations are the same. However, this causes HOG to be unable to discern some local structures that are different from each other. It is possible for 2 different structures to have the similar feature representation. This is illustrated in Fig. 2. The problem with HOG is that gradients of opposite directions *in the same cell* are mapped to the same bin. In Fig. 2(a), the first pair of structures represent a slightly bent human torso against a background (edge) and human limbs against a background (ridge). HOG produces the same feature for these very different 2 structures. Similarly, in Fig. 2(b) and Fig. 2(c), it can be seen that, for each pair of structures, they are represented as the same by HOG.

### B. The Proposed Extended Histogram of Gradients

Consider an unnormalized HG cell feature, $b_k$ where $k$ is the $k^{th}$ cell in the block. Let $i$ denote the bin of quantized gradient direction $\theta$, $h_{g_k}(i)$ the HG bin value and $L$ the even number of HG bins. We find that HOG can be created simply from HG as follows:

$$h_{og_k}(i) = h_{g_k}(i) + h_{g_k}(i + \tfrac{L}{2}), \quad 1 \leq i \leq \tfrac{L}{2} \qquad (1)$$

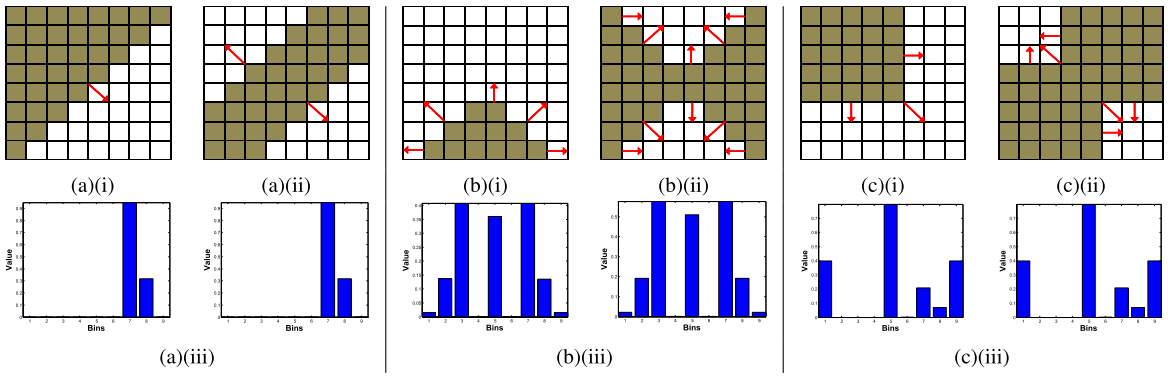



Fig. 2. Problem of HOG. 6 local structures are shown. HOG makes the different structures in (a) (i) and (ii) similar as shown in (a) (iii) and different structures in (b) (i) and (ii) similar as shown in (b) (iii). A similar situation can be observed in (c).

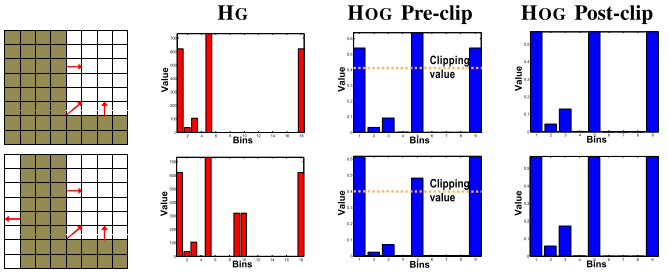



Fig. 3. Example of 2 structures that have similar HOG representations after clipping and normalization.

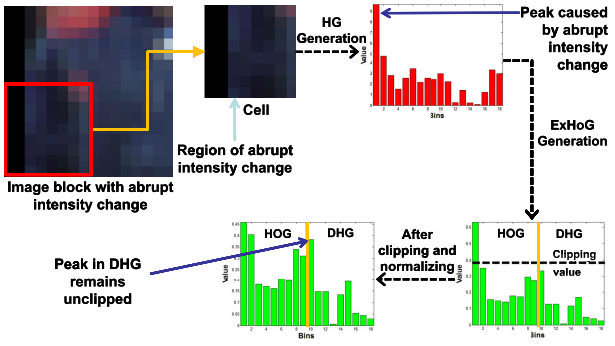



Fig. 4. Problem of normalization directly on ExHoG [Best viewed in colour]. HG is created with a large peak due to the abrupt intensity change. One of the peaks remain unclipped in the formation of ExHoG.

where $h_{og_k}(i)$ is the $i^{th}$ HOG bin value. We see that HOG, in fact, is the sum of two corresponding bins of HG.

Now, consider the absolute difference between $h_{g_k}(i)$ and $h_{g_k}(i + L/2)$ to form a Difference of HG (DHG) as follows:

$$h_{dg_k}(i) = |h_{g_k}(i) - h_{g_k}(i + \tfrac{L}{2})|, \quad 1 \leq i \leq \tfrac{L}{2} \tag{2}$$

where $h_{dg_k}(i)$ is the $i^{th}$ DHG bin value. DHG produces the same feature as HOG for patterns that contain no gradients of opposite directions. It differentiates these patterns from the ones that contain opposite gradients by assigning small values to the mapped bins. The concatenation of these 2 histograms produces the *Extended Histogram of Gradients* (ExHoG).

In [3], [8], [10], [25], HOG is clipped and renormalized after creation. This reduces the illumination variations and noise. However, it presents a problem for some structures in different cells as their features may become similar. An example is illustrated in Fig. 3. The HOG features before clipping are different for the 2 different structures. After clipping, they become the same.

Consider the same normalization procedure in [3], [8], [10], [25] for ExHoG. The magnitudes of the bins of HOG are much larger compared to DHG since creation of HOG involves summation of two positive values while creation of DHG involves an absolute difference of them. Hence, if there are noisy gradient pixels with large magnitudes or very abrupt intensity changes in the image, these large gradient magnitude peaks, which are captured in HG (Fig. 4), are propagated into HOG and DHG. These peaks are larger in HOG than in DHG.

If we perform normalization of the ExHoG feature as illustrated in Fig. 4 similar to [3], these large gradient magnitude peaks are only clipped in the HOG component of ExHoG and remain unclipped in the DHG component of ExHoG.

In our work, we propose that the normalization be performed directly after the HG block feature is created and before the summation and subtraction of HG. The normalization steps are described as follows:

$$h_{gn_k}(i) = \frac{h_{g_k}(i)}{\sqrt{\sum_{k=1}^{N}\sum_{i=1}^{L}\left(h_{g_k}(i)\right)^2}} \tag{3}$$

$$h_{gc_k}(i) = \begin{cases} h_{gn_k}(i), & h_{gn_k}(i) < T \\ T, & h_{gn_k}(i) \geq T, \end{cases} \tag{4}$$

$$h_{gcn_k}(i) = \frac{h_{gc_k}(i)}{\sqrt{\sum_{k=1}^{N}\sum_{i=1}^{L}\left(h_{gc_k}(i)\right)^2}} \tag{5}$$

where $N$ is the number of cells in the block and $T$ is the clipping threshold. The HOG and DHG features are then generated from this normalized HG feature, $h_{gcn_k}(i)$.

Fig. 5 shows the effect of this proposed normalization scheme on the structures in Fig. 3. It is seen that the resulting HOG features for the 2 structures remain different. In [3], the bins of HG are first merged to form HOG and then clipped. In the proposed normalization scheme, the bins of HG are first clipped and then merged to form HOG. This allows the differentiation of some structures to remain after clipping and normalization. Furthermore, it also clips the large gradient

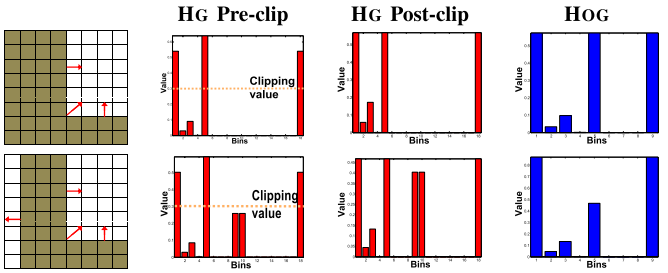



Fig. 5. Effect of the proposed normalization scheme on the 2 structures in Fig. 3. The resulting HOG features for the 2 structures are different.

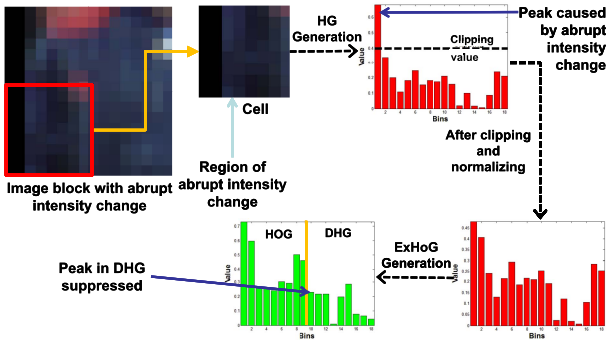



Fig. 6. Proposed normalization scheme for ExHoG [Best viewed in colour]. The HG feature is clipped as shown. It can be observed that the large peak has been clipped. The peak in the histogram of DHG has been suppressed.

peaks before they can be propagated into the HOG and DHG features. This allows ExHoG to be more robust to noise and abrupt image intensity changes (Fig. 6).

The proposed ExHoG of a cell is constructed from the clipped L2-norm normalized HG by:

$$h_{eg_k}(i) = \begin{cases} h_{gcn_k}(i) + h_{gcn_k}(i + \frac{L}{2}), & 1 \le i \le \frac{L}{2} \\ |h_{gcn_k}(i) - h_{gcn_k}(i - \frac{L}{2})|, & \frac{L}{2} + 1 \le i \le L, \end{cases} \quad (6)$$

where $h_{eg_k}(i)$ is the $i^{th}$ ExHoG bin value.

Unlike HOG, the proposed ExHoG differentiates gradients of opposite directions from those of same direction in the same cell. Using the same local structures in Fig. 2, it is clearly illustrated in Fig. 7 that the ExHoG representations of each local structure is unique. Furthermore, ExHoG also resolves the larger intra-class variation of humans caused by the brightness reversal of human and background in Fig. 8. Hence, ExHoG represents the human contour more discriminatively than HOG and has less intra-class variation than HG.

## III. Quadratic Classification in Subspace

### A. Linear Versus Non-linear Classification

Support Vector Machine (SVM) classifiers are most widely used for human detection with HOG [3], [8]–[10], [22], [25], [29], [30], [36], [37], [48], [50], [52], [59]. Non-linear kernels can be used with SVM classifiers for classification. However, the degree of kernel non-linearity is unknown and not easy to be controlled. Using an inappropriate degree of kernel non-linearity could lead to an overfitting problem with SVM classifiers.

Furthermore, the computational complexity of non-linear SVM classifiers during classification heavily depends on the

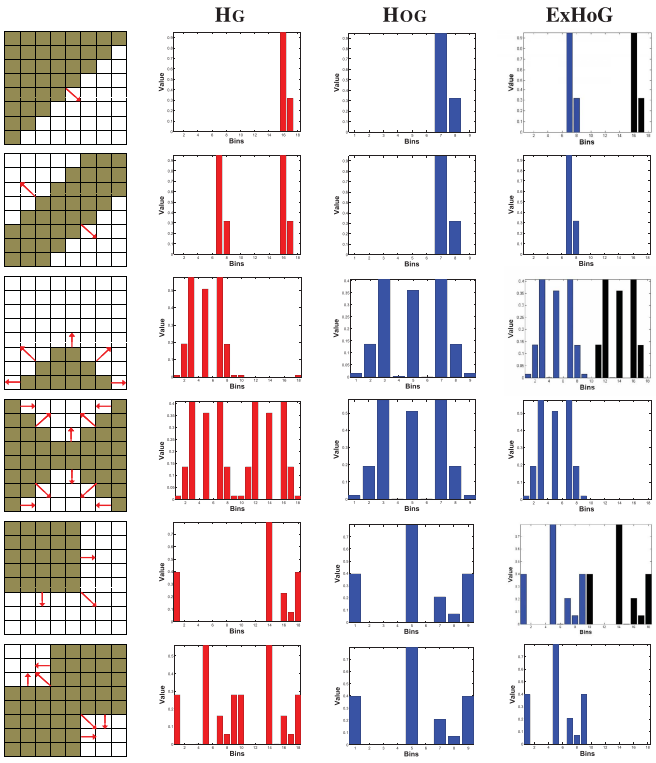



Fig. 7. ExHoG representations of local structures in Fig. 2. ExHoG differentiates the local structure pairs in Fig. 2 misrepresented by HOG.

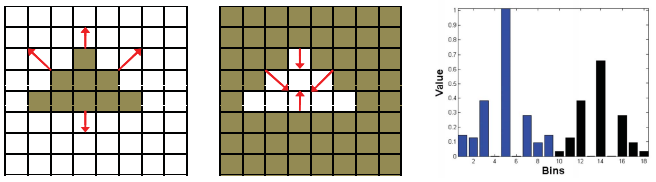



Fig. 8. Same ExHoGs are produced for patterns in Fig. 1.

number of support vectors, the dimensionality of the features and the kernel used. Let $N_s$ the number of support vectors and $l$ the dimension of the samples. The complexity is $O(O(l)N_s)$ where $O(l)$ is the number of operations required to evaluate the non-linear kernel. For a human detection problem, $l$ and $N_s$ can be in values of several thousands. Hence, the complexity is extremely high.

In order to mitigate the overfitting problem and for speed, linear SVM classifier is most widely used [3], [8]–[10], [22], [25], [35]–[37], [48], [50], [52], [59]. However, linear SVM classifiers only work well for *distinct* classification problems like cats versus dogs. For asymmetrical classification problems like human detection where it is one object versus all *other* objects, linear SVM may not perform well.

Boosting-based classifiers are also employed in human detection problems. Compared to SVM classifiers, cascade structured boosting-based classifiers enable very fast detection and the resultant strong classifiers are non-linear. There are many types of boosting-based classifiers used in human detection such as AdaBoost [5]–[7], [16], [32], [38], [46], [60], RealBoost [11], [13], [54], LogitBoost [14], [27], [45], MILBoost [4], [24], [47], Cluster Boosting Tree [53], [55], GentleBoost [26], [44], [58] and MPLBoost [49], [51].

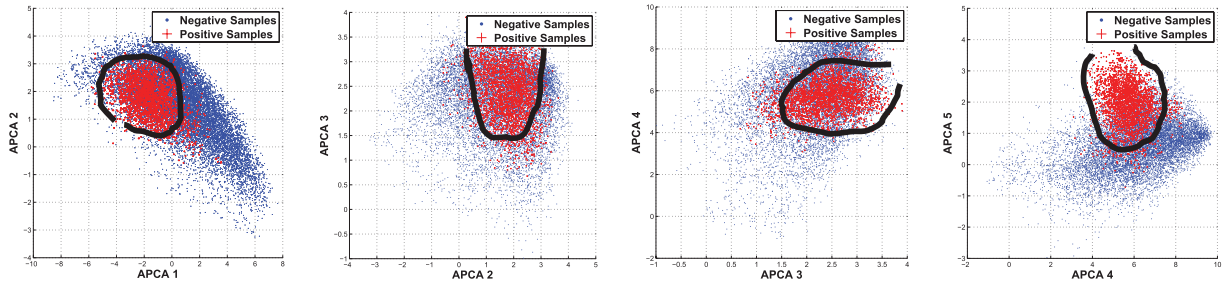



Fig. 9. Distribution of ExHoG features in the APCA subspace [Best viewed in colour]. The first 5 dimensions (APCA 1, APCA 2, APCA 3, APCA 4 and APCA 5) after projection are chosen for the scatter plots. (a) Plot of APCA 1 vs APCA 2. (b) Plot of APCA 2 vs APCA 3. (c) Plot of APCA 3 vs APCA 4. (d) Plot of APCA 4 vs APCA 5. In (a) – (d), it can be clearly seen the positive samples occupy a small space while being surrounded by the negative samples. The thick black curve indicates the ideal quadratic boundary for separation.

However, boosting-based classifiers are based on a subset of features selected from a huge number of all possible features (usually numbering in hundreds of thousands or higher) by learning from the database. The feature subset selected by the classifiers may or may not yield a good classification result [20], [43], especially if the feature pool for selection is small. This work focuses on the further development of a single type of feature based on the widely used HG. It is not feasible to use a boosting-based classifier to select a feature subset from this limited number of specific features as the merit and the main strength of the boosting-based approach is to select a subset of features from a huge number of possible features.

In human detection, the negative class comprises of all other classes that are not human. In the feature space, the positive class usually occupies a small space surrounded by the negative class. In order to illustrate this, the ExHoG features of the initial training set of 2416 positive samples and 12180 negative samples from INRIA [3] are projected to a lower-dimensional Asymmetric Principal Component Analysis (APCA) (Section III-B) subspace. Fig. 9 shows the 2-D scatter plots of the first 5 dimensions of the projected ExHoG features in the APCA subspace. Clearly, from the 4 scatter plots in Fig. 9, it is observed that a linear boundary is not optimal for separating the two classes. A hyper-quadratic boundary can be used to separate the two classes much better than a linear boundary. It is not difficult to understand the roughly quadratic surface boundary between the human class and the non-human class. Human class is one object while non-human class include all other objects. Therefore, the human samples are surrounded by the non-human samples in the feature space.

An example of a quadratic classifier is the Minimum Mahalanobis Distance classifier (MMDC) whose decision rule is given as follows:

$$(X-\mu_n)^T\boldsymbol{\Sigma}_n^{-1}(X-\mu_n)-(X-\mu_p)^T\boldsymbol{\Sigma}_p^{-1}(X-\mu_p) > b, \quad (7)$$

where $X$ is the feature vector, $\boldsymbol{\Sigma}_n$ is the negative covariance matrix, $\boldsymbol{\Sigma}_p$ is the positive covariance matrix, $\mu_n$ and $\mu_p$ are the negative and positive class means and $b$ is a user-defined classification threshold. The MMDC is the minimum error Bayes classifier for Gaussian distribution of the positive and negative data with arbitrary means and covariance matrices.

In general, the class means and covariance matrices of the human and non-human data population are unknown. It can only be estimated from the limited number of training samples. Hence, if the estimated variances deviate greatly from those of the data population, the MMDC will overfit the data as it uses the inverse of the covariance matrices. This will result in a poor generalization. This problem becomes very severe if the feature dimensionality is high [17], [18]. Therefore, in order to improve the classification performance, there is a need to reduce the feature dimensions so that the unreliable dimensions are removed before classification.

### *B. Dimensionality Reduction*

Most dimensionality reduction approaches apply discriminant analysis (DA) [2], [15], [19], [31], [56], [57] to extract as few features as possible with minimum loss of discriminative information. They produce a portable feature vector for fast classification but are not directly targeted at solving generalization problem of the classifier in the high dimensional feature space. These approaches may improve the classification generalization if the highly nonlinear nearest neighbour (NN) classifier is applied. For human detection, there are only two classes and many training samples in each class. Obviously, the NN classifier is not feasible. As analyzed before, a quadratic classifier is devised in our system. The classification requires the inverse of the class-conditional covariance matrices. In the high dimensional space, the inverse of the class-conditional covariance matrices will cause a big classification problem as the unreliable small eigenvalues of the estimated class-conditional covariance matrices can largely deviate from the true values. Asymmetric Principal Component Analysis (APCA) [17], [18] is a dimensionality reduction technique that directly targets at solving such problem.

Suppose there are $q_p$ $l$-dimensional samples belonging to the positive class $\omega_p$ and $q_n$ samples belonging to the negative class $\omega_n$. It is studied in [17], [18] how Principal Component Analysis (PCA) can be used to enhance classification accuracy. The total scatter matrix, $\boldsymbol{\Sigma}_t$, for the 2 classes is in fact a weighted linear combination of covariance matrices. If $\boldsymbol{\Sigma}_t$ is decomposed such that $\boldsymbol{\Sigma}_t = \boldsymbol{\Psi}\boldsymbol{\Upsilon}\boldsymbol{\Psi}^T$ where $\boldsymbol{\Psi}$ is the eigenvector matrix and $\boldsymbol{\Upsilon}$ is the diagonal matrix containing the eigenvalues, the transformation matrix, $\boldsymbol{\Psi}_m$, $\boldsymbol{\Psi}_m \in \mathbb{R}^{l\times m}$, $m < l$, of PCA keeps $m$ eigenvectors corresponding to the $m$ largest eigenvalues. Hence, PCA removes the unreliable dimensions of small eigenvalues [17], [18].

However, PCA does not remove the unreliable dimensions for classification. As $\mathbf{\Sigma}_t$ is not constructed from the classifier point of view, PCA removes unreliable dimensions from the class more well-represented by the training samples. However, unreliable dimensions from the class *less* well-represented by the training samples should be removed. APCA solves this issue by weighting $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_n$ differently. APCA proposes to define a covariance mixture to replace $\mathbf{\Sigma}_t$ as follows:

$$\mathbf{\Sigma}_{t'} = \delta_p \mathbf{\Sigma}_p + \delta_n \mathbf{\Sigma}_n + \mathbf{\Sigma}_m \tag{8}$$

where $\delta_p + \delta_n = 1$, $\delta_p, \delta_n$ are the empirically estimated user-defined weights and $\mathbf{\Sigma}_m$ is the covariance matrix of the class means. Typically, the less well-represented covariance matrix should have a larger weight so that unreliable dimensions from this class will be removed. This also addresses the asymmetry in the training data. In human detection, usually, the number of negative training samples far exceeds the number of positive training samples. Hence, a weight proportional to the number of negative training samples is assigned to the positive covariance matrix and vice-versa for the negative covariance matrix. The weights can then be fine-tuned using cross-validation. However, in the experiments of this paper, the weights are not fine-tuned so that a unified parameter setting is used for all data sets. $\delta_p$ is simply chosen to be proportional to the number of negative training samples and $\delta_n$ to be proportional to the number of positive training samples as follows:

$$\mathbf{\Sigma}_{t'} = \frac{1}{q_p + q_n}(q_n \mathbf{\Sigma}_p + q_p \mathbf{\Sigma}_n) + \mathbf{\Sigma}_m \tag{9}$$

Eigen-decomposition is performed on $\mathbf{\Sigma}_{t'}$ as:

$$\mathbf{\Sigma}_{t'} = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}^T \tag{10}$$

and $m$ eigenvectors $\hat{\mathbf{\Phi}}$ are extracted from $\mathbf{\Phi}$ corresponding to $m$ largest eigenvalues in $\mathbf{\Lambda}$. The projected covariance matrices are found as $\hat{\mathbf{\Sigma}}_p = \hat{\mathbf{\Phi}}^T \mathbf{\Sigma}_p \hat{\mathbf{\Phi}}$ and $\hat{\mathbf{\Sigma}}_n = \hat{\mathbf{\Phi}}^T \mathbf{\Sigma}_n \hat{\mathbf{\Phi}}$.

APCA removes the subspace spanned by the eigenvectors corresponding to the smallest eigenvalues of $\mathbf{\Sigma}_{t'}$. By doing so, APCA removes the unreliable dimensions of both classes (more from the less reliable class) and keeps the large inter-class distinction in the subspace spanned by the eigenvectors of the large eigenvalues of $\mathbf{\Sigma}_{t'}$. As such, APCA alleviates the overfitting problem which lead to better generalization for the unknown query data [17], [18].

### C. Quadratic Classification in APCA Subspace

After APCA, the eigenvalues in the APCA subspace are generally biased upwards. The bias is higher for the less well-represented class [17]. Hence, regularization of the covariance matrices are required for better classification. Classification is performed using a modified MMDC [17] which uses the regularized covariance matrices in the APCA space as follows:

$$(\hat{X} - \hat{\mu}_n)^T \hat{\mathbf{\Sigma}}_{n'}^{-1} (\hat{X} - \hat{\mu}_n) - (\hat{X} - \hat{\mu}_p)^T \hat{\mathbf{\Sigma}}_p^{-1} (\hat{X} - \hat{\mu}_p) > b, \tag{11}$$

where $\hat{X} = \hat{\mathbf{\Phi}}^T X$, $\hat{\mu}_n = \hat{\mathbf{\Phi}}^T \mu_n$, $\hat{\mu}_p = \hat{\mathbf{\Phi}}^T \mu_p$ and $\hat{\mathbf{\Sigma}}_{n'} = \beta \hat{\mathbf{\Sigma}}_n$. $\beta$, $0.5 \leq \beta \leq 2$, is a regularization parameter for the negative matrix. Compared to [17], the upper bound for $\beta$ is increased. For human detection training sets, the number of positive samples is usually far smaller than the number of negatives samples. Therefore, the positive covariance matrix is less reliable than the negative covariance matrix. After APCA, the large eigenvalues of the positive covariance matrix are hence typically biased upwards more than the eigenvalues of the negative covariance matrix. As we need to suppress the positive covariance matrix, the weight of the negative covariance matrix, $\beta$, needs to be larger than 1.

## IV. Experimental Results and Discussions

We perform experiments on three data sets - INRIA [3], Caltech Pedestrian Data Set [8] and Daimler Pedestrian Classification Benchmark [34]. The performance of the proposed approach is compared against some of the state-of-the-art methods on these given data sets. Results are reported for the INRIA and Caltech data sets using the per-image methodology as it is a better evaluation method [8]. The *per-window* evaluation methodology is used for the Daimler data set as the uncropped positive images for testing are unavailable. 2 classifiers are used - linear SVM on ExHoG and modified MMDC on the APCA projected ExHoG. They will be referred to as ExHoG and ExHoG APCA respectively in the discussions.

For the INRIA and Caltech data sets, a cell size of $8 \times 8$ pixels is used with a block size of $2 \times 2$ cells. The number of bins for each cell for ExHoG is 18. A 50% overlap of blocks are used in the construction of the feature vectors. The HG block feature is normalized using a clipping value of 0.08. For the Daimler data set, a cell size of $3 \times 3$ pixels is used with a block size of $2 \times 2$ cells. The number of bins for each cell for ExHoG is 18. A 66.6% overlap of blocks are used in the computation of the feature vectors. The normalization procedure is the same as INRIA and Caltech.

### A. Training of Classifiers

For INRIA and Caltech data sets, the INRIA training set is used to train the classifiers. The training set contains 2416 cropped positive images and 1218 uncropped negative images. The sliding window size is $128 \times 64$ pixels. For Daimler data set, the training data set contains 3 sets of cropped positive and negative samples. In each set, there are 4800 positive samples and 5000 negative samples of $36 \times 18$ pixels.

*1) INRIA and Caltech Data Sets*: We randomly take 10 samples from each negative image to obtain a total of 12180 negative samples for training the linear SVM classifier. Bootstrapping is performed across multiple scales at a scale step of 1.05 to obtain 89400 hard negatives which are combined with the original training set to retrain the classifier.

To estimate $\beta$ for the modified MMDC and $m$ for APCA, a 4-fold cross-validation is performed on the training set. The negative images are scanned across different scales at a scale step of 1.2. At each scale, 7 samples are randomly selected from each negative image. 47499 negative samples are obtained.

$\beta = 1.7$ and $m = 200$ give the best results. $\beta$ is larger than 1 which indicates that eigenvalues of the positive covariance matrix are biased upwards more than the eigenvalues of the negative covariance matrix in the APCA subspace. This result

TABLE I
PERFORMANCE OF EXHOG AGAINST HOG AND/OR HG ON INRIA AND DAIMLER

| Feature + Classifier | INRIA | | Daimler | |
|---|---|---|---|---|
| | LAMR (%) | 0.1 FPPI MR (%) | Accuracy (%) | 0.05 FPPW MR (%) |
| **ExHoG** | **37.00** | **36.39** | **91.22 ± 2.17** | **14.91** |
| HG | 48.00 | 47.94 | 90.28 ± 2.00 | 15.50 |
| HOG | 46.00 | 49.65 | 90.75 ± 2.24 | 16.08 |
| **ExHoG APCA** | **36.00** | **33.36** | **93.06 ± 1.78** | **8.78** |
| HG APCA | 43.00 | 41.11 | 91.84 ± 1.73 | 9.55 |
| HOG APCA | 40.00 | 37.69 | 91.39 ± 2.29 | 11.20 |
| **ExHoG+HIKSVM** | **36.00** | **34.16** | **91.22 ± 1.85** | **14.04** |
| HIKSVM (HOG) | 43.00 | 44.24 | 89.03 ± 1.39 | 18.00 |

TABLE II
PERFORMANCE OF APCA AGAINST MFA WITH MMDC ON INRIA

| Dimension Reduction Method | LAMR (%) | 0.1 FPPI MR (%) |
|---|---|---|
| **APCA** | **36.00** | **33.36** |
| MFA [56] | 50.00 | 50.42 |

verifies our analysis earlier whereby it was discussed that the positive covariance matrix will be less reliable due to the smaller number of training samples available. Hence, its large eigenvalues will be heavily biased upwards.

Using these parameters, the modified MMDC is trained. Bootstrapping is performed on the negative images across multiple scales at a scale step of 1.05 to obtain 89400 hard negatives. They are combined with the original training set to retrain the modified MMDC. $\beta$ and $m$ remain unchanged during classifier retraining to simplify the training process.

*2) Daimler Data Set*: Three linear SVM classifiers are trained by choosing 2 out of 3 training sets at a time. To estimate $\beta$ for the modified MMDCs and $m$ for APCA, 3-fold cross-validation is performed on the training set. From the cross-validation experiments, $\beta = 0.9$ and $m = 400$ give the best results. $\beta$ is close to 1 which indicates that the eigenvalues of both classes have almost the same amount of bias. For this data set, the number of training samples for each class is almost the same. However, the positive class only consist of humans while the negative class consist of all other non-humans. Thus, even though both classes have almost the same number of training samples, the negative class is actually less well-represented than the positive class. This leads to the optimal value of $\beta$ smaller than 1. Using these parameters, the 3 modified MMDCs are trained.

### *B. Comparison of Classifiers and ExHoG Against* HG *and* HOG

The performance of ExHoG against HOG and/or HG with linear SVM, HIKSVM [29], [30] and APCA+MMDC is tested on INRIA and Daimler. After cross-validation on INRIA, $\beta = 1.4$ and $m = 200$ give the best results for HG. For HOG, $\beta = 1.25$ and $m = 200$ give the best results. After cross-validation on Daimler, $m = 400$ and $\beta = 0.9$ gave best results for both.

The INRIA test set contains 288 images. We scan the images using the classifiers over multiple scales at a scale step of 1.05. The window stride is 8 pixels in the $x$ and $y$ directions. To compare between different detectors, the miss rate (MR) against false positives per image (FPPI) (using log-log plots) is plotted. To summarize the detector performance, the *log-average miss rate* [8] (LAMR) is used which is computed by averaging the MRs at nine FPPI rates evenly spaced in log-space in the range $10^{-2}$ to $10^{0}$. If any of the curves end before reaching $10^{0}$, the minimum miss rate achieved is used [8].

The Daimler test set contains 2 sets of positive and negative samples. In each set, there are 4800 positive images and 5000 negative images of $36 \times 18$ pixels. Following the evaluation process in [34], we run the classifiers on both test sets. 6 ROC curves are obtained. Under the assumption that each test follows a Gaussian distribution and is independent, a 95% confidence interval of the true mean MR, which is given by the $t$-distribution, is taken [34].

The results are shown in Table I. Rows 1 to 3 show the results of ExHoG against HOG and HG using linear SVM classifiers. Rows 4 to 6 show those using APCA+MMDC. Rows 7 to 8 show those using HIKSVM. It can be seen that ExHoG consistently outperforms HOG and/or HG for a particular classifier for both data sets in each section. This demonstrates that the proposed feature is better suited for human description compared to HG and HOG. Comparing between sections, the effectiveness of APCA+MMDC can be compared against linear SVM and HIKSVM. On both INRIA and Daimler, APCA+MMDC outperforms linear SVM for all 3 features. It also outperforms the nonlinear HIKSVM.

In addition, we have also compared the performance of APCA against a state-of-the-art dimension reduction approach, Marginal Fisher Analysis (MFA) [56] on INRIA. ExHoG is used as the feature. Modified MMDC is used for classification as it is not possible to apply the NN classifier for this huge number of training samples. We use cross-validation experiments to determine the best parameters for MFA, its pre-PCA and the modified MMDC. They are 200 for PCA, $k_1 = 10$ and $k_2 = 180$ for MFA and $\beta = 0.8$ for MMDC. Table II shows that APCA performs significantly better than MFA.

### *C. Comparison With State-of-the-Art on INRIA*

We compare the performance of ExHoG and ExHoG APCA with VJ [46], SHAPELET [38], POSEINV [22], LATSVM-V1 [9], HIKSVM, HOG [3] and LATSVM-V2 [10]. In order to keep the comparisons clearly within the domain of single type of features, performance comparisons with methods that use hybrid features are omitted. The results of all compared detectors are given in [8]. These detectors are trained and optimized by their respective authors and tested in [8].

From Fig. 10, ExHoG and ExHoG APCA achieve a LAMR of 37% and 36% respectively which are significantly lower than all the state-of-the-art methods being compared with except LATSVM-V2. HOG has a LAMR of 46% while HIKSVM and LATSVM-V1 have a LAMR of 43% and 44% respectively. ExHoG and ExHoG APCA do not perform

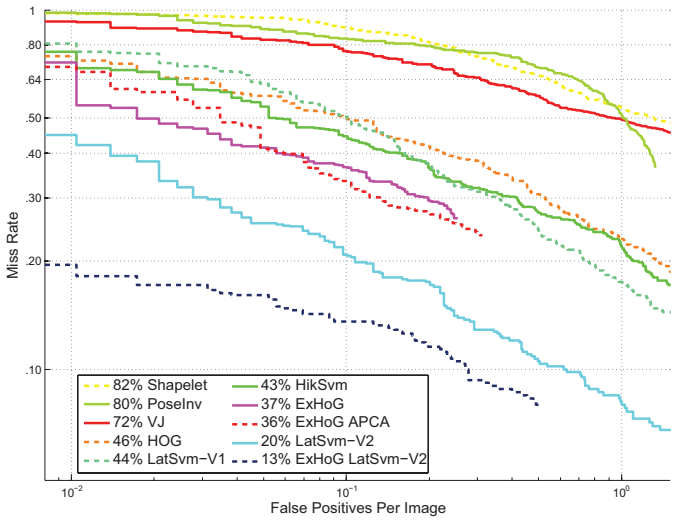



Fig. 10. Performance of ExHoG and ExHoG APCA against existing state-of-the-art methods [Best viewed in colour]. ExHoG APCA outperforms almost all other methods except for LATSVM-V2.

as well as LATSVM-V2. In LATSVM-V2, the latent SVM classifier has been modified to a better objective function compared to LATSVM-V1 and use mixture models which enables a more robust parts-based representation. Since INRIA consist of very clean high-resolution images (minimum size of human for detection is 128 × 64 pixels), these improvements significantly improves performance. Since LATSVM-V2 uses HOG features, we also tested ExHoG with LATSVM-V2 by replacing HOG with ExHoG. Using ExHoG with LATSVM-V2, we achieve the lowest LAMR of 13% on INRIA to date. However, the parts-based approach may not work well for low-resolution or low-quality or complex-background images. This is shown in the medium and reasonable scales of Caltech data set (Fig. 11(c) and (f)) in Section IV-D where LATSVM-V2 underperforms the proposed ExHoG and ExHoG APCA.

### *D. Comparison With State-of-the-Art on Caltech*

The Caltech data set [8] contains color video sequences and pedestrians with a wide range of scales and more scene variations. It has been created from a recorded video on a moving car through densely populated human areas. As such, it contains artifacts of motion, blur and noise, and has various stages of occlusion (from almost complete to none). The data set is divided into 11 sessions. The first 6 sessions are the training set while the remaining 5 are the test set.

In [8], the authors reported results whereby they used detectors trained on other data sets like INRIA for classification on their test set. We also present our results in a similar manner where our detectors are trained using the INRIA data set and tested on the test sessions. The scale step used is 1.05. The window stride is 8 pixels in the $x$ and $y$ directions. Same as [8], in order to detect humans at smaller scales, the original images are upscaled. Only every $30^{th}$ frame is evaluated so that our comparisons is consistent with those in [8].

Detailed results are presented in Fig. 11. The detectors we compare with ExHoG and ExHoG APCA are the same as those in Section IV-C (except ExHoG LATSVM-V2). The results of the compared detectors are given in [8]. These detectors are trained and optimized by their respective authors and tested in [8]. The performance is analyzed under six conditions as in [8]. Fig. 11 show the overall performance on the test set, on near and medium scales, under no and partial occlusions and on clearly visible pedestrians (reasonable). As in [8], the MR versus FPPI is plotted and LAMR is used as a common reference value for summarizing performance. The results are discussed under each condition in more details as follows.

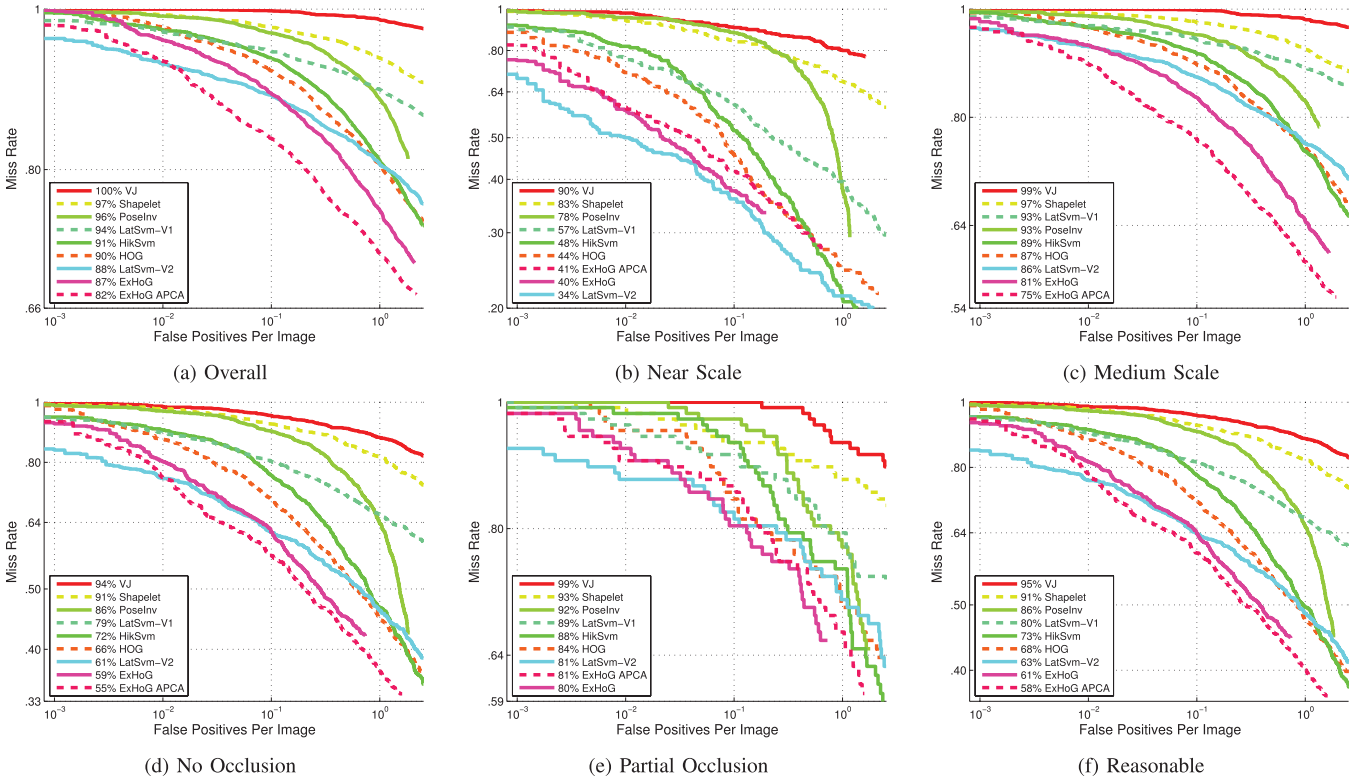
**Overall:** Fig. 11(a) plots the performance on all test sessions for *every* annotated pedestrian. ExHoG APCA ranks first at 82% followed by ExHoG at 87% and LATSVM-V2 at 88%.

**Scale:** Fig. 11(b) plots the performance on unoccluded pedestrians of heights over 80 pixels. Here, ExHoG APCA does not perform as well as LATSVM-V2 and is marginally worse than ExHoG. ExHoG APCA has a LAMR of 41% with ExHoG at 40% and LATSVM-V2, performing the best, at 34%. At this scale, the resolution of pedestrians is high for the parts-based LATSVM-V2 to perform better than ExHoG and ExHoG APCA. However, good performance at this scale is ***not*** crucial for pedestrian detection applications [8]. The pedestrians will be too close to the vehicle and the driver or the automated driving system will not have sufficient time to react to avoid accidents. It is more important to detect pedestrians at a distance much further away from the vehicles.

Fig. 11(c) plots the performance on unoccluded pedestrians of heights between 30 - 80 pixels. ExHoG APCA ranks first at 75% followed by ExHoG at 81% and LATSVM-V2 at 86%. At this scale, ExHoG APCA outperforms LATSVM-V2 by a large margin. This highlights that at low- and medium-resolutions, using quadratic classifier in the APCA subspace is more robust in detecting pedestrians than other approaches. This is an important aspect as in pedestrian detection problems, it is necessary to detect humans further away from the vehicle more accurately so that there is ample time to react to prevent an accident. 30-80 pixel height of pedestrians is the most appropriate image resolution for pedestrian detection [8].

**Occlusion:** Fig. 11(d) plots the performance on unoccluded pedestrians of heights over 50 pixels. ExHoG APCA ranks first at 55% and LATSVM-V2 ranks third at 61%. Fig. 11(e) plots the performance on partially occluded (1 - 35% occluded) pedestrians of heights over 50 pixels. ExHoG ranks first at 80% and LATSVM-V2 ranks third at 81%. As occlusion has degraded the performance significantly, the parts-based LATSVM-V2 should outperform the whole human detection methods. However, it does not outperform ExHoG and ExHoG APCA. The reason for the good performance of ExHoG and ExHoG APCA is due to the robustness of our approach to detection of humans in low- and medium-resolutions.

**Reasonable:** Fig. 11(f) plots the performance on reasonable condition that evaluates performance on pedestrians that are over 50 pixels tall under no or partial occlusion. ExHoG APCA ranks first at 58% and LATSVM-V2 ranks third at 63%. ExHoG APCA is able to handle low- and medium-resolutions more robustly compared to the other methods. This accounts for its performance under this condition.




(a) Overall (b) Near Scale (c) Medium Scale

(d) No Occlusion (e) Partial Occlusion (f) Reasonable

Fig. 11. Evaluation results under six different conditions on the test set of the Caltech Pedestrian Data Set [Best viewed in colour]. (a) ExHoG APCA ranks first in performance on all annotated pedestrians. (b) ExHoG ranks second in performance on unoccluded pedestrians over 80 pixels (near scale) followed by ExHoG APCA. (c) ExHoG APCA ranks first in performance on unoccluded pedestrians between 30–80 pixels. (d) ExHoG APCA ranks first in performance on unoccluded pedestrians over 50 pixels tall. (e) Even under partial occlusion, ExHoG performs the best among all other methods. ExHoG APCA ranks second. (f) ExHoG APCA ranks first in performance on 50-pixel or taller, unoccluded or partially occluded pedestrians (reasonable).

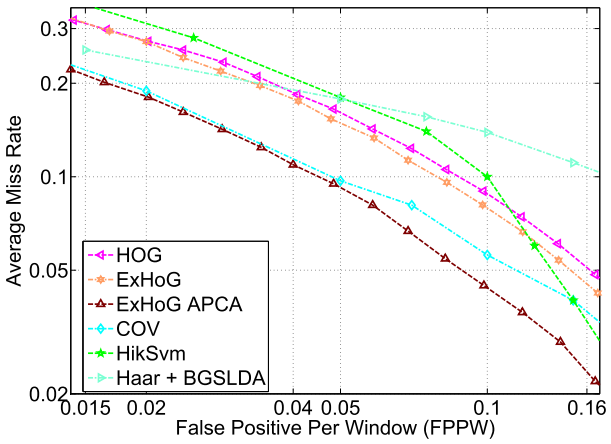



Fig. 12. Performance of ExHoG APCA against existing state-of-the-art methods [Best viewed in colour]. ExHoG APCA outperforms all other methods.

### *E. Comparison With State-of-the-Art on Daimler*

Fig. 12 shows the performance of ExHoG and ExHoG APCA against HOG with linear SVM classifier, Covariance Descriptor (COV) [45], HIKSVM [29], [30] and Haar features with BGSLDA (Haar + BGSLDA) [42]. The results of COV, HIKSVM and Haar + BGSLDA were obtained from their respective papers. ExHoG only achieves a 1% improvement in MR compared to HOG at 0.05 FPPW. ExHoG outperforms HIKSVM. ExHoG also performs much better than Haar + BGSLDA which uses Linear Discriminant Analysis as its criteria for feature selection with boosting. At 0.05 FPPW, ExHoG has a MR of 14.91% while HIKSVM and Haar + BGSLDA have a MR of 18%. However, in comparison to COV, ExHoG performance is inferior. We attribute this less-than-desired performance to the linear classification approach of ExHoG in comparison to the non-linear classification approach of COV. Dimensionality reduction with APCA and classification with modified MMDC improves the MR of ExHoG by about 6% to a MR of 8.78% at 0.05 FPPW compared with ExHoG with Linear SVM classifiers. ExHoG APCA outperforms all other methods.

TABLE III
SUMMARY OF RUNTIME PERFORMANCE OF EXHOG (INRIA)

| Feature | Dimension | Extraction Speed (ms) | Classification Speed (ms) | | |
|---|---|---|---|---|---|
| | | | Linear SVM | HIKSVM | APCA+ MMDC |
| **ExHoG** | 7560 | 0.19 | 0.07 | 46.73 | 6.63 |
| **HOG** | 3780 | 0.13 | 0.04 | 23.37 | 4.58 |
| **HG** | 7560 | 0.19 | 0.07 | 46.73 | 6.63 |

### *F. Runtime Performance*

Table III summarizes the extraction and classification time per window of ExHoG versus HG and HOG for INRIA.

The off-line training time of APCA+MMDC is approximately 12 hours which is slower than linear SVM but much faster than HIKSVM which takes around a day to train. Classification using MMDC is slower than linear SVM but is much faster than HIKSVM which is a *fast* non-linear kernel SVM classifier. For online human detection, both the computational and memory complexity is $O(l \cdot m)$ for the dimensionality reduction and $O(m^2)$ for MMDC. Overall, both the computational and memory complexity of APCA+MMDC is $O(l \cdot m + m^2)$.

## V. CONCLUSION

This paper proposes a quadratic classification approach on the subspace of Extended Histogram of Gradients (ExHoG) for human detection. ExHoG is derived by observing the inherent weaknesses of Histogram of Gradients (HG) and Histogram of Oriented Gradients (HOG). HG differentiates a bright human against a dark background and vice-versa which increases the intra-class variation of humans. HOG maps gradients of opposite directions into the same histogram bin. Hence, it is unable to differentiate some local structures and produces the same feature. ExHoG alleviates these weaknesses by considering both the sum and absolute difference of HG with the opposite gradients.

Furthermore, we propose to exploit a quadratic classifier, a Minimum Mahalanobis Distance classifier (MMDC) which uses the inverse of the covariance matrices estimated from the training samples. When the estimated eigenvalues of some feature dimensions deviate from those of the data population, the classifier overfits the training data. Hence, feature dimensionality reduction is proposed to remove the unreliable dimensions and alleviate the poor classifier generalization. The asymmetry issue in human detection training sets is also considered where there are much fewer images available for human than non-human. This results in a difference in the reliability of the estimated covariance matrices which makes Principal Component Analysis ineffective to remove unreliable dimensions. In order to solve this, we propose using Asymmetric Principal Component Analysis (APCA) which asymmetrically weighs the covariance matrices. Furthermore, a modified MMDC is also employed which regularizes the covariance matrices in the APCA subspace.

We present results of the proposed framework on 3 data sets - INRIA, Caltech and Daimler - and compare them with some of the state-of-the-art human detection approaches. Results demonstrate that the proposed framework outperforms compared state-of-the-art holistic human detection approaches.

## REFERENCES


[1] A. Bar-Hillel, D. Levi, E. Krupka, and C. Goldberg, "Part-based feature synthesis for human detection," in *Proc. 11th Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 127–142.

[2] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 4–13, Jan. 2005.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.

[4] P. Dollar, B. Babenko, S. Belongie, P. Perona, and Z. Tu, "Multiple component learning for object detection," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 211–224.

[5] P. Dollar, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 1–11.

[6] P. Dollar, Z. Tu, H. Tao, and S. Belongie, "Feature mining for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[7] P. Dollar, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–11.

[8] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.

[9] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[10] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[11] W. Gao, H. Ai, and S. Lao, "Adaptive contour features in oriented granular space for human detection and segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1786–1793.

[12] D. Gavrila, "A bayesian, exemplar-based approach to hierarchical shape matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1408–1421, Aug. 2007.

[13] D. Geronimo, A. Sappa, A. Lopez, and D. Ponsa, "Adaptive image sampling and windows classification for on-board pedestrian detection," in *Proc. 5th Int. Conf. Comput. Vis. Syst.*, Mar. 2007, pp. 1–10.

[14] G. Gualdi, A. Prati, and R. Cucchiara, "Multi-stage sampling with boosting cascades for pedestrian detection in images and videos," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 196–209.

[15] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[16] M. Hiromoto and R. Miyamoto, "Cascade classifier using divided cohog features for rapid pedestrian detection," in *Proc. IEEE 7th Int. Conf. Comput. Vis. Syst.*, Oct. 2009, pp. 53–62.

[17] X. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

[18] X. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 16–26, Mar. 2011.

[19] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.

[20] H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou, "An efficient fuzzy classifier with feature selection based on fuzzy entropy," *IEEE Trans. Syst., Man, Cybern. Part B, Cybern.*, vol. 31, no. 3, pp. 426–432, Jun. 2001.

[21] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 259–289, May 2008.

[22] Z. Lin and L. Davis, "A pose-invariant descriptor for human detection and segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 423–436.

[23] Z. Lin, L. Davis, D. Doermann, and D. DeMenthon, "Hierarchical part-template matching for human detection and segmentation," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[24] Z. Lin, G. Hua, and L. Davis, "Multiple instance fFeature for robust part-based object detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 405–412.

[25] Z. Lin and L. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 604–618, Apr. 2010.

[26] X. Liu, T. Yu, T. Sebastian, and P. Tu, "Boosted deformable model for human body alignment," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[27] Y. Liu, S. Shan, W. Zhang, X. Chen, and W. Gao, "Granularity-tunable gradients partition descriptors for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1255–1262.

[28] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[29] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[30] S. Maji, A. C. Berg, and J. Malik, "Efficient classification for additive kernel svms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 66–77, Jan. 2013.

[31] A. Martinez and A. Kak, "Pca versus lda," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[32] K. Mikolajczyk, C. Schmid, and A. Zisserman, "Human detection based on a probabilistic assembly of robust part detectors," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 69–82.
[33] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 349–361, Apr. 2001.
[34] S. Munder and D. Gavrila, "An experimental study on pedestrian classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1863–1868, Nov. 2006.
[35] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *Int. J. Comput. Vis.*, vol. 38, no. 1, pp. 15–33, Jun. 2000.
[36] D. Park, D. Ramanan, and C. Fowlkes, "Multiresolution models for object detection," in *Proc. 11th Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 241–254.
[37] M. Pedersoli, J. Gonzàlez, A. Bagdanov, and J. Villanueva, "Recursive coarse-to-fine localization for fast object detection," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 280–293.
[38] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[39] A. Satpathy, X. Jiang, and H.-L. Eng, "Extended histogram of gradients feature for human detection," in *Proc. IEEE Int. Conf. Image. Process.*, Sep. 2010, pp. 3473–3476.
[40] A. Satpathy, X. Jiang, and H.-L. Eng, "Extended histogram of gradients with asymmetric principal component and discriminant analyses for human detection," in *Proc. IEEE Canad. Conf. Comput. Robot. Vis.*, May 2011, pp. 64–71.
[41] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2009, pp. 24–31.
[42] C. Shen, S. Paisitkriangkrai, and J. Zhang, "Efficiently learning a detection cascade with sparse eigenvectors," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 22–35, Jan. 2011.
[43] R. Thawonmas and S. Abe, "A novel approach to feature selection based on analysis of class regions," *IEEE Trans. Syst., Man, Cybern. Part B, Cybern.*, vol. 27, no. 2, pp. 196–207, Apr. 1997.
[44] A. Torralba, K. Murphy, and W. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jul. 2004, pp. 762–769.
[45] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
[46] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 153–161, 2005.
[47] P. A. Viola, J. C. Platt, and C. Zhang, "Multiple instance boosting for object detection," in *Proc. Adv. Neural Inf. Process. Syst. Conf.*, 2005, pp. 1417–1426.
[48] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1030–1037.
[49] S. Walk, K. Schindler, and B. Schiele, "Disparity statistics for pedestrian detection: Combining appearance, motion and stereo," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 182–195.
[50] T. Wang, X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Jun. 2009, pp. 32–39.
[51] C. Wojek, S. Walk, and B. Schiele, "Multi-cue onboard pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 794–801.
[52] C. Wojek and B. Schiele, "A performance evaluation of single and multi-feature people detection," in *Proc. Pattern Recognit. Symp.*, 2008, pp. 82–91.
[53] B. Wu and R. Nevatia, "Cluster boosted tree classifier for multi-view, multi-pose object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[54] B. Wu and R. Nevatia, "Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[55] B. Wu and R. Nevatia, "Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 185–204, Apr. 2009.
[56] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[57] J. Ye, R. Janardan, C. Park, and H. Park, "An optimization criterion for generalized discriminant analysis on undersampled problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 982–994, Aug. 2004.
[58] J. Zhang, K. Huang, Y. Yu, and T. Tan, "Boosted local structured hog-lbp for object localization," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1393–1400.
[59] W. Zhang, G. Zelinsky, and D. Samaras, "Real-time accurate object detection using multiple resolutions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[60] Q. Zhu, M. Yeh, K. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1491–1498.

**Amit Satpathy** received the B.Eng. degree in Electrical and Electronics Engineering from Nanyang Technological University (NTU), Singapore, in 2007, and is currently pursuing the Ph.D. degree at NTU under the supervision of Associate Professor X. Jiang of NTU and H.-L. Eng with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. He is a recipient of the A*STAR Graduate Scholarship. His current research interests include feature development and extraction for object detection and recognition, image/video processing, pattern recognition, computer vision, and machine learning. Currently, he is with the Institute for Infocomm Research as a Scientist.

**Xudong Jiang** (M'02–SM'06) received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree from Helmut Schmidt University, Hamburg, Germany, in 1997, all in electrical engineering. From 1986 to 1993, he was a Lecturer with UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1993 to 1997, he was a Scientific Assistant with Helmut Schmidt University. From 1998 to 2004, he was with the Institute for Infocomm Research, A*Star, Singapore, as a Lead Scientist and the Head of the Biometrics Laboratory, where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition in 2000. He joined Nanyang Technological University (NTU), Singapore, as a Faculty Member, in 2004, and served as the Director of the Centre for Information Security from 2005 to 2011. Currently, he is a Tenured Associate Professor with the School of EEE, NTU. He has published over 100 papers, where 15 papers in IEEE journals, including TPAMI (4), TIP (4), TSP (2), SPM, TIFS, TCSVT, TCS-II, and SPL. He holds seven patents. His current research interests include signal/image processing, pattern recognition, computer vision, machine learning, and biometrics.

**How-Lung Eng** (M'03) received the B.Eng. and Ph.D. degrees in Electrical and Electronics Engineering from Nanyang Technological University, Singapore, in 1998 and 2002, respectively. Currently, he is with the Institute for Infocomm Research, Singapore, as a Research Scientist and Programme Manager of Video Behavioral Analytics Programme. His current research interests include real-time vision, pattern classification, and machine learning for abnormal event detection. He has made several PCT fillings related to video surveillance applications and has actively published. He was a recipient of the Tan Kah Kee Young Inventors' Award in 2000 (Silver, Open Section) for his Ph.D. study, and a recipient of the TEC Innovator Award in 2002, the IES Prestigious Engineering Awards in 2006 and 2008 and IWA PIA Asia Pacific Regional Award in 2012 for his works in the areas of video surveillance and video monitoring to ensure safe drinking water.