

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/256661993>

Pedestrian Detection in Far Infrared Images

Article *in* Integrated Computer Aided Engineering · September 2013

DOI: 10.3233/ICA-130441

CITATIONS

25

READS

608

5 authors, including:



[Cristiano Premebida](#)

University of Coimbra

29 PUBLICATIONS 505 CITATIONS

[SEE PROFILE](#)



[Urbano Nunes](#)

University of Coimbra

245 PUBLICATIONS 3,270 CITATIONS

[SEE PROFILE](#)



[J.M. Armingol](#)

University Carlos III de Madrid

135 PUBLICATIONS 1,889 CITATIONS

[SEE PROFILE](#)



[Arturo de la Escalera](#)

University Carlos III de Madrid

134 PUBLICATIONS 1,931 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



AutoCITS (PT) [View project](#)



Urban Information Assistant for Safer Driving - URBANITA-Safe [View project](#)

Pedestrian detection in far infrared images

Daniel Olmeda^{a,*}, Cristiano Premebida^b, Urbano Nunes^b, Jose Maria Armingol^a and Arturo de la Escalera^a

^aDepartment of Systems Engineering, Intelligent Systems Lab, Universidad Carlos III de Madrid, Madrid, Spain

^bDepartment of Electrical and Computer Engineering, Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal

Abstract. This paper presents an experimental study on pedestrian classification and detection in far infrared (FIR) images. The study includes an in-depth evaluation of several combinations of features and classifiers, which include features previously used for daylight scenarios, as well as a new descriptor (HOPE – Histograms of Oriented Phase Energy), specifically targeted to infrared images, and a new adaptation of a latent variable SVM approach to FIR images. The presented results are validated on a new classification and detection dataset of FIR images collected in outdoor environments from a moving vehicle. The classification space contains 16152 pedestrians and 65440 background samples evenly selected from several sequences acquired at different temperatures and different illumination conditions. The detection dataset consist on 15224 images with ground truth information. The authors are making this dataset public for benchmarking new detectors in the area of intelligent vehicles and field robotics applications.

Keywords: Pedestrian detection, far infrared, advanced driver assistance systems

1. Introduction

Image analysis and computer vision is lately being incorporated in civil engineering [13,28,47]. The applications are plenty and include freeway work zone analysis [1,27,36,37], automatic image search [34], human detection and modeling [12,69] and face recognition [7]. Object recognition in images has become a very important topic in the fields of traffic infrastructure and driving assistance system [30,54]. Applications such as traffic signs recognition [32,53,56,65], obstacle avoidance [19] and traffic surveillance [59] have gotten the attention of the industry for some time now. The case of people detection is an exceptionally relevant case, as it leads to a number of important applications, some of which strive for saving lives.

Pedestrian recognition in images is geared toward a variety of applications, which include safety focused road infrastructures [35], driver assistance sys-

tems [25] and autonomous robotic vehicles [55]. It is also useful in security, be it for automatic surveillance or people counting [63,69]. Those applications can be extended to low visibility conditions by using FIR images. Another application that could benefit of automatic recognition of people in low light conditions is unmanned aerial vehicles [70].

Most of the recent research in this topic is based on visible light (VL) images. FIR images share some key characteristics with their VL images counterparts. They both are 2D representations of a scene captured by redirecting electromagnetic waves by means of a lens, light in the first case and infrared radiation, which is proportional to the objects temperature, in the second. Some of the key ideas on pedestrian classification in VL images can be extended to work on FIR images, exploiting common characteristics of both, or adapt them to take benefit of the different kinds of information provided by FIR images.

Regarding pedestrian detection in VL images, there exists a reasonable number of benchmark datasets publicly available, such as: MIT [52], CVC [26], TUD-det [8], INRIA [15], DC [46], ETH [21] and Cal-

*Corresponding author: Daniel Olmeda, Department of Systems Engineering, Intelligent Systems Lab, Universidad Carlos III de Madrid, Madrid, Spain. E-mail: dolmeda@ing.uc3m.es.

tech [17]. For an overview of recent work on pedestrian detection on these datasets, the authors refer to [18,20, 25,63]. In the case of FIR images, the authors found a lack of a complete pedestrian dataset that could serve as a tool to benchmark new features and methods.

In this paper the authors present an in-depth analysis of several well-known VL pedestrian classifiers applied to FIR images: Principal Component Analysis (PCA), Local Binary Patterns (LBP) [48] and Histogram of Oriented Gradients (HOG) [15]. Moreover, experimental results are compared with the proposed descriptor specifically targeted to FIR images: Histograms of Oriented Phase Energy (HOPE) [50].

The results derived from this study were obtained from a new pedestrian dataset, that the authors are making public.¹ Our dataset is divided in two parts, classification and detection. The Classification Dataset contains a preset of cropped images of positives (pedestrians) and negatives (background), rescaled to the same dimensions. The Detection Dataset contains full size images and labels indicating the position and dimensions of each pedestrian.

In short, there are three main contributions in this paper: (i) a new pedestrian dataset in far infrared images; (ii) an extensive study on the pedestrian descriptors using FIR images, and (iii) an adaptation of a latent variable SVM approach [23] to FIR images. This paper is structured as follows. Section 2 includes a brief discussion on pedestrian datasets and the relevant state of the art in pedestrian detection. It also covers an overview of descriptors and FIR image-based classification methods. Characteristics of the FIR image-based pedestrian dataset are discussed in Sections 3 and 4, including the methodology of acquisition and sample selection, as well as useful statistics. Sections 5 and 6 focus on the features and methods used for classification and detection performance assessment. In Sections 7 and 8 experiments for pedestrian classification and detection, respectively, are presented and further discussed in Section 9.

2. Related work

2.1. Pedestrian datasets

The availability of publicly released datasets for pedestrian classification has been a key element that

helped advances in this area. It provides a way for researchers to test and benchmark new classification algorithms in a way that can be directly compared with other works. It is also useful for replicating experiments performed by other research groups.

In this domain, datasets are usually divided into two types: classification and detection datasets. In the first one, a fixed set of cropped windows containing pedestrians and background is provided, while detection datasets consist on full images with annotated locations of pedestrians. Usually, a subset of full-frames, with no positives (pedestrians), is provided for negative examples extraction. The method for background sample extraction varies from one author to the other, so the classifiers are not really trained on the same data.

A classification dataset is useful for approaches based on the sliding window paradigm. This detection technique consists on analyzing an image by shifting a fixed sized window in the horizontal and vertical axis. This approach can be extended to a multi resolution search by incrementally resizing the original image. Each window analysis becomes independent from all the others and, as such, the detection turns into a classification problem. Improving the classifier performance would also improve detection performance. The classification performance is usually expressed in terms of miss rate vs. false negative rate per window, while per frame is more suitable for detection performance.

In [46] Munder and Gavrila introduced the DC classification dataset. It consists of 4000 up-right pedestrian and 25000 background samples captured in outdoor urban environments. All of them are resized to 18×36 pixels. In their work, the authors evaluate Haar, Principal Component Analysis (PCA) [29] and Local Receptive Fields (LRF) in combination with neural networks and Support Vector Machine (SVM) classifiers [14,40]. From their results it can be concluded that the size of the dataset is a key element in improving the classification performance. For the extraction of a large number of background images they apply bootstrapping [64] techniques. The dataset is split into 3 train and 2 test subsets, for cross-validation purposes.

In [15] Dalal et al. presented the INRIA dataset, which is still widely used nowadays. It consists on 2478 128×64 cropped images of people for training, and 566 for testing, along with full images for negative extraction. The images were selected from a collection of photographs acquired in urban and rural scenes, and not initially thought to serve as a dataset for driving assistance systems.

More recently, Dollar et al. introduced in [17] the Caltech Detection Dataset, as well as a benchmark of

¹<http://www.uc3m.es/islab/repository>



Fig. 1. Example cropped images of the classification dataset. The upper row contains examples of pedestrians acquired under different temperatures and illumination conditions. The lower row contains randomly selected windows from images containing no pedestrians. For visualization purposes the contrast has been enhanced.

several pedestrian detection algorithms. Their results were further extended in [18]. This dataset contains approximately 250 k labeled pedestrians within several video sequences acquired from a moving vehicle in urban traffic. In their work, the authors directly compare results with other publicly available pedestrian datasets, as the Daimler detection dataset [20].

2.2. Image based descriptors

In [52] Papageorgiou et al. introduced a sliding window detector based on SVM trained with a Haar wavelet feature vector. This work inspired Viola and Jones face detector, presented in [66], which accelerates the classification step by applying a cascade approach.

Along with the INRIA dataset, Dalal et al. introduced the HOG descriptor [15]. It became a benchmark for pedestrian classification, due to their excellent results. Since then, there have been many new features using local histograms [63], including classifiers targeted to FIR images [50].

Color or gray level value has also been exploited as a descriptor of the shape of a person. In [12] Ciarelli et al. create a live color model with a modified Mean Shift algorithm. More recently, Self Similarity (SS) has been proposed as a descriptor for pedestrian classification in visible light images in [67]. This method encodes the distribution of color as repetition across the image. In [45] this approach was adapted to work in mono channel images, and tested in a FIR pedestrian classification problem.

2.3. FIR image-based methods

Given that FIR images represent the heat emitted by

people, and do not need external illumination, some authors have developed classification methods based on the temperature distribution of the human body. Most systems take advantage of this feature and select regions of interest based on the distribution of the warm parts of the image [9,10]. A review of techniques for pedestrian automotive infrared detection can be found in [51]. In these systems, the discriminating feature of pedestrians is the body shape, and the regions of interest are validated by correlation with predefined probabilistic models [49].

3. Classification dataset

One of the contributions of this work is our pedestrian classification dataset, which consists of FIR images collected from a vehicle driven in outdoors urban scenarios. The dataset was recorded in Leganés, Spain and Coimbra, Portugal. Images were acquired with an Indigo Omega imager, with a resolution of 164×129 pixels, a grey-level scale of 14 bits, and focal length of 318 pixels. The camera was mounted on the exterior of the vehicle, to avoid infrared filtering of the windshield.

Recorded images were manually annotated, where each pedestrian is labeled as a bounding box. To prevent bias introduced by border artifacts their height is subsequently upscaled by 5%. Figure 1 shows some cropped-image examples of positives and negatives of the classification dataset. The pedestrians appear in an up-right position. In order to not introduce a pose bias, some tolerance is allowed. Specifically, images where head or torso is at a slant due to the normal walking gate are not discarded.

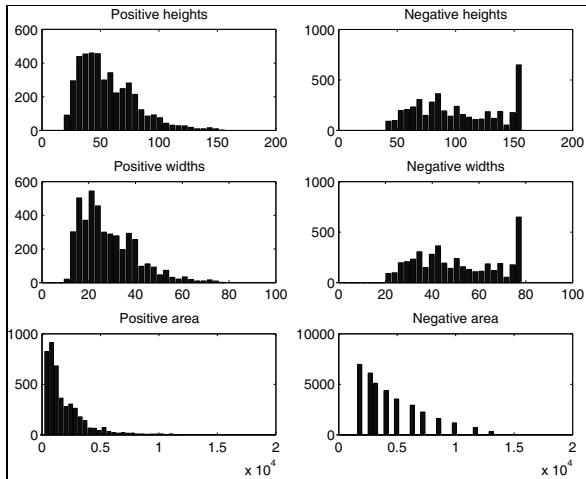


Fig. 2. Histograms of bounding boxes sizes and areas for positive and negative samples of the train dataset. X axis represent the measure (positive and negative height, width and area) and Y axis the number of samples having those value.

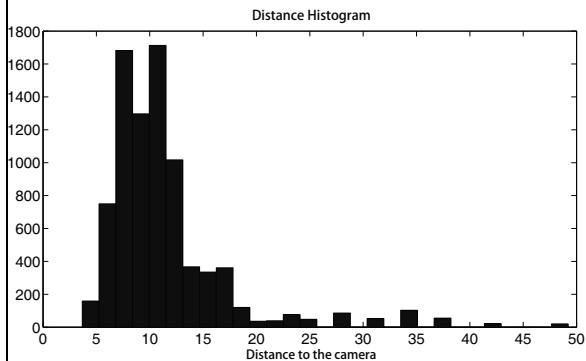


Fig. 3. Histograms of pedestrian's distance to the camera for the train and test dataset.

3.1. Dataset statistics

3.1.1. Number of samples

The dataset comprises 81592 14-bit one-channel images, divided in 16152 positives and 65440 negatives. The train set contains 10208 positives and 43390 negatives, while the test set contains 5944 positives and 22050 negatives.

3.1.2. Aspect ratio

Out of the annotated images, the bounding boxes are resized to a constant aspect ratio ($w/h = 0.5$) by changing their width (w) and height (h) appropriately. Figure 2 contains histograms for heights, widths and areas of positive and negative bounding boxes. The height of positive bounding boxes has a mean of 40 pixels. Those bounding boxes refer to pedestrians standing at

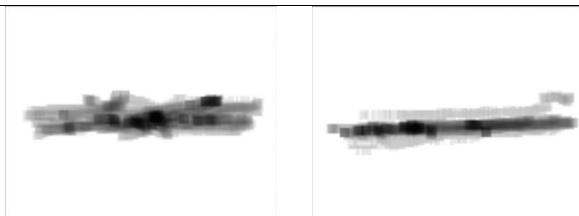


Fig. 4. Centers of bounding boxes for positives of the train and test dataset on a logarithmic scale. Left image: Positives in train dataset. Right image: Positives in test dataset.

approximately 10 m from the camera as seen in the histogram of Fig. 3. However, less frequent cases appear up to 50 m. Any bounding box below 10 pixels in height is ignored. The remaining bounding boxes are resized to 64×32 pixels using bilinear interpolation.

3.1.3. Density

Images were acquired from the usual point of view of the driver. As such, pedestrians appear more often in the center of the image as shown in Fig. 4, which represents the logarithmic density of the centers of the bounding boxes. In the case of negative samples, the bounding boxes are randomly selected, so the centers appear all over the image, with less density near the borders.

4. Detection dataset

The detection dataset contains the full frames from which the classification dataset was extracted, along with manual annotations of the pedestrian's positions.

The detection dataset was acquired in 13 different sessions, each containing a varying number of images. It comprises 15224 14-bit one-channel images, with dimension 164×129 pixels. The train set contains 6159 images, and the test set contains 9065 images. Only non-occluded pedestrians are considered for future evaluation, therefore images containing pedestrians with more than 20% of the area of the original bounding box occluded behind other obstacles are disregarded.

Each session occurred at a different location and with different illumination and temperature conditions. Out of those sessions 6 were used to compose the train set, leaving the remaining 7 for test set. This ensures that Train and Test are independent from one another. The temperature at which they were shot, which in turn affects the grey level and the histogram spread, causes the most important difference in appearance between sequences. Figure 5 contains the histogram of the mean grey level value of the train and test detection datasets.

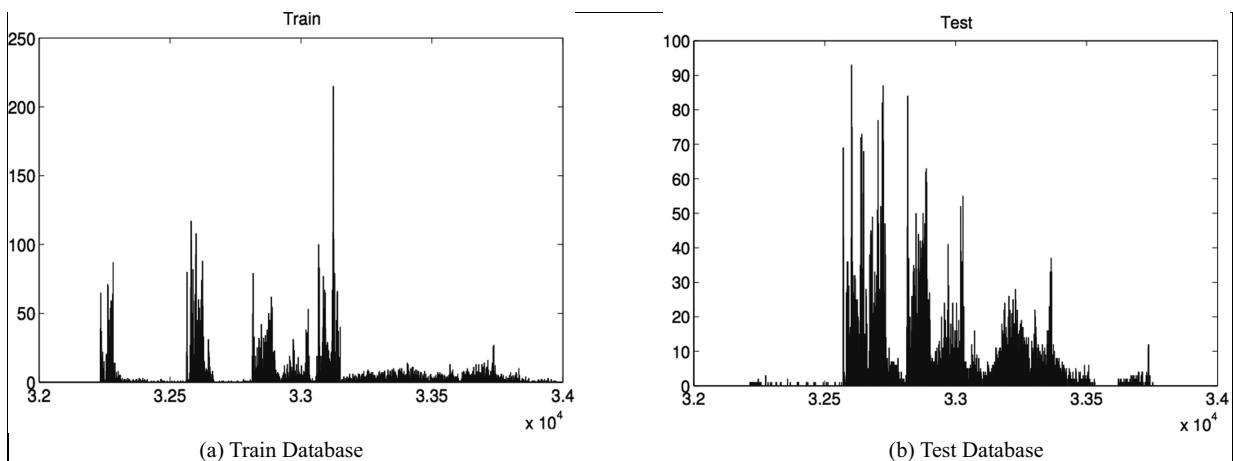


Fig. 5. Histogram of mean gray level of the images in the train and test datasets. X axis all possible values of gray level and Y axis the number of pixels having those levels. Left image: Train dataset. Right image: Test dataset.



Fig. 6. First five eigenpedestrians.

5. Image-based features

In this section the feature selection is discussed, along with implementation details.

LBP. Local Binary Patterns (LBP), as introduced in [48], represents the image as a similarity vector of each pixel with their surroundings. This descriptor encodes information as a binary number. For each pixel, the neighbors with a gray value higher or equal contribute with one in their position in the binary number, otherwise with zero. Each sample is divided in 3×3 pixel non-overlapping cells.

HOG. In this work we have used 5×5 pixel non-overlapping cells. In our 64×32 dataset this means 10×4 cells per image, once removed the border blocks. Within each cell a 9 bins histogram of orientation between 0 and 2π radians is calculated.

HOPE. Defining the shape of a pedestrian in FIR images is challenging using one dimensional point derivatives. This is due to the much wider infrared spectrum, compared with visible light. Another difficulty is that the sensitivity curve of an uncooled microbolometer sensor changes very quickly with minimum changes of its temperature [31]. To overcome

these challenges, we proposed in [50] a contrast invariant descriptor for pedestrian classification in FIR images called HOPE. Basically, the HOPE descriptor encodes a grid of local oriented histograms extracted from the phase congruency of the images, which is computed from a joint of Gabor filters.

The histograms are calculated in 5×5 pixel non-overlapping cells with 9 bins of orientation between 0 and 2π radians, for a total of 10×4 cells per image. No normalization step is applied. Phase congruency was calculated out of a set of 30 complex Gabor filters, divided in 5 scales ranging between a minimum wavelength of 2 pixels and a maximum of 10, and 6 orientations, ranging from 0 to 2π radians.

PCA. We treat PCA [29,44] eigenvectors as a grey-level feature vector. The initial motivation for applying this approach is that PCA tends to disregard small details at high frequency, as seen in Fig. 6, while FIR images usually have poor levels of detail, as they present softness due to motion blur, especially at low resolutions. We retain the 30 most significant eigenvectors, that is, those with the largest eigenvalues.

Feature concatenation. Descriptor fusion is explored as feature vector concatenation, resulting in a new higher dimension feature vector with different kinds of complementary information, which can improve the overall performance.

Selecting those features in the train dataset with a higher score in Welch's t-test shortens high-dimensional feature vectors. The minimum number of features selected is set so that the resulting feature vector has an accuracy on the test dataset within 1% of the unabridged vector. This usually results in feature vectors with half the dimensions of the original.

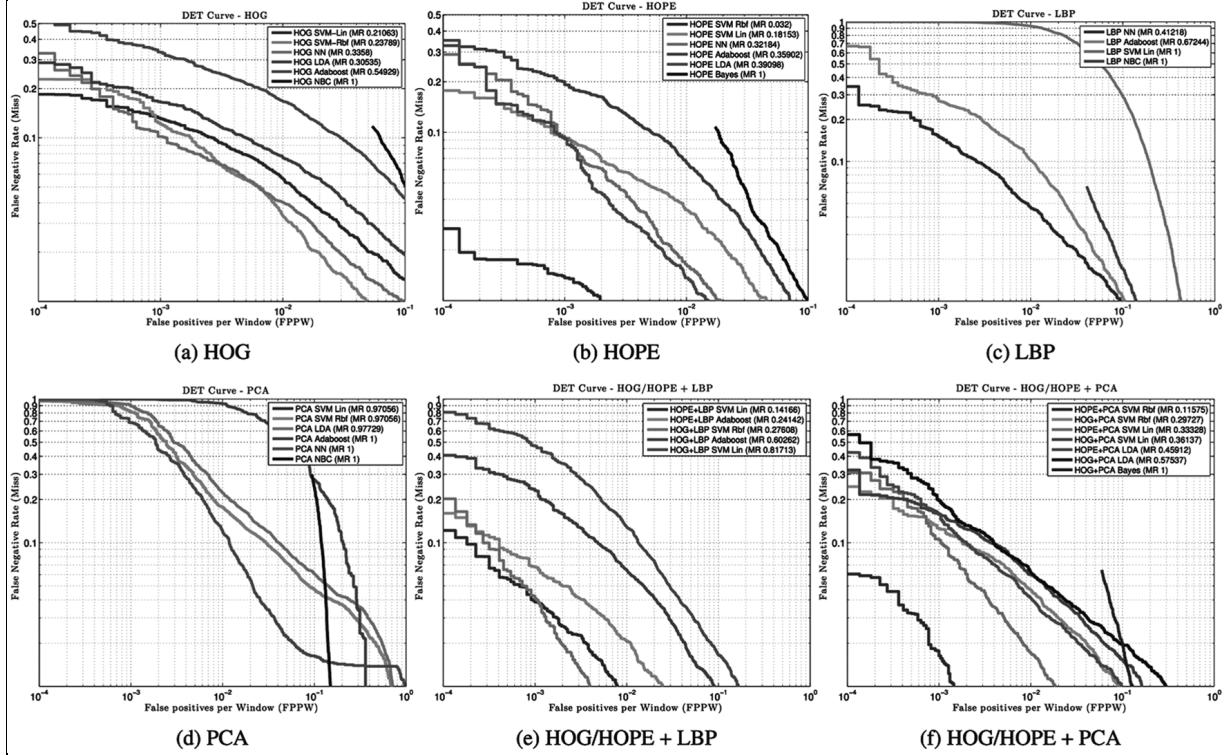


Fig. 7. DET curves for each feature vector in the classification dataset. Horizontal axis represent the number of false positives per window analyzed. Vertical axis represent the miss rate (MR) at 10^{-4} FPPW.

6. Classification methods

Pedestrian classification is treated as a supervised pattern recognition problem [16], where the set $x \in D = \{\{x_1, y_1\}, \dots, \{x_n, y_n\}\}$, is the collection of manually labeled examples, $x_i \in R^d$ is a feature vector and $y_i \in \{\pm 1\}$ is a binary label. In this section five kinds of classification methods have been used: SVM, Naïve Bayes Classifier (NBC), Quadratic Discriminant Analysis (QDA), Neural Networks (NN) and Adaboost. The parameters selected for the different classifiers are discussed in the sequel.

Support vector machines. Concerning SVM [11], two different kernels were used for benchmarking: a linear classifier, hereafter called SVM-Lin, and a radial basis function kernel (RBF) [2,41], designated by SVM-Rbf. In this implementation the radial Gaussian function kernel $K(x, y) = e^{-\gamma||x-y||^2}$ has a scale parameter $\gamma = 1$. Both linear and RBF kernel have a regularization trade-off parameter $C = 0.05$.

Naïve bayes classifier. NBC [39] is a member of the bayesian classifiers family [38,57]. It is designed for use when features are independent of one another within each class, but it appears to work well in prac-

tice in other circumstances. Naïve Bayes classification is based on estimating the conditional probability of the feature vector given the class.

Discriminant analysis. Linear Discriminant Analysis [44,58,68] is used as a linear classification model in terms of dimensionality reduction. Considering a two-class separation problem the D-dimensional input vector x can be projected down to one dimension as $y = w^T x$, where w is the components weight vector. Selecting appropriate weights the projection can be done over the dimension that maximally separates both classes, avoiding the exceeding overlapping that can occur due to dimensionality reduction. Over this projection a threshold w_0 is selected, where values $y \leq -w_0$ are classified as pedestrians, whereas values $y > -w_0$ are classified as background. In this implementation the coefficient matrix of the boundary equation is quadratic thus, the discriminant analysis takes a quadratic form, designated QDA, assuming normally distributed classes. The multivariate normal densities are fitted with covariance estimates stratified by group.

Neural network. A neural network pattern recognition scheme [33] is used with a two-layer feed-forward network, with ten hidden and one output sigmoid neu-

335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358

359 rons. The network is trained with scaled conjugate
 360 gradient backpropagation. The overall network function
 361 follows Eq. (1), where σ is the sigmoid function, rnk
 362 is the output ranking, N is the number of inputs and M
 363 is the maximum number of linear combinations of the
 364 N inputs.

$$rnk(x, w) = \sigma \left(\sum_{j=1}^M w_j^{(2)} h \left(\sum_{i=1}^N w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_0^{(2)} \right) \quad (1)$$

365 Neural Networks are a very popular pattern recogni-
 366 tion technique in the field of ITS and infrastructure
 367 engineering [3,26,42], and many variations of it have
 368 been developed in these fields [5,6,44,61,62].

369 *Adaboost.* We use Real AdaBoost as described
 370 in [60]. The key idea is that the combined response of a
 371 set of weak classifiers can build a strong one, improv-
 372 ing the performance that a complex classifier alone
 373 would have. Iteratively, Adaboost selects a threshold
 374 that best separates each feature set x_i in one of the
 375 classes y_i , applying a higher weight to misclassified
 376 samples. In this implementation the maximum number
 377 of iterations is set to 50. The final ranking of each fea-
 378 ture vector is $rnk = \sum_{i=1}^N x_i(f_i)$. In the case of HOG
 379 and HOPE, each bin in the orientation histograms is
 380 treated as a weak feature.

381 7. Evaluation on the classification dataset

382 In this section we compare the performance of each
 383 feature discussed in Section 5 by applying the classifi-
 384 cation techniques described in Section 6.

385 The train dataset is used to train the feature-classifier
 386 combinations. Likewise, testing is performed on the
 387 whole test dataset. Classification performance is eval-
 388 uated by means of Detection-Error Trade-off (DET)
 389 curves, which quantify the trade-off between miss rate
 390 and false positive rate.

391 Results for different ensembles of features and clas-
 392 sifiers are shown in Fig. 7. From these curves it can
 393 be observed that approaches based on local orientated
 394 histograms, such as HOG and HOPE, get better results
 395 than PCA or LBP. The best performing feature seems
 396 to be HOPE, with a miss rate of 0.3% at 10^{-4} false
 397 positives (FP) for the SVM-Lin classifier, followed by
 398 HOG with a miss rate of 0.38% at 10^{-4} FP. With an
 399 RBF kernel performance improves up to 0.06% miss
 400 rate at 10^{-4} FP in the case of HOPE and 0.25% at 10^{-4}
 401 FP in the case of HOG. DET curves of the SVM-Lin

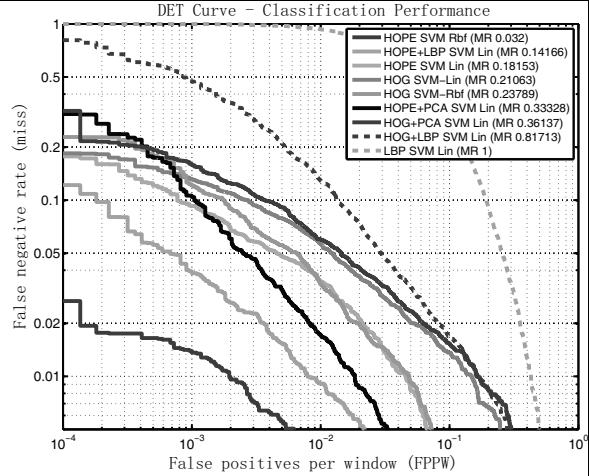


Fig. 8. DET curves of the best performing SVM-Lin classifiers. Results with Rbf kernels are plotted where they represent a significant gain. Legend states Miss Rate (MR) at 10^{-4} FPPW.

402 results are shown in Fig. 8. SVM-Rbf results are plot-
 403 ted in case of a significant gain.

404 Regarding feature combination, we have used an
 405 SVM-Lin to assess the impact of the features in the
 406 classification performance. LBP features combined
 407 with HOPE significantly reduce the miss rate, by 18%
 408 at 10^{-4} FP. PCA, although getting better results as an
 409 independent classifier than LBP, does not improve sig-
 410 nificantly classification when merged with other fea-
 411 tures. When merged with HOG the miss rate is reduced
 412 by 6% at 10^{-4} FP, while the improvement of merging
 413 it with HOPE is negligible.

414 Concerning the classification methods, SVM-Rbf
 415 generally has the best performance followed by SVM-
 416 Lin. LDA classifier performs almost as well, or bet-
 417 ter than Linear SVM for the HOG and HOPE descrip-
 418 tors. NN showed some promising results, though per-
 419 formance could be improved by using a robust training
 420 scheme, as the one presented in [4]. The NBC showed
 421 the worst performance, except for LBP features.

422 Additionally we performed some experiments to as-
 423 sess the impact on classification performance by vary-
 424 ing the number of negative examples on the train set.
 425 Figure 9 shows that, for the HOPE SVM-Rbf classifier,
 426 the performance gets significantly better by increasing
 427 the number of train negatives and no saturation is ap-
 428 preciated.

429 7.1. Statistical significance of the results

430 Statistical significance is assessed with McNemar's
 431 approximate test [22]. It is used to compare two classi-

Table 1

Results of the McNemar's approximate significance test for every pair of classifiers. The value expressed in the table's fields is χ^2 , as stated in Eq. (2)

	HOGLin	HOGRbf	HOPELin	HOPERbf	LBP	PCA
HOGLin	0	247.1	81.5	342.1	1852.9	1508.6
HOGRbf	247.1	0	51.6	11.5	2825.5	2443.7
HOPELin	81.5	51.6	0	108.0	2427.6	2056.1
HOPERbf	342.1	11.5	108.0	0	2980.8	2596.7
LBP	1852.9	2825.5	2427.6	2980.8	0	23.8
PCA	1508.6	2443.7	2056.1	2596.7	23.8	0

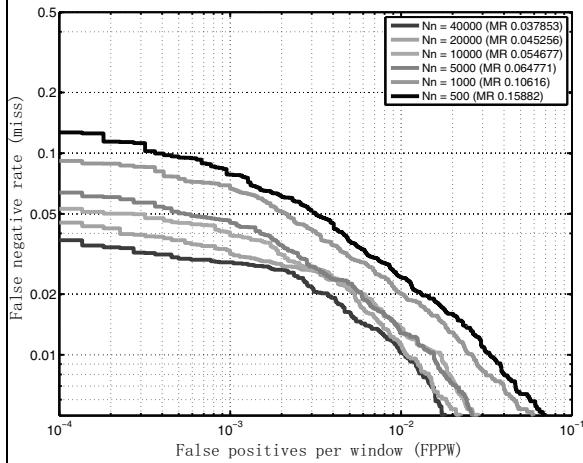


Fig. 9. DET curve of the HOPE SVM-Rbf classifier trained with an increasing number of negatives. Legend states Miss Rate (MR) at 10^{-4} FPPW.

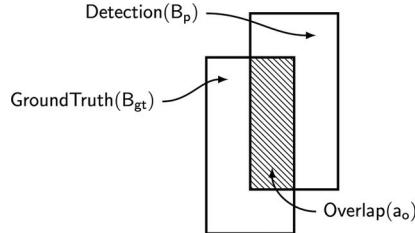


Fig. 10. Overlapping area of ground truth and detection.

432 fiers at a particular value of bias. To determine whether
433 classifier (C1) is significantly better than (C2), the χ^2
434 statistic is used Eq. (2).

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \quad (2)$$

435 Where n_{01} is a number of cases misclassified by
436 C1 and classified correctly by C2, and n_{10} is a num-
437 ber of cases misclassified by C2 and classified cor-
438 rectly by C1. The null hypothesis H_0 states that the
439 performance of both classifiers is the same. H_0 hy-
440 pothesis may be rejected if χ^2 falls below a proba-
441 bility of 5%, i.e. $\chi^2_{1,0.95} \geq 3.841$. If that is the case, it can

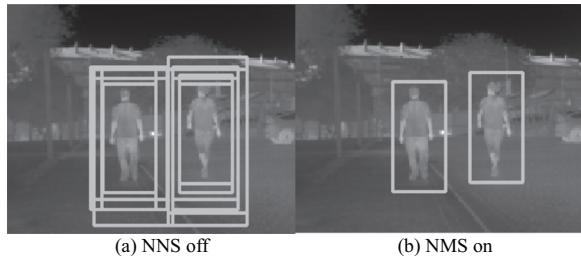


Fig. 11. Example of non maximum suppression of multiple detections for each pedestrian.

be assumed that one classifier performs significantly
442 better than the other.
443

444 Table 1 contains χ^2 values for every pair of classi-
445 fiers. The bias of all classifiers has been $b = 0$,
446 after rescaling, as it is the value that maximally sep-
447 arates both classes. From these results it can be con-
448 cluded that the null hypothesis can be rejected for all
449 classifier pairs.

8. Performance on the detection dataset

451 Pedestrian detection resumes on finding the position
452 and scale of an a priori unknown number of pedes-
453 trian on a set of full images. Detection is evaluated
454 by applying the best performing feature-classifier con-
455 structs explained in Sections 5 and 6, in a sliding win-
456 dows approach. For brevity, only best performing en-
457 sembles (HOG and HOPE in combination with SVM-
458 Lin, SVMRbf and Adaboost) are further discussed.

459 For multi-resolution detection purposes the input
460 image is resized to 6 different scales per octave, with an
461 average scaling step between scales of $ss = 1.1$. This
462 results in a scanning window that searches for pedes-
463 trians between 120 and 24 pixels in height. At each scale,
464 the detection window is moved sideways so that each
465 one has an 80% overlap with the previous one. Like-
466 wise, it is moved from top to bottom so that the over-
467 lapping is 90% in the vertical axis. This results in 2377
468 windows per image.

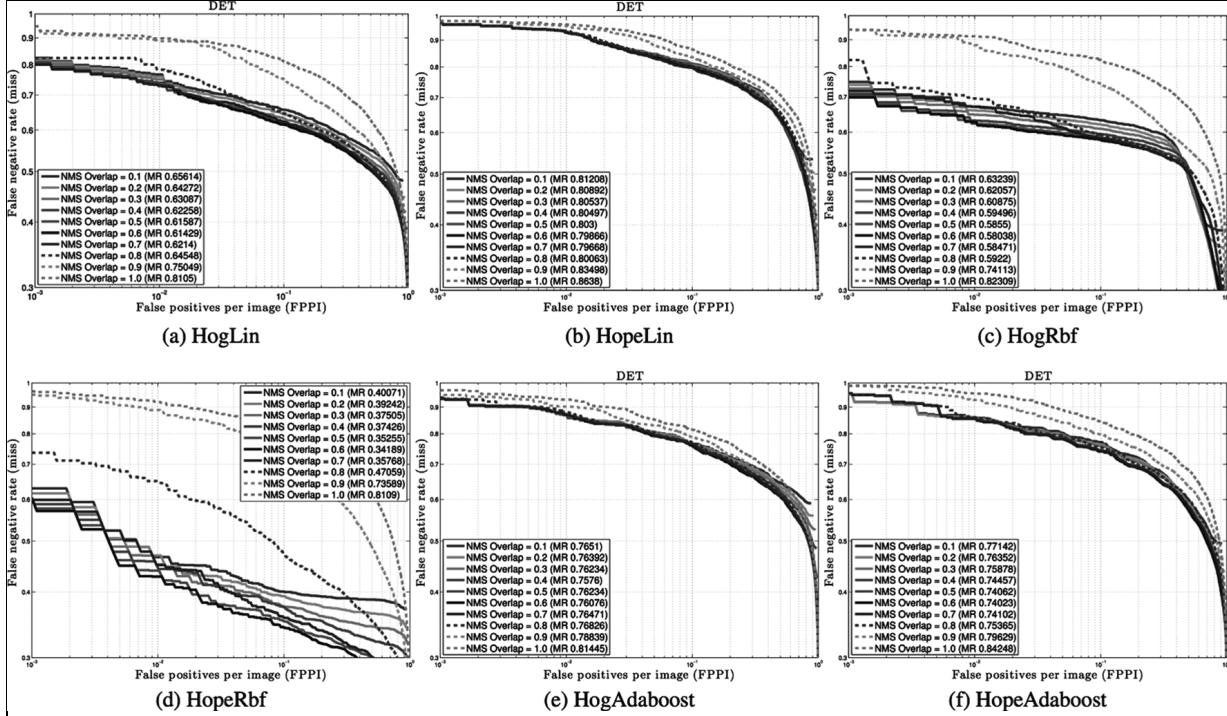


Fig. 12. DET curves for each feature vector in the detection database. For each non-maximum suppression overlap area threshold, one curve is plotted. Horizontal axis represent the number of false positives per image analyzed. Vertical axis represent the miss rate. Legend states miss rate (MR) at 0.1 FPPI.

469 8.1. Evaluation methodology

470 The detection task is evaluated by the Pascal Criteria [14], plotting results in DET curves. Detect ions are
 471 considered true or false positives based on the area of
 472 overlapping with ground truth bounding boxes. To be
 473 considered a correct detection, the area of overlap a_o
 474 between the predicted bounding box B_p and ground-truth
 475 bounding-box B_{gt} must exceed 50% by the Eq. (3), as
 476 depicted in Fig. 10.

$$477 \quad a_o = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})} \quad (3)$$

478 For each pedestrian, it is usual that many detections appear in the neighborhood around the ground-
 479 truth bounding-box. If two or more detections match the same ground-truth bounding-box, only the one with
 480 the higher score would be considered a true positive.
 481 Other overlapping detections are considered false positives. To minimize the number of repeated detections,
 482 a greedy non-maximum suppression (NMS) algorithm,
 483 pairwise max (PM) suppression [24], is applied to all
 484 bounding boxes. It selects iteratively detections with
 485 higher scores than their neighborhood, discarding de-

486 tections with lower scores over an overlapping percentage.
 487 This overlap is again calculated with Eq. (3). Figure
 488 11 shows an example of multiple detections for
 489 the same pedestrians, and the result after applying the
 490 NMS algorithm.

491 In Fig. 12 one curve is plotted for each overlapping
 492 percentage tested, in the range $a_o = \{0.1, \dots, 1.0\}$ for
 493 every detector considered. The legend states Miss Rate
 494 at 0.1 FPPI.

495 8.2. Results

496 In Fig. 13 the DET curves of the best performing de-
 497 tectors are plotted. Best results were obtained with the
 498 HOPE-Rbf detector, with a 35% miss rate at 0.1 False
 499 Positives per Image (FPPI), followed by the HOGRbf,
 500 with a 58% miss rate at 0.1 FPPI. Using a linear SVM,
 501 the HOG descriptor get an slightly higher miss rate of
 502 76% at 0.1 FPPI while for the HOG feature results de-
 503 grade up to 80% at 0.1 FPPI. With an Adaboost ap-
 504 proach the miss rates are 76% at 0.1 FPPI for the HOG
 505 descriptor and 74% at 0.1 FPPI for the HOPE descrip-
 506 tor. Based on these results it might seem that there is
 507 a correlation between classification and detection re-
 508 sults. Per image results present a correlation with the
 509 per window results.

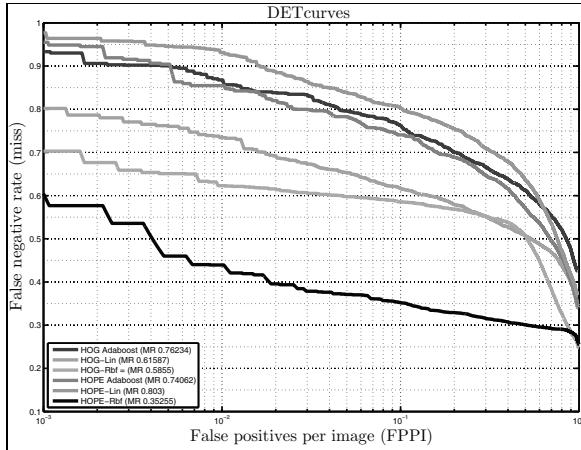


Fig. 13. Detection DET Curves after applying the PM NMS algorithm with an overlap threshold of $a_0 = 0.5$. Legend states miss rate (MR) at 0.1 FPPI.

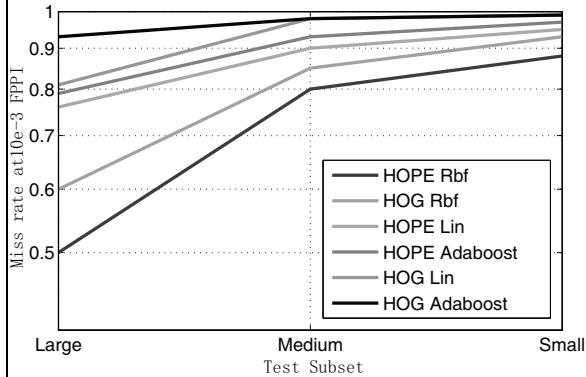


Fig. 14. Miss rate at 10^{-3} FPPI for all detectors in the Small, Medium and large test subsets.

8.2.1. Impact of pedestrian size

Pedestrian size has a big impact on detection results. Pedestrians located at a long distance from the vehicle appear at a lower resolution on the image and, as such, have lower detection rates. Some experiments are performed to assess the impact on detection performance by subdividing the test set into several subsets, i.e. small pedestrians ([10–40] pixels in height), medium ([40–80]) and large ([80–120]). While evaluating a subset, ground-truth bounding boxes not belonging to a given subset are assigned to class 0. A detection window overlapping any of the subsets is not counted as a true positive or as a false positive. Likewise, a missed detection of a class 0 bounding box is not counted as a false negative. Figure 14 shows the miss rate at 10^{-3} FPPI for the six applied detectors. Higher resolution pedestrians get the lowest miss rate in all cases.

Scanning the image looking for very small pedestrians also have an impact on computational demands of the algorithm in a sliding window detector.

8.3. Latent-SVM HOPE

One of the limitations of classifiers based on dense histograms of orientations is that performance degrades for very small cells [15]. Intuitively, finer grained descriptors retain more details, so classification should benefit from it. However, due to non-rigid deformations of pedestrians the overall classification scores for high-resolution descriptors are low. To overcome these limitations Felzenszwalb et al. introduced in [23,24] their latent SVM detector. This classification method relies on a set of filters: a low-resolution root filter and a set of high resolution part filters that define a hidden or latent structure. The locations of the parts of the pedestrian that best define its presence on the image are the latent variables z . While training, the exact location of the ground-truth bounding-box of positive examples is also a latent variable. This allows for auto-correcting mistakes made while labeling the dataset.

The detection is treated as a binary classification problem in a sliding window approach. Given a training set $D = \{\{x_1, y_1\}, \dots, \{x_n, y_n\}\}$, where $x_i \in R_d$ is a feature vector and $x_i \in \{\pm 1\}$ is a binary label, each region of interest of the image is assigned a score,

$$f_{\beta, Z}(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \quad (4)$$

Where x is the region of interest, $Z(x)$ is the set of all possible positions of the parts, and β is a vector of model parameters. The function $\Phi(x, z)$ is the feature vector assembled from the root filter and the best latent parts. β should then minimize Eq. (5).

$$\frac{1}{2} \left\| \max_{i=1, \dots, k} \beta \right\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i \cdot f_{\beta, Z}(x_i)) \quad (5)$$

Where k is the number of components in the model, and the second term determines the softness of the SVM margin.

To assess the impact on performance of part based detection, two descriptors have been trained using the Latent-SVM approach: HOG and HOPE. Detection performance is evaluated using the same methodology explained in Section 8.1, except for the following ex-

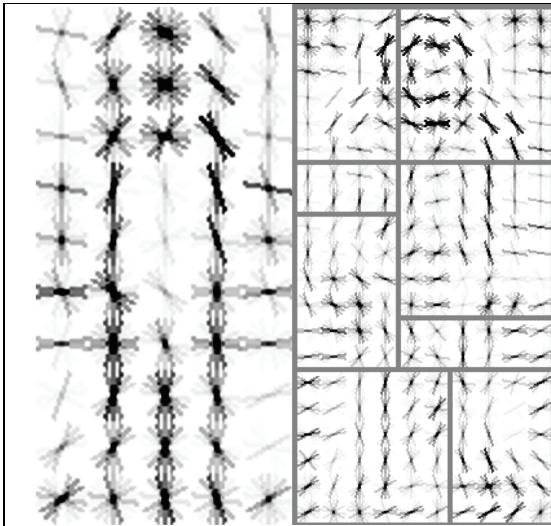


Fig. 15. Latent SVM filters using HOG feature.

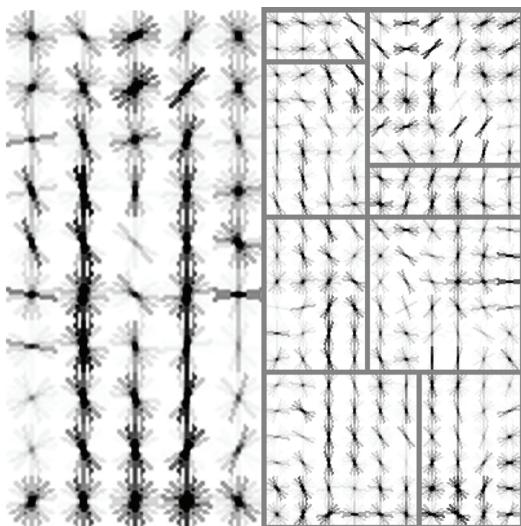
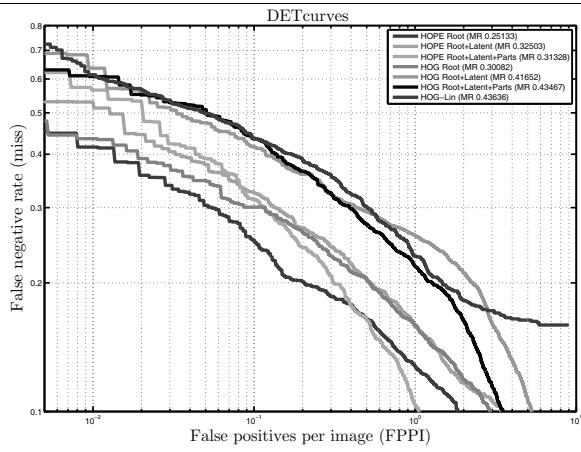


Fig. 16. Latent SVM filters using HOPE feature.

ception: the minimum overlap threshold between B_p and B_{gt} is set to a more restrictive value of $a_o = 0.7$.

The latent-SVM models for both the HOG and HOPE descriptors have been trained using the same parameters as their standalone counterparts. They have been initialized for $k = 8$ latent parts and (6×6) cells. The descriptors are extracted from a feature pyramid, where the parts filter has double the resolution as the root filter. Figure 15 represents the filters trained with the HOG descriptor and Fig. 16 the filters trained with the HOPE descriptor.

The latent parts are computed at twice the resolutions as the root filter. In low-resolution images, such

Fig. 17. Detection DET Curves of Latent-SVM HOG and HOPE after applying the PM NMS algorithm with an overlap threshold of $a_0 = 0.7$. Legend states Miss Rate (MR) at 0.1 FPPI.

as the ones presented in this database, latent parts are hard to find in the smallest objects. Because of this limitation, the following results are based on a subsample of the testing database. Only images in the Medium and Large subsets are used. For a pedestrian with less than 40 pixels in height the size of a cell of the root filter would have (3×3) pixels, therefore it is not possible to construct the part filters at double that resolution. All other images are previously resized to double their size in order to satisfy minimum resolution restriction mentioned above.

The DET curves in Fig. 17 compares the performance of root, latent and parts detectors for the HOG and HOPE descriptors. Root filters are the original low-resolution descriptors trained with an extended set of warped positives. There is a notable decrease in miss rate when compared with their standalone counterparts, due to the extended training set, allowing the detector to cope with subtle pose variations. Introducing latent detections (i.e. automatic estimations of the pedestrians correct position) seems to degrade slightly the performance. The overall best results for Large pedestrians derive from searching for pedestrian parts. This seems to help specially in the case of semi-occluded pedestrians. As the pedestrians get smaller in the images, adding the parts filter loses its edge, due to greater resolution demands. For comparing purposes the DET curve of the original HOG-Lin tested under the same conditions is also included in the figure.

Finally, a qualitative inspection of misclassified samples suggests that ambient temperature has a determinant impact on performance. Sequences collected at a high environment temperature or under direct sun



Fig. 18. Misclassified samples. Upper row. False negatives due to low resolution, motion blur and pose variation lower row: False positives in areas with a high vertical symmetry.

light present the most false positives, and also the highest miss rate. A qualitative evidence of this issue is shown in Fig. 18 where we present examples of misclassified positives and negatives. Other sources of misclassification are motion blur, which in FIR images appear frequently, and pose variation. False positives appear mostly in negative examples with a high vertical symmetry.

9. Conclusions and discussion

This paper presents an experimental study on pedestrian classification in FIR images. Several combinations of descriptor and classification methods have been tested in a new FIR dataset. By our best knowledge this is the first complete FIR based pedestrian classification and detection dataset publicly available for benchmarking.

From the experimental results reported in the previous sections it can be concluded that histogram based features perform best than LBP or PCA features. Among the features, HOPE performs better both for classification and detection problems. LBP and PCA features get worse overall performance, though some gains can be achieved by merging them with HOG or HOPE. In terms of classification methods, SVM achieved the best performance. The RBF kernel can significantly reduce misclassifications compared with

a linear kernel, but is more computationally demanding. This is a critical factor in computer aided transportation applications.

Pedestrian resolution has an important impact on performance, with all detectors having lower hit rates for small pedestrians. For the best performing detector, the miss rate at 10^{-3} goes from 50% for *Large* pedestrians to 88% for *Small* ones. In the case of Latent SVM approach, the detectors are not able to detect any pedestrian in the *Small* subsample. From these results it may be concluded that current pedestrian detectors in FIR images suffer from the same limitations as VL detectors when it comes to detect small pedestrians. This is specially relevant in this case, as FIR cameras tend to have low resolution sensors.

The results presented in this paper suggests that FIR images are a very useful source of information for pedestrian classification and detection, having similar performance to that found in state of the art in VL images, with advantage in low visibility applications.

Acknowledgments

This work was supported by the Spanish Government through the Cicyt projects FEDORA (GRANT TRA2010-20225-C03-01) and Driver Distraction Detector System (GRANT TRA2011-29454-C03-02), and the Comunidad de Madrid through the project SEGVAUTO (S2009/DPI-1509).

References

- [1] H. Adeli and S. Ghosh-Dastidar, Mesoscopic-wavelet freeway work zone flow and congestion feature extraction model, *Journal of Transportation Engineering* **130**(1) (2003), 94–103.
- [2] H. Adeli and A. Karim, Fuzzy-wavelet rbnn model for freeway incident detection, *Journal of Transportation Engineering* **126**(6) (2000), 464–471.
- [3] H. Adeli and A. Panakkat, A probabilistic neural network for earthquake magnitude prediction, *Neural Networks* **22**(7) (2009), 1018–1024.
- [4] M. Ahmadi and H. Adeli, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integrated Computer-Aided Engineering* **17**(3) (2010), 197–210.
- [5] S.U. Ahmed, M. Shahjahan and K. Murase, A lempel-ziv complexity-based neural network pruning algorithm, *International Journal of Neural Systems* **21**(5) (2011), 427–441.
- [6] M. Akhand and K. Murase, Ensembles of neural networks based on the alteration of input feature values, *International Journal of Neural Systems* **22**(1) (2012), 77–87.
- [7] M. Al-Naser and U. Söderström, Reconstruction of occluded facial images using asymmetrical principal component analysis, *Integrated Computer-Aided Engineering* **19**(3) (2012), 273–283.

- [8] M. Andriluka, S. Roth and B. Schiele, People tracking-by-detection and people-detection-by-tracking, *Computer Vision and Pattern Recognition*, IEEE (2008), 1–8.
- [9] M. Bertozi, A. Broggi, M.D. Rose, M. Felisa, A. Raktamamonjy and S. Frédéric, A pedestrian detector using histograms of oriented gradients and a support vector machine classifier, *Intelligent Transportation Systems Conference*, IEEE (2007), 143–148.
- [10] E. Binelli, A. Broggi, A. Fascoli, S. Ghidoni, P. Grisleri, T. Graf and M. Meinecke, A modular tracking system for far infrared pedestrian recognition, *Intelligent Vehicles Symposium*, IEEE (2005), 759–764.
- [11] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* **2**(2) (1998), 121–167.
- [12] P.M. Ciarelli, E.O.T. Salles and E. Oliveira, Human automatic detection and tracking for outdoor video, *Integrated Computer-Aided Engineering* **18**(4) (2011), 379–390.
- [13] S. Cruz-Ramírez, Y. Mae, T. Arai, T. Takubo and K. Ohara, Vision-based hierarchical recognition for dismantling robot applied to interior renewal of buildings, *Computer-Aided Civil and Infrastructure Engineering* **26**(5) (2011), 336–355.
- [14] H. Dai, H. Zhang, W. Wang and G. Xue, Structural reliability assessment by local approximation of limit state functions using adaptive markov chain simulation and support vector regression, *Computer-Aided Civil and Infrastructure Engineering* **27**(9) (2012), 676–686.
- [15] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition*, IEEE **1** (2005), 886–893.
- [16] T.G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation* **10**(7) (1998), 1895–1923.
- [17] P. Dollár, C. Wojek, B. Schiele and P. Perona, Pedestrian detection: A benchmark, *Computer Vision and Pattern Recognition*, IEEE (2009), 304–311.
- [18] P. Dollar, C. Wojek, B. Schiele and P. Perona, Pedestrian detection: An evaluation of the state of the art, *Pattern Analysis and Machine Intelligence* **34**(4) 743–761, 2012.
- [19] V. Enescu, G. de Cubber, K. Cauwerts, H. Sahli, E. Demeerster, D. Vanhooydonck and M. Nuttin, Active stereo vision-based mobile robot navigation for person tracking, *Integrated Computer-Aided Engineering* **13**(3) (2006), 203–222.
- [20] M. Enzweiler and D.M. Gavrila, Monocular pedestrian detection: Survey and experiments, *Pattern Analysis and Machine Intelligence* **31**(12) (2009), 2179–2195.
- [21] A. Ess, B. Leibe and L. van Gool, Depth and appearance for mobile scene analysis, *The International Computer Vision Conference*, IEEE (2007), 1–8.
- [22] M. Everingham, L. van Gool, C.K.I. Williams, J. Winn and A. Zisserman, The pascal visual object classes (voc) challenge, *Proceedings of the International Journal of Computer Vision* **88**(2) (2010), 303–338.
- [23] P. Felzenszwalb, D. McAllester and D. Ramanan, A discriminatively trained, multiscale, deformable part model, *Computer Vision and Pattern Recognition*, IEEE (2008), 1–8.
- [24] P. Felzenszwalb, R.B. Girshick, D. McAllester and D. Ramanan, Object detection with discriminatively trained part-based models, *Pattern Analysis and Machine Intelligence* **32**(9) (2010), 1627–1645.
- [25] D. Geronimo, A.M. Lopez, A.D. Sappa and T. Graf, Survey of pedestrian detection for advanced driver assistance systems, *Pattern Analysis and Machine Intelligence* **32**(7) (2010), 1239–1258.
- [26] D. Geronimo, A. Sappa, A.M. López and D. Ponsa, Adaptive image sampling and windows classification for onboard pedestrian detection, *Proceedings of the International Conference on Computer Vision Systems*, Bielefeld, Germany, (2007).
- [27] S. Ghosh-Dastidar and H. Adeli, Wavelet-clustering neural network model for freeway incident detection, *Computer-Aided Civil and Infrastructure Engineering* **18**(5) (2003), 325–338.
- [28] S. Ghosh-Dastidar and H. Adeli, Neural network wavelet microsimulation model for delay and queue length estimation at freeway work zones, *Journal of Transportation Engineering* **132**(4) (2006), 331–341.
- [29] S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Transactions on Biomedical Engineering* **55**(2) (2008), 512–518.
- [30] R. Harrison, R. Birchall, D. Mann and W. Wang, Novel consensus approaches to the reliable ranking of features for seabed imagery classification, *International Journal of Neural Systems* **22**(6) (2012).
- [31] M.K. Hinders and W.L. Fehlman, Passive infrared thermographic imaging for mobile robot object identification, *AIP Conference Proceedings* **1211**(2010) (2010).
- [32] S. Hooshdar and H. Adeli, Toward intelligent variable message signs in freeway work zones: Neural network model, *Journal of Transportation Engineering* **130**(1) (2003), 83–93.
- [33] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences* **79**(8) (1982), 2554–2558.
- [34] J. Hou, Z. Chen, X. Qin and D. Zhang, Automatic image search based on improved feature descriptors and decision tree, *Integrated Computer-Aided Engineering* **18**(2) (2011), 167–180.
- [35] X. Jiang and H. Adeli, Freeway work zone traffic delay and cost optimization model, *Journal of Transportation Engineering* **129**(3) (2003), 230–241.
- [36] X. Jiang and H. Adeli, Freeway work zone traffic delay and cost optimization model, *Journal of Transportation Engineering* **129**(3) (2003), 230–241.
- [37] X. Jiang and H. Adeli, Object-oriented model for freeway work zone capacity and queue delay estimation, *Computer-Aided Civil and Infrastructure Engineering* **19**(2) (2004), 144–156.
- [38] X. Jiang, S. Mahadevan and H. Adeli, Bayesian wavelet packet denoising for structural system identification, *Structural Control and Health Monitoring* **14**(2) (2007), 333–356.
- [39] G.H. John and P. Langley, Estimating continuous distributions in bayesian classifiers, *Proceedings of the Uncertainty in Artificial Intelligence Conference* (1995), 338–345.
- [40] V. Jumutc, P. Zayakin and A. Borisov, Ranking based kernels in applied biomedical diagnostics using a support vector machine, *International Journal of Neural Systems* **21**(6) (2011), 459–473.
- [41] A. Karim and H. Adeli, Comparison of fuzzy-wavelet radial basis function neural network freeway incident detection model with california algorithm, *Journal of Transportation Engineering* **128**(1) (2002), 21–30.
- [42] C.M. Lin, A.B. Ting, C.F. Hsu and C.M. Chung, Adaptive control for mimo uncertain nonlinear systems using recurrent wavelet neural network, *International Journal of Neural Systems* **22**(1) (2012), 37–50.

- [43] F. Liu and M.J. Er, A novel efficient learning algorithm for self-generating fuzzy neural network with applications, *International Journal of Neural Systems* **22**(1) (2012), 21–35.
- [44] A.M. Martínez and A.C. Kak, Pca versus lda, *Pattern Analysis and Machine Intelligence* **23**(2) (2001), 228–233.
- [45] A. Miron, B. Besbes, A. Rogozan, S. Ainouz and A. Ben-srhair, Intensity self similarity features for pedestrian detection in far-infrared images, *Intelligent Vehicles Symposium*, IEEE (2012), 1120–1125.
- [46] S. Munder and D.M. Gavrila, An experimental study on pedestrian classification, *Pattern Analysis and Machine Intelligence* **28**(11) (2006), 1863–1868.
- [47] T. Nishikawa, J. Yoshida, T. Sugiyama and Y. Fujino, Concrete crack detection by multiple sequential image filtering, *Computer-Aided Civil and Infrastructure Engineering* **27**(1) (2012), 29–47.
- [48] T. Ojala, M. Pietikainen and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *Pattern Analysis and Machine Intelligence* **24**(7) (2002), 971–987.
- [49] D. Olmeda, A. de la Escalera and J. Armingol, Far infrared pedestrian detection and tracking for night driving, *Robotica* **29**(4) (2011), 495–505.
- [50] D. Olmeda, A. de la Escalera and J. Armingol, Contrast invariant features for human detection in far infrared images, *Intelligent Vehicles Symposium*, IEEE (2012), 117–122.
- [51] R. O’Malley, M. Glavin and E. Jones, A review of automotive infrared pedestrian detection techniques, *Signals and Systems Conference* (2008), 168–173.
- [52] C. Papageorgiou and T. Poggio, A trainable system for object detection, *International Journal of Computer Vision* **38**(1) (2000), 15–33.
- [53] G. Pelaez, M. Romero, J. Armingol, A. de la Escalera, J. Muñoz, W. van Bijsterveld and J. Bolaño, Detection and classification of road signs for automatic inventory systems using computer vision, *Integrated Computer-Aided Engineering* **19**(3) (2012), 285–298.
- [54] F. Petitjean, F. Masseglia, P. Gancarski and G. Forestier, Discovering significant evolution patterns from satellite image time series, *International Journal of Neural Systems* **21**(6) (2011), 475–489.
- [55] C. Premebida and U. Nunes, Fusing lidar, camera and semantic information: A context-based approach for pedestrian detection, *The International Journal of Robotics Research* **32**(3) (2013), 371–384.
- [56] R. Putha, L. Quadrifoglio and E. Zechman, Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions, *Computer-Aided Civil and Infrastructure Engineering* **27**(1) (2012), 14–28.
- [57] S. Saliminejad and N.G. Gharaibeh, A spatial bayesian technique for imputing pavement network repair data, *Computer-Aided Civil and Infrastructure Engineering* **27**(8) (2012), 594–607.
- [58] A. Samant and H. Adeli, Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis, *Computer-Aided Civil and Infrastructure Engineering* **15**(4) (2000), 241–250.
- [59] A. Sánchez, E.O. Nunes and A. Conci, Using adaptive background subtraction into a multi-level model for traffic surveillance, *Integrated Computer-Aided Engineering* **19**(3) (2012), 239–256.
- [60] R.E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, *Machine Learning* **37**(3) (1999), 297–336.
- [61] F.M. Schleif and T. Villmann, Barbara hammer and petra schneider, efficient kernelized prototype based classification, *International Journal of Neural Systems* **21**(06) (2011), 443–457.
- [62] L. Shi, Y. Shi, Y. Gao, L. Shang and Y.B. Yang, Xcsc: A novel approach to clustering with extended classifier system, *International Journal of Neural Systems* **21**(1) (2011), 79–93.
- [63] D. Simonnet, S.A. Velastin, E. Turkbeyler and J. Orwell, Backgroundless detection of pedestrians in cluttered conditions based on monocular images: A review, *Computer Vision, IET* **6**(6) (2012), 540–550.
- [64] K.K. Sung and T. Poggio, Example-based learning for view-based human face detection, *Pattern Analysis and Machine Intelligence* **20**(1) (1998), 39–51.
- [65] Y. Tsai and Y. Huang, A generalized framework for parallelizing traffic sign inventory of video log images using multi-core processors, *Computer-Aided Civil and Infrastructure Engineering* **27**(7) (2012), 476–493.
- [66] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *Computer Vision and Pattern Recognition*, IEEE **1** (2001), I–511.
- [67] S. Walk, N. Majer, K. Schindler and B. Schiele, New features and insights for pedestrian detection, *Computer Vision and Pattern Recognition*, IEEE **1** (2010), 1030–1037.
- [68] A.P. Worth and M.T.D. Cronin, The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects, *Journal of Molecular Structure: THEOCHEM* **622**(1) (2003), 97–111.
- [69] T. Xiong, M. Zhang, C.W. Shu, S.C. Wong and P. Zhang, High-order computational scheme for a dynamic continuum model for bi-directional pedestrian flows, *Computer-Aided Civil and Infrastructure Engineering* **26**(4) (2011), 298–310.
- [70] C. Zhang and A. Elaksher, An unmanned aerial vehicle-based imaging system for 3d measurement of unpaved road surface distresses, *Computer-Aided Civil and Infrastructure Engineering* **27**(2) (2012), 118–129.