Machine Vision and Applications (2019) 30:645–655
https://doi.org/10.1007/s00138-019-01016-w

ORIGINAL PAPER

Affordable person detection in omnidirectional cameras using radial integral channel features

Barış Evrim Demiröz[1] · Albert Ali Salah[1,2] · Yalin Bastanlar[3] · Lale Akarun[1]

Received: 20 May 2018 / Revised: 13 December 2018 / Accepted: 12 February 2019 / Published online: 18 March 2019

Abstract
Omnidirectional cameras cover more ground than perspective cameras, at the expense of resolution. Their comprehensive field of view makes omnidirectional cameras appealing for security and ambient intelligence applications. Person detection is usually a core part of such applications. Conventional methods fail for omnidirectional images due to different image geometry and formation. In this study, we propose a method for person detection in omnidirectional images, which is based on the integral channel features approach. Features are extracted from various channels, such as LUV and gradient magnitude, and classified using boosted decision trees. Features are pixel sums inside annular sectors (doughnut slice shapes) contained by the detection window. We also propose a novel data structure called *radial integral image* that allows to calculate sums inside annular sectors efficiently. We have shown with experiments that our method outperforms the previous state of the art and uses significantly less computational resources.

Keywords Omnidirectional camera · Object detection · Human detection · Person detection · Integral channel features · Integral image

1 Introduction

We are entering an era where humans and intelligent systems will coexist in factory and office environments, smart cities and roads and home environments. Intelligent systems need to detect humans in order to protect them, serve them and communicate with them. Person detection is a crucial step for surveillance, autonomous vehicles and assisted living applications. Depending on the application domain, detecting the body of a human being is called human detection, pedestrian detection or person detection. Many applications require the detection of multiple persons using video sensors, as well as the tracking of detected persons, their re-identification in different camera views, and at later times, the classification of their actions. The detection of persons in an environment using minimal computational resources is a challenging first step.

Many person detection studies use conventional perspective cameras. In that case, multiple cameras are needed to cover the ground of interest [1]. Omnidirectional cameras have a very wide field of view and might reduce, if not eliminate, the need to use multiple perspective cameras. However, the use of omnidirectional cameras for object detection has been limited. This is partly because conventional camera approaches are not directly applicable and need to be modified in a theoretically correct and practical manner to be used with omnidirectional cameras. In this work, we propose a method to perform person detection directly on images obtained by omnidirectional cameras. Our method requires minimal computational resources to achieve the state-of-the-art person detection performance.

Our contribution in this paper is twofold. First, we introduce a novel integral image scheme for omnidirectional

✉ Barış Evrim Demiröz
baris.evrim.demiroz@gmail.com

Albert Ali Salah
salah@boun.edu.tr

Yalin Bastanlar
yalinbastanlar@iyte.edu.tr

Lale Akarun
akarun@boun.edu.tr

[1] Computer Engineering Department, Boğaziçi University, 34342 Istanbul, Turkey

[2] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

[3] Department of Computer Engineering, Izmir Institute of Technology, 35430 Urla, Izmir, Turkey

Fig. 1 Person detection in an omnidirectional camera setting, with the detection results of the proposed approach superimposed

images to speed up feature extraction. Integral images have been extremely useful for speeding up detection problems and were used in the rapid face detection study proposed by Viola and Jones [37]. However, integral images work with rectangular bounding boxes, an assumption which no longer holds in omnidirectional images. As Fig. 1 illustrates, the bounding box around the detected person is much narrower in its base. Our proposed solution makes it possible to use integral images directly on omnidirectional camera images.

As a second contribution, we advance the state of the art in omnidirectional camera-based person detection. Using the new integral image structure and the *integral channel features* (ICF) approach [15], we outperform recent omnidirectional camera person detection algorithms [10]. We also compare our method to converting the omnidirectional image to a panoramic image and then applying the standard ICF method, and experimentally show the superior performance of our approach.

This paper is structured as follows: In Sect. 2, a brief overview of the feature extraction and person detection methods using omnidirectional cameras is given. In Sect. 3, our novel person detection scheme using omnidirectional cameras is outlined, including our camera model in Sect. 3.2, and a novel data structure, namely the radial integral image, to rapidly extract feature vectors from omnidirectional images in Sect. 3.3. In Sect. 4, the experiments we have conducted to validate our approach are reported with comparisons to the state of the art. Section 5 concludes the paper.

2 Related work

2.1 Camera-based person detection

In a classical work of object detection [11], Dalal and Triggs extracted histogram of oriented gradient (HOG) features from overlapping rectangular regions from the detection window. They used these features with support vector machine (SVM) classifiers to perform human detection. Consulting such gradient-based features is a well-established idea in person detection. Later, Zhu et al. [41] used integral image histograms to speed up the feature extraction step in the HOG detector. Felzenszwalb et al. [19] developed a similar model with multiple body parts, where the positions of these parts were inferred as the latent variables of an SVM. In [36], authors used the covariance matrix of different image features (i.e., covariance features) for pedestrian detection. Conventional classifiers do not perform well for covariance matrices, because they do not form a vector space. Instead they proposed a method to effectively do classification with covariance features on a Riemannian manifold.

In [2], slanted integral images are used to approximate the Laplacian of Gaussian filter to detect key points in the image. With the slanted integral images, the area of a right trapezoidal can be computed in constant time. In a similar vein, [30] used rotated integral images to extract Haar-like features from images. Although the time complexity is constant, the space complexity of both methods increases with each added rotation angle, because a separate integral image needs to be computed for that angle.

Performance is a major concern in person detectors, as the applications typically require real-time computations. In [15], authors combined the idea of boosting multiple simple features, as in the Viola–Jones detector [37] throughout multiple channels, including the HOG channel. Their work, namely, integral channel features (ICF), has been very successful due to its simplicity, low computational cost and detection performance. ICF forms the backbone of our approach. To deal with computational performance issues, Dollar et al. [14] proposed that feature responses can be used to approximate feature responses at nearby scales. They accelerated person detection by avoiding the building of the full-scale space. Following a similar line of work, Benenson et al. [7] investigated every component of a rigid (i.e., not part based) detector and improved HOG + SVM miss rate by more than 30%, through adjustment of system components such as feature pooling and normalization. In addition to HOG, local binary patterns (LBP) are also used for detection [35].

In [40], authors investigated the failure cases of top performing pedestrian detectors (most of them being from

the ICF family) in detail and suggested ways to design and improve existing detectors. We refer the readers to [17] for a comprehensive review of methods on person detection using perspective cameras.

Introduction of deep learning changed the scene for object detection [21, 27, 33]. Although some of the person detection approaches rely on deep neural networks [31] until recently, convolutional neural networks (CNN) failed to catch up with ICF-based methods [22, 39]. In a recent work, Cao et al. [9] used handcrafted feature channels (LUV, HOG, etc.) and features from inner layers of a CNN to perform classification with AdaBoost. This can be considered a hybrid approach, as neural networks are not used in an end-to-end fashion. It is possible to fuse CNN features with other detectors, but this represents a trade-off between accuracy and speed. Furthermore, ICF-based methods are still attractive due to computational advantage gained by the simplicity of these features and good detection performance.

2.2 Detection in omnidirectional cameras

Regarding object detection studies with omnidirectional cameras, some previous approaches first transform the omnidirectional image into a panoramic image and then apply conventional detection methods on this image [23, 25, 38, 42]. However, this transformation introduces extra parameters for tuning and brings additional computational effort. Panoramic transformation also distorts objects. Therefore, objects in transformed images differ from those in perspective cameras. Especially for tall objects, this distortion results in a decreased detection performance [10].

Geyer and Daniilidis [20] have shown that every central projection system can be modeled as a projection to a sphere, followed by a projection to the image plane. Based on this sphere model, researchers recently proposed methods to compute features directly on omnidirectional images. Puig and Guerrero [32] proposed using differential operators on the sphere to construct a scale space for omnidirectional images. Arıcan and Frossard [6] used the same idea to construct a feature detection and extraction method for catadioptric omnidirectional cameras. The features they used are similar to Lowe's [29] popular scale invariant feature transform (SIFT). Lourenço et al. [28] proposed a similar framework, called sRD-SIFT, for images with radial lens distortions. Their method corrects the gradient using lens distortion coefficient, provided that it is available.

Tracking people with omnidirectional cameras is relevant for both indoor and outdoor settings. Saito et al. [34] used template matching in a Bayesian framework to detect and track multiple people in omnidirectional cameras. They generated different templates for people standing at different distances from the camera. Alahi et al. [3, 4] used omnidirectional cameras along with perspective cameras for person detection in a basketball game. They used a dictionary of binary human silhouettes for each discrete location and inferred the actual occupancies using binary foreground detection as the input. They formulated the problem as a linear inverse problem and added a constraint on the maximum number of people to enforce the sparsity of the solution. In [12], authors used the same basketball game data to generate ground occupancy maps by backprojecting the foreground maps to ground plane. In [13], authors used, silhouette-based approach similar to [4], along with a hierarchical hidden Markov model to track a person in a room and detect falls.

Features can be tailored for omnidirectional images. Cinaroglu and Bastanlar [10] took the traditional HOG approach and modified the features according to the Riemannian metric on the sphere camera model. In that way, object detection was done directly on the omnidirectional image. They also proposed rotating annular sectors (doughnut slice shapes) to improve the performance over rotating rectangular windows. Their approach is computationally expensive, since the transformation of HOG features is done separately for each sliding-rotating window. In this study, we propose a faster and more accurate approach.

In [26], a circular grid scheme was proposed instead of a rectangular grid to calculate HOG responses over annular sectors. In this way, cyclic shifts of the final descriptor represent image rotations, which helps achieve some rotation invariance, especially if no assumption can be made regarding the object's orientation. However, in omnidirectional images, standing humans are typically aligned with lines diverging from the image center. Incorporating full rotation invariance in such scenarios would reduce the discriminative power.

To our knowledge, the method we propose in this paper is the first to compute the integral image for omnidirectional images and therefore, it is the first to apply the state-of-the-art ICF method [17] for object detection with omnidirectional cameras. We describe our approach in the next section.

3 Methodology

3.1 Integral channel features method

This subsection briefly describes the integral channel features method [15], followed by a description of the proposed *radial integral channel features* (RICF).

The integral channel features (ICF) method first computes multiple channels from a single input image. These include individual color channels (e.g., RGB, LUV), gradient magnitude, HOG and the difference of Gaussian filtered image. Then, summarizing features are extracted from each of these

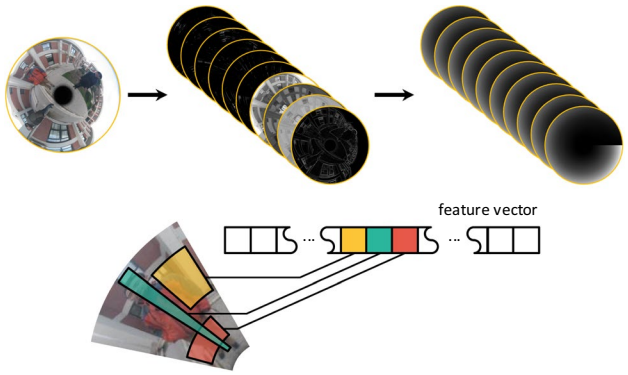

Fig. 2 Visualization of the radial integral channel features. Top: First, the input image is transformed into various channels (LUV, gradient magnitude etc). Then, radial integral images are formed. Bottom: Lastly, for a given annular sector shaped window, the sum inside the annular sectors (corresponding to a channel) is calculated. Features are shown on the original image for clarity

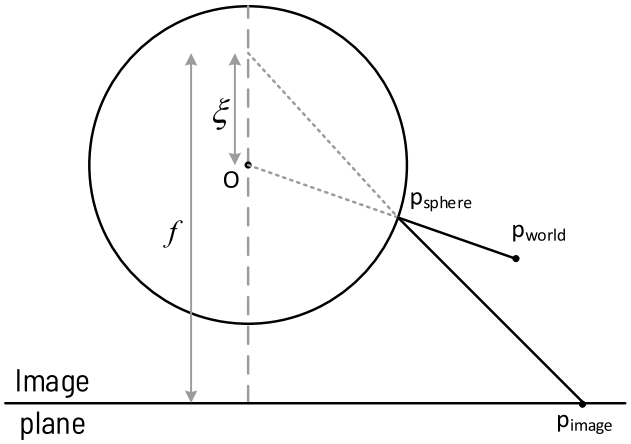

Fig. 3 In the spherical camera model, a point is projected onto the unit sphere first and then projected onto the image plane. For cameras using parabolic mirrors, $\xi = 1$

channels. These are pixel sums over rectangular regions on the channels, which can be computed rapidly using integral images. Finally, boosting is used for classification. Using shallow decision trees as weak classifiers also serves as feature selection.

ICF is not directly applicable to omnidirectional images, because the rectangular feature extraction scheme and the sliding window approach are designed for perspective images and fail for the omnidirectional image geometry. The irregular distribution of pixels in the omnidirectional image makes it impossible to apply the efficient recursive approaches used for rectangular areas [18]. Our proposed method replaces the sliding window with a rotating annular sector (ring/doughnut slice shape) as in [10], and rectangular regions with annular sectors. To calculate pixel sums inside annular sectors rapidly, we propose a novel structure, *radial integral image* (Sect. 3.3). The idea is similar to the conventional integral image, but instead of querying points in the Cartesian coordinate system, polar coordinates are used. See Fig. 2 for the illustration of the radial integral channel features method.

3.2 Camera model

In this work, we use the sphere camera model [20], according to which all central catadioptric (mirror+lens) systems can be modeled as projection to a sphere, followed by a secondary projection from the sphere surface to the image plane via a projection point.

The projection point, which acts as the camera center of a virtual camera inside the sphere, is located on the diameter that is perpendicular to the image plane and ξ units away from the center of the sphere (Fig. 3). We can assume that the sphere is a unit sphere, and by changing the position of the image plane, we can scale the image.

Let the z axis be perpendicular and pointing toward the image plane; f be the distance of the image plane to the projection point; (X, Y, Z) be the coordinates of an arbitrary 3D point in the world, denoted by p_{world}; (x, y, z) be the coordinates of p_{sphere}, which is the projection of p_{world} on the sphere, and $(x_{\text{im}}, y_{\text{im}})$ be the projection of p_{sphere} on the image. A graphical depiction is shown in Fig. 3. Then, $r = \sqrt{x^2 + y^2 + z^2}$ and the projection from world coordinates to image coordinates can be expressed as:

$$(x_{\text{im}}, y_{\text{im}}) = \left(\frac{fx}{\xi + z}, \frac{fy}{\xi + z} \right)$$

For cameras using parabolic mirrors, $\xi = 1$. In other words, the projection point is located on the sphere. This is a typical situation and also known as stereographic projection. We use a dataset collected with a parabolic mirror and use this model in the rest of the paper.

3.3 Radial integral image

An annulus is a region bounded by two concentric circles. Annular sector (a.k.a. circular ring sector or doughnut slice shape) is a cut from the annulus, which is bordered by two straight lines from its center.

Annular sector features are very similar to the rectangular features used in integral channel features (ICF) [15]. An annular sector feature is the sum of the pixels inside that annular sector. Instead of using Cartesian coordinates to obtain sums inside rectangular regions, it uses polar coordinates to obtain sums inside circular sectors. Annular

sector features can be calculated rapidly using the proposed radial integral image.

Radial integral image, $\tilde{I}$, is defined as:

$$\tilde{I}(p) = \sum_{q:\theta_q \le \theta_p, r_q \le r_p} I(q)$$

where I is the input image, p and q are pixels, θ_p and r_p are angular and radial coordinates of pixel p.

According to this definition, we can calculate the sum inside an annular sector $(\theta_{\min}, \theta_{\max}, r_{\min}, r_{\max})$ as:

$$\begin{aligned} S(\theta_{\min}, \theta_{\max}, r_{\min}, r_{\max}) = &\tilde{I}(p_{\theta_{\max}, r_{\max}}) - \tilde{I}(p_{\theta_{\max}, r_{\min}}) \\ &- \tilde{I}(p_{\theta_{\min}, r_{\max}}) + \tilde{I}(p_{\theta_{\min}, r_{\min}}) \end{aligned}$$

where $p_{\theta,r}$ is the pixel that has the polar coordinates (θ, r). See Fig. 4a for an illustration. Once the radial integral image is computed, calculating each sum has $O(1)$ complexity (four lookups and three operations).

Since a given (θ, r) usually corresponds to fractional pixel coordinates in the actual image plane, we have used bilinear interpolation to calculate the pixel values in the integral image. We have observed that using interpolation instead of rounding the coordinates is a crucial part of the sum calculation. Rounding leads to inclusion of unwanted pixel values in the sum, whereas interpolation provides a value much closer to the true sum inside the given range.

If the annular sector crosses the $\theta = 0$ angle, the sum can not be calculated directly, but it can be decomposed into two sums (Fig. 4b):

$$\begin{aligned} S(\theta_{\min}, \theta_{\max}, r_{\min}, r_{\max}) = &S(\theta_{\min}, 2\pi, r_{\min}, r_{\max}) \\ &+ S(0, \theta_{\max}, r_{\min}, r_{\max}) \end{aligned}$$

This requires looking up pixel values corresponding to $\theta = 2\pi$, which can be achieved by storing an extra row for $\theta = 2\pi$ in addition to the radial integral image.

When n denotes the number of pixels in the image, a naïve algorithm to compute the radial integral image has $O(n^2)$ complexity, because for each pixel, θ and r should be compared with a subset of pixels in the image (on average half of the pixels). In the following section, we adapt a way to compute the radial integral image in $O(n \log n)$ time.

3.3.1 Fast computation of the radial integral image

Like rectangular integral image, computation of radial integral image is a domination problem. In this application, the question is to find which other pixels are dominated, given a pixel and the non-rectangular image structure. We adopt a multidimensional divide-and-conquer approach [8] to solve this 2D domination problem. In this approach, at each recursive step, the problem of input size n is converted to two subproblems of input size $n/2$, plus a merge step that is solved in linear time. Let $T(n)$ denote the complexity of calculating the radial integral image, the corresponding recurrence becomes

$$T(n) = 2T(n/2) + O(n)$$

which is solved in $O(n \log n)$ time.

We apply this approach to our case as follows: Along with each pixel value, we store radius r (distance from image center) and angle θ. Note that, domination is computed in 2D (r, θ) space but pixels do not have a grid structure in (r, θ) space (Fig. 5a). We say that the point p *dominates* q if all the coordinates (r and θ) of p are greater than or equal to the corresponding coordinates of q. At each step, the set of all the pixels is divided into two subsets A and B using the median radius value (Fig. 5b). This means that the radius of every pixel in B is greater than the radius of every pixel in A and no pixel in A is dominating a pixel in B. The algorithm

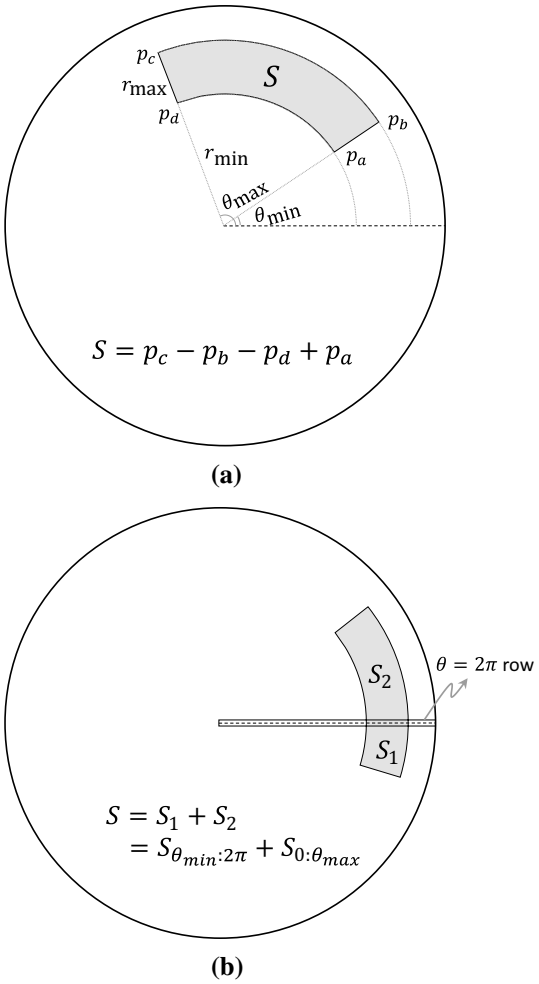

Fig. 4 **a** Illustration of rapid computation of the sum inside an annular region using radial integral image. Four lookups and three operations are sufficient to calculate the sum. **b** If an annular sector query crosses $\theta = 2\pi$ angle, it can be broken up into two annular sectors. In this case, an extra column is needed to look-up values that correspond to $\theta = 2\pi$

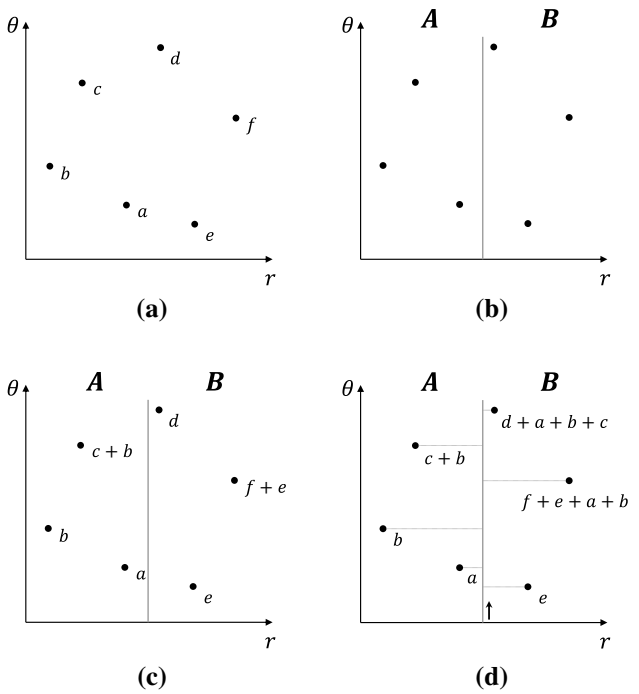

Fig. 5 Steps of computing the radial integral image. **a** Polar coordinates and values of the pixel are given as the input. **b** Pixels are split from the median radius to form two equally sized sets, A and B. **c** The algorithm is recursively called for each set. **d** Pixels are iterated according to increasing angle value, and two sets are merged by updating the values in B

is recursively called for A and B. These are the two half-size subproblems.

Upon completion of each recursive step, we are at a point that every pixel in A gives the desired sum, and every pixel in B gives the sum of the dominated pixels in B (Fig. 5c). Pixels in B might be dominating some pixels in A; thus, values in set B need to be updated. This update is the merge step of the divide-and-conquer approach. Here, we assume that pixels were sorted by θ in a pre-processing step. In this sorted list of n points (moving along vertical axis, Fig. 5d), the algorithm keeps track of the sum of values in A so far. Each time a point in B is observed, the current sum is added to that point. Therefore, the time spent for the merge operation is $O(n)$.

When the number of pixels in the set is less than a particular value, we stop the recursion and switch to a naïve implementation. This strategy avoids the recursion overhead for small sets and increases CPU cache utilization. We observed that the critical set size is 16 for our hardware architecture. We also make the implementation of the radial integral image open source.[1]

[1] The code is publicly available at https://github.com/barisdemiroz/radial_integral_image.

3.3.2 Gradient correction

It has been shown that modifying gradient magnitudes according to the Riemannian metric on the sphere improves human detection performance using omnidirectional cameras [10]. For cameras with parabolic mirror, the gradient magnitude channels are updated with:

$$|\nabla_{S^2} I| = \frac{(4 + x^2 + y^2)}{4} |\nabla_{\mathcal{R}^2} I|$$

where $|\nabla_{S^2} I|$ and $|\nabla_{\mathcal{R}^2} I|$ are the gradient magnitudes on the sphere and the image, respectively. We observe that at the center of the omnidirectional image, $(x, y) = (0, 0)$, gradients are the same. As we move from the center to the periphery of the omnidirectional image, gradients on the sphere are the magnified versions of the gradients on the image.

4 Experimental setup

In our experiments, we followed a setting that is very similar to the one described in the original ICF paper [15]. For each input image, we extracted 10 channels in total: LUV for color, gradient magnitude and gradient histogram for 6 equally spaced orientations. We generated 20,000 random features and trained an AdaBoost classifier with 1024 decision trees with a depth of two. This gives a good balance between classification accuracy and computational load.

We present three sets of experiments: For the first experiment, we have generated artificial omnidirectional images using the images in the INRIA dataset [11]. We report false positive per window versus miss rate for this experiment and show the validity of our approach. The second experiment was conducted on real omnidirectional images and compares a state-of-the-art approach with the proposed method, illustrating the improvement in accuracy and speed. Lastly, we compare the performance of our method against transforming omnidirectional images to panoramic images and using a conventional person detection method and show that working directly on omnidirectional images is a better approach.

4.1 Evaluation on the INRIA person dataset

The first set of experiments is intended as a validation of the method using artificially generated omnidirectional images. For this purpose, we have created a virtual camera, placed images such that the persons feet are on the ground plane, and projected the INRIA images onto the omnidirectional camera's image plane (Fig. 6). For each image, we have applied a random rotation to the camera around the vertical axis. We have trained our classifier with all of the 2416

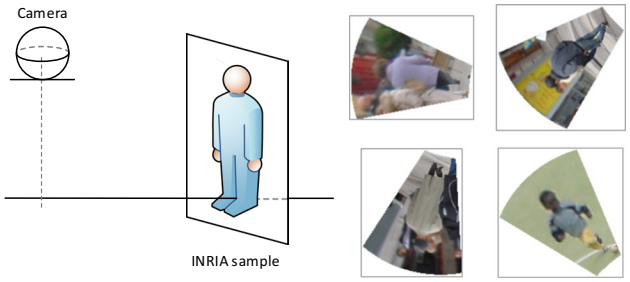

Fig. 6 Left: Each sample in the INRIA dataset is placed on the ground plane, and a virtual omnidirectional image is formed. (Camera is shown as a sphere since we used the sphere camera model.) Right: Some examples of the resulting images

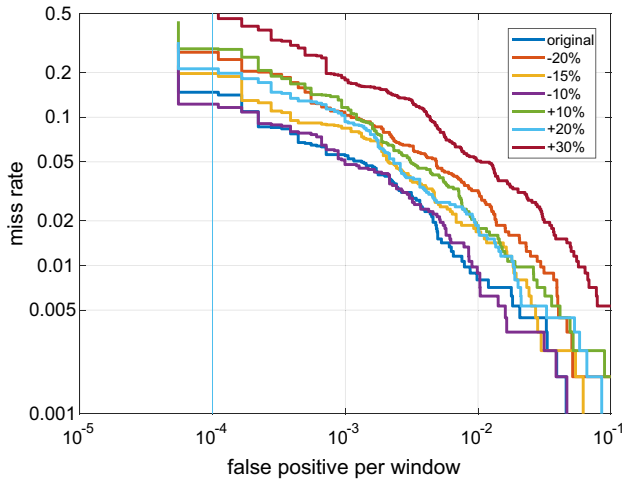

Fig. 7 The performance of the proposed method (RICF) on the INRIA dataset, where each window is transformed into an omnidirectional image. The performance decreases rapidly as the test images get further away from the camera. Best viewed in color

positive samples and 5000 random windows from negative samples from the INRIA training set.

Using these artificial omnidirectional images, we have trained a boosted classifier utilizing features extracted using radial integral images.[2] Training the final AdaBoost classifier only takes approximately 2 min on our hardware, thanks to early pruning of underachieving features [5]. For performance, we have used a custom radial integral image implementation that processes a part of the full omnidirectional image (which corresponds to the projected INRIA image). Feature extraction, training and classification require minimal resources. All of the experiments are run on a PC with an Intel i7 CPU and 4 GB RAM.

[2] The code is publicly available at https://github.com/barisdemiroz/adaboost_cpp.

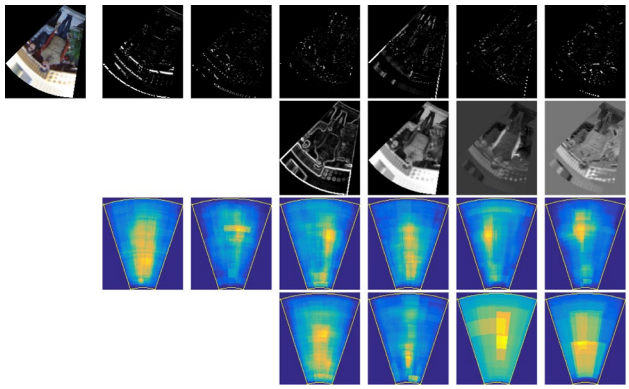

Fig. 8 Top left: An example image synthesized from the INRIA dataset. Top two rows: Channels generated from the image. Six gradient orientations (starting from zero degrees), gradient magnitude and L, U, V channels, respectively. Bottom two rows: Visualization of learned features from the training data. Each feature region is painted and averaged for visualization. Images correspond to the aforementioned channels in the same order

We plot false positive per window (fppw) versus miss rate (Fig. 7), and we observe 85.3% detection rate at the reference point of 10^{-4} fppw. Considering that the detection rate for perspective cameras is around 90% [15], this result shows that our approach is plausible for semi-synthetic omnidirectional images. In Fig. 7, we have also plotted the performance of the same classifier by changing the distance of the INRIA samples to the camera by -20%, -15%, -10%, $+10\%$, $+20\%$ and $+30\%$. We have trained a single classifier and run on the different samples without scaling the detection window or the test image. The performance decreases rapidly as the samples are placed away from the distance it was trained on. This is due to the image of the INRIA samples being significantly different at different distances and the classifier trained for a particular distance cannot be applied directly to the human shapes at different distances. For detection in real omnidirectional images, we run the detector at multiple scales to detect people at different distances (see Figs. 1 and 10). In the next section (Sect. 4.2), we report the performance on real omnidirectional images. The visualization of the channels and selected features after boosting can be seen in Fig. 8. Note that for related gradient orientations, features are clustered around shoulders. For the gradient magnitude and color channels, most of the learned features are from regions covered by the human body.

The false positive per window criterion is better suited for evaluating a binary classifier. On the other hand, for evaluating an end-to-end detection system, using false positive per image (fppi) is a better criterion, because in such a setting ideally all possible windows are considered, resulting potentially in much more false positives than the per-window approach [16]. Besides, not all detectors work on a per-window basis. Using fppi allows us to compare results of

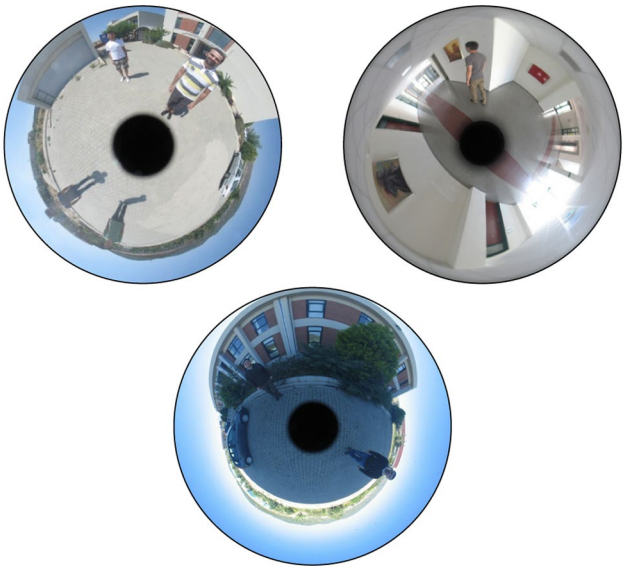

Fig. 9 Three samples from the IYTE dataset

different type of detectors. However, in this section, we used fppw because we generate human centered omnidirectional images which leaves the rest of the image unrealistic, if not empty. We report experiments using the false positive per image (fppi) criterion in Sect. 4.2, which uses real omnidirectional images.

4.2 Evaluation on the IYTE omnidirectional image dataset

For the second experiment, we have compared our method with Cinaroglu and Bastanlar's recent approach, which is the state of the art in omnidirectional person detection [10]. We trained a detector using artificial omnidirectional images using the INRIA dataset as described in the previous section. For testing, we used the same dataset[3] with [10]. This dataset contains images taken with a real omnidirectional camera (Fig. 9), and humans are manually annotated for each segmented person as annular sectors. While testing, we use a scale step of 1.04 and window step size of 6 pixels. Note that for rotating annular sectors, the window step size corresponds to varying angle step size for different radii. We have also experimented with different step sizes. We have observed that increasing step size affects the performance only slightly because, in IYTE dataset the humans are large and the detection windows are always dense enough on multiple scales to capture humans. In Fig. 1, an example result of our RICF method can be seen. In Fig. 10, example instances can be seen where our RICF method fails.

[3] IYTE dataset—available at http://cvrg.iyte.edu.tr.

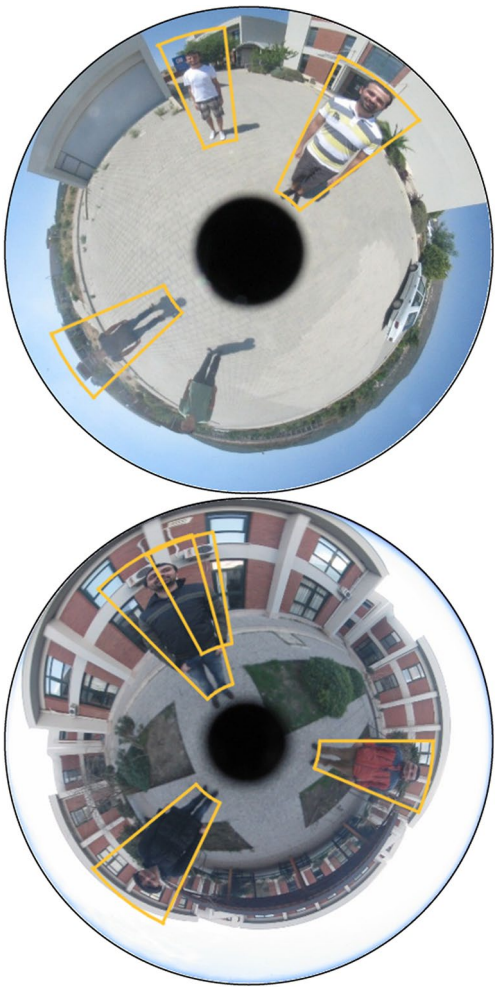

Fig. 10 Example instances where our RICF method fails. Top: false negative example. Bottom: false positive example

In [10], precision–recall curves are reported for their OmniHOG method. To keep the comparison fair, we report our results using the same metrics, namely precision and recall. We have selected a similar ratio of the negative samples as well. We also provide a false positive per image vs. miss rate plot, which is better suited for the person detection task.

The precision–recall curve shows that we have surpassed the detection performance of OmniHOG (Fig. 11a). We have also obtained 11.59% log-average miss rate on the IYTE dataset where the miss rate at fppi = 1 is 4.5% (Fig. 11b). In [17], it is reported that 10% miss rate at fppi = 1 is the best result for the datasets evaluated. Although our dataset is different and not as challenging as some of those datasets, our results show that using semi-synthetic omnidirectional images is a viable way to train person detectors for omnidirectional cameras.

Also note that, our method RICF is much faster than OmniHOG, since in OmniHOG, the transformation of HOG

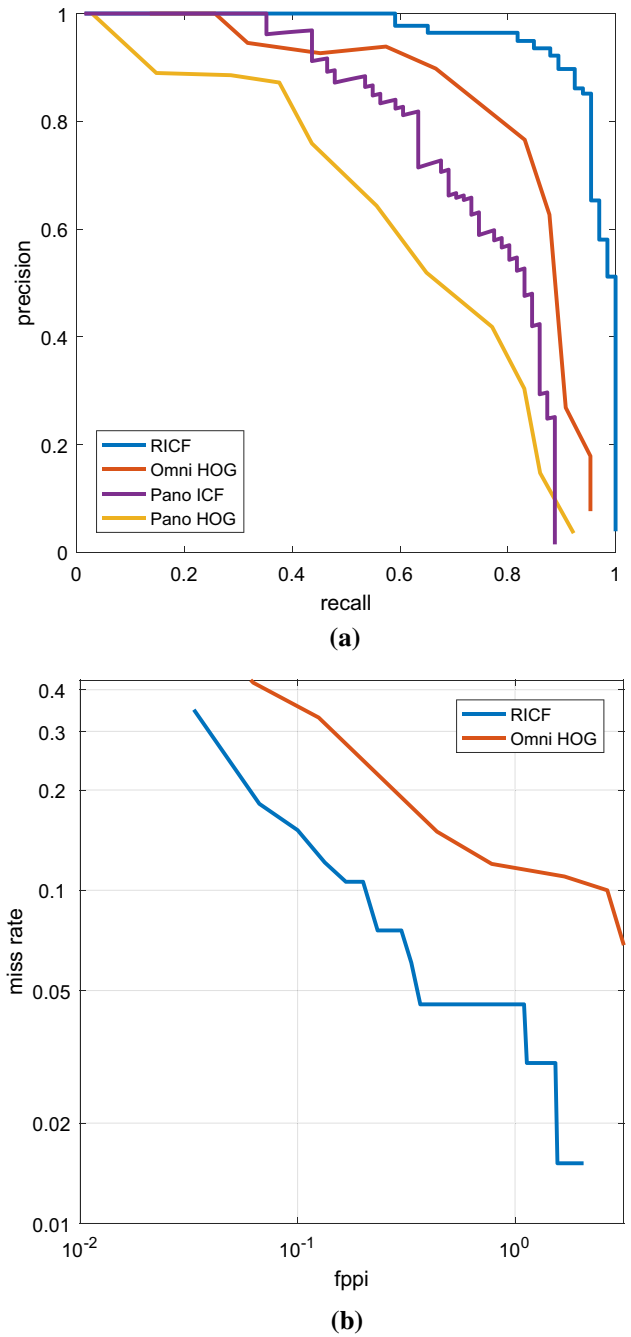

Fig. 11 Comparison of our method with [10]. Best viewed in color. **a** Working directly on omnidirectional images outperforms transforming image to panoramic and using conventional method (Pano-ICF). OmniHOG and Pano-HOG results are taken from [10], and higher precision values are better. **b** Miss rate versus false positive per image plot of our method and OmniHOG method, where lower values are better

features is done separately for each sliding-rotating window. For a given input image, OmniHOG takes about 17 ms per window, where our lightly optimized OmniIntegral implementation takes 1.6 ms per window on a similar hardware. In other words, our method is more than 10 times faster than the previous state of the art.

4.3 Evaluation on panoramic images

Lastly, we compare the performance of our method against transforming the omnidirectional image to a panoramic image and using a conventional integral image sliding window approach [15]. We chose to compare our method against this baseline approach, because this is usually the most straightforward way to work with omnidirectional images [23, 25, 38, 42].

We have used spherical projection to convert omnidirectional images to panoramic images. Spherical projection provides equiangular representation in the vertical direction of panoramic images, and it was shown to provide better performance over cylindrical projection [24]. We also set the image height so that it preserves the 2:1 aspect ratio for humans. By doing this, we actually gave the panoramic method an advantage. Nonetheless, our proposed radial integral channel features method outperformed the panoramic method (Fig. 11a). We conclude that working directly on omnidirectional images instead of transforming them into panoramic images has clear benefits for person detection. Besides, since the omnidirectional integral image is computed only once for the input image, the computational complexity of detection on the omnidirectional image is not higher than converting to a panoramic image and applying the standard perspective camera method.

5 Conclusions

In this paper, we have presented a novel method, called radial integral channel features (RICF), to detect people in images acquired by omnidirectional cameras. We have presented a new data structure called *radial integral image* to speed up feature extraction in omnidirectional images. RICF beats the current state of the art for person detection in omnidirectional cameras and demands less computational resources.

Our experiments illustrate that working directly on native omnidirectional images is better than converting them to panoramic images, followed by applying traditional approaches. The distortions caused by such rectification are problematic. If efficient native versions of useful algorithms are introduced, omnidirectional cameras will be more accessible to system developers.

Efficient omnidirectional image processing for detecting humans has great potential for many applications, including indoor scenarios such as smart environments and mobile robot-based applications, as well as outdoor scenarios, such as pedestrian detection. We believe advances such as

proposed in this paper will result in more widespread use of omnidirectional cameras.

Acknowledgements The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). This study has been funded by the Turkish Ministry of Development under the TAM Project number DPT2007K120610. Part of the work was performed when B. E. Demiröz was with NVIDIA and A. A. Salah was with Nagoya University, Future Value Creation Research Center.

References

1. Aghajan, H., Cavallaro, A.: Multi-camera Networks: Principles and Applications. Academic Press, Cambridge (2009)
2. Agrawal, M., Konolige, K., Blas, M.R.: Censure: center surround extremas for realtime feature detection and matching. In: European Conference on Computer Vision, pp. 102–115. Springer (2008)
3. Alahi, A., Boursier, Y., Jacques, L., Vandergheynst, P.: Sport players detection and tracking with a mixed network of planar and omnidirectional cameras. In: International Conference on Distributed Smart Cameras (ICDSC), pp. 1–8. IEEE (2009). https://doi.org/10.1109/ICDSC.2009.5289406
4. Alahi, A., Jacques, L., Boursier, Y., Vandergheynst, P.: Sparsity driven people localization with a heterogeneous network of cameras. J. Math. Imaging Vis. **41**(1–2), 39–58 (2011). https://doi.org/10.1007/s10851-010-0258-7
5. Appel, R., Fuchs, T., Dollár, P., Perona, P.: Quickly boosting decision trees–pruning underachieving features early. In: International Conference on Machine Learning, pp. 594–602 (2013)
6. Arican, Z., Frossard, P.: OmniSIFT: Scale invariant features in omnidirectional images. In: International Conference on Image Processing, pp. 3505–3508. IEEE (2010)
7. Benenson, R., Mathias, M., Tuytelaars, T., Van Gool, L.: Seeking the strongest rigid detector. In: Computer Vision and Pattern Recognition, pp. 3666–3673 (2013). https://doi.org/10.1109/CVPR.2013.470
8. Bentley, J.L.: Multidimensional divide-and-conquer. Commun. ACM **23**(4), 214–229 (1980). https://doi.org/10.1145/358841.358850
9. Cao, J., Pang, Y., Li, X.: Learning multilayer channel features for pedestrian detection. IEEE Trans. Image Process. **26**(7), 3210–3220 (2017). https://doi.org/10.1109/TIP.2017.2694224
10. Cinaroglu, I., Bastanlar, Y.: A direct approach for object detection with catadioptric omnidirectional cameras. Signal Image and Video Process. **10**(2), 413–420 (2016). https://doi.org/10.1007/s11760-015-0768-2
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, vol. I, pp. 886–893. IEEE, San Diego (2005). https://doi.org/10.1109/CVPR.2005.177
12. Delannay, D., Danhier, N., De Vleeschouwer, C.: Detection and recognition of sports (wo) men from multiple views. In: Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on, pp. 1–7. IEEE (2009)
13. Demiröz, B.E., Salah, A.A., Akarun, L.: Coupling fall detection and tracking in omnidirectional cameras. In: International Workshop on Human Behavior Understanding (HBU@ECCV), pp. 73–85 (2014)
14. Dollar, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. In: British Machine Vision Conference, pp. 68.1–68.11 (2010). https://doi.org/10.5244/C.24.68
15. Dollár, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: British Machine Vision Conference, pp. 1–11 (2009). https://doi.org/10.5244/C.23.91
16. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: a benchmark. In: Computer Vision and Pattern Recognition, pp. 304–311 (2009). https://doi.org/10.1109/CVPRW.2009.5206631
17. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012). https://doi.org/10.1109/TPAMI.2011.155
18. Ehsan, S., Clark, A.F., ur Rehman, N., McDonald-Maier, K.D.: Integral images: efficient algorithms for their computation and storage in resource-constrained embedded vision systems. Sensors **15**(7), 16804–16830 (2015)
19. Felzenswalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. PAMI **32**(9), 1627–1645 (2010). https://doi.org/10.1109/MC.2014.42
20. Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical implications. In: European Conference on Computer Vision, pp. 445–461. Springer (2000). https://doi.org/10.1007/3-540-45053-X_29
21. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
22. Hu, Q., Wang, P., Shen, C., van den Hengel, A., Porikli, F.: Pushing the Limits of Deep CNNs for Pedestrian Detection. IEEE Transactions on Circuits and Systems for Video Technology pp. 1–1 (2017). https://doi.org/10.1109/TCSVT.2017.2648850
23. Kang, S., Roh, A., Nam, B., Hong, H.: People detection method using graphics processing units for a mobile robot with an omnidirectional camera. Opt. Eng. **50**(12), 127204 (2011). https://doi.org/10.1117/1.3660573
24. Karaimer, H.C., Baştanlar, Y.: Car detection with omnidirectional cameras using Haar-like features and cascaded boosting. In: IEEE Signal Processing and Communications Applications Conference, pp. 301–304 (2014). https://doi.org/10.1109/SIU.2014.6830225
25. Kobilarov, M., Sukhatme, G., Hyams, J., Batavia, P.: People tracking and following with mobile robot using an omnidirectional camera and a laser. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 557–562 (2006)
26. Liu, B., Wu, H., Su, W., Zhang, W., Sun, J.: Rotation-invariant object detection using sector-ring hog and boosted random ferns. Vis. Comput. **34**(5), 707–719 (2018)
27. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37. Springer (2016)
28. Lourenço, M., Barreto, J.P., Vasconcelos, F.: sRD-SIFT: keypoint detection and matching in images with radial distortion. IEEE Trans. Robot. **28**(3), 752–760 (2012)
29. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
30. Messom, C.H., Barczak, A.L.: Stream processing for fast and efficient rotated haar-like features using rotated integral images. Int. J. Intell. Syst. Technol. Appl. **7**(1), 40–57 (2009)
31. Ouyang, W., Wang, X.: Joint deep learning for pedestrian detection. In: International Conference on Computer Vision, pp. 2056–2063 (2013). https://doi.org/10.1109/ICCV.2013.257
32. Puig, L., Guerrero, J.J.: Scale space for central catadioptric systems: Towards a generic camera feature extractor. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1599–1606 (2011). https://doi.org/10.1109/ICCV.2011.6126420
33. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
34. Saito, M., Kitaguchi, K., Kimura, G., Hashimoto, M.: People detection and tracking from fish-eye image based on probabilistic appearance model. SICE **1**, 435–440 (2011)
35. Trichet, R., Bremond, F.: LBP channels for pedestrian detection. In: Winter Conference on Applications of Computer Vision. Lake Tahoe (2018)
36. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on Riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell. **30**(10), 1713–1727 (2008). https://doi.org/10.1109/TPAMI.2008.75
37. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. Comput. Vis. Pattern Recognit. **1**, 511–518 (2001). https://doi.org/10.1109/CVPR.2001.990517
38. Wang, M.L., Lin, H.Y.: Object recognition from omnidirectional visual sensing for mobile robot applications. In: IEEE International Conference on Systems, Man and Cybernetics, October 2009, pp. 1941–1946 (2009). https://doi.org/10.1109/ICSMC.2009.5345895
39. Zhang, L., Lin, L., Liang, X., He, K.: Is faster R-CNN doing well for pedestrian detection? In: Lecture Notes in Computer Science, 9906 LNCS, pp. 443–457 (2016). https://doi.org/10.1007/978-3-319-46475-6_28
40. Zhang, S., Benenson, R., Omran, M., Hosang, J., Schiele, B.: How far are we from solving pedestrian detection? In: Computer Vision and Pattern Recognition (2016). https://doi.org/10.1109/CVPR.2016.141
41. Zhu, Q., Avidan, S., Yeh, M.C., Cheng, K.T.: Fast human detection using a cascade of histograms of oriented gradients. Comput. Vis. Pattern Recognit. **2**, 1491–1498 (2006). https://doi.org/10.1109/CVPR.2006.119
42. Zivkovic, Z., Krose, B.: Part based people detection using 2D range data and images. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 214–219 (2007)

Baris Evrim Demiröz is a Ph.D. candidate at the Department of Computer Engineering, Bogaziçi University. He is a computer vision software engineer at NVIDIA working in the autonomous vehicle group. His main research interests are object detection and tracking, discrete optimization and Bayesian inference.

Albert Ali Salah is an associate professor at the Department of Information and Computing Sciences, University of Utrecht, and at the Dept. of Computer Engineering, Bogaziçi University. His research interests are social and affective computing and computer analysis of human behavior.

Yalin Bastanlar is an associate professor at the Department of Computer Engineering, Izmir Institute of Technology, Izmir, Turkey. He received his Ph.D. degree from the Informatics Institute, Middle East Technical University, Ankara, Turkey, in 2001. His major research interests are omnidirectional vision and object detection.

Lale Akarun is a professor of computer engineering at Bogaziçi University, Istanbul, Turkey. She is the director of TETAM, the Center for Research in Informatics and Telecommunication. Her research interests are in analysis of human movements, biometrics, gesture-based interaction and sign language recognition.