

RESEARCH

Open Access

Shadow extraction and application in pedestrian detection

Junqiu Wang^{1*} and Yasushi Yagi²

Abstract

We find shadows in many images and videos. Traditionally, shadows are considered as noises because they make hurdles for visual tasks such as detection and tracking. In this work, we show that shadows are helpful in pedestrian detection instead. Occlusions make pedestrian detection difficult. Existing shape-based detection methods can have false-positives on shadows since they have similar shapes with foreground objects. Appearance-based detection methods cannot detect heavily occluded pedestrians. To deal with these problems, we use appearance, shadow, and motion information simultaneously in our method. We detect pedestrians using appearance information of pedestrians and shape information of shadow regions. Then, we filter the detection results based on motion information if available. The proposed method gives low false-positives due to the integration of different features. Moreover, it alleviates the problem brought by occlusions since shadows can still be observable when foreground objects are occluded. Our experimental results show that the proposed algorithm provides good performance in many difficult scenarios.

1 Introduction

Shadows can be found in many images and videos. Shadows are formulated when direct light from a light source cannot reach due to obstruction by an opaque object. Traditionally, shadows are regarded as noises for vision tasks such as detection and tracking. In this work, we propose a detection algorithm considering shadows as helpful information in pedestrian detection.

Pedestrian detection in images and videos is a key issue for many applications such as autonomous vehicles and visual surveillance. A large number of pedestrian detection algorithms have been proposed in recent years. Among them, various features, descriptors, and classification methods have been investigated. Despite of the good performance achieved by many detection methods, pedestrian detection is still an open problem. For example, pedestrian detection methods usually have a low detection rate when many pedestrians are occluded. Appearance variations due to viewpoint or illumination also bring problems to pedestrian detection.

We classify image features into four categories: shape, appearance, motion, and depth features. Shape features

are employed in many detection algorithms due to their invariance to viewpoint changes [1]. Shape features are very sparse in detection and modeling processes. Therefore, shape-based detection methods can be efficient. However, it is rather difficult to extract accurate shape features because of background clutters. Appearance features are successfully applied in sliding window-based detection systems [2,3]. They compute contrast information and describe such information using various descriptors. Histograms of gradient orientations (HOGs) have achieved good performance in pedestrian detection [3]. Unfortunately, it is rather difficult to detect heavily occluded objects using appearance-based methods.

In visual surveillance scenarios, stationary cameras are widely used. Background subtraction is the first step to understand the scene. Object detection and tracking can be performed based on background subtraction results. In a crowded environment, motion blobs found by comparison with the learned background is informative [4]. In this work, we also perform background subtraction in video sequences.

Although background modeling and subtraction is helpful, pedestrian detection is still difficult since subtraction result can contain many errors. Background subtraction is far from perfect: one blob in the subtraction

*Correspondence: jerywangjq@gmail.com

¹Aviation Industry Cooperation China, Beijing, China

Full list of author information is available at the end of the article

results may merge several objects; one object may be split into a few blobs. Incorrect background model updating tends to introduce incorrect models because of motion ambiguities. The problem becomes more difficult when foreground objects have similar appearance with its background.

In surveillance and many other scenarios, we can find cast shadows easily [5]. Shadows are regions where direct light cannot reach due to obstruction by an object. The space behind an opaque object is occupied by the shadow. The shape and position of the shadow are determined by the shape of the object and the position of the light sources. We can calculate the position and the shape of shadows approximately if we know the light source and the rough shape of the object. Therefore, shadows are informative in telling the existence of an object. We will detect shadows in background subtraction results based on the properties of shadows. To be specific, we compute the position of the Sun using the location and timing information. Then, we predict the orientation, position, and shape of an object in images.

Detection-based appearance information has achieved great success in the last decades. However, detection of an occluded object is still very difficult. For example, in Figure 1, there are five people in two groups in the image. The two girls in the left group have certain occlusion. Detection is still possible in this case. In the right group, the last two people are heavily occluded. It is rather difficult to reason the number and the position of the persons based on appearance information. Fortunately, the problem becomes easy if we consider the shadow information in Figure 1c. Shadow information is not noise. It is helpful in visual tasks such as tracking and detection.

The rest of this paper is organized as follows. After a brief review of important related works in Section 2, we discuss feature extraction in our algorithm in Section 3. Then, we will give the geometric transforms for feature

extraction in Section 4. Detection method using foreground and shadow information is given in Section 5. Motion filtering is described in Section 6. Experimental results on real image sequences are demonstrated in Section 7. Section 8 explains the difficulty in applying shadow information in indoor environments. Section 9 concludes this work.

2 Related work

Pedestrian detection has been intensively investigated in the last decades. Gavrila [1] proposed a shape-based pedestrian detection method. Dalal and Triggs [3] proposed an appearance-based object detection method using histogram of gradient orientations. Their approach is very effective for detecting articulated objects such as pedestrians. Tuzel et al. [6] found that covariance description has nice properties for object detection. All of the above detection methods are carried out based on appearance information only. Bertozzi et al. [7] detect pedestrians in infrared images using active contours and neural networks.

Motion information has been noticed to be helpful in detecting objects. Dalal et al. [8] normalized optical flow in video frames and applied motion information into pedestrian detection. Other works also looked at pedestrian detection from video sequences. Actually, motion information has been used in a few previous works before Dalal et al. [8]. Viola and Jones [9] detect pedestrians using patterns of motion and appearance. They model both of the motion and appearance using Haar-like features. Cutler et al. [10] proposed a detection method using long-term periodic motion information. In order to find periodicity, they analyzed long video sequences. Their system can be applied in image sequences with very low resolutions. Jones and Snow [11] extended the original pedestrian detection algorithm. They analyzed a moderate number of frames in a batch processing.

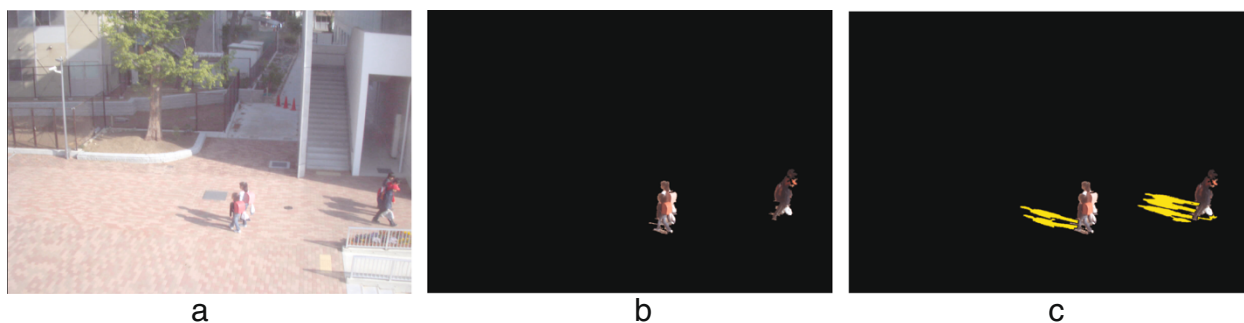


Figure 1 Shadow information is important. (a) Input image. (b) Foreground of our background subtraction result. (c) Foreground and shadows of our background subtraction result. It is difficult to discriminate how many persons in the right group in (a) with appearance information only. The shadows of the persons provide helpful information to remove the ambiguities.

Depth information has also been adopted for pedestrian detection. Gavrila and Munder [12] designed a detection algorithm for driving aid. This algorithm integrates various techniques for finding pedestrians including the use of stereo cameras. Ess et al. [13] explicitly use depth information and projected two-dimensional (2D) detection results onto the three-dimensional (3D) space.

Part-based detection can be successfully applied in pedestrian detection under the condition that the resolution of images is sufficiently high. Felzenszwalb et al. [14] presented a general object detection algorithm that is able to detect objects in partial occlusions. Lempitsky [15] applied a similar idea in object detection using HOGs.

There are some works combining detection and tracking in an integrated framework, e.g., Leibe et al. [16] presented a detection and tracking algorithm in which object detection and trajectory estimation are coupled.

We detect pedestrians using appearance information of pedestrians and shape information of shadow regions. Motion information is used if available. The focus of our work is to show the power of shadow information. Therefore, we do not combine all the features in our work.

3 Feature extraction

We model the background using Gaussian mixture models for each pixel. In addition, we also model possible shadows using available time and region information. We segment input images into three kinds of regions: foreground, shadows, and background.

3.1 Background modeling

We represent each pixel using one Gaussian mixture model. Other representations such as texture or non-parametric representations [17,18] can also be used.

We update the background using a recursive filter [19,20]. We assume that $\eta(t)$ is a learning rate set for our recursive filter. We calculate the parameter of each pixel using this learning rate:

$$\mu(t) = (1 - \eta(t)) \times \mu(t-1) + (I(t) - \mu(t-1)) \times \eta(t) \quad (1)$$

where $I(t)$ is the pixel value in the input image, and $\mu(t-1)$ and $\mu(t)$ are the mean values calculated at $t-1$ and t . Here, we set $\eta(t)$ to 0.03. The learning of other parameters follows the similar approach in [19].

3.2 Shadow detection

One of the difficulties in using shadow information is detecting shadows in input images. Invariant color properties have been used in shadow detection, e.g., normalized RGB color space and a lightness measure are employed in shadow detection [17]. Pixels with similar hue and saturation values and lower luminosity in hue-saturation-value (HSV) color space are classified as cast shadows [21].

In surveillance scenarios, an object casts shadows on surfaces. Shadow regions tend to have lower intensities due to the obstruction of the direct light source. Given a color vector without cast shadows, many shadow detection algorithms assume that the vector under cast shadows keeps the original vector direction. This assumption is not correct in outdoor environments because the ambient light source is blue. The values in different color channels are attenuated differently.

Background subtraction results include foreground objects and shadows. To separate shadows from foreground blobs, we apply a morphological close filter on background subtraction results to fill the gaps. Then, we convert the input images into HSV space which explicitly separates chromaticity and luminosity channels. A pixel in background subtraction results is considered as a possible shadow pixel when it has lower luminosity and similar hue values compared with the mode in the background model. After the classification, we calculate pixels that can be confidently classified into shadows. We use the Canny edge detection algorithm to find edges. The edges on shadow boundaries are found by comparing the hue and luminosity values. When shadows are projected on textured background, many edges are found including texture edges. The gradient orientations of such pixels are similar to those in the background model.

4 Geometry properties of shadows

Shadows are helpful for pedestrian detection. However, shadows tend to vary according to the relative position between a pedestrian and the Sun. The Sun angle varies according to timing, latitude, and aptitude of the camera [22].

4.1 Shadows in the 3D world coordinate

It is possible to infer time based on shadow direction and length. The reverse inference is much easier since we can get precise timing and location information.

The setting of the coordinates related to shadows is shown in Figure 2a. The Sun zenith angle θ_S is calculated by [22-24]

$$\theta_S = 90 - e_0 - \frac{P}{1010} \times \frac{283}{273 + T} \times \frac{1.02}{60 \tan(e_0 + \frac{10.3}{e_0 + 5.11})}, \quad (2)$$

where P is the local pressure, T the time, and e_0 the Sun's topocentric elevation angle without atmospheric refraction correction. e_0 is calculated by

$$e_0 = \arcsin(\sin \phi_o \sin \delta' + \cos \phi_o \cos \delta' \cos H'), \quad (3)$$

where ϕ_o is the observer geometric latitude calculated using the local latitude; δ' , the topocentric sun declination calculated using the geocentric sun declination from the

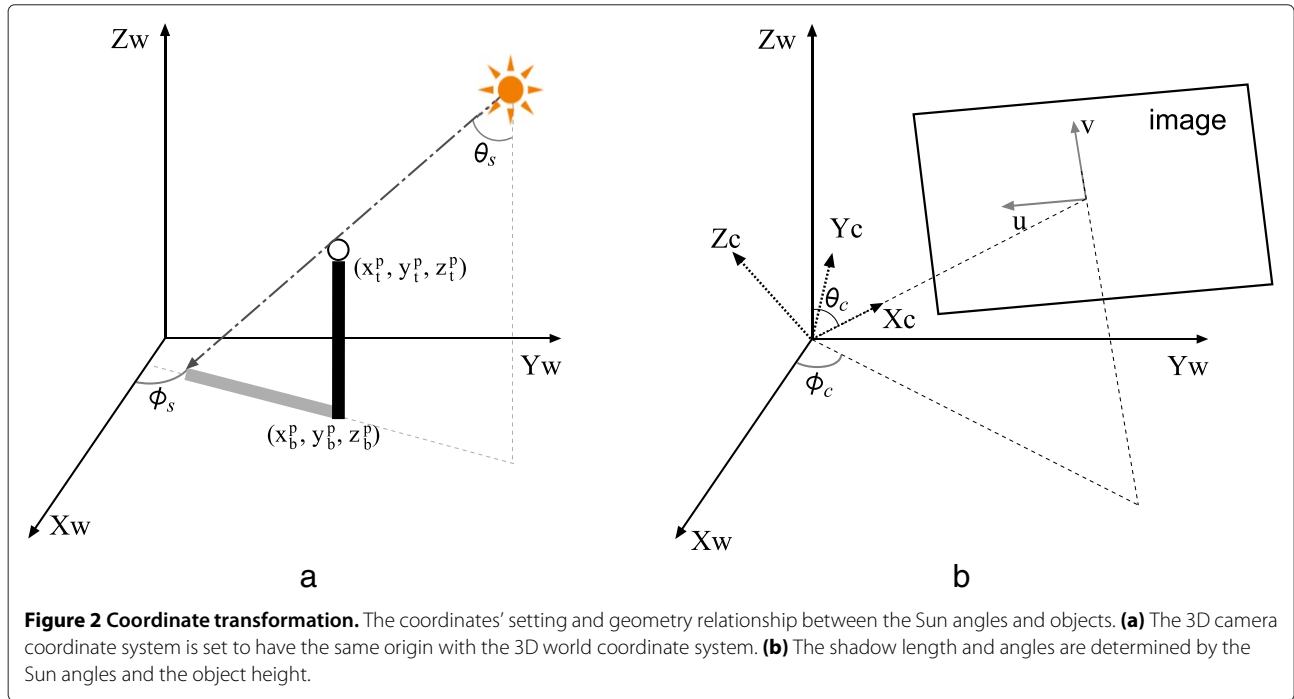


Figure 2 Coordinate transformation. The coordinates' setting and geometry relationship between the Sun angles and objects. **(a)** The 3D camera coordinate system is set to have the same origin with the 3D world coordinate system. **(b)** The shadow length and angles are determined by the Sun angles and the object height.

local longitude and current time; H' , the topocentric local hour angle from the current time.

The Sun topocentric azimuth angle is calculated using

$$\phi_S = \arctan\left(\frac{\sin H'}{\cos H' \sin \phi_o - \tan \delta' \cos \phi_o}\right) + 180. \quad (4)$$

4.2 Camera projection matrix

The setting of the coordinates of the camera is shown in Figure 2b. The projected coordinates of 3D points in the image can be obtained by multiplying its 3D coordinates with the camera projection matrix $\mathbf{u} = \mathbf{M}\mathbf{x}$. We calibrate the cameras that are used for video capture. Both camera intrinsic and extrinsic parameters are known. For single images, we obtain camera lens length from the EXIF of these images. Then, we calibrate the image using the method introduced in [25].

The camera projection matrix \mathbf{M} is calculated by multiplying the camera intrinsic matrix \mathbf{A} and the extrinsic matrix

$$\mathbf{M} = \mathbf{A} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (5)$$

where \mathbf{R} is the rotation matrix; \mathbf{t} , the translation vector. According to the setting in Figure 2, the translation vector $\mathbf{t} = [0 \ 0 \ 0]^T$.

Similarly, according to the setting of the coordinates, we calculate the rotation matrix by

$$\mathbf{R} = \begin{bmatrix} \cos \phi_C \cos \theta'_C & \sin \phi_C \cos \theta'_C & -\sin \theta'_C \\ -\sin \phi_C & \cos \phi_C & 0 \\ \cos \phi_C \sin \theta'_C & \sin \phi_C \sin \theta'_C & \cos \theta'_C \end{bmatrix}, \quad (6)$$

where $\theta'_C = \theta_C - \frac{\pi}{2}$.

We use a simple camera model which has no skewness. The pixels obtained are assumed as squares. The camera intrinsic matrix is described by

$$\mathbf{A} = \begin{bmatrix} f_c & 0 & 0 \\ 0 & f_c & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (7)$$

4.3 Shadows in images

To estimate the shape of a shadow, we need the height of the obstruction object and the Sun angle. We calculate the Sun angle based on the all sky model [22-24]. We define a world coordinate $(\mathbf{x}_w, \mathbf{y}_w, \mathbf{z}_w)$. We assume the 3D coordinates of the Sun, \mathbf{s} . The Sun position is determined by its zenith angle θ_S and azimuth angle ϕ_S . The two angles decide the shadow projection in the image. We denote the camera local frame by $(\mathbf{x}_c, \mathbf{y}_c, \mathbf{z}_c)$, which is rotated by angles (θ_C, ϕ_C) .

We calculate the length of the shadow of an object by

$$L_p^s = z_t^p \tan \theta_S. \quad (8)$$

The 3D coordinates of the shadow of the head is $\mathbf{x}_t = [x_t \ y_t \ z_t]^T$. It is calculated by

$$\begin{aligned} x_t^s &= x_b^p - L_p^s \cos \phi_S \\ y_t^s &= y_b^p - L_p^s \sin \phi_S, \end{aligned}$$

and $z_t^s = -Z_C$ since shadows are on the ground.

5 Detection

Appearance, shadow, and motion information are used simultaneously in our method. We detect pedestrians using appearance information of pedestrians and shape information of shadow regions. We also filter the detection results based on motion information if available. The flow charts of our method and typical traditional detection methods are illustrated in Figure 3.

5.1 Detection in foreground regions

We compute detection probabilities in foreground regions using appearance information. First, we train a Hough forest to model pedestrians. The Hough forest consists of many Hough trees that are efficient in matching descriptors. In testing stages, we calculate histograms of orientations for images in different scales. After that, we accumulate voting probabilities in a Hough space similar to the approach in [15]. The probabilistic formulation in [15] fits into our framework quite well.

There are many pedestrian detection approaches in the literature. We select the Hough forest due to a few merits of this approach. First, the Hough forest can detect

multiple pedestrians under heavy occlusions. According to the survey by Dollar et al. [26], occlusion is one of the major difficulties for pedestrian detection. Second, the Hough forest detection model has a probabilistic nature. It can be easily integrated with other knowledge. Pedestrian detection has been considered in a Hough-based framework using object segmentation and an MDL prior [27]. The implicit shape model (ISM) interleaves pedestrian detection and segmentation. Therefore, the probabilistic aspect of this work is not very clear because of the interleaving. DPM [14] is a multi-scale sliding window object detector. It is good at dealing with pose variations and small occlusions. It usually gets multiple overlapping detections for a pedestrian. Non-maximum suppression has to be carried out on the initial detection results. A greedy procedure is adopted in DPM for discarding repeated pedestrian detections. Some of the true detection can be eliminated in this procedure. In contrast, the non-maximum suppression in the Hough forest detector is more reasonable since it accumulates detection probabilities according to the comparison of the voting in the iterations. This strategy can lead to a good performance for occluded pedestrians. There are other approaches for pedestrian detection. However, most of them perform non-maximum suppression as the DPM method. Therefore, they are not very good at dealing with heavy occlusions. The Hough forest is better in this aspect. Our approach improves such ability by incorporating shadow information.

Let $\mathbf{g} = \{g_i\}$ be random variables describing correspondences between voting elements in the Hough spaces and hypothesis and $\mathbf{f} = \{f_{ii}\}$ be binary variables representing

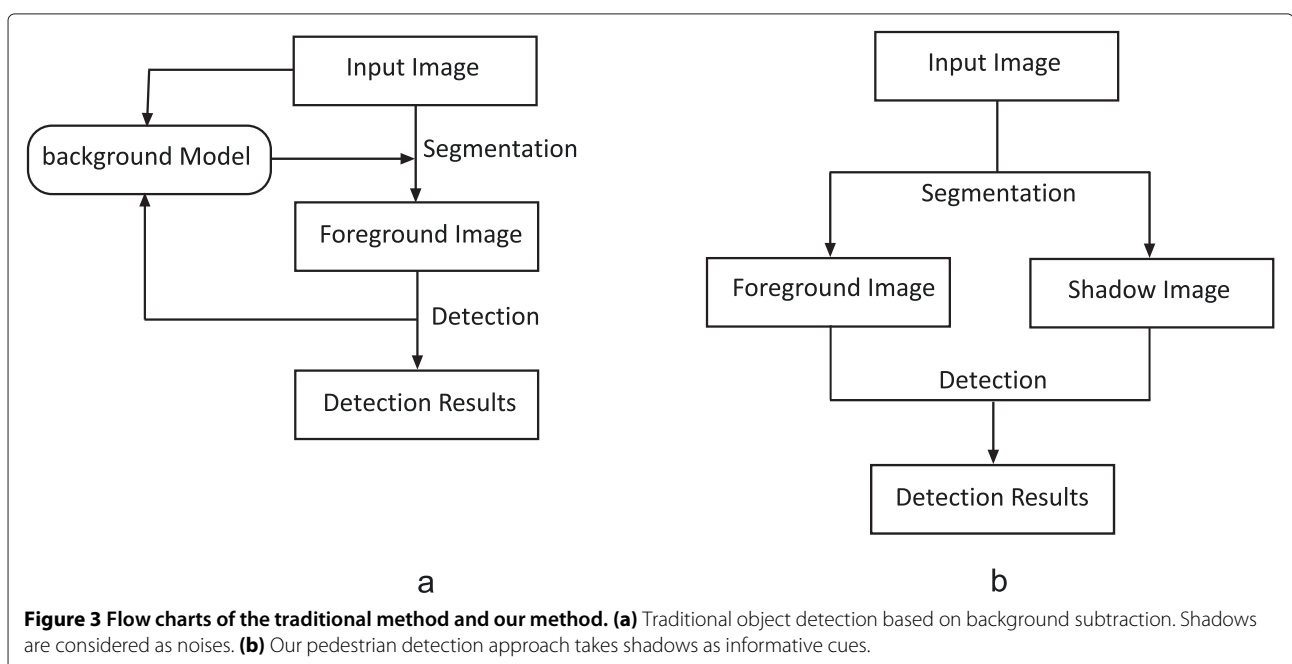


Figure 3 Flow charts of the traditional method and our method. (a) Traditional object detection based on background subtraction. Shadows are considered as noises. (b) Our pedestrian detection approach takes shadows as informative cues.

whether the hypotheses h actually correspond to a real pedestrian. We calculate

$$p(\mathbf{g}, \mathbf{f}|\mathbf{L}_A) \propto p(\mathbf{L}_A|\mathbf{g}, \mathbf{f})p(\mathbf{g}, \mathbf{f}), \quad (9)$$

where \mathbf{L}_A denotes HOG descriptors obtained from the appearance information. The details of the calculation can be found in [15].

5.2 Shape representation and matching in shadows

We construct a simple 3D model [4,28] for pedestrian detection. Since we can calculate geometry properties of shadows, we can generate specific shape templates according to different timing and locations. To be specific, we project the 3D model onto 2D space based on shadow geometries. The shape templates generated are matched with shadow silhouettes.

We perform matching based on the chamfer distance function. We match contours of the shadow mask for two reasons. First, we can use distance transform to accelerate the matching process. Shape matching is very efficient using distance transform results [29]. Second, the contours detected around shadows contain similar information with the region. We have a set of templates described by points $\Upsilon_M = \{\alpha_M^i\}_{i=1}^{N_{\Upsilon M}}$. We detect shadow boundaries consisting sets of points $\Lambda_S = \{\beta_S^i\}_{i=1}^{N_{\Lambda S}}$. We calculate the average of the minimum distances between each points of the templates and the edge detection results:

$$d(\Upsilon, \Lambda) = \frac{1}{N_{\Upsilon}} \sum_{\alpha \in \Upsilon} \min \|\alpha - \beta\|^2. \quad (10)$$

We can accelerate the matching process using a distance transform for the chamfer function. This transformation takes the set of points on the detected edges as input. The nearest boundary point to each location is calculated, and the minimum distance is assigned to the locations. The chamfer function (Equation 10) for a single template can be obtained by assigning the distance directly on the transformed results. To increase the robustness against partial occlusion, the distance is limited to a predefined threshold d_{\min} .

$$d_{\Upsilon, \Lambda} = \frac{1}{N_{\Upsilon}} \sum_{i=1}^{N_{\Lambda}} \sum_{\alpha \in \Upsilon} \min(\min \|\alpha - \beta\|^2, d_{\min}). \quad (11)$$

We define the probability of an object by

$$p(\mathbf{g}, \mathbf{f}|\mathbf{L}_S) = \exp(-\lambda \|d_{\Upsilon, \Lambda}\|) \quad (12)$$

where $\exp(-\lambda \|d_{\Upsilon, \Lambda}\|)$ considers the overlapping between the modeling and shadow regions.

5.3 Fusing detection probabilities

We fuse the detection probabilities calculated based on appearance and shadow information. We check the probabilities in Equations 9 and 12. We favor large probabilities

based on both shadow and appearance information. However, since pedestrians can be occluded in many cases, we calculate the fused probability using a maximization procedure,

$$p(\mathbf{g}, \mathbf{f}|\mathbf{L}_S, \mathbf{L}_A) = \max(p(\mathbf{g}, \mathbf{f}|\mathbf{L}_S), p(\mathbf{g}, \mathbf{f}|\mathbf{L}_A)). \quad (13)$$

When the appearance cue of a pedestrian is available and the shadow cue is unavailable (e.g., a pedestrian's shadow is in another large shadow region), the probability calculated based on the shadow cue is zero (or a very small value). We get an extremely low probability (or zero) if we use a simple product probability fusion method. In fact, we can infer the existence of this pedestrian based on the appearance cue only. The probability using a product of probabilities from the shadow and the appearance can be misleading. We meet a similar problem when the shadow cue is available and the appearance cue is unavailable (e.g., a pedestrian walks outside of the image, but his/her shadow is still in the image). The probability using a max of probabilities can detect pedestrians successfully when either an appearance cue or a shadow cues are available. We select the cue which is more informative when both of the cues are available.

In case of multiple pedestrians having overlaps in the image, the probabilities of the appearance given the state cannot be simply considered as a probability of a single pedestrian. Instead, a joint likelihood of the whole image, given all pedestrians, needs to be considered. The exact normalization of the probabilities' distributions based on appearance and shadow is not easy. We carry out the normalization using a heuristic way. First, We discard those probabilities in both distributions less than p_{TA} or p_{TS} . Then, we normalize the remaining probabilities in $[0, 1]$.

6 Filtering based on motion information

We improve detection performance by filtering detection results using motion information. In images with low resolutions, a pedestrian is roughly a blob moving following a curve. It is difficult to discriminate motion of arms in this resolution. Motion information is more complicated in high resolutions. Despite of the complexity, we found that false-positive detection results usually have different motion patterns with true-positives. Many false-positive detection results are due to the appearance similarity with the object models. However, they do not have any motion in long time durations. We apply the motion filters described in [9] in the blobs that possible hypothesis exists. We learn a pattern library off-line using real pedestrian motion. Then, we compare the motion patterns of possible detections with the modeling motion patterns. We discard those detections when they are very different from the pattern models.



Figure 4 Detection results in outdoor environments. The detection results using the part-based method [11] (a), the voting method [1] (b), and the proposed method (c).

7 Experimental results

We implemented the proposed method and tested it on a data set collected from several video sequences and single images. We have 4,230 images in the data set. Among them, we randomly select 126 images. We label the subset of the images to make the ground truth for quantitative analysis. We detect shadows in the images captured by stationary cameras using the method described in Section 3.2. We estimate shadow regions in single images using the method described in [30].

We compare our algorithm with two detection methods. The first one is a part-based detection algorithm [14] that integrates appearance and spatial information using part representation and assembly. One of the nice properties of their method is that it can detect objects in partial occlusions. The second one is the Hough transform-based method that is also very good at detecting objects in partial occlusions [15].

The qualitative analysis of our experimental results are shown in Figure 4. The detection results of the part-based [14], Hough transform-based [15], and our detectors are show in the first, second, and third columns, respectively. The detection task in the first row is not very difficult since there is almost no occlusion. However, the part-based method gives a few false-positives. The Hough transform-based method provides better results. Our detector gives better performance in this simple detection task. In the second row, one person is partially occluded. The part-based detector [14] gives correct detection results. However, it also has many false-positives. The Hough transform-based detector [15] merges the two persons into one object and provides a wrong detection result. The input image in the third row is relatively more difficult since several persons walk in a crowd. The part-based detector gives four correct detections and a few false-positives. The Hough transform-based detector

gives three correct detections and one false-positive. Our detector misses one object because the information is incomplete. The input in the fourth and fifth rows are single images. We detect shadows using the method in [30]. The motion filtering is not applicable in these images since motion information is not available. The proposed method gives fewer false-positives than the part-based and the Hough transform-based methods in the third and fourth rows. The Hough transform-based method provides a bad performance in the fourth row due to invalid scale estimation. The problem can be partially solved in our formulation because the shadows give hints for scale estimation.

We show the recall-precision curves of the three methods in Figure 5. To demonstrate the power of the shadow and motion cues, We calculate multiple ROC curves for (a) detection results with only shadow cues, (b) with motion cue, and (c) with both shadow and motion cues. We found that shadow cue plays an important role in pedestrian detection. The detection results using shadow cues are better than the results without shadow and motion cues. The part-based and Hough transform-based methods fail to achieve high recall curve values in the data set. Our detector outperforms the other two detectors because the other two detectors omit shadow and motion information. This confirms our expectation that fusing different kinds of information is important for pedestrian detection.

8 Detection using shadow information in indoor environments

We have demonstrated the power of shadow information for pedestrian detection in outdoor environments. It seems that detecting people in indoor environments is simpler. However, it is much more difficult to apply shadow information in indoor environments. The major

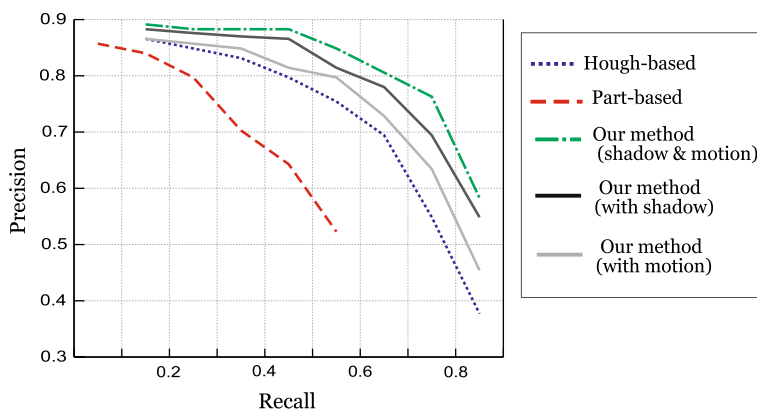


Figure 5 Detection precision-recall curves. Precision-recall curves for the proposed method (the dash-dot green curve), the improved Hough transform method (the dashed blue curve), and the part-based method (the red solid curve). Our method achieves better performance on the data set.

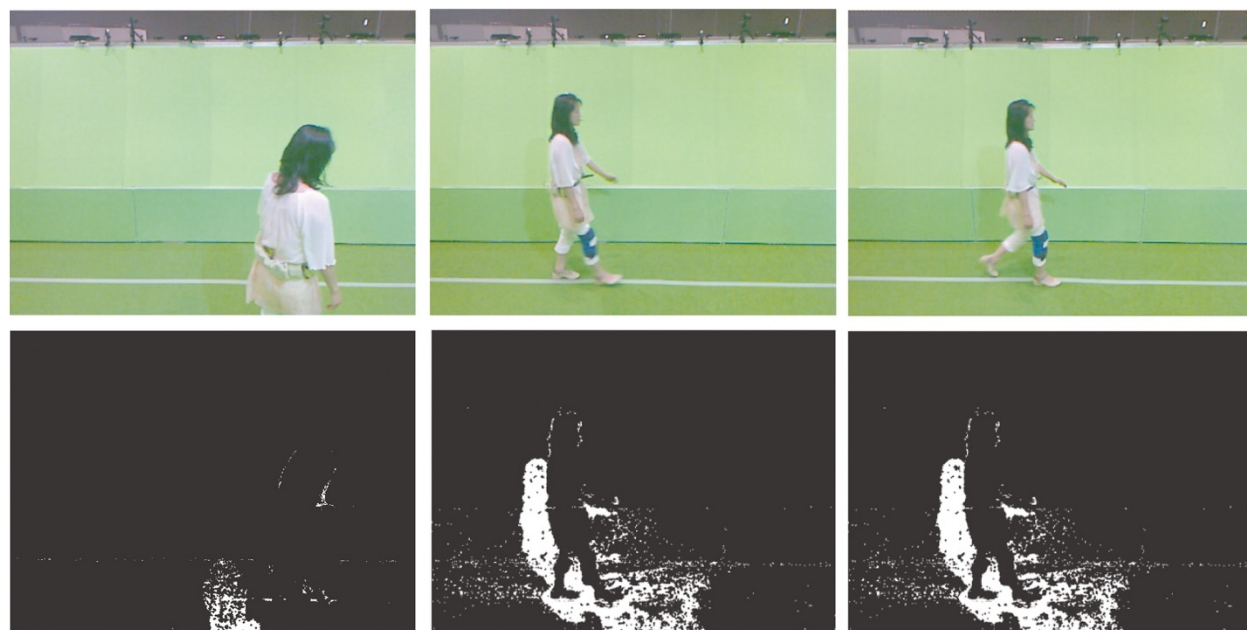


Figure 6 Shadow information in indoor environments.

difficulty for using shadow information is due to the complicated lighting conditions in indoor environments. There can be many different kinds of light sources in an indoor environment. Moreover, inter-reflections are common in indoor environments. The inter-reflections can be very strong in many cases. Due to these reasons, shadows formulated in an indoor environment can be very complicated. We add a few examples in Figure 6. In these examples, shadows are detected. However, it is not easy to apply the shadow information in the detection. The examples are captured in the environments with relatively 'simple' lighting conditions. Although we believe shadow information cannot be easily applied in object detection in indoor environments, we did not say that such application is not possible. There are two ways to solve this problem. First, we can apply shadow information in an indoor environment if the lighting conditions and the geometry are known. We use the similar strategy in the detection. Second, in most cases, there are fewer objects in indoor environments. We can claim a detection if we can relax the detection condition to 'moving object with shadows might be walking people'. Basically, detecting pedestrians in outdoor environments is more difficult in general. We introduce shadow information in the detection, which is helpful in improving the detection performance.

9 Conclusions

We show that integration of multiple cues is helpful in designing an effective object detection system. To be specific, we found that shadow information should be

considered as informative instead of noise. In addition, motion information-based filtering process finds false-positives and improves the performance of our detection system. The experimental results confirm our expectation that fusing multiple information is important for object detection.

Our method has a few limitations. First, it is rather difficult to improve its performance in overcast or raining days. Second, a pedestrian's shadow cannot be extracted reliably when his/her shadow is merged into a large shadow formulated by a large object. Third, shadow cues are not very informative when the zenith angle is very small. We consider shadow as informative features in good weather.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Aviation Industry Cooperation China, Beijing, China. ²The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan.

Received: 7 January 2013 Accepted: 31 January 2014

Published: 24 February 2014

References

1. D Gavrilu, V Philomin, Real-time object detection for smart vehicles, in *Proc. Int. Conf. Computer Vision* (IEEE, Corfu, 1999), pp. 87–93
2. P Viola, M Jones, Rapid object detection using a boosted cascade of simple features, in *Proc. of Conf. on Computer Vision and Pattern Recognition* (IEEE, Kauai, 2001), pp. 511–518
3. N Dalal, B Triggs, Histograms of oriented gradients for human detection, in *Proc. of Conf. on Computer Vision and Pattern Recognition* (IEEE, San Diego, 2005), pp. 886–893

4. T Zhao, R Nevatia, Tracking multiple humans in complex situations. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1208–1221 (2004)
5. J Wang, Y Yagi, Pedestrian detection based on appearance, motion, and shadow information, in *Proc. of Int. Conf. on Systems, Man, Cybernetics* (IEEE, Seoul, 2012)
6. O Tuzel, F Porikli, P Meer, Human detection via classification on Riemannian manifolds, in *Proc. of Conf. on Computer Vision and Pattern Recognition* (IEEE, Minneapolis, 2007), pp. 1–8
7. M Bertozzi, P Cerri, M Felisa, S Ghidoni, MD Rose, Pedestrian validation in infrared images by means of active contours and neural networks. *EURASIP J. Adv. Signal Process.* **2010**(5), (2010)
8. N Dalal, B Triggs, C Schmid, Human detection using oriented histograms of flow and appearance, in *Proc. of European Conf. on Computer Vision* (Springer, Graz, 2006), pp. 428–441
9. PA Viola, MJ Jones, D Snow, Detecting pedestrians using patterns of motion and appearance. *Int. J. Comput. Vis.* **63**(2), 153–161 (2005)
10. R Cutler, L Davis, Robust real-time periodic motion detection: analysis and applications. *IEEE Trans. Patt. Anal. Mach. Intell.* **22**(7), 781–796 (2000)
11. M Jones, D Snow, Pedestrian detection using boosted features over many frames, in *Proc. of Int. Conf. on Pattern Recognition* (IEEE, Tampa, 2008), pp. 1–4
12. DM Gavrilu, S Munder, Multi-cue pedestrian detection and tracking from a moving vehicle. *Int. J. Comput. Vis.* **73**, 41–59 (2007)
13. A Ess, B Leibe, LJV Gool, Depth and appearance for mobile scene analysis, in *Proc. of Int. Conf. on Computer Vision* (IEEE, Rio de Janeiro, 2007), pp. 1–8
14. PF Felzenszwalb, RB Girshick, McDA Allester, D Ramanan, Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010)
15. O Barinova, V Lempitsky, P Kohli, On detection of multiple object instances using hough transforms, in *Proc. of Conf. on Computer Vision and Pattern Recognition* (IEEE, San Francisco, 2010), pp. 2233–2240
16. B Leibe, K Schindler, L Gool, Coupled detection and trajectory estimation for multi-object tracking, in *Proc. of Int. Conf. on Computer Vision* (IEEE, Rio de Janeiro, 2007), pp. 1–8
17. AM Elgammal, R Duraiswami, D Harwood, LS Davis, Background and foreground modeling using non-parametric Kernel density estimation for visual surveillance. *Proc. IEEE.* **10**(7), 1151–1163 (2002)
18. AM Elgammal, D Harwood, LS Davis, Non-parametric model for background subtraction, in *Proc. of European Conf. on Computer Vision* (Springer, Marseille, 2000), pp. 751–767
19. N Friedman, S Russell, Image segmentation in video sequences: a probabilistic approach, in *Proc. 13th Conf. on Uncertainty in Artificial Intelligence* (AUAI, Providence, 1997), pp. 175–181
20. B Stenger, V Ramesh, N Paragios, F Coetzee, JM Buhmann, Topology free hidden Markov models: application to background modeling, in *Proc. of Int. Conf. on Computer Vision* (IEEE, Vancouver, 2001), pp. 294–301
21. R Cucchiara, C Grana, M Piccardi, A Prati, Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(10), 1337–1342 (2003)
22. M Blanco-Muriel, DC Alarcon-Padilla, T Lopez-Moratalla, M Lara-Coira, Computing the solar vector. *Solar Energy.* **70**(5), 431–441 (2001)
23. AJ Preetham, P Shirley, B Smits, A practical analytic model for daylight, in *Proceedings of ACM SIGGRAPH* (ACM, Los Angeles, 1999), pp. 91–100
24. I Reda, A Andreas, *Solar position algorithm for solar radiation applications*. Technical report NREL/TP-560-34302, National Renewable Energy Laboratory, USA, (2005)
25. D Hoiem, A Efros, M Hebert, Putting objects in perspective. *Int. J. Comput. Vis.* **80**, 3–15 (2008)
26. P Dollar, C Wojek, B Schiele, P Perona, Pedestrian detection: an evaluation of the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(4), 743–761 (2012)
27. B Leibe, A Leonardis, B Schiele, Robust object detection with interleaved categorization and segmentation. *Int. J. Comput. Vis.* **77**(3), 259–289 (2008)
28. T Zhao, R Nevatia, B Wu, Segmentation and tracking of multiple humans in crowded environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(7), 1198–1211 (2008)
29. K Toyama, A Blake, Probabilistic tracking in a metric space, in *Proc. of Int. Conf. on Computer Vision, Corfu* (IEEE, Corfu, 2001), pp. 50–59
30. R Guo, Q Dai, D Hoiem, Single-image shadow detection and removal using paired regions, in *Proc. of Conf. on Computer Vision and Pattern Recognition* (IEEE, Colorado Springs, 2011), pp. 2033–2040

doi:10.1186/1687-5281-2014-12

Cite this article as: Wang and Yagi: Shadow extraction and application in pedestrian detection. *EURASIP Journal on Image and Video Processing* 2014 **2014**:12.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com