

Research Article

Pedestrian Detection in Crowded Environments through Bayesian Prediction of Sequential Probability Matrices

Javier Hernández-Aceituno, Leopoldo Acosta, and José D. Piñeiro

Departamento de Ingeniería Informática y de Sistemas, Universidad de La Laguna, Avenida Astrofísico Fco. Sánchez s/n, 38204 Islas Canarias, Spain

Correspondence should be addressed to Javier Hernández-Aceituno; javier@isaatc.ull.es

Received 19 December 2014; Accepted 28 March 2015

Academic Editor: Yong Zhang

Copyright © 2016 Javier Hernández-Aceituno et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to safely navigate populated environments, an autonomous vehicle must be able to detect human shapes using its sensory systems, so that it can properly avoid a collision. In this paper, we introduce a Bayesian approach to the Viola-Jones algorithm, as a method to automatically detect pedestrians in image sequences. We present a probabilistic interpretation of the basic execution of the original tool and develop a technique to produce approximate convolutions of probability matrices with multiple local maxima.

1. Introduction

Being able to detect and avoid pedestrians is an essential feature of autonomous vehicles, if they are to guarantee a safe behavior in populated environments. However, automatically detecting human shapes in images is a very complex procedure for a computer vision system, and it has been widely studied before.

One of the most usual frameworks in literature is Viola-Jones [1], based on feature training and classifier cascades, which is explained in detail in Section 2.1. This technique has been improved by its authors by considering object motion [2, 3] and also by applying several classifiers simultaneously [4] or RealBoost to improve weak classifiers [5].

The main contributions of this paper are the introduction of a Bayesian approach to pedestrian detection methods—exemplified by, but not limited to, the Viola-Jones framework—, by creating a statistical interpretation of the basic execution of the original algorithm and developing a technique to produce approximate convolutions of probabilistic matrices with multiple local maxima. This aims to increase the precision of the framework for its usage on autonomous vehicles, in order to more efficiently detect and avoid obstacles and pedestrians in image sequences.

Furthermore, the method we present can be used with both preprocessed binary results and unaltered probabilistic

elements. As the latter are commonly returned by the sensors of a robot, this allows for greater flexibility and a more accurate management of the uncertainty of the available data.

1.1. Related Work. Another important algorithm for detecting pedestrians consists of using Histograms of Oriented Gradients (HOG) to define the features on an image [6]. This algorithm has been implemented for FPGA-based accelerators [7] and GPUs [8] and combined with Support Vector Machine (SVM) classifiers [9, 10]. Variations of histogram-based detection methods, such as Co-occurrence HOG [11] and combinations with wavelet methods [12] also exist. Bayesian methods have also been applied to the problem of pedestrian detection [13].

Both HOG and Viola-Jones algorithms are included in the official release of OpenCV [14]. Although the former usually provides very precise detection results, as studied in [15], it has been proved to perform slightly slower than the latter and is therefore less suitable for a real-time operation like pedestrian detection for a moving vehicle.

2. Materials and Methods

2.1. Viola-Jones Framework. The Viola-Jones object detection framework uses object features which, similarly to Haar-like

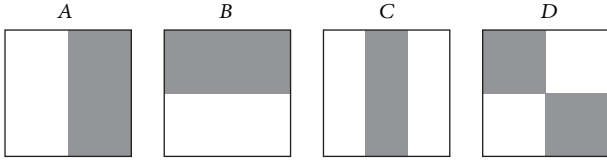


FIGURE 1: Features used by the Viola-Jones framework. The value of each feature is the sum of the pixels in the white area minus the sum of the pixels in the gray area.

features [16], are defined by additions and subtractions of the sums of pixel values within rectangular, nonrotated areas of an image. The different types of features used by Viola-Jones are shown in Figure 1.

Thanks to the usage of integral images, such that

$$J_P = \sum_{x=1}^{P_x} \sum_{y=1}^{P_y} I_{xy}, \quad (1)$$

where J is the integral of image I , these operations can be done in constant time. For example, the sum of all the pixels of the rectangle in Figure 2 would be calculated as

$$\sum_{x=A_x}^{D_x} \sum_{y=A_y}^{D_y} I_{xy} = J_D - J_B - J_C + J_A, \quad (2)$$

since each J_P value is the sum of all the pixels in the rectangle defined by the opposite corners O and P .

A set of classifiers are then trained using AdaBoost [17], and a cascade architecture allows the result to be used in real-time, by immediately discarding a sample as soon as one classifier rejects it, as shown in Figure 3.

2.2. *Bayesian Model.* Let x and z be two random variables.

- (i) x expresses the existence or absence of objects of interest (in our case, pedestrians) within an image, for each pixel location.
- (ii) z shows an equivalent value, as returned by the Viola-Jones detection when applied to an image.

It is possible to use z as evidence to evaluate the degree of belief of proposition x (i.e., $p(x | z)$), by applying Bayes' theorem:

$$\underbrace{p(x | z)}_{\text{posterior}} \propto \underbrace{p(z | x)}_{\text{likelihood}} \cdot \underbrace{p(x)}_{\text{prior}}. \quad (3)$$

The common use of a Bayesian model is to weed out wrong positive detections by comparing them to previous observations. However, when detecting pedestrians this decision could be damaging to the procedure, since false positives are preferable to false negatives, a missed detection involves immediate danger, whereas a false detection would only cause a less efficient route.

Therefore, we propose a reverse application of Bayes' theorem, which filters absences of objects rather than detections, by considering the reverse values of the presented variables:

$$p(\bar{x} | \bar{z}) \propto p(\bar{z} | \bar{x}) \cdot p(\bar{x}), \quad (4)$$

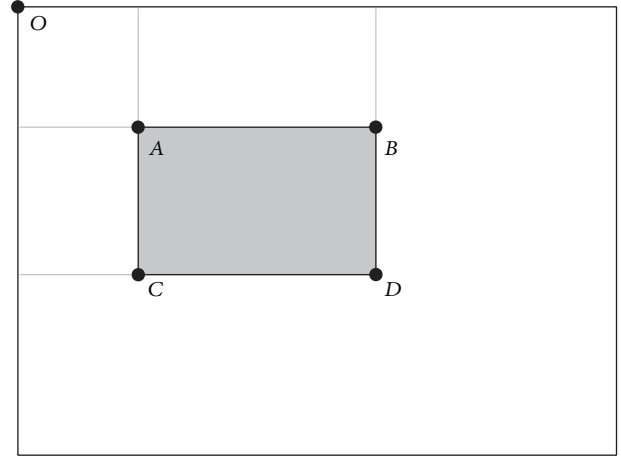


FIGURE 2: Example of a rectangle in an integral image. The sum of its pixels would be calculated as $J_D - J_B - J_C + J_A$.

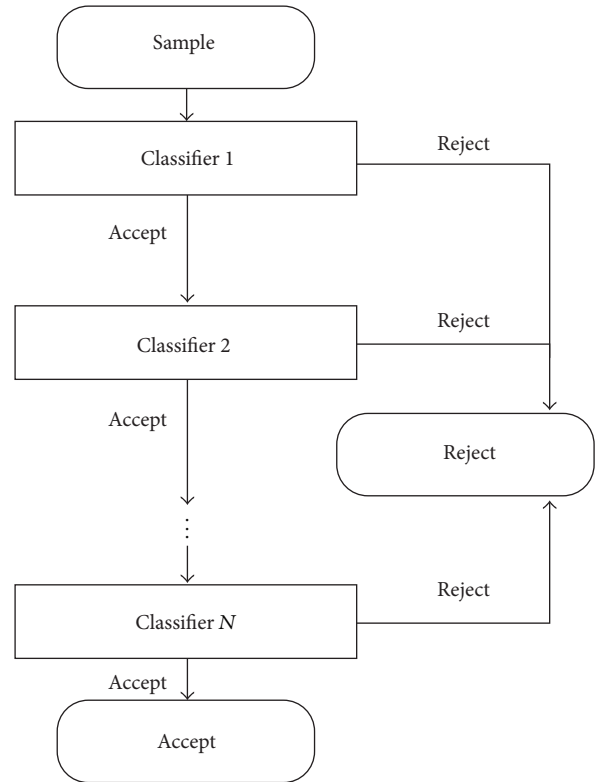


FIGURE 3: Classifier cascade architecture.

where $p(\bar{z} | \bar{x})$ and $p(\bar{x})$ are calculated as explained in the following subsections.

2.2.1. *Likelihood.* The default behavior of the Viola-Jones detection method, for a given image, is to return a set of rectangles within which objects of interest have been found.

A binary matrix can be produced from these areas, such that each cell is set to 1 if it belongs to one of them, and 0 otherwise. In our work, the binary matrix corresponding to the i th rectangle is named B_i .

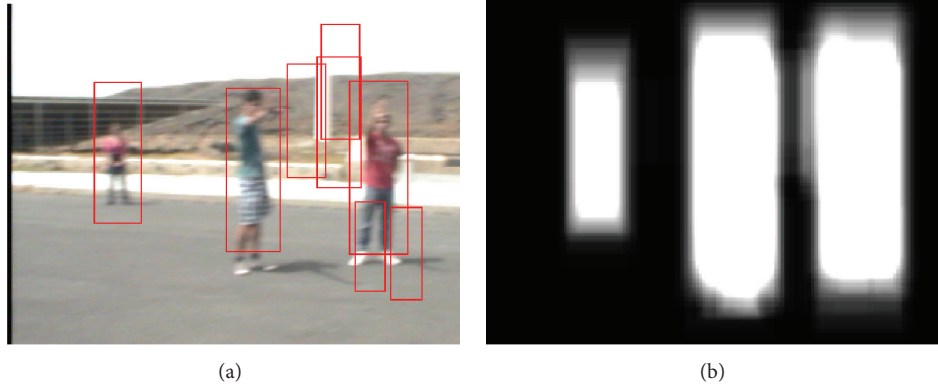


FIGURE 4: Unaltered Viola-Jones result for a minimum of three overlapping detections (a) and corresponding likelihood probability function (b). Brighter areas represent a higher probability of presence of objects.

Some of these marked areas may be superfluous (false positives), and others may overlap. The more rectangles that overlap over a group of pixels, the more likely it will be to contain an actual object of interest.

The original Viola-Jones algorithm allows for a minimum overlap restriction: a rectangle would only be valid if it can be computed as the intersection of a given number of overlapping detections.

Instead, we suggest to produce a detection matrix, such that the value of each one of its cells is equal to the number of rectangles that overlap over its corresponding pixel (Figure 4). This matrix is equal to the sum of the binary matrices of all the observed detections.

The likelihood matrix for the probability of *absence* of objects of interest within an image is proportional to the opposite value of the detection matrix; for N detections, this would be

$$p(\bar{z}_t | \bar{x}_t) \propto N - \sum_{i=1}^N B_i. \quad (5)$$

The concept of associating a weight value to each detection was also presented in the Soft Cascade method [18]. Its results are returned as rectangular areas, but unlike Viola-Jones, these are isolated and as such cannot be processed into probabilistic matrices. Preliminary tests showed that, because of this restriction, the accuracy of this technique is noticeably inferior to that of the probabilistic interpretation of Viola-Jones that we present in this work. Therefore, we chose not to use Soft Cascade in our experiments.

2.2.2. Prior. The usage of Bayes' theorem involves an evolution of the resulting posterior probability function, in order to produce the prior probability function for the following iteration of the algorithm (typically a convolution is applied).

Ideally, at each time step t , the location of an object is determined by a certain probability distribution. The distribution of the appearance of objects of interest in our experiments is extracted from the normalized addition of overlapping binary rectangular distributions, which is asymmetrical and has a flat top. A new probability distribution was developed to approximate this behavior.

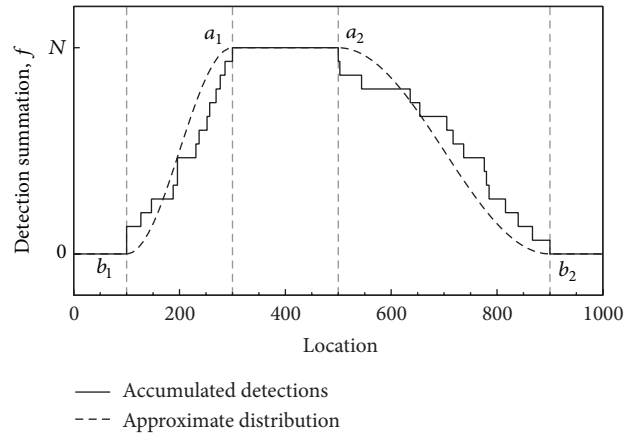


FIGURE 5: Example of theoretical cross section of the approximate probability distribution for an object of interest.

Let D be a set of detections as returned by the Viola-Jones method for a particular object of interest. An object can be represented as a $\langle |D|, A, B \rangle$ tuple, such that

- (i) $|D|$ is the number of elements in set D ,
- (ii) A is the minimal rectangle area that holds the intersection of all the elements in D , and
- (iii) B is the minimal rectangle area that holds the union of all the elements in D .

Using these data, a two-dimensional function which simulates the summation of all the elements in D was modeled:

$$F(r, \langle |D|, A, B \rangle) = \begin{cases} 0 & \text{if } r \notin B \\ |D| & \text{if } r \in A \end{cases} \quad (6)$$

$$\forall r \in \mathbb{R}^2.$$

If considering a single dimension, rectangles A and B can be seen as two segments $\overline{a_1 a_2}$ and $\overline{b_1 b_2}$, respectively, where $b_1 < a_1 < a_2 < b_2$ (Figure 5).

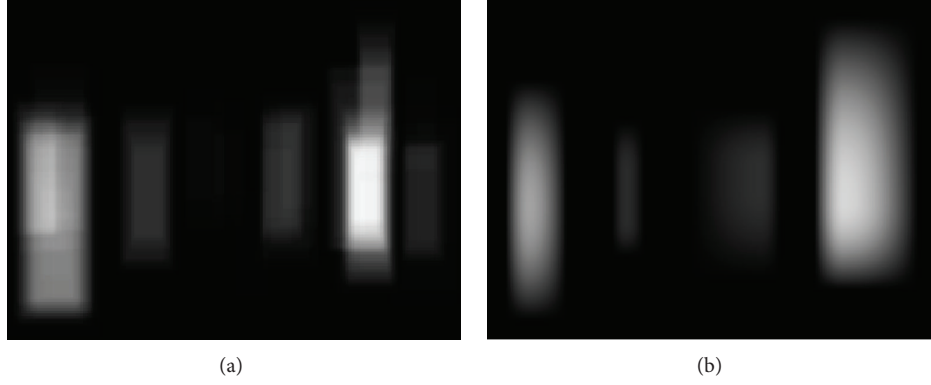


FIGURE 6: Analytical reconstruction of a probability matrix (b copies a).

Consider the following function:

$$f(c, b_1, a_1, a_2, b_2) = \begin{cases} 0 & \text{if } c < b_1 \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \cdot \frac{a_1 - c}{a_1 - b_1}\right) & \text{if } b_1 \leq c < a_1 \\ 1 & \text{if } a_1 \leq c < a_2 \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \cdot \frac{c - a_2}{b_2 - a_2}\right) & \text{if } a_2 \leq c < b_2 \\ 0 & \text{if } b_2 \leq c. \end{cases} \quad (7)$$

The shape of f suits our needs, but its height is scaled down so that, for two dimensions, the summation of the detections of a single object can be calculated as

$$F\left(\begin{bmatrix} r_x \\ r_y \end{bmatrix}, \langle |D|, A, B \rangle\right) = |D| \cdot f(r_x, b_L, a_L, a_R, b_R) \cdot f(r_y, b_D, a_D, a_U, b_U) \quad (8)$$

for $[r_x, r_y]^T \in \mathbb{R}^2$, and where a_L , a_R , a_D , and a_U are, respectively, the leftmost, rightmost, upper, and lower limits of area A , and b_L , b_R , b_D , and b_U are the corresponding limits of area B .

A probability matrix can therefore be generated, using the tuples which define the detected objects of interest. For M objects

$$P(x) \propto \sum_{i=1}^M F(r, \langle |D|, A, B \rangle_i) \quad \forall r \in \mathbb{R}^2. \quad (9)$$

In order to isolate each object of interest among the added distributions of all the detections in an image, we locate the maximum value in the probability matrix and analyze its adjacent cells to define a tuple, such that

- (i) area A contains all the cells that share a maximum probability value $|D|$, caused by the overlapping of all the involved detection rectangles, and

- (ii) area B contains all the cells that are delimited by local minima and zero values, so that we can assume that all nonzero cells that are not contained in B belong to unrelated detections.

After an object is located, its data are stored and it is removed from the probability matrix. This procedure is repeated until the matrix is empty.

Once all objects are extracted, they are matched to those of previous time steps to study their relative movement. When the objects involved are clearly individual, their movements can be analyzed and predicted separately. In our case, their number and their correspondences between frames are unknown.

Using a minimum mean square error estimation, each object is then added to a previously stored trajectory, which is used to predict new values for the following time step, using a linear regression over the tuple values.

The prediction values are finally used to generate the prior probability matrix using (9) (Figure 6).

3. Results and Discussion

Our method was tested over twelve image sequences, described in Table 1 and exemplified by Figure 7. Dataset *ETSII* was recorded in the parking lot of the Computer Engineering School of Universidad de La Laguna. Datasets *ITER1* and *ITER2* were filmed in the outer limits and in the parking lot of the Institute of Technology and Renewable Energy (ITER) facilities in Tenerife (Spain), respectively.

These three image sequences were captured by the visual sensors of the VERDINO prototype (Figure 8), a modified EZ-GO TXT-2 golf cart equipped with computerized steering, braking, and traction control systems. Its sensor system consists of a differential GPS, an Inertial Measurement Unit (IMU), an odometer, three Sick LMS221-30206 laser range finders, two thermal stereo cameras, and two Santachi DSP220x optical cameras.

Datasets *BAHNHOF*, *JELMOLI*, and *SUNNY DAY* were downloaded from Andreas Ess' *Robust Multi-Person Tracking from Mobile Platforms* website at the Swiss Federal Institute of Technology. These image sequences were recorded using a pair

TABLE 1: Features of the image datasets.

Dataset	Environment	Robot trajectory	Pedestrian behavior
(a) ETSII	Urban	Slow, straight	Static or erratic
(b) ITER1	Rural	Fast, straight	Static
(c) ITER2	Rural	Fast, erratic	Static
(d) BAHNHOF	Urban	Slow, straight	Parallel to robot
(e) JELMOLI	Urban	Fast, erratic	Several directions
(f) SUNNY DAY	Urban	Fast, straight	Parallel to robot
(g) CAVIARI	Indoors	Static	Erratic
(h) CAVIAR2	Indoors	Static	Static or erratic
(i) CAVIAR3	Indoors	Static	Static or erratic
(j) CAVIAR4	Indoors	Static	Erratic, crowded
(k) DAIMLER	Urban	Fast, erratic	Several directions
(l) CALTECH	Urban	Fast, straight	Parallel to robot



FIGURE 7: Example frames for all datasets referenced in Table 1.



FIGURE 8: VERDINO prototype.

of AVT Marlins F033C and have been used in publications [19–22].

Datasets *CAVIARI* to *CAVIAR4* belong to the *Context Aware Vision using Image-based Active Recognition (CAVIAR)* project [23] and were recorded in a shopping center in Portugal using a static camera. The selected image sequences correspond to the corridor views of clips *Walk-ByShop1* (*CAVIARI*), *OneShopOneWait1* (*CAVIAR2*), *OneShopOneWait2* (*CAVIAR3*), and *ThreePastShop1* (*CAVIAR4*).

Dataset *DAIMLER* corresponds to the Daimler pedestrian detection benchmark dataset, introduced in [24], and dataset *CALTECH* corresponds to sequence *V002* from testing set *seq06* of the Caltech pedestrian detection benchmark [15, 25]. Both datasets were recorded from a vehicle driving through regular traffic in an urban environment.

Ten tests were conducted over each image dataset; the average results are shown in Figures 10 and 9. As explained in Section 2.2, the main goal of our detection enhancement

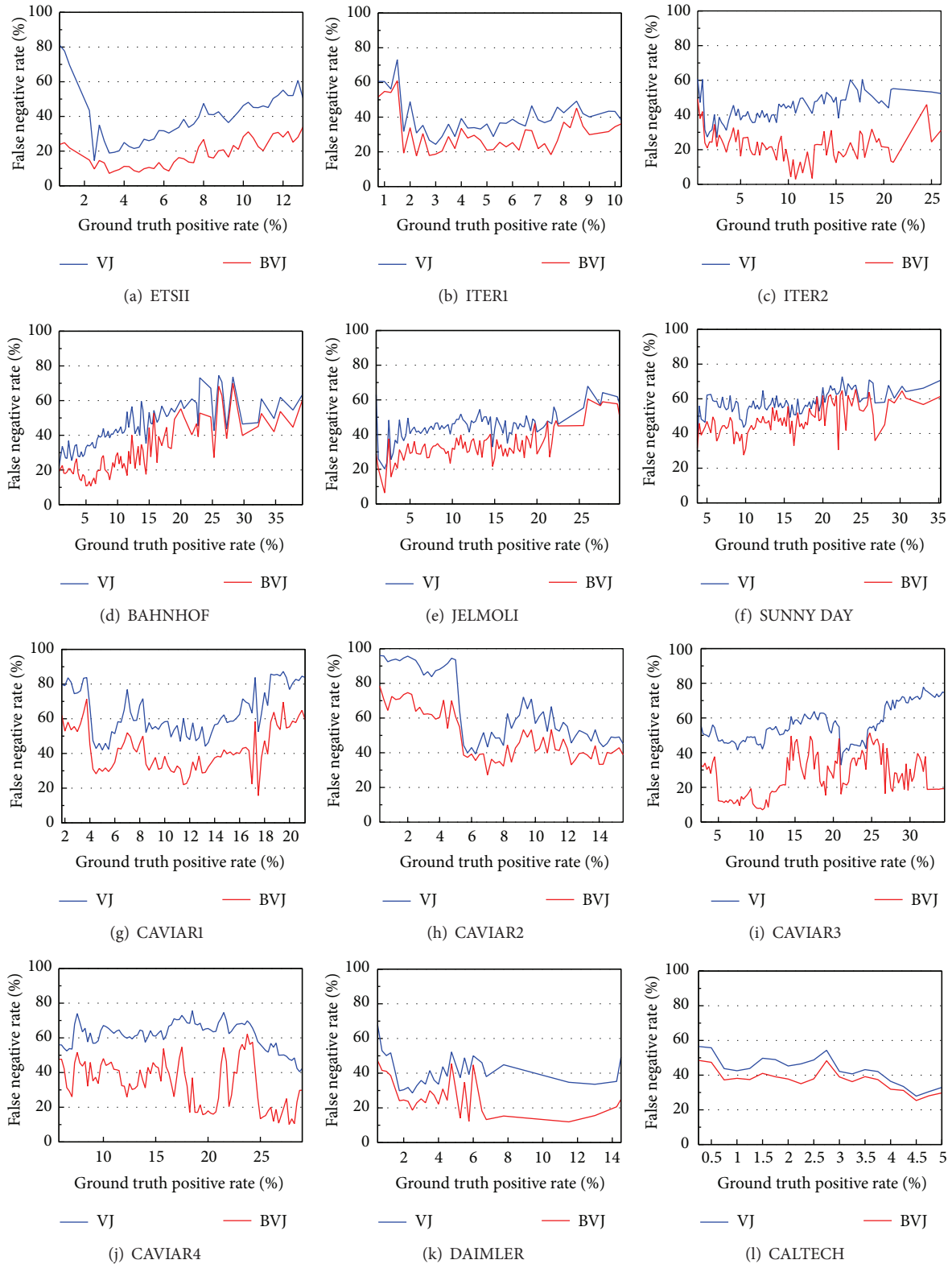


FIGURE 9: Comparison of the performances of the unaltered Viola-Jones tool and the presented Bayesian method.

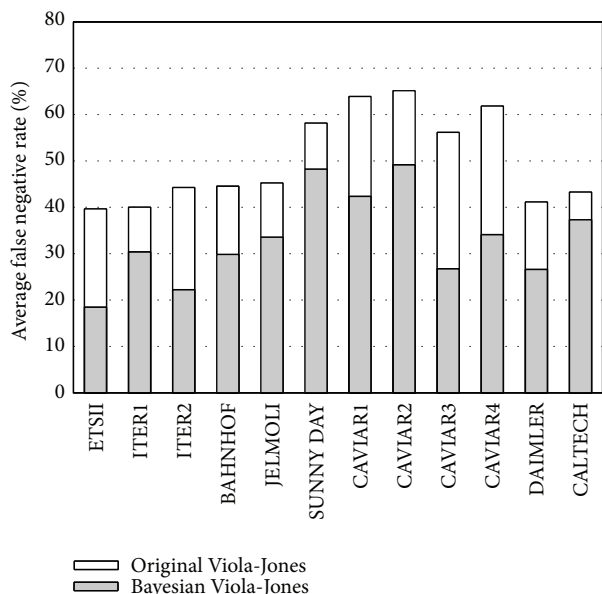


FIGURE 10: Average false negative rate for each complete image dataset.

method is to reduce the amount of false negatives returned by the Viola-Jones framework. As such, classic analysis techniques such as *receiver operating characteristic* (ROC) and *detection error tradeoff* (DET) curves, which depend on the amount of false positives of the results, do not properly display the improvement introduced by our approach. We instead present the average ratio between the amount of false positives returned by both the original and the enhanced detection methods, and the amount of true positives found in the input frames.

We observed that our Bayesian approach always provides less conservative detection rates than Viola-Jones, successfully lowering the rate of false positives for all datasets. Results were especially good for the *ETSII*, *ITER*, *CAVIAR*, and *DAIMLER* datasets. The sequences for these sets have good visibility, which results in more accurate detections by the original method and, consequently, a higher improvement introduced by our approach.

The rest of the datasets have higher occlusion rates and feature pedestrians in poses and locations that complicate their detection, thus lowering the enhancement of a Bayesian processing. This effect was especially noticeable for the *CALTECH* dataset, which features very few clearly visible pedestrians.

4. Conclusions

We have developed a Bayesian approach to the Viola-Jones detection method and applied it to a real case where pedestrians must be located and avoided by a self-guided device. Our method describes a statistical modification of the original tool, which is combined with a form of approximate convolution of two-dimensional probability matrices with multiple local maxima.

Our algorithm has been proved to improve the precision of the results, by restricting a probabilistic matrix returned by the original method to the area where objects are expected to appear, according to their previously observed movements.

It was found that our method behaves best when pedestrians are clearly visible, so that the detections by the original method can be properly enhanced by a Bayesian processing. More accurate detection algorithms are expected to improve the results of our approach in situations of high visual occlusion. This proposal serves as grounds for further research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the contribution of the Spanish Ministry of Economy and Competitiveness (<http://www.mineco.gob.es/>) under Project STIRPE DPI2013-46897-C2-1-R. Javier Hernández-Aceituno's research is supported by a FPU Grant (Formación de Profesorado Universitario) FPU2012-3568, from the Spanish Ministry of Science and Innovation (<http://www.micinn.es/>). The authors gratefully acknowledge the funding granted to the Universidad de La Laguna by the *Agencia Canaria de Investigación, Innovación y Sociedad de la Información*; 85% was cofunded by the European Social Fund.

References

- [1] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [2] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 153–161, 2005.
- [3] M. J. Jones and D. Snow, "Pedestrian detection using boosted features over many frames," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, December 2008.
- [4] T. Gao and D. Koller, "Active classification based on value of classifier," in *Advances in Neural Information Processing Systems*, vol. 24, pp. 1062–1070, 2011.
- [5] B. Rasolzadeh, L. Petersson, and N. Pettersson, "Response binning: improved weak classifiers for boosting," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 344–349, Tokyo, Japan, 2006.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, June 2005.
- [7] Y. Zhu, Y. Liu, D. Zhang, S. Li, P. Zhang, and T. Hadley, "Acceleration of pedestrian detection algorithm on novel C2RTL HW/SW co-design platform," in *Proceedings of the 1st International Conference on Green Circuits and Systems (ICGCS '10)*, pp. 615–620, June 2010.

- [8] V. Prisacariu and I. Reid, "fastHOG—a real-time GPU implementation of HOG," Tech. Rep. 2310/09, Department of Engineering Science, Oxford University, 2009.
- [9] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian detection using infrared images and histograms of oriented gradients," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 206–212, Tokyo, Japan, 2006.
- [10] M. Bertozzi, A. Broggi, M. D. Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC 2007)*, pp. 143–148, October 2007.
- [11] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," in *Advances in Image and Video Technology*, T. Wada, F. Huang, and S. Lin, Eds., vol. 5414 of *Lecture Notes in Computer Science*, pp. 37–47, Springer, Berlin, Germany, 2009.
- [12] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '2000)*, pp. 746–751, June 2000.
- [13] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 1, pp. 90–97, October 2005.
- [14] G. Bradski, "The OpenCV library," *Dr. Dobbs Journal of Software Tools*, 2000.
- [15] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: a benchmark," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 304–311, June 2009.
- [16] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian detection using wavelet templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 193–199, June 1997.
- [17] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*, P. Vitányi, Ed., vol. 904 of *Lecture Notes in Computer Science*, pp. 23–37, Springer, Berlin, Germany, 1995.
- [18] L. Bourdev and J. Brandt, "Robust object detection via soft cascade," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 236–243, June 2005.
- [19] A. Ess, B. Leibe, and L. van Gool, "Depth and appearance for mobile scene analysis," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, October 2007.
- [20] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "A mobile vision system for robust multi-person tracking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, IEEE, Anchorage, Alaska, USA, June 2008.
- [21] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "Moving obstacle detection in highly dynamic scenes," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09)*, pp. 56–63, May 2009.
- [22] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "Robust multiperson tracking from a mobile platform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1831–1846, 2009.
- [23] R. Fisher, J. Santos-Victor, and J. Crowley, "Context aware vision using image-based active recognition," EC's Information Society Technology's Programme Project IST2001-3754, 2001.
- [24] M. Enzweiler and D. M. Gavrilu, "Monocular pedestrian detection: survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, 2009.
- [25] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

