Robust System for Partially Occluded People Detection in RGB Images

Marcos Baptista-Ríos, Marta Marrón-Romera, Cristina Losada-Gutiérrez, José Angel Cruz-Lozano and Antonio del Abril

Department of Electronics, University of Alcalá, Campus Universitario s/n, Alcalá de Henares, Spain
marcos.baptista@depeca.uah.es, {marta.marron, cristina.losada}@uah.es

Keywords: People Detector, Partial Occlusion, Histogram of Oriented Gradients (HOG), Support Vector Machine (SVM).

Abstract: This work presents a robust system for people detection in RGB images. The proposal increases the robustness of previous approaches against partial occlusions, and it is based on a bank of individual detectors whose results are combined using a multimodal association algorithm. Each individual detector is trained for a different body part (full body, half top, half bottom, half left and half right body parts). It consists of two elements: a feature extractor that obtains a Histogram of Oriented Gradients (HOG) descriptor, and a Support Vector Machine (SVM) for classification. Several experimental tests have been carried out in order to validate the proposal, using INRIA and CAVIAR datasets, that have been widely used by the scientific community. The obtained results show that the association of all the body part detections presents a better accuracy that any of the parts individually. Regarding the body parts, the best results have been obtained for the full body and half top body.

1 INTRODUCTION

In recent decades, topics such as the analysis of video sequences and image interpretation have become increasingly important because of its potential applications (Poppe, 2010) as security, video surveillance (Reid et al., 2013; Arroyo et al., 2015), smart spaces or oriented marketing. More specifically, the detection of people for recognition of their activities (Martínez et al., 2016) is a topic that arouses a great interest in the scientific community. Within this context, in this paper we present a work that implements a robust solution for people detection in RGB video sequences, specifically focused on video-surveillance scenarios. This proposal has been validated using the well known CAVIAR (cav, 2005) and INRIA (Weinland et al., 2006) datasets, and the main results are shown in this paper.

There are several works in the literature whose aim is the robust detection of people in RGB images. The main works can be divided into three groups depending on the features used in order to extract information from the images. The first group includes the alternatives based on segmentation (Gu et al., 2009). These works use a prior knowledge of the background in order to separate it from the foreground corresponding to people that have to be detected. These proposals are not robust against lighting changes, dynamic backgrounds or camera movements. On the other hand, there are detectors based on the Implicit Shape Model (ISM) (Leibe et al., 2005; Seemann et al., 2005; Wohlhart et al., 2012), but these alternatives only work properly in high resolution images. Finally, there are several works based on a sliding window approach, that obtain a feature vector for a local region (window) that moves along the image. The extracted features are classified using a previously trained classifier (Papageorgiou and Poggio, 2000; Viola and Jones, 2004), in order to determinate if they correspond to a person or not.

Within the sliding window detectors, the selection of a suitable feature vector is essential in order to obtain correct results. The features used in the state of the art can be divided into different groups depending on the provided information. Thus, the main approaches are the ones based on Histograms of Oriented Gradients (HOG) (Dalal and Triggs, 2005; Zhu et al., 2006), shape features (Gavrila and Philomin, 1999; Gavrila, 2007), movement features (Viola et al., 2005), and features based on joints and human body parts. Among these features, the ones that show a better performance for people detection are the HOG (Dalal and Triggs, 2005). However, many of them only work properly if there are not partial occlusions of people to be detected. The proposal in this paper avoids this problem, improving the robustness of the people detector against partial occlusions.

In the last years, they have been published several

Baptista-RÃ■os M., MarrÃşn-Romera M., Losada-GutiÃĺrrez C., Angel Cruz-Lozano J. and del Abril A.
Robust System for Partially Occluded People Detection in RGB Images.
DOI: 10.5220/0006165005320539
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 532-539
ISBN: 978-989-758-225-7

approaches that are able to cope with partial occlusions for objects (Wohlhart et al., 2012; Bera, 2015) or people detection (Wu and Nevatia, 2005; Shu et al., 2012; Li et al., 2014; Chan et al., 2015) and tracking (Wu and Nevatia, 2006; Shu et al., 2012). These approaches are all based on parts detection using different alternatives, such as AdaBoost detectors (Wu and Nevatia, 2005), AdaBoost combined with HOG features (Li et al., 2014) or Implicit Shape Models (Wohlhart et al., 2012).

Additionally, and in order to complete the introduction of this work, it has to be pointed out that in the field of computer vision it is common to use a public dataset in order to quantitatively compare the obtained results within the different ones reached by the scientific community proposals. In this context, there are several datasets that are widely used for people detection and activity recognition, such as KTH (kth, 2011; Schuldt et al., 2004), Muhavi (Singh et al., 2010) or MSR (msr, ; Wang et al., 2012), all of them described in (Chaaraoui et al., 2012). As mentioned before, the proposal hereby presented is validated using the well known CAVIAR (cav, 2005) and INRIA (Weinland et al., 2006).

The rest of the paper is organized as follows: Section 2 presents the proposal for robust people detection; detailed experimental results are presented in Section 3; and finally Section 4 describes the main conclusions achieved.

2 ROBUST PEOPLE DETECTION

The aim of this stage is determining if there are people in the image. If so, the detector provides both the position and size in the image plane for each detected person. Two steps are included in the people detection proposal in order to accomplish the partial occlusion's robustness pursuit, both of which are described in the following sections.

2.1 Basic Detector

As it has been explained in the introduction, there are several approaches for people detection, being the most widely used those based on a sliding window and HOG descriptors (Dalal and Triggs, 2005).

The proposal in this work, also uses HOG descriptors (which have demonstrated if efficacy for people detection) and a sliding window based detector, but it has been modified in order to increase its robustness against partial occlusions. In order to do that, the proposal is composed by five different part detectors for the full body, as well as for the half top, half bottom, half left and half right parts or the body, as shown in Figure 1. Any of these detectors have the classical structure for a detector, including two different modules: the feature extraction and the classification.

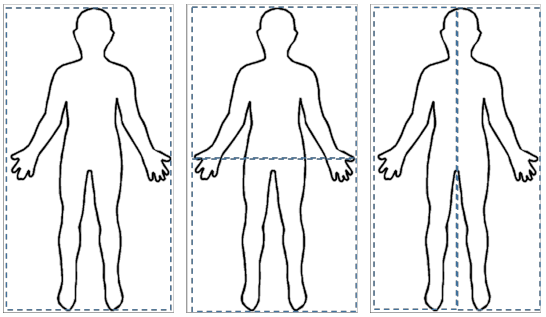

Figure 1: Full body and body parts considered in the detection stage.

The first stage of each part detector is the feature extraction. In this stage, the HOG descriptors are obtained for a given input image. This is carried out using different sizes for the sliding window, depending on their aspect ratio (related to the full body window size: 64×128 pixels). Moreover, HOG descriptors are obtained for 64 different scales of the input image, in order to detect people with different sizes in the scene.

- Full body descriptor: obtained with a 64×128 pixels window, 16×16 pixels blocks, and 8×8 pixels cells.
- Half top body descriptor: obtained with a 64×64 pixels window, 16×16 pixels blocks, and 8×8 pixels cells.
- Half bottom body descriptor: obtained with a 64×64 pixels window, 16×16 pixels blocks, and 8×8 pixels cells.
- Half left body descriptor: obtained with a 32×128 pixels window, 16×16 pixels blocks, and 8×8 pixels cells.
- Half right body descriptor: obtained with a 32×128 pixels window, 16×16 pixels blocks, and 8×8 pixels cells.

Then, a different binary Support Vector Machine (SVM) classifier is trained for each body part, using images from INRIA dataset (Weinland et al., 2006), which includes a set of 64x128 pixels images, each of them presenting a properly centered person. These images have been cropped in order to have only a body part instead of a full person. An example of the cropped images used for training is shown in Figure 2.

Regarding the SVM parameters, they have been adjusted experimentally, using k-fold cross validation. Moreover, different kernels have been consid-

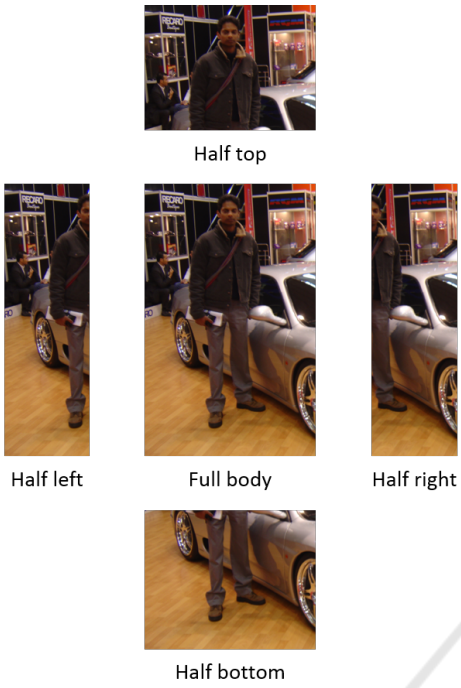

Figure 2: Example of cropped images used for the training of each body part SVM.

ered: lineal, polynomial and Radial Basis Function (RBF), obtaining results for all of them that are presented and analysed in detail in Section 3.

Once the trained has been performed, the extracted features are classified using the corresponding SVM. So, each of the SVM provides, for each detected person, its position and size (width and height) in the image, as well as the distance of the detection to the classification hyperplane, which is therefore related to the reliability of the detection. These detections are then associated using a multimodal approach, that is explained in Section 2.2. A general block diagram of the proposal, including all the explained stages is shown in Figure 3.

2.2 Multimodal Detector

The term *multimodal* is used to qualify a particular type of the global people detector architecture proposed, based on the detection of different parts of the same person. Thus, the multimodal detector groups the detections of different parts, in order to increase the robustness in detecting partially occluded persons.

To describe the multimodal detector proposed, the following definitions have to be taken into account: a set of m detection sub-windows in the image $B_r = \{\mathbf{b}_{r_i}\}_{i=0}^{m-1} = \{(u_{r_i}, v_{r_i}, w_{r_i}, h_{r_i})\}_{i=0}^{m-1}$ related to the general window $\mathbf{b} = (\frac{w}{2}, \frac{h}{2}, w, h)$, including all mentioned sub-windows i.e. $\mathbf{b}_{r_i} \subseteq \mathbf{b}, \forall \mathbf{b}_{r_i} \in B_r$, where $\mathbf{b}^* = (u^*, v^*, w^*, h^*)$ referees the size of the window $w^* \times h^*$ centered in the image point (u^*, v^*), as graphically described in Figure 4.

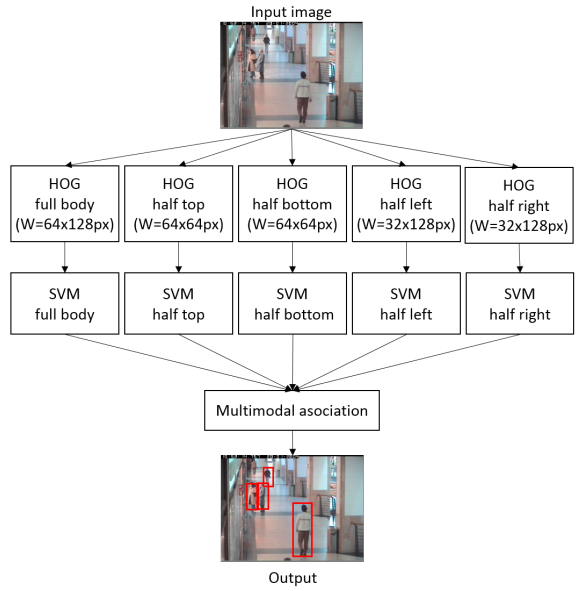

Figure 3: General block diagram of the proposal.

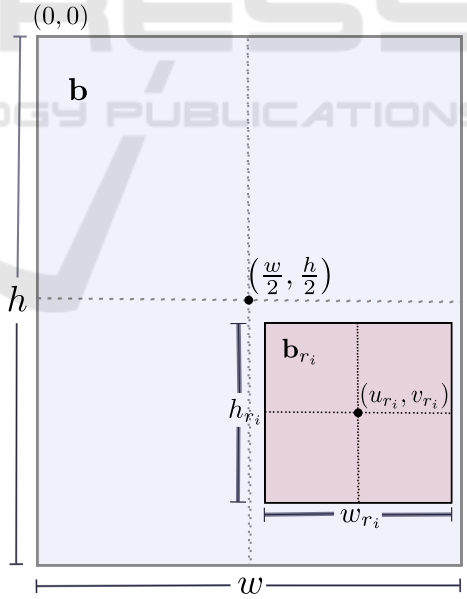

Figure 4: Geometrical relation of the i sub-window and the general window in the multimodal detection proposal, and definition of its most important terms.

Each of the k partial detections performed $\mathbf{d}_{i\,k} = (u_k, v_k, f_k : \mathbf{b}_{r_i})$ is thus related to the i classifier, and the $\mathbf{b}_{r_i}$ region. In order to perform the grouping process that allows the multimodal detection, a nor-

malised version $\mathbf{d}^n_{i\,k}$, of each $\mathbf{d}_{i\,k}$ partial detection, is obtained through equation (1).

$$\mathbf{d}^n_{i\,k} = (u_{nk}, v_{nk}, f_{nk} : \mathbf{b}) \tag{1}$$

where

$$\begin{aligned} u_{nk} &= u_k + \frac{1}{f_k}\left(\frac{w}{2} - u_{r_i}\right) \\ v_{nk} &= v_k + \frac{1}{f_k}\left(\frac{h}{2} - v_{r_i}\right) \\ f_{nk} &= f_k \end{aligned} \tag{2}$$

The grouping process proposed to obtain the multimodal associated people detector is incrementally performed from all the $i = 1,...,m-1$ partial detections in matrices $D^n_i = \{\mathbf{d}^n_{i\,k}\}^{m_i-1}_{k=0}$ in the corresponding $m\,B_r$ sub-windows. Thus a new global multimodal detections matrix D^n is created, initially being just a copy of the first partial detections one D^n_0, and incrementally including those elements from the next partial detections matrix, with low similarity $S(\mathbf{d}^n_{i\,k}, \mathbf{d}^n_h)$ with all m_T increasing number of elements already in the multimodal detections matrix $D^n = \{\mathbf{d}^n_h\}^{m_T-1}_{h=0}$, and modifying those with high similarity, till the last partial detection matrix D^n_m is reached.

The association process mentioned is based on the Hungarian Algorithm (Kuhn, 1955) and described in 1, in which the similarity function S for every $\mathbf{d}^n_{i\,k}$ with $k = 0...m_i$ detection element with all $D^n = \{\mathbf{d}^n_h\}^{m_T-1}_{h=0}$, is also defined.

3 EXPERIMENTAL RESULTS

Once the global people detector proposed is detailed presented, in this section a complete set of results and its analysis is included, in order to determine the contribution of the different parts of the proposal, i.e. partial detectors and multimodal one and their behaviour within different types of SVM classifiers.

First, the different kernels considered for the SVM classifier have been evaluated using a set of test images belonging to INRIA dataset (Weinland et al., 2006). 1 presents the precision (*ACC*) obtained using the expression in equation 3 where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives and FN is the number of false negatives.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

As it can be seen in table 1, the highest accuracy is obtained using a RBF Kernel. However, the computational cost for this kernel increases notably. For linear and polynomial Kernels, the multimodal detector has the highest accuracy. Regarding the body parts detectors, the full body and half top body detectors present a better performance than the other body parts. Once the different alternatives have been analysed, the linear kernel is selected because of its balance between the accuracy and the computational cost.

Table 1: Precision obtained for each body part detector, and the multimodal one, using different kernels in the SVM classifier, for the test set of INRIA images dataset.

ACC	Linear	Polynomial	RBF
Full body	0.9978	0.9931	1.0000
Half top	0.9850	0.9858	0.9980
Half bottom	0.9781	0.9772	1.0000
Half left	0.9844	0.9858	1.0000
Half right	0.9854	0.9850	1.0000
Multimodal	0.9982	0.9951	0.9988

The F_1 score, defined in equation 4 has also been computed, and the obtained results are shown in table 2. This metric gives information about both, the precision and recall of the classification algorithm, and as in the previous case, the body part detectors for full and half top body are the ones that present a better performance. Regarding the multimodal detector (associating information from all partial detectors) it shows a F_1 score similar to the full body detector one.

$$F_1 = \frac{2TP}{2TP+FN+FP} \tag{4}$$

Table 2: F_1 score obtained for each body part detector, and the multimodal one, using different kernels in the SVM classifier, for the test set of INRIA dataset images.

F1	Linear	Polynomial	RBF
Full body	0.9726	0.9755	0.9767
Half top	0.9382	0.9499	0.9534
Half bottom	0.8800	0.9176	0.9212
Half left	0.9178	0.9489	0.9462
Half right	0.9258	0.9468	0.9500
Multimodal	0.9686	0.9825	0.9644

The error rates of the algorithm have also been analysed bay using DET (Detection Error Trade-off) graphs, where the x-axis represents the False Acceptance Rate (*FAR*), and the y-axis represents the False Rejection Rate (*FRR*) defined in equations 5 and 6.

$$FAR = \frac{FP}{TN+FP} \tag{5}$$

$$FRR = \frac{FN}{TP+FN} \tag{6}$$

As in the previous results, the DET graphs are shown for each single detector, for the multimodal

Input : All $i = 1, \ldots, m-1$ partial detections in matrices $D_i^n = \{\mathbf{d}_{ik}^n\}_{k=0}^{m_i-1}$
$D_n = D_0^n$ **for** *i=1:m* // For all D_i^n partial detections matrices
do
  for $k = 0 : m_i$ // For every $\mathbf{d}_{ik}^n$ detection in partial detection matrix
  do
    for $h = 0 : m_T$ // Compute S with all $\mathbf{d}_h^n$ in D^n multimodal detection matrix
    do
      $S(\mathbf{d}_{ik}^n, \mathbf{d}_h^n) = e^{-\left[\left(\frac{u_{nk}-u_{nh}}{w\sqrt{f_k f_h}}\right)^2+\left(\frac{v_{nk}-v_{nh}}{h\sqrt{f_k f_h}}\right)^2+\left(\frac{f_{nk}-f_{nh}}{\sqrt{f_k f_h}}\right)^2\right]}$
    end
    $S(\mathbf{d}_{ik}^n) = \arg\max_{h \in m_T}(S(\mathbf{d}_{ik}^n, \mathbf{d}_h^n))$ //
    if $S(\mathbf{d}_{ik}^n) \geq \tau$ // Modify the similar partial detection $\mathbf{d}_{ik}$ to be included in D^n
    then
      $\mathbf{d}_h^n = \mathbf{d}_{associated\ k,h}^n = \left[\frac{u_{nk}+u_{nh}}{2}, \frac{v_{nk}+v_{nh}}{2}, \sqrt{f_{nk} f_{nh}}\right]^T$
    end
    else
      $\mathbf{d}_{h+1}^n = \mathbf{d}_{ik}^n$ // Include the dissimilar partial detection $\mathbf{d}_{ik}^n$ in D_n
      $m_T = m_T + 1$
    end
  end
end
Output: Final multimodal detection matrix $D_n = \{\mathbf{d}_h^n\}_{h=0}^{m_T-1}$

Algorithm 1: Association algorithm for the multimodal detector.

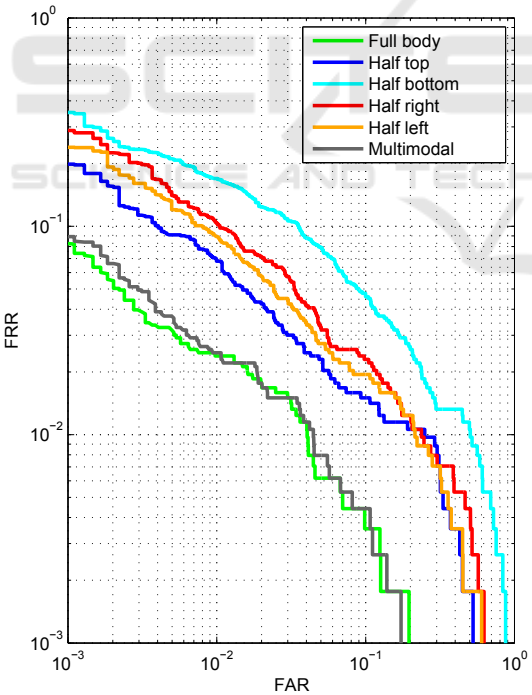

Figure 5: DET graphs for each single body part detector and the multimodal detector obtained for the lineal SVM.

detector and for each kernel: lineal (Figure 5), polynomial (Figure 6) and RBF (Figure 7). As it was expected, for the lineal and polynomial SVM, the lowest error rates are obtained for the full body detector and the multimodal one, followed by the half top body detector. On the other hand, the highest error rates correspond to the half bottom body detector. Thus, this results confirm that the full body and the half top body detectors are the part detectors that provide the highest precision.

Next, figure 8 shows the DET graph for the multimodal detector for the three kernels. In this figure, it can be observed that for the multimodal detector, the polynomial kernel is the one with the lowest error.

Finally, an example of the different stages of the detection process is shown. The presented results are obtained for an image belonging to CAVIAR dataset (cav, 2005). First, Figure 9. Figure 10 shows the detections for full body (left) and the top half body (right) detectors. The detections are then grouped obtaining the results shown in fig 11. Finally, all the results are combined using the multimodal association algorithm described in Section 2.2, in order to get the final result shown in Figure 12.

4 CONCLUSIONS AND FUTURE WORK

As stated in the introduction, the detection of people in images and video sequences is still a challenging task in artificial vision. There are, furthermore, numerous reasons for its scientific and technological

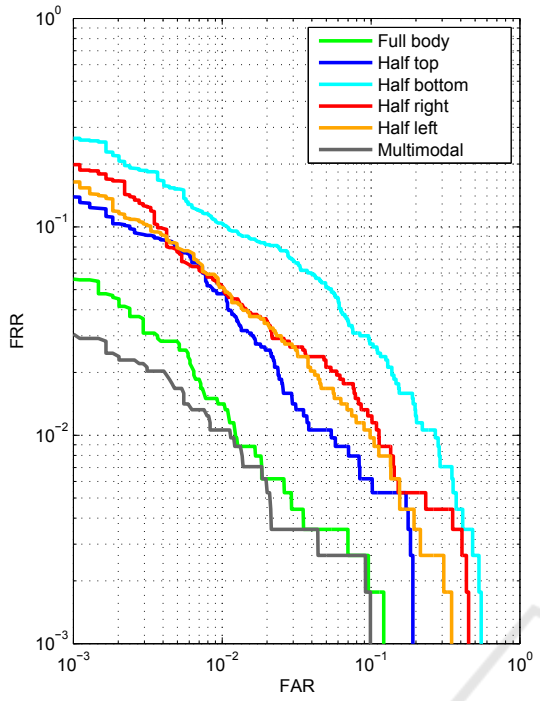

Figure 6: DET graphs for each single body part detector and the multimodal detector obtained for the polynomial SVM.

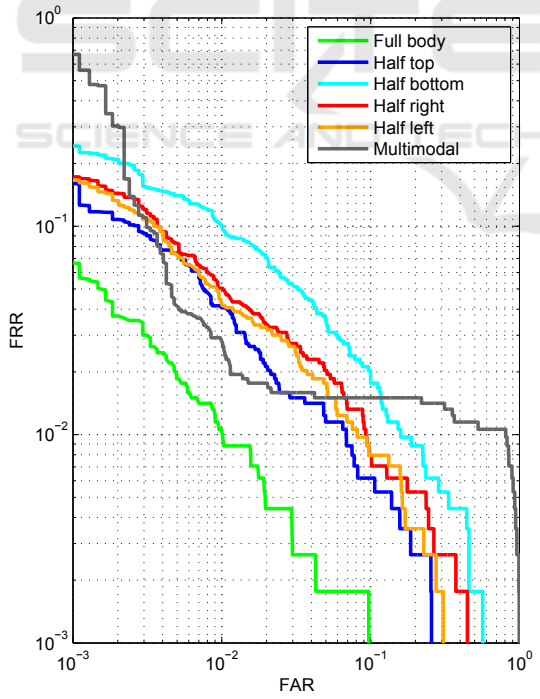

Figure 7: DET graphs for each single body part detector and the multimodal detector obtained for the RBF SVM.

interest, being the main one his big and day by day increasing applicability. Among all the possible applications for this kind of technology it is in video

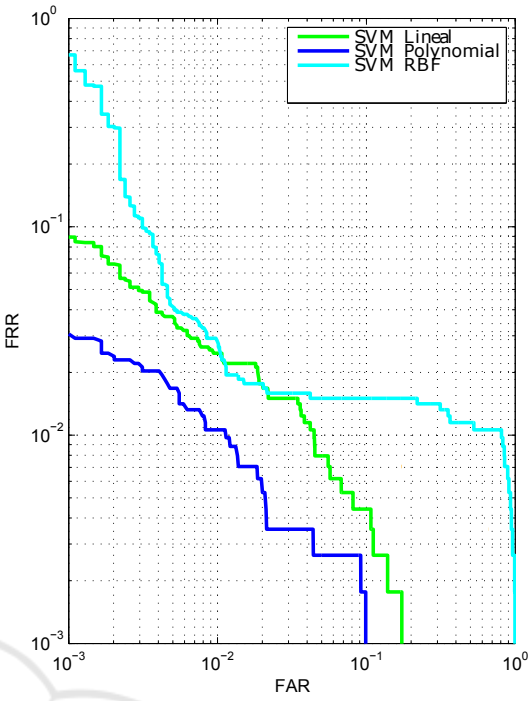

Figure 8: DET graphs for the multimodal detector for the three SVM Kernels.

Figure 9: Input image belonging to CAVIAR dataset.

Figure 10: Example of detections obtained for the full body (left) and the top half body (right) detectors.

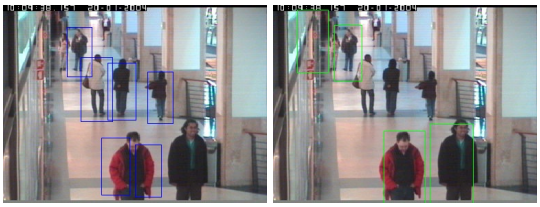

Figure 11: Example of detections obtained for the full body (left) and the top half body (right) detectors once grouped.

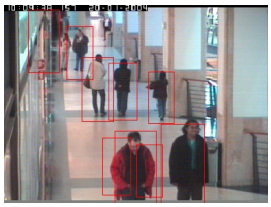

Figure 12: Example of final result obtained after combining the body part detectors.

surveillance where it presents a greater (mainly economic) interest, and thus, this is the focus of the work hereby presented.

In it, it has been evaluated the feasibility of an automatic people detection system with monocular images and without information about the background. Besides, a pretty good set of results and analysis of different and new techniques has been included.

This work also makes some important contributions that keep open some lines of research and development related to improvements in people detection systems. Within these, the following lines must be pointed out to be taken in consideration as main conclusions of the paper:

- New multimodal detectors with basic HOG descriptors. This proposal provides additional information to basic detectors that allows a more robust behaviour in complex situations.
- Mapping of reliability. In addition, this work has presented a simple way to build reliability maps of objects detection that have proven to improve surveillance task.

On the other hand, as a future work already in process, an improvement in the algorithm computational efficiency is needed to be applied to video-surveillance applications, in order to make it run in real time. With this focus in mind, the proposal presented can be easily computationally parallelized and programmable on GPU.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish Ministry of Economy and Competitiveness under projects SPACES-UAH (TIN2013-47630-C2-1-R) and HEIMDAL (TIN2016-75982-C2-1-R), and by the University of Alcalá under projects SCALA (CCG2016/EXP-010), DETECTOR (CCG2015/EXP-019) and ARMIS (CCG2015/EXP-054).

REFERENCES

MSR - Action Recognition Datasets and Codes. http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/. (Accesed July 2016).

(2005). EC Funded CAVIAR project/IST 2001 37540. http://homepages.inf.ed.ac.uk/rbf/CAVIAR/. (Accesed July 2016).

(2011). KTH-Recognition of human actions . http://www.nada.kth.se/cvap/actions/. (Accesed July 2016).

Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G., and Almazn, J. (2015). Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Systems with Applications*, 42(21):7991 – 8005.

Bera, S. (2015). Partially occluded object detection and counting. In *2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT)*, pages 1–6.

Chaaraoui, A. A., Climent-Pérez, P., and Flórez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications*, 39(12):10873 – 10888.

Chan, K. C., Ayvaci, A., and Heisele, B. (2015). Partially occluded object detection by finding the visible features and parts. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 2130–2134.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893.

Gavrila, D. M. (2007). A bayesian, exemplar-based approach to hierarchical shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1408–1421.

Gavrila, D. M. and Philomin, V. (1999). Real-time object detection for "smart" vehicles. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, volume 1, pages 87–93. IEEE.

Gu, C., Lim, J. J., Arbelaez, P., and Malik, J. (2009). Recognition using regions. In *CVPR*, pages 1030–1037. IEEE Computer Society.

Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97.

Leibe, B., Seemann, E., and Schiele, B. (2005). Pedestrian detection in crowded scenes. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 878–885, Washington, DC, USA. IEEE Computer Society.

Li, W., Ni, H., Wang, Y., Fu, B., Liu, P., and Wang, S. (2014). Detection of partially occluded pedestrians by an enhanced cascade detector. *IET Intelligent Transport Systems*, 8(7):621–630.

Martínez, C., Baptista, M., Losada, C., Marrón, M., and Boggian, V. (2016). Human action recognition in realistic scenes based on action bank. In *International*

Work-conference on Bioinformatics and Biomedical Engineering, pages 314–325, Granada.

Papageorgiou, C. and Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, 28(6):976 – 990.

Reid, D., Samangooei, S., Chen, C., Nixon, M., and Ross, A. (2013). Soft biometrics for surveillance: an overview. *Machine learning: theory and applications. Elsevier*, pages 327–352.

Schuldt, C., Laptev, I., and Caputo, B. (2004). Recognizing human actions: A local svm approach. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03*, ICPR '04, pages 32–36, Washington, DC, USA. IEEE Computer Society.

Seemann, E., Leibe, B., Mikolajczyk, K., and Schiele, B. (2005). An evaluation of local shape-based features for pedestrian detection. In *Proceedings of the British Machine Vision Conference*, pages 5.1–5.10. BMVA Press. doi:10.5244/C.19.5.

Shu, G., Dehghan, A., Oreifej, O., Hand, E., and Shah, M. (2012). Part-based multiple-person tracking with partial occlusion handling. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1815–1821. IEEE.

Singh, S., Velastin, S. A., and Ragheb, H. (2010). Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 48–55. IEEE.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.

Viola, P., Jones, M. J., and Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161.

Wang, J., Liu, Z., Wu, Y., and Yuan, J. (2012). Mining actionlet ensemble for action recognition with depth cameras. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1290–1297.

Weinland, D., Ronfard, R., and Boyer, E. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 104(2):249–257.

Wohlhart, P., Donoser, M., Roth, P. M., and Bischof, H. (2012). Detecting partially occluded objects with an implicit shape model random field. In *Asian Conference on Computer Vision*, pages 302–315. Springer.

Wu, B. and Nevatia, R. (2005). Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 90–97. IEEE.

Wu, B. and Nevatia, R. (2006). Tracking of multiple, partially occluded humans based on static body part detection. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 951–958. IEEE.

Zhu, Q., Yeh, M.-C., Cheng, K.-T., and Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1491–1498. IEEE.