Proc. of the Third Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering- ABBE 2015

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

Bio-Inspired Pedestrian Detection and Tracking

Miguel Farrajota, João M.F. Rodrigues and J.M.H. du Buf

***Abstract*— Pedestrian detection and tracking remains a popular issue in computer vision, spawning many applications in robotics, surveillance and security, biometrics and human-computer interaction. In this paper we present a biological framework for detecting and tracking pedestrians by using a monocular moving camera. This framework is based on visual cortex cells, namely complex and end-stopped cells, the last being used to extract keypoints. By employing a modified HOG descriptor combined with the responses of complex cells and a linear SVM, pedestrians can be detected. By combining the above information with keypoints, motion information and tracked features, persons can be tracked in complex scenarios where partial occlusions exist.**

***Keywords*— Cortical cells, tracking, keypoints, pedestrians.**

I. Introduction

Pedestrian detection is a very challenging task due to the large variability caused by different clothing and poses, abundant partial occlusions, complex/cluttered backgrounds and frequent changes in illumination. If it is also pretended to track persons with a moving camera in a dynamic environment, the problem is even more difficult because of the combined effects of scene activity and egomotion.

In recent years, considerable progress in the development of approaches and applications has been obtained concerning object detection and class-specific segmentation in tracking scenarios, pedestrian detection being of particular interest [5][13][15]. Several of those approaches have used either global models with full-body appearance [18], silhouettes [10] and assembly of local features [27], or part detectors [16]. Many descriptor-based detectors have been used in this problem. Among the popular ones is the Dalal and Triggs Histograms of Oriented Gradients (HOG) descriptor [3], which extends the idea of the popular local Scale Invariant Feature Transform (SIFT) descriptor [14] to represent entire objects.

Other authors have proposed additional features to improve the visual representation of the descriptor, such as the use of color through self-similarity features (CSS) [28], texture through block-based Local Binary Patterns (LBP) [29], and the design of efficient gradient-based features via integral channels [4]. On the other hand, cortical cells [19] have been used for several applications [8][21][23][24], including face [19][20] and hand [22] detection.

Miguel Farrajota, João M.F. Rodrigues and J.M.H du Buf
Vision Laboratory, LARSyS (ISR – Lisbon),
FCT, ISE, University of the Algarve
Faro, Portugal

This paper presents an extension of a biologically inspired model based on cortical cells to pedestrian detection and tracking in crowded environments with partial occlusions and using a moving camera. The model employs responses of cortical complex cells as HOG-like descriptors to code a pedestrian's shape, within a detection window to be classified as either pedestrian or non-pedestrian with a linear SVM. After detection, the detection window and features (keypoints) from detected pedestrians are tracked using the camera's and keypoints' motions during a few frames. This two-step process helps to reduce the tracking problems due to classification errors of the classifier, and alleviates the computational burden of detecting pedestrians in every frame, speeding up the overall process. The proposed method was trained and evaluated on the well-known INRIA pedestrian dataset [3], and compared to Dalal and Triggs' popular HOG-based pedestrian detector for performance comparison.

The main contribution of the paper is the bio-inspired model, i.e., the model for the detection and tracking of pedestrian in crowded environments.

II. Model overview

The model is divided into two main steps: detection and tracking. For detection, HOG-like features based on cortical complex cell responses [19] are used and coded in a multiscale fashion, and then classified by a linear SVM inside a detection window. A non-maximum suppression (NMS) filter is applied to discard multiple detections. For tracking, the detection windows with several scales are tracked by estimating the windows' trajectories. This is achieved by estimating the pedestrians' motions using keypoints extracted from cortical cells inside the detection windows (regions) and by eliminating the camera's egomotion effect, also by projecting a window's location in a frame to the next one.

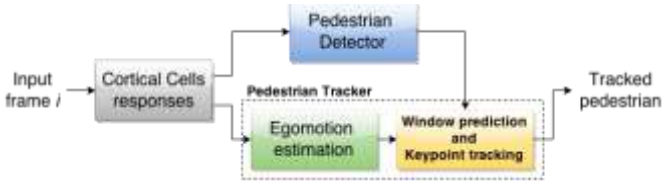

Figure 1. Model overview. The detection and motion estimation blocks are independent from each other, and later combined for tracking.

Figure 1 illustrates the model scheme. At the top, the pedestrian detector consists of three main consecutive steps: (a) HOG-like feature detection; (b) classification with a linear SVM; and (c) NMS filtering in detection windows (Pedestrian Detector block). At the bottom, the tracker is composed of four processing steps: (a) keypoint extraction and (b) matching, (c) egomotion estimation (Egomotion estimation

Proc. of the Third Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering- ABBE 2015

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

block) and (d) window prediction and keypoint tracking. Note that the pedestrian detector and egomotion estimation blocks are independent, which permits parallel processing to speed up the system. The diagram blocks will be explained in detail in the following sections.

Cortical cell responses: Responses of cortical cells provide the basis of the model. The principle for multi-scale processing is based on Gabor quadrature filters which provide a model of cortical simple cells [19]. In the spatial domain (x, y) they consist of a real cosine and an imaginary sine, both with a Gaussian envelope. Responses of even and odd simple cells, which correspond to real and imaginary parts of a Gabor filter, are obtained by convolving the input image with the filter kernel, and are denoted by $R^E_{s,i}(x, y)$ and $R^O_{s,i}(x, y)$, s being the scale, i the orientation ($\theta_i = i\pi/N_\theta$) and N_θ the number of orientations (here 8) with $i = [0, N_\theta - 1]$. Responses of complex cells are modeled by the modulus $C_{s,i}(x, y)$. In addition, there are two types of end-stopped cells, single and double. These are applied to $C_{s,i}$ and combined with tangential and radial inhibition schemes in order to obtain precise keypoint maps $K_s(x, y)$. For a detailed explanation with illustrations see [19]. Here, only one Gabor filter scale (s) is used, which is given by $\lambda = 5$, λ being the spatial wavelength of the Gabor filters expressed in pixels. Figure 2 (middle) illustrates the responses of complex cells in 8 orientations.

III. Pedestrian Detection

As mentioned above, detection of pedestrians is achieved by coding responses of cortical complex cells within a region into a feature vector, using a classifier (linear SVM) to predict if a pedestrian is inside the detection region or not. The detection process employs a single scale ($\lambda = 5$), but several scales of the HOG-like features (several sizes) over the entire image, and then a sliding window (7×18 blocks) to scan all blocks inside its region. At each layer, the pooling cell size is increased, but the detection window size and the block's size remain the same. To this purpose, we use between 6×6 and 20×20 pixels per cell with a stride of 2 pixels. Finally, NMS filtering is applied to eliminate multiple detections of the same person. By coding the entire image as HOG-like features but using only a single scale, a quick pedestrian search is performed with a small loss in accuracy (less than 5%).

We use the INRIA dataset to train our detector and to evaluate its performance. Although this dataset is rather small, it is widely used in the literature [5] and for comparing the performance of different algorithms. The dataset comprises 2416 positive samples for training and 1126 positive samples for testing. For negative samples, a set of 12180 patches randomly sampled from a 1218 person-free training image dataset provided the initial negative set for training, having an additional set comprised of 1126 positive images and 4530 negative samples from 453 images for testing. We bootstrap the detector with additional 10K random false positives as hard negatives to train the final detector.

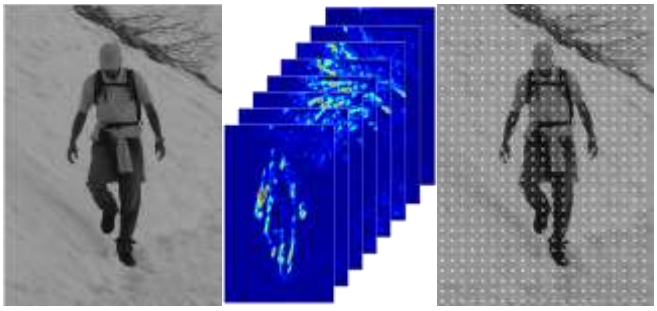

Figure 2. A pedestrian input image in grayscale (left) coded by complex cell responses in 8 orientations (middle), which are then combined into HOG-like features (right).

HOG-like features: HOG features have clear advantages for object recognition [7]. A modified version of Dalal and Triggs HOG features [3] is used here for pedestrian detection. In their implementation [3] they used an RGB color space with no gamma correction, 1D gradient filters, linear gradient voting into 9 orientation bins, 16×16 pixel blocks of four 8×8 pixel cells, a Gaussian spatial window with $\sigma = 8$ pixels, L2-Hys block normalization, a block spacing stride of 8 pixels and a 64×128 detection window for training the linear SVM classifier.

Here we use an adapted and slightly modified version of the previous procedure: (a) we use only grayscale information for speed purposes, (b) complex cell responses are used as gradient information, (c) no Gaussian window is applied, and (d) the L2-norm is used instead of L2-Hys. Complex cell responses provide a good alternative for gradient information since they are also robust to noise. In addition, the linear gradient voting can be skipped because of the cells' oriented responses. We use 24×24 pixel blocks of sixteen 6×6 pixel cells with 75% block overlap; see below for parameter assessment and evaluation. This ensures optimal performance while complying also with Dalal and Triggs' recommendations [3] of having many orientation bins, and moderately sized, strongly normalized, overlapping descriptor blocks for good performance. See Figure 2 (right) for the bio-inspired HOG-like features.

Classification: To detect a person, features in a detection window are classified using a linear SVM. The classifier is trained using the INRIA dataset with a soft-margin linear SVM (C=0.01) using LIBSVM [2]. Similarly to [3], we use a 64×128 pixel detection window for training the classifier. This results in 7×18 blocks to be used by the classifier across the image at all scales. We found that a detection window of 64×128 pixels for training constitutes a good trade-off between performance and running speed. Increasing the detection window's size beyond 64×128, although increasing the detection performance, has a higher computational cost due to more features being used for classification. This has a significant impact on processing time and also on the overall size of detectable pedestrians (smaller pedestrians cannot be detected).

Proc. of the Third Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering- ABBE 2015

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

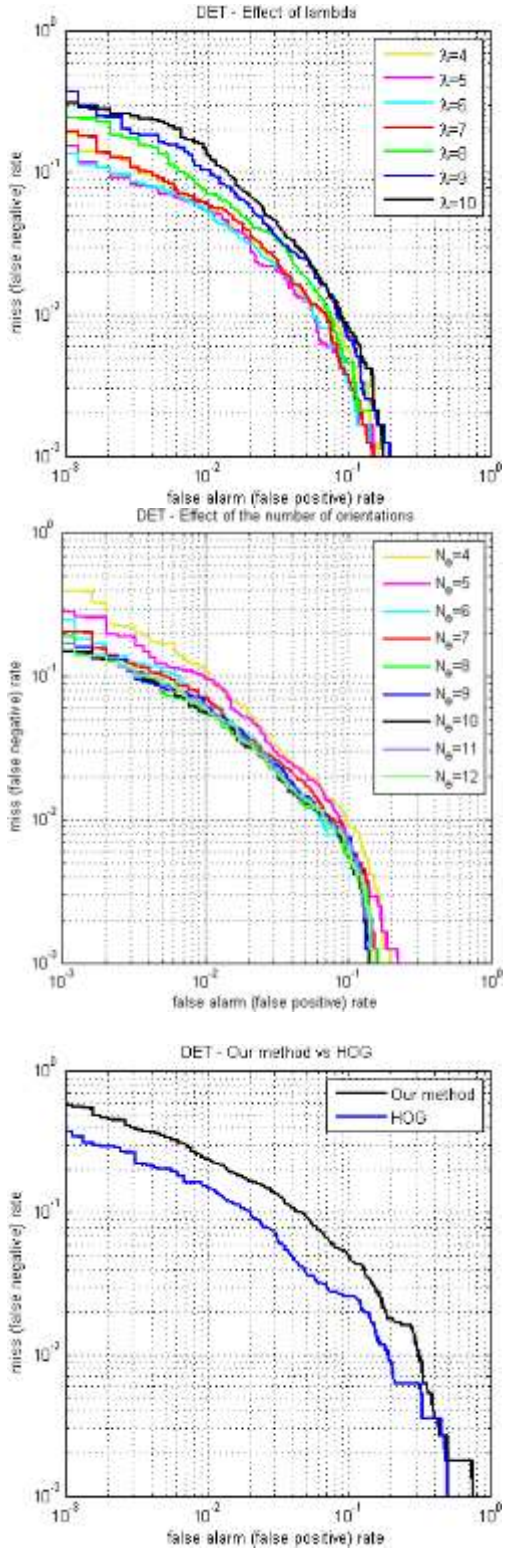

Figure 3. HOG-like feature parameter performance. Top: effects of λ. Middle: effect of number of orientations. Bottom: plot comparing our method (HOG-like features) with HOG [3].

Performance evaluation. Several factors have been evaluated in the classifier training stage, namely Gabor filter scale (λ), number of orientations (N_θ), cell size, block size and overlap. A smaller set of 500 positive and 500 negative random samples for training, and 200 positive and 200 negative random samples for testing were used for cross-validation and parameter optimization. We used detection error trade-off (DET) and miss rate (better: 1.0 - Recall) measures to quantify the performance.

First, the scale of the cortical cells was analyzed in order to determine the optimal λ. Figure 3 (top) shows the overall performance of seven different scales $\lambda = [4,12]$. Smaller scales yield better performance than bigger scales, mainly due to lines and edges being better encoded by smaller filters. This complies with the findings of [3]. Moreover, by choosing a smaller λ the processing time decreases. Here, $\lambda = 5$ performed best. Figure 3 (middle) shows the performance impact of the number of orientations used in the HOG-like feature bins. From the nine orientations tested, increasing the number of orientations beyond 8 does not improve performance significantly. Therefore, for the final classifier we chose $N_\theta = 8$ orientations which gave a 2.69% miss rate. Figure 3 (bottom) shows our method (HOG-like features) vs.\ HOG [3] performance comparison. HOG performed better than our method: a 4.35% miss rate difference at a false alarm rate of 10^{-1}. For comparison purposes, the HOG model was trained with the INRIA dataset using grayscale information only, with the same parameters as in [3].

Three other key factors taken into account were the block size vs. cell size vs. block overlap. From all tested combinations of 0%, 25%, 50% and 75% block overlap with block sizes ranging from 1×1 to 5×5 and pooling cell sizes from 2×2 to 12×12, block sizes of 4×4 with 6×6 pooling cells and 75% overlap gave the best performance with a 3.8% miss rate.

Non-maximum suppression: Now, the detection window is applied to the entire image using a sliding window approach, where multiple detections of the same person often occur. To remove multiple detections due to the sliding window, a non-maximum suppression technique is used to discard overlapping windows: when two windows overlap at least 50%, the window with the weakest classification response is discarded. To this purpose we use the unsigned SVM classification output in order to determine the window's classification response.

IV. Pedestrian Tracker

Hardware processing speed has greatly increased over the past years and many pedestrian detectors have become quite efficient [5]. However, tracking after detection remains a key element in any real-time application. Here, pedestrian tracking after detection is achieved by tracking keypoints inside the detection window and by predicting the window's location using both camera and pedestrian motion information.

Periodically (every 5th frame), the detector is used over the image to detect new pedestrians to track and to update the existing window positions. In the following sections, we address the estimation of the camera's egomotion for the prediction of the detection window's shift from one frame to

Proc. of the Third Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering- ABBE 2015

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

the next, and the tracking of keypoints within the detection window for estimating a pedestrian's trajectory.

Egomotion estimation: This process is done by two main modules: (a) extraction, classification and matching of keypoints; and (b) egomotion estimation from keypoint correspondences.

(a) Keypoint extraction and matching. As already mentioned, keypoints are based on cortical end-stopped cells (Section II). They provide important information because they code local image complexity and have many applications [19][20][22]. In our application, keypoints are used for tracking features of pedestrians inside the detection window and to determine camera and pedestrian motion. We refer to [19] and [25] for a detailed explanation of the multi-scale keypoint detection. Figure 4 (top) shows one example of detected keypoints (+) in two consecutive frames at $\lambda = 5$.

For each keypoint position we use the FREAK descriptor [1], which provides a fast and robust binary descriptor for complex scenarios where pedestrians move and tend to occlude others and themselves. Concerning keypoint descriptor matching, we use the Muja and Lowe matching algorithm [17], which provides good correspondence results (Figure 4 bottom-left).

Figure 4. Egomotion estimation. Detected keypoints (green +) in frames $i-1$ (top-left) and i (top-right) are matched (bottom-left) and egomotion is estimated using epipolar geometry (bottom-right). Here, only 30 sampled keypoints (red boxes) plus epipolar lines (red lines) are displayed for illustration.

(b) Egomotion. In order to predict the correct location of a pedestrian's detection window in a frame in scenarios where a moving camera is used, it is necessary to remove the egomotion effect, i.e., the global motion of the camera needs to be estimated so that the rigid transformation (rotation and/or translation) between two frames can be obtained. Here, camera egomotion is determined by using keypoint correspondences between two frames (current and previous) to estimate the fundamental matrix F. The fundamental matrix provides a general representation of the egomotion captured in two views by a projective camera, without knowledge of the camera's calibration [9][11]. To compute the fundamental matrix we use the 8-point algorithm (see [12] for details). When computing the fundamental matrix, outliers typically arise from gross errors (such as correspondence mismatches) or from movement which is inconsistent with the real egomotion (moving objects or shadows, occluding contours, etc.). To cope with such impairments we use the random sample consensus (RANSAC) algorithm to filter out outliers. Here, RANSAC is used not only for its simplicity and speed, but also for its robustness against large amounts of outliers [26], which often appear in moving scenes populated by pedestrians. We use a residual error threshold of $\tau = 0.2$ for outlier detection, and for speed purposes we set the RANSAC's maximum number of iterations to 100.

Finally, from the fundamental matrix, the rotational and translational parameters are estimated and egomotion is eliminated by mapping detected keypoint positions to a new set of locations according to the transformation. Figure 4 (bottom-right) shows the focus of expansion (FoE) of the estimated camera egomotion while moving forward. The residual error ϵ measure used for inlier estimation corresponds to the distance between the detected keypoints (represented by red squares) and the epipolar lines (red lines) computed using the fundamental matrix. Large distances (high residuals) denote outliers which are not consistent with the global egomotion, while small distances denote motion in line with the camera's egomotion.

Figure 5. Sequence showing motion correction.

(c) Pedestrian tracking: Having a person detected and the egomotion effect removed, person tracking is summarized by tracking both the detection window (bounding box) and the keypoints inside. The procedure is as follows: first, (a) the new coordinates of the detection window in frame i are computed by transforming the previous frame's coordinates (frame $i-1$) with the estimated egomotion parameters; next, the trajectory of the pedestrian is computed where (b) if, at least, 2 or more positive keypoint-descriptors correspondences exist inside the detection windows in both frames i and $i-1$. Then (c) the vector for each pair of corresponding keypoints (optic flow) is

Proc. of the Third Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering- ABBE 2015

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

determined, and (d) the average flow of all vectors is computed (without egomotion); see Figure 5, right column. This corresponds to the average trajectory of a person in the scene.

In case that the average motion is not available, i.e., if no keypoint correspondence between the two consecutive frames i and $i-1$ exists (no known trajectory information), it is assumed that the window has the same trajectory (average flow) in frame i as in the previous frame $i-1$. To keep validating if a person exists inside the detection window, the detection of the person will still be repeated every 5th frame (small changes of this number does not affect the final result). Detection windows (and the person inside) stop being tracked and are discarded from the tracking list when the person detector classifier (which is also applied every 5th frame) fails twice to predict a positive pedestrian.

V. Results

Some results on pedestrian tracking are presented using the ETH dataset [6]. This dataset contains sequences of moving pedestrians using a moving camera with occlusions. For display purposes, the calibration information of the cameras used was not applied in the final results. The reason concerns the generalization of the use of any type of data of pedestrians tracking which may not contain information about the camera's parameters (e.g., YouTube videos). Figure 6 shows results from three sequences of the ETH dataset, namely the BAHNHOF (top row), JELMOLI (second row) and LOEWENPLATZ (third and fourth rows) sequences. In the first two rows, detection of pedestrians in several frames is displayed. Here some false detections occur due to training the classifier with a small dataset. In addition, missed detections also occur due to the fact of the dataset not including training images with partially occluded persons. However, more pedestrians are correctly detected than missed. The bottom rows show the tracking of several pedestrians in four consecutive frames, where the tracker correctly followed the pedestrians.

VI. Conclusions

In this paper we presented a biologically inspired method for pedestrian detection and tracking. The model was optimized concerning execution time versus accuracy. The method works in real time using a 2.4GHz quad core CPU and it yields good results despite the fact that it does not use color information (as used in [3]). By limiting the number of frames where a thorough search for pedestrians is performed with the use of keypoint descriptors for tracking, the method achieves real-time (5 frames per second) results on 640×480 pixel images, since the method switches regularly between tracking and detection, also achieving robustness to errors in detections (false negatives) by keeping tracking records of positive detections. The tracker was also able to recover from failures due to sudden changes in either shape, motion or partial occlusion. Nevertheless, the model is still in the stage of prove-of-concept.

In the future CPU time consumption can decrease even more by also using the same information (keypoints, complex cells, etc.) for constructing saliency maps for Focus-of-Attention (FoA) [19], and with this information applying the sliding windows only in regions with more saliency. Also by combining and coding cortical cells into small overlapping blocks (similarly to HOG's features) and varying its size in a multiscale way.

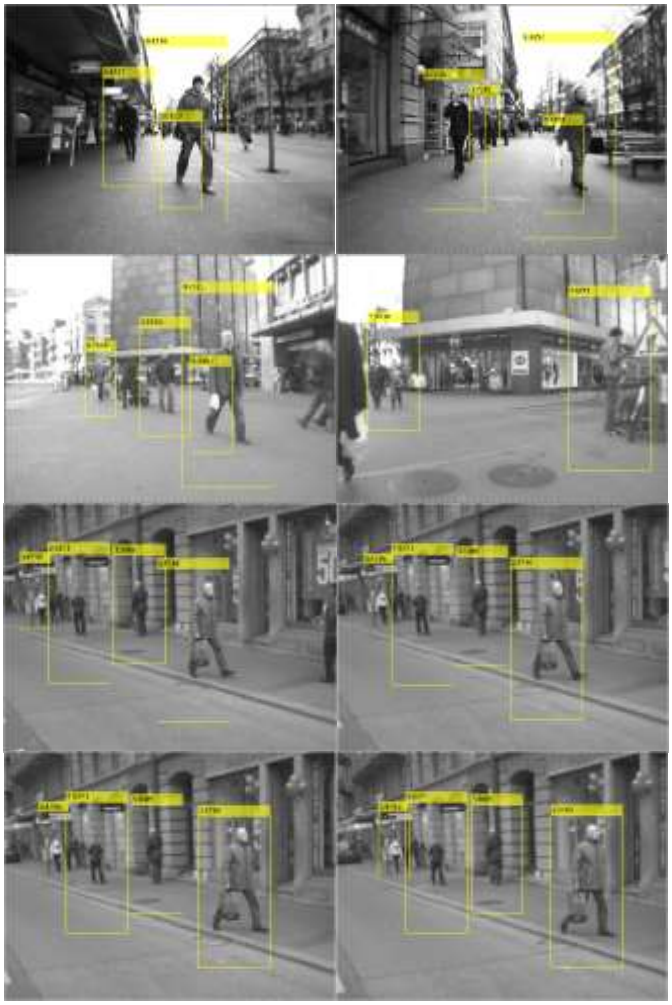

Figure 6. Results on the ETH dataset. Top to bottom: BAHNHOF sequence, JELMOLI sequence and LOEWENPLATZ sequence results. The top two rows show results of pedestrian detection and the bottom rows shows the tracking of pedestrians in four consecutive frames.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology (FCT) projects LARSyS: UID/EEA/50009/2013, CIAC: PEstOE/EAT/UI4019/2013, SparseCoding: EXPL/EEI-SII/1982/2013 and FCT PhD grant to author MF (SFRH/BD/79812/2011).

ISBN: 978-1-63248-060-6 doi: 10.15224/ 978-1-63248-060-6-07

References

[1] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.

[2] C. Chang and C. Lin, LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, pp. 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[4] P. Dollár, R. Appel, and W. Kienzle. Crosstalk cascades for frame-rate pedestrian detection. *Computer Vision–ECCV 2012*, 2012, pp. 645–659.

[5] 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, 2012, pp. 743–761.

[6] A. Ess, B. Leibe, K. Schindler, , and L. van Gool. A mobile vision system for robust multi-person tracking. *In Proc. Computer Vision and Pattern Recognition*, IEEE, 2008.

[7] Han Feng et al. A two-stage approach to people and vehicle detection with hog-based SVM. *Workshop Performance Metrics for Intelligent Systems*, 2006, pp. 133–140.

[8] M. Farrajota, J.M.F. Rodrigues, and J.M.H. du Buf. Optical flow by multi-scale annotated keypoints: A biological approach. *In Proc. Int. Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS 2011), Rome, Italy*, 2011, pp. 307–315.

[9] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? *Computer Vision ECCV'92*, 1992, pp. 563–578.

[10] P. F. Felzenszwalb. Learning models for object recognition. *In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. I–1056.

[11] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. *Computer Vision ECCV'92*, 1992, pp. 579–587.

[12] Richard I Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, 1997, pp. 580–593.

[13] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. *In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 878–885.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, vol. 60, 2001, pp. 91–110.

[15] J. Marin, D. Vázquez, A. M López, J. Amores, and B. Leibe. Random forests of local experts for pedestrian detection. *In Proc. IEEE Int. Conf. on Computer Vision*, 2013, pp. 2592–2599.

[16] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. *Computer Vision-ECCV 2004*, 2004, pp. 69–82.

[17] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *In Proc. VISAPP*, 2009, pp. 331–340, 2009.

[18] C. Papageorgiou and T. Poggio. A trainable system for object detection. *Int. J. of Computer Vision*, vol. 38, 2000, pp.15–33.

[19] J. Rodrigues and J.M.H. du Buf. Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. *BioSystems*, vol. 2, 2006, pp. 75–90.

[20] J.M.F. Rodrigues, R. Lam, and J.M.H. du Buf. Cortical 3D face and object recognition using 2d projections. *Int. J. of Creative Interfaces and Computer Graphics*, vol. 3, 2012, pp.45–62.

[21] J.M.F. Rodrigues, J.A. Martins, R. Lam, and J.M.H. Buf. Fast cortical keypoints for real-time object recognition. *In A. Campilho and M. Kamel (Eds.), Image Analysis and Recognition SE*, vol. 7324, 2013, pp.296-303.

[22] M. Saleiro, M. Farrajota, K. Terzic, J.M.F. Rodrigues, and J.M.H du Buf. A biological and realtime framework for hand gestures and head poses. *In C. Stephanidis and M. Antona (Eds.) Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion SE*, vol. 8009, 2013, pp. 556–565.

[23] M. Saleiro, K. Terzic, D. Lobato, J.M.F. Rodrigues, and J.M.H. du Buf. Biologically inspired vision for indoor robot navigation. *In Campilho, Aurlio, Kamel, Mohamed (Eds.): Image Analysis and Recognition*, vol. 8815, 2014, pp. 469–477.

[24] K. Terzic, J.M.F. Rodrigues, and J.M.H du Buf. Fast cortical keypoints for real-time object recognition. *In Proc. IEEE Int. Conf. on Image Processing, Melbourne, Australia, 15-18 Sept.*, 2013, pp. 3372–3376.

[25] K. Terzic, J.M.F. Rodrigues, and J.M.H du Buf. Bimp: A real-time biological model of multi-scale keypoint detection in v1. *Neurocomputing*, vol. 150, 2015, pp. 227–237.

[26] P. HS Torr and D. W Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int. J. of Computer Vision*, vol.24, no.3, 1997, pp. 271–300.

[27] P. Viola, M.J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance *In Proc. 9th IEEE Int. Conf. on Computer Vision*, 2003, pp. 734–741.

[28] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010, pp. 1030–1037.

[29] X. Wang, T.X. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. *In Proc. IEEE 12th Int. Conf. on Computer Vision*, 2009, pp. 32–39.

About Author (s):

Miguel Farrajota graduated in Electric and Eletronic Engineering in 2008, he got his M.Sc. in 2011, and is currently finishing his PhD degree on Signal Processing. He is a researcher for the University of the Algarve, member of LARSyS and the ISR in Lisbon. His major research interests lies on computer and human vision, pattern recogniton, motion and human-computer interaction.

Professor João Rodrigues graduated in Electrical Engineering in 1993, he got his M.Sc. in Computer Systems Engineering in 1998 and Ph.D. Electronics and Computer Engineering in 2008 from University of the Algarve, Portugal. He is Adjunct Professor at Instituto Superior de Engenharia, also in the University of the Algarve, where he lectures Computer Science and Computer Vision since 1994. He is member of associative laboratory LARSyS (ISR-Lisbon), CIAC and the Associations APRP, IAPR and ARTECH. He participated in 13 financed scientific projects, and he is co-author more than 100 scientific publications. His major research interests lies on computer and human vision, assistive technologies and human-computer interaction.

Professor Hans du Buf got his M.Sc. in Electrical Engineering in 1983 and Ph.D. in 1987 at the Technical University of Eindhoven, Netherlands. He worked for 7 years at the Swiss Federal Institute of Technology in Lausanne (EPFL) before moving to the University of the Algarve in Faro, Portugal, in March 1994. He is an associate Professor at Dept. of Electronics and Computer Science, also in the University of the Algarve, where he lectures Digital Systems, Computer Architecture, Computer Graphics as well as Image Processing. He is a member of the LARSyS and the ISR in Lisbon, the IEEE Computer Society, as well as the Portuguese Association for Pattern Recognition APRP and the International Association for Pattern Recognition IAPR. He is co-author of over 40 journal papers and 110 conference papers. His major research interests lies on computer and human vision.