Tracking Pedestrians with Bacterial Foraging Optimization Swarms

Hoang Thanh Nguyen, Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside
nthoang@cs.ucr.edu, bhanu@cris.ucr.edu

***Abstract*—Pedestrian tracking is an important problem with many practical applications in fields such as security, animation, and human computer interaction (HCI). In this paper, we introduce a previously-unexplored swarm intelligence approach to multi-object monocular tracking by using Bacterial Foraging Optimization (BFO) swarms to drive a novel part-based pedestrian appearance tracker. We show that tracking a pedestrian by segmenting the body into parts outperforms popular blob-based methods and that using BFO can improve performance over traditional Particle Swarm Optimization and Particle Filter methods.**

***Index Terms*—monocular pedestrian tracking, swarm intelligence, bacterial foraging optimization, uncalibrated cameras.**

I. Introduction

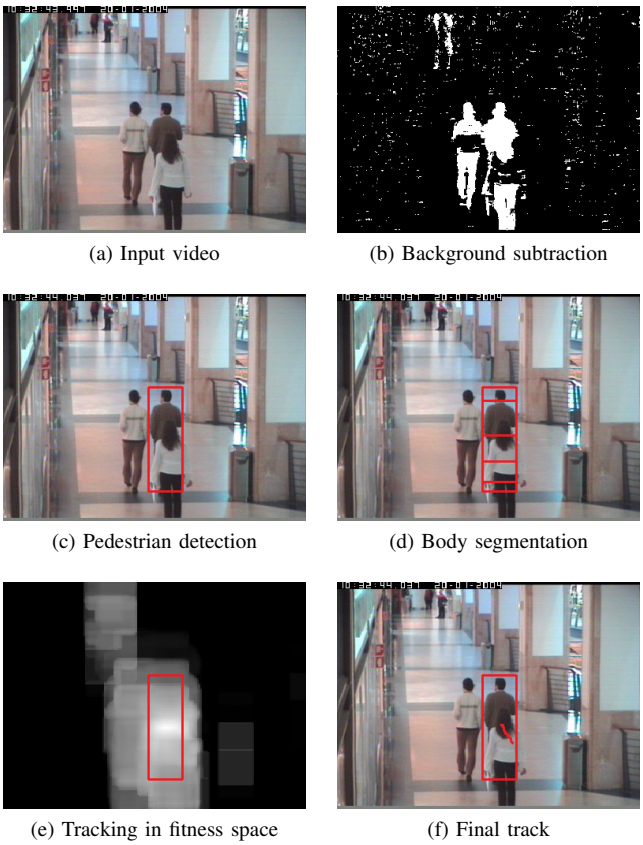
Tracking is a classic computer vision problem of significant importance where the objective is to continuously locate an object (e.g., a cell, pedestrian, vehicle, or crowd) in sequential frames of a video. Pedestrian tracking in particular poses considerable interest in fields such as surveillance, scene analysis, human-computer interaction, pervasive computing, and computer animation. A complete tracking system generally consists of 4 main components: object detection (detecting the objects to track), object tracking (tracking the objects across frames), track association (joining short-term "tracklets" into long-term tracks), and analysis (utilizing the data for high-level understanding). Since much of the computation effort is spent on tracking, making robust tracking approaches faster helps to move tracking systems into the real-time domain and more accurate low-level tracking improves the performance of higher-level track association.

In order to further explore algorithms with less exposure, this paper extends the general object tracking approach proposed in [11] to the real-world problem of pedestrian tracking and asks the question: can Bacterial Foraging Optimization outperform more popular search algorithms for real-world tracking applications? This paper is organized as follows: Section II discusses related work and contributions, Section III details the technical components of the proposed tracking approach, Section IV presents experimental results on the CAVIAR [5] dataset, and Section V offers closing remarks.

II. Related Work and Contributions

Traditional approaches toward pedestrian tracking focus around Particle Filters [8], Mean Shift [2], or detection-based tracking [15]. An alternative approach considers a family of biologically-inspired evolutionary computation algorithms known as swarm intelligence. In this category, the most popular approaches are Particle Swarm Optimization (PSO) [9] or a combination of Ant Colony Optimization with the above approaches [6]. These trackers are often used to generate short-term "tracklets" which are then used for methods such as Data Association Tracking (DAT) [10], [14] to produce long-term inter or intra-camera tracks.

This paper focuses on a less-widely known swarm intelligence algorithm and is the first paper to propose Bacterial Foraging Optimization for pedestrian tracking. We show that



(a) Input video (b) Background subtraction (c) Pedestrian detection (d) Body segmentation (e) Tracking in fitness space (f) Final track

Fig. 1. Bacterial Foraging Optimization-based pedestrian tracker. (a) Input video, (b) background subtraction to extract areas to perform detection, (c) pedestrian detection on blobs not already tracked, (d) segment the bounding box using statistical body ratios, and (e) track in the fitness space using Bacterial Foraging Optimization, (e) output final tracks.

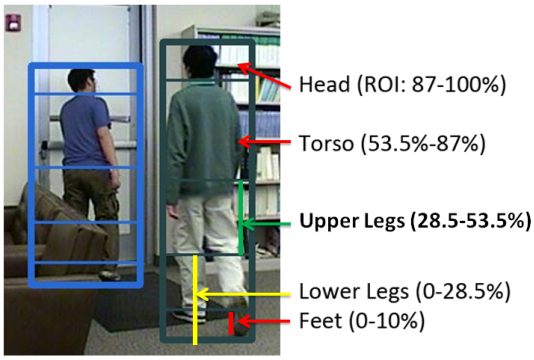


Fig. 2. Body segmentation of two pedestrian rectangles of interest using fixed statistical ratios.

it is a viable approach for fast appearance-based tracking and we make the following claims:

- Segmenting the body into separate regions of interest (ROIs) improves tracking performance as opposed to popular single ROI-based approaches.
- Bacterial Foraging Optimization offers improved performance over traditional Particle Filter and over Particle Swarm Optimization utilizing the same resources.

This paper assumes tracking is to be performed using uncalibrated monocular cameras. This is a reasonable assumption as it is the characteristic of most of the existing camera systems. Figure 1 outlines the proposed approach.

III. Technical Approach

A. Pedestrian detection

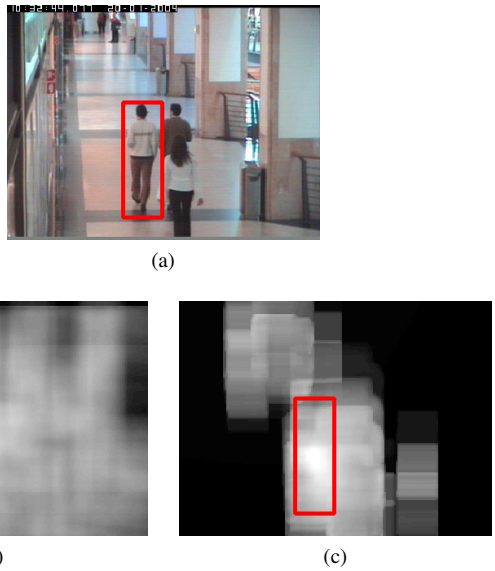
Tracking requires an initial estimate of an object's location in order to begin. Background subtraction is first used to extract a rough foreground mask for every frame. This paper uses the modified Gaussian mixture model (GMM) proposed in [16] to segment the foreground by dynamically learning the background as the video progresses.

Labeling the connected components of the foreground mask gives a set of blobs which can be input into a pedestrian detector. Since detection algorithms are often resource-intensive, executing them only on the blobs not already covered by any current tracker (as opposed to the entire image) provides a significant speed-up in performance. This paper uses the Viola-Jones Haar wavelet-based detector [3], [15] trained to detect the head and shoulders of pedestrians. Performing detection on upper bodies (as opposed to full or lower bodies) makes detection more robust to occlusion, e.g., when facing crowds where only the head/top portion of the body is visible and when pedestrians enter/leave through the bottom of the video's field of view. The ROIs of detected upper bodies are then extended to encompass for the full body to match the upper body's ratio. This paper takes a square detection box of a person's head and shoulders and extends it down to a rough location of the feet by increasing the box height by a fixed ratio and preserving the top-left coordinates and width of the original detection box:

$$full\_body_{height} = upper\_body_{height} \times R$$

$$full\_body_{x,y,width} = upper\_body_{x,y,width}$$

The value of ratio R is dependent on the detection classifier used and experimentally selected to be 3.1 in this paper.



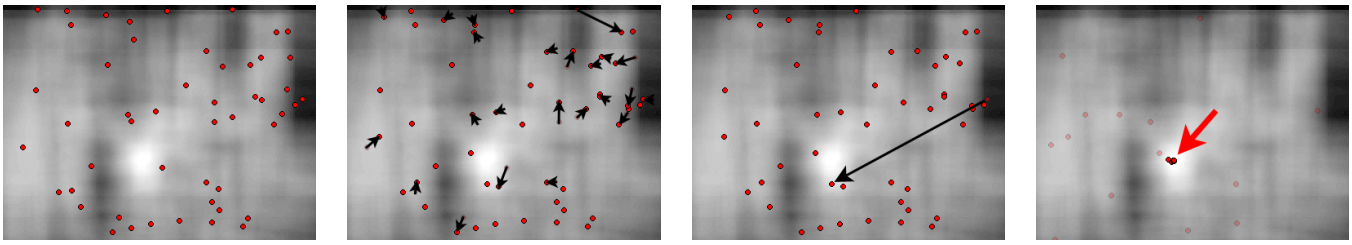
(a) (b) (c)

Fig. 3. Fitness space for tracking a pedestrian using color similarity in the HSV color space (brighter = higher fitness). (a) Detected pedestrian, (b) fitness space for the pedestrian's signature in the original image, (c) fitness space for the pedestrian in the background-subtracted image.

B. Body segmentation

In most cases, it is reasonable to assume that most people do not dress from head to toe in a single solid color (though exceptions exist, e.g., students at a graduation). In addition, people in a single video's field of view can often be distinguished based on general appearance alone. To make use of these observations, we take a parts-based approach to identifying an individual in the short-term by segmenting a pedestrian's ROI into a number of distinct horizontal segments based on statistical percentages of body ratios [1], [4], [7]. Figure 2 shows the resulting body segments overlaid on a pedestrian.

The foreground mask is then applied to the image in order to prevent background pixels from affecting a pedestrian's signature. Once the body has been segmented into 5 main parts (head, torso, upper legs, lower legs, and feet), signature extraction (e.g., generating color histograms) is performed on each body part separately (the more common alternative is to extract a single signature for the entire ROI). This is performed by sampling every foreground pixel in each body part's ROI, though further speedups can be achieved though the use of subsampling. This paper uses color histograms extracted in the HSV color space to generate signatures, extracting a fixed $N = 32$ bins from the hue and saturation components and normalizing to sum. Calculating the similarity between two



(a) (b) (c) (d)

Fig. 4. Behavior of a single Bacterial Foraging Optimization swarm searching for a pedestrian. (a) Random initialization, (b) gradient-hill climbing in random directions, (c) death/rebirth of agents with poor fitness to location of agents with best fitness, (d) target location based on consensus of the best agents.

histograms A and B of length N bins is performed using histogram intersection:

$$hist\_intersect(A, B) = \sum_{i}^{N} min(A_i, B_i)$$

The average percentage intersection of the body part histograms can then be used to compute a fitness between a query location and the original initialized target signature. This fitness function can then be used by any search algorithm to track the pedestrian. Figure 3 shows the exhaustively-generated fitness space for an initialized ROI.

C. Tracking with Bacterial Foraging Optimization

Bacterial Foraging Optimization (BFO) [12] is a stochastic evolutionary swarm intelligence search algorithm designed to model the movement and feeding behavior of *E.coli* bacteria. A swarm consists of a number of particles or "agents" which move or "swim" and "tumble" through an environment searching for concentrations of food (or regions of high fitness from a feature space point of view). Given an image, a swarm of agents is first randomly initialized on the image. The algorithm consists of R "reproduction" loops which execute a number of C "chemotaxis" or movement loops. In each chemotaxis loop, all agents "tumble" (choose a random direction) and are allowed to "swim" (or sample) up to S times in steps of size $Step$ in a gradient hill-climbing manner. At the end of each reproduction step, the bottom T agents with the worst fitness scores die off and an equal number of agents are born at the locations of the T best agents. In this manner, resources are quickly allocated to regions of higher fitness. At the end of the algorithm, the agents undergo an elimination/dispersal step which randomly relocates agents with probability P. This step helps to simulate a changing environment such that the swarm does not fully converge and cease to track in succeeding frames. Figure 4 shows the behavior of a BFO swarm in a fitness space.

BFO has never been used previously for pedestrian tracking, yet possesses traits which make it suitable to the problem. The near-uniform coverage of the search space is useful for overcoming occlusion (whereas many other approaches lose track once they converge). In addition, the fast propagation of agents to regions of high fitness reduces overhead of having the agents gradually making their way toward global-best fitness regions. This paper utilizes BFO with the enhancements proposed in [11] for additional characteristics such as early termination, lookahead, and elitism. Algorithm 1 summarizes the full procedure.

D. Track smoothness

In the event that there are multiple people in a single frame who both have similar appearance signatures (e.g., similar clothing such as uniforms), it is helpful to add an additional smoothness constraint to the fitness function. This allows a tracker to prefer a location which more closely resembles a pedestrian's current trajectory rather than, for instance, a location of slightly higher fitness that is clear across the frame. From [13], smoothness at a current trajectory point P_i can be computed as the difference between the vectors V of the previous trajectory point P_{i-1} and a proposed point P_{i+1}:

$$V_i = P_{i+1} - P_i$$

$$smoothness_i = W \times \frac{V_{i-1} \cdot V_i}{|V_{i-1}||V_i|} + (1 - W) \times \frac{2\sqrt{|V_{i-1}||V_i|}}{|V_{i-1}| + |V_i|}$$

where W is a weight value between 0 and 1 (a higher value favors smoothness in direction, a lower value favors smoothness in velocity). W is set to 0.50 in this paper to give both smoothness components equal weight. After normalizing the scores of the histogram intersections, the smoothness function can then be integrated into the final fitness function:

$$fitness = W_s \times smoothness + (1 - W_s) \times hist\_intersect$$

where W_s is experimentally set to 0.01 for all videos in this paper to favor smoother locations when deciding between locations of similar fitness.

IV. Experimental Results

The proposed tracking system is tested on 7 videos from the CAVIAR dataset [5] which are considered to be the most challenging [14]. The CAVIAR videos show the following challenges: 1) relatively low resolution CIF video (382×288), 2) pedestrian size changes dramatically depending on the position in the corridor (e.g., ROI sizes change by up to 450%), and 3) pedestrians occlude each other not only in the middle of a pedestrian's path, but also when pedestrians first enter as well as leave the field of view. Table I shows frame, pedestrian,

Algorithm 1 BFO algorithm modified for tracking [11]

$I \leftarrow$ image to search
$Target \leftarrow$ hist and prev. location of object to search for
$R \leftarrow$ number of reproduction steps
$C \leftarrow$ number of chemotaxis steps per reproduction
$S \leftarrow$ max number of swims per chemotaxis step
$Step \leftarrow$ swim step size in pixels
$T \leftarrow$ number of agents to relocate per reproduction
$P \leftarrow$ probability a non-immune agent gets relocated
$Thresh \leftarrow$ minimum fitness to trigger early termination

procedure BACTERIALFORAGING
  if I is first frame of target **then** ▷ Init first frame
    Initialize agent locations on I
  end if ▷ Early termination?
  if $fitness(Target_{loc}, I, Target_{hist} \geq Thresh$ **then**
    return $Target_{loc}$
  end if
  for R reproduction steps **do** ▷ Begin search
    for C chemotaxis steps **do**
      for all agents A **do**
        $d \leftarrow$ random direction
        for up to S swims **do**
          $l \leftarrow$ new location $Step$ pixels from A toward direction d
          $f \leftarrow fitness(l, I, Target_{hist})$
          if $f > A_{current\_fitness}$ **then**
            $A_{current\_fitness} \leftarrow f$
            $A_{current\_location} \leftarrow l$
          else
            Break
          end if
        end for
      end for
    end for
    for all top T agents A with best fitness **do**
      $A_{immunity} \leftarrow true$
    end for ▷ Death/rebirth
    Move the T agents with worst fitness to locations of the T agents with best fitness
  end for
  ▷ Elimination/dispersal
  for all agents A where $A_{immunity} \neq true$ **do**
    Relocate A to random position with probability P
  end for
  ▷ Return updated location
  $Target_{loc} \leftarrow$ best location based on all agents A where $A_{immunity} = true$
  return $Target_{loc}$
end procedure

Fig. 5. Sample frames from the CAVIAR dataset [5]. (a) OneShopOneWait1cor, (b) OneStopMoveEnter1cor, (c) ThreePastShop1cor, (d) ThreePastShop2cor, (e) TwoEnterShop1cor, (f) TwoEnterShop2cor, (g) TwoEnterShop3cor.

Video	# Frames	# Pedestrians	# ROIs
OneShopOneWait1cor	1,377	6	4,496
OneStopMoveEnter1cor	1,590	19	13,691
ThreePastShop1cor	1,650	8	9,642
ThreePastShop2cor	1,521	9	9,452
TwoEnterShop1cor	1,645	11	7,190
TwoEnterShop2cor	1,605	15	7,930
TwoEnterShop3cor	1,149	14	6,856
Total	10,537	82	59,257

TABLE I
VIDEOS USED FROM THE CAVIAR DATASET.

and ROI information about the CAVIAR videos and Figure 5 shows sample frames from the videos.

For the results, tracking accuracy is defined as the percentage of groundtruth ROIs covered by the tracker initialized for that pedestrian. An ROI is considered to be tracking a target if its intersection with the groundtruth ROI exceeds at least 50% of their union:

$$is\_tracked(ROI_{query}, ROI_{gt}) = \frac{ROI_{query} \cap ROI_{gt}}{ROI_{query} \cup ROI_{gt}} > 0.50$$

i.e., a tracking accuracy rate of "44%" means 26,000 of

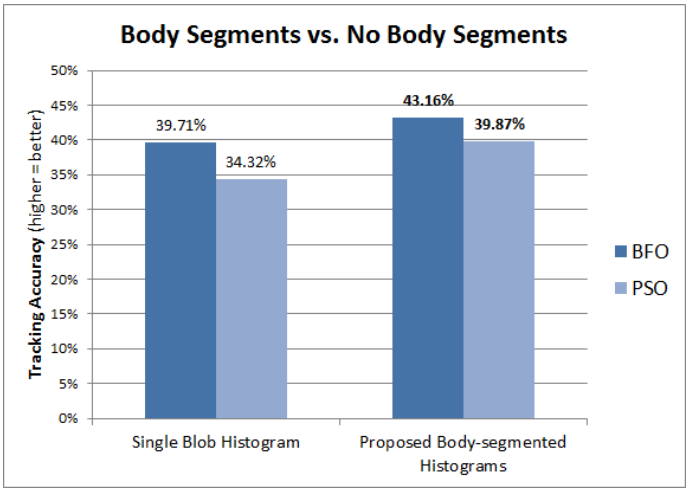


Fig. 6. Tracking accuracy of traditional single blob-based tracking vs. body segments, averaged over 10 runs across the 7 CAVIAR videos seen in Table I.

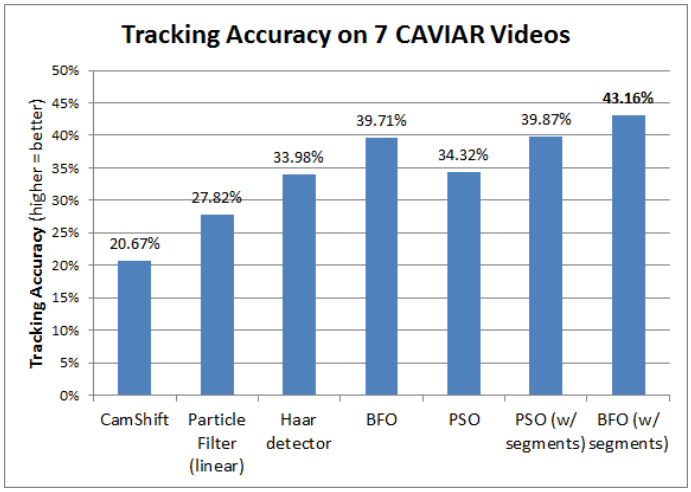


Fig. 7. Tracking accuracy of multiple trackers, averaged over 10 runs across the 7 CAVIAR videos seen in Table I.

the 59,000 ROIs were correctly located with at least 50% groundtruth intersection.

Figure 6 shows the tracking accuracy results on the CAVIAR dataset when histograms are extracted for each body segment versus traditional single blob approaches. All tests are averaged across 10 runs. For BFO and PSO, the parameters were manually optimized such that each algorithm made on average 300 calls to the fitness function (a number which allows the tracker to perform at a reasonable speed of 20 frames per second on the test machine, a 2.40GHz Intel Q6600 CPU), i.e., PSO uses 30 particles with 10 iterations and BFO uses $A = 10$ particles, $R = 12$ reproductions, $C = 1$ chemotaxis step, $S = 5$ max swims per chemotaxis, $Step = 5$px, $T = 1$ death/rebirth per reproduction, and $P = 90\%$ dispersal rate. Segmenting the body achieves roughly 5% increase in tracking performance over conventional single-blob approaches. Speed-wise, since the total ROI area for an object is equivalent between the two approaches, there is only the negligible overhead of computing the segmented ROIs and 4 more histogram intersection comparisons.

Figure 7 shows a tracking accuracy comparison of the proposed BFO approach with PSO using the same body segmentation scheme, Particle Filter, CamShift, and the Viola-Jones based detection-based tracker.

V. Conclusions

We proposed using Bacterial Foraging Optimization swarms for pedestrian tracking and showed that using a parts-based segmentation approach can outperform both PSO and traditional Particle Filter methods. Future work involves addressing noisy histogram initialization caused by occlusion and bad foreground segmentation and using the low-level tracker designed in this paper to improve the performance of higher-level DAT-based trackers and on investigating alternative pedestrian signatures for improved monocular and multi-camera association.

References

[1] M. Altab Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43:2281–2291, June 2010.

[2] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal Q2 1998*, 1998.

[3] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, pages 130–140, April 2007.

[4] W. T. Dempster and G. R. L. Gaughran. Properties of body segments based on size and weight. *American Journal of Anatomy*, January 1967.

[5] R. B. Fisher. The PETS04 surveillance ground-truth data sets. In *Proceedings of the Sixth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2004)*, pages 1–5, May 2004. http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.

[6] Z. Hao, X. Zhang, P. Yu, and H. Li. Video object tracing based on particle filter with ant colony optimization. In *Proceedings of the Second IEEE International Conference on Advanced Computer Control (ICACC 2010)*, volume 3, pages 232 –236, 2010.

[7] P. Hewitt and D. Dobberfuhl. The science and art of proportionality. *Science Scope*, 27(4):30–31, January 2004.

[8] M. Isard and A. Blake. CONDENSATION: Conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29:5–28, 1998.

[9] J. Kennedy and R. Eberhart. Particle swarm optimization. *IEEE International Conference on Neural Networks*, 4:1942–1948 vol.4, 1995.

[10] Y. Li, C. Huang, and R. Nevatia. Learning to associate: HybridBoosted multi-target tracker for crowded scene. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 2953 –2960, 2009.

[11] H. T. Nguyen and B. Bhanu. Tracking multiple objects in non-stationary video. In *GECCO '09: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, pages 1561–1568, July 2009.

[12] K. M. Passino. Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Systems Magazine*, Vol. 22, No. 3:52–67, 2002.

[13] L. G. Shapiro and G. Stockman. *Computer Vision*. Prentice Hall, January 2001.

[14] B. Song, T.-Y. Jeng, E. Staudt, and A. K. Roy-Chowdhury. A stochastic graph evolution framework for robust multi-target tracking. In *Proceedings of the 11th European conference on Computer vision (ECCV 2010)*, pages 605–619, Berlin, Heidelberg, 2010. Springer-Verlag.

[15] P. Viola and M. Jones. Robust real-time object detection. *Second International Workshop on Statistical and Computational Theories of Vision*, 2001.

[16] Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27:773–780, May 2006.