



Fast Multi-scale fHOG Feature Extraction Using Histogram Downsampling

Mihai Polceanu^(✉), Fabrice Harrouet, and Cédric Buche

LAB-STICC, ENIB, Brest, France
{polceanu,harrouet,buche}@enib.fr

Abstract. Object detection is crucial for autonomous robotic systems to interact with the world around them but, in robots with low computational resources, deep learning is difficult to take advantage of. We develop incremental improvements to related work on feature approximation and describe an adaptive fHOG feature pyramid construction scheme based on histogram downsampling, together with a SVM classifier. Varying the pyramid level to which the scheme is applied gives control over the trade-off between precision or speed. We evaluate the proposed scheme on a modern computer and on a NAO humanoid robot in the context of the RoboCup competition, *i.e.*, robot and soccer ball detection, in which we obtain significant increase (1.57x and 1.68x on PC and robot respectively) in pyramid construction speed relative to our baseline (the dlib library) without any loss in detection performance. The scheme can be adapted to increase speed while trading off precision until it reaches the conditions of a state-of-the-art power law feature scaling method.

Keywords: Vision for robotics · Object detection · Feature approximation · Histogram of Oriented Gradients

1 Introduction

Object detection has seen tremendous progress in the past years, which stemmed both from hand-crafted features and the relatively recent convolutional neural networks (CNN). As CPUs improved and with the advent of cheaper GPU processing, some subfields witnessed super-human image recognition accuracy.

Our focus is however on conditions where the system is required to function with low computational resources. Such conditions can be found in the Standard Platform League (SPL) of the RoboCup soccer competition. Here, teams must use the commercially available SoftBank NAO humanoid robot without hardware modifications. The resources available in this setup are an Intel Atom 1.6 GHz processor with one thread and 1GB of RAM. Although sufficient for vision, these resources must be shared by several processes to perform locomotion, localization and strategic behavior necessary for the soccer match, hence in reality, object

detection can only account for a small amount of resources (roughly 5% CPU in the implementation we used for this work) when the robot is in motion.

The main tasks for computer vision in this competition are the detection of lines, goal posts, the ball and other robots (teammates and opponents). For humanoid detection we turn to the subfield of pedestrian detection, where Histograms of Oriented Gradients (HOG) continue to play an important role since their invention in 2005 [5]. HOG consists in dividing an image into cells, computing the gradient of each cell, binning gradients into a histogram of main orientations and finally normalizing over blocks of cells to produce features (originally 36-dimensional) that are used for training a Support Vector Machine (SVM) classifier over a sliding window. Later, [8] introduced a refined version of HOG features (commonly referred to as Felzenszwalb HOG or fHOG) which proved more robust for pedestrian but also generic object detection. Despite its high popularity, HOG-like (HOG, fHOG, or other variants) feature extraction is known to be slow, although more energy efficient than more accurate CNNs [19].

While many other flavors of object detectors have been proposed, most combine HOG-like features with other types to obtain better results under different conditions. Our main focus is to accelerate fHOG feature extraction in order for it to run on a robotic platform with low computational resources, while maximizing detection precision.

We first discuss existing work on feature approximation and how our work differs from the state-of-the-art method [6]. We then present a series of evaluations on a modern computer (Intel Core i7, 2.60 GHz) to observe how our choices have an impact on detection quality and execution speed. All results are obtained using a single execution thread and averaged over multiple runs to ensure validity. From these evaluations we make several observations that lead to an adaptive scheme that gives control over the trade-off between average precision and execution speed. Finally we validate the results on the chosen robotic platform, where we obtain higher execution time compared with the modern computer as expected, due to the lower quality processor, but still observe significant improvement relative to the baseline.

2 Related Work

The Histogram of Oriented Gradients is a widely used feature descriptor, first proposed for human detection by [5]. The intuition behind this approach is that the shape and appearance can be characterized by the distribution of local edge directions (intensity gradients). The main steps in obtaining a HOG-based detector, as originally described by [5], are to compute gradients that are binned in histogram of 9 dominant directions (bins), which are then normalized over blocks of 4 cells and finally concatenated to obtain the HOG features ($9 \times 4 = 36$ dimensional). These features are then fed as input to a SVM classifier that is trained with the corresponding labels such as pedestrians. Afterwards, [8] improved upon the result with insights gained from applying Principal Component Analysis

(PCA) to reduce the dimension of the initial features while retaining the same performance and used additional contrast sensitive components to improve overall performance.

Our work falls in the concept of feature approximation, notably used by [7] to significantly increase the speed of the feature pyramid construction for non scale invariant features like HOG. The feature pyramid is constructed by downsampling the image by a factor, *i.e.*, a smaller image is obtained by downscaling the original by a chosen fraction. For example, downsampling a 640×480 image by a $4/5$ factor means that the obtained image is $640 * 4/5 = 512$ pixels wide and $480 * 4/5 = 384$ pixels high. The main intuition is that instead of downscaling the image and extracting features at each level of the pyramid, intermediary levels could be *approximated* using nearby feature maps. Approximating intermediary pyramid levels is complementary to other works focusing on optimized variants of HOG, such as faster cell-based scanning [18], heavy use of parallelism [12] and different approaches to computing gradients [4].

Successive image down-sampling and calculating features at each pyramid level are computationally expensive. Taking advantage of the generally fractal structure of natural images, it is possible to obtain similar performance by only downsampling and computing the features for each halved image; *i.e.*, at each octave [6]. In-between, the features are approximated (upsampled and downsampled) from the ones directly calculated at each octave, which directly leads to higher execution speed that is crucial for embedded systems [11].

Regarding the context of application, object detection in the RoboCup Soccer competition (standard platform league) has been achieved with (generally manually tuned) detectors based on color segmentation [13, 16] or color histograms [17], statistical modeling [2], line detection, rough shapes [3], or simply as non-green patches which differentiate themselves from the green football field [10]. We note that some approaches [16] make use of smaller image sizes and regions of interest (instead of sliding windows) to limit the search space in the image to improve performance; this technique is complementary to our work on analyzing the effects of feature approximation on performance.

In our experiments, we use the object detector provided by Dlib [14] which uses the well known 31-dimensional feature extraction method described by [8] (fHOG) together with Max-Margin Object Detection (MMOD) [15] which improves training efficiency. While other implementations exist (such as OpenCV and others), our focus is on studying the impact of feature approximation, regardless of particular implementation. Final detections are obtained by applying non-maximum suppression on the ensemble of overlapping detection boxes. Dlib represents one of the most efficient implementations of state-of-the-art classical object detection and is used by both academia and industry. Throughout this work, our contribution is compared to the current implementation of fHOG-based SVM classifier in Dlib as primary baseline.

3 Image and Histogram Downsampling

The first important observation is that the main bottleneck in computing fHOG features resides in calculating the gradient histogram which happens before computing the final feature map. Most modern processors provide Single Instruction, Multiple Data (SIMD) instructions, which have the same execution time and electricity consumption as their scalar counterparts, but handle 16 bytes of data simultaneously which, for our purposes, enables 4 floating point operations instead of one. This is also the case for the Intel Atom 1.6 GHz processor of the NAO robot used in our research. However, computing histograms involves the decision of which bin is associated with each data point, and therefore cannot be fully vectorized. This observation led us to consider avoiding successive histogram computations, similar to how other authors avoid direct feature extraction.

Another important aspect to consider is that the bulk of computational expense rests in the first few levels of the pyramid that use large scale images. For example, for a 10-level pyramid that uses 4/5 downscaling factor on a 640×480 image, computing features for each level costs 36.70%, 23.23%, 14.81%, 9.54%, 6.02%, 3.91%, 2.53%, 1.52%, 1.07% and 0.67% of the total time, respectively.

We can identify three aspects of the fHOG feature extraction algorithm that can be accelerated: (a) The level to which the scheme is applied, (b) Image downsampling strategy and (c) Histogram downsampling strategy. Here we describe each of these aspects, and then put them to the test in the following section, in comparison with related state-of-the-art schemes.

All tests are performed on a dataset that contains images from the publicly available SPQR dataset [1] and also includes new frames coming from robots during test matches in different lighting conditions and of lower resolution. Lower resolution images were upscaled to 640×480 which is the chosen resolution of our evaluation, as this is the real image size that is usable from the NAO robot camera; in fact, the output of the camera is 960p/30fps (1280×960) and is provided in YUYV format (also known as Y'UV422), but processing the full sized image exhausts much of the resources available on the robot [9]. For training and testing, we only consider the Y value which can very efficiently be read directly from the raw camera output. The images in the enhanced dataset have been randomized and divided into 100 training images (50 with horizontal flip added) and 98 test images, amounting to 190 and 185 positive examples respectively. The training set was kept small to avoid excess robot pose variation which makes training a HOG detector inefficient (*i.e.* the resulting vignette becomes blurred); this proves to obtain acceptable accuracy without the need to train multiple detectors for several poses which drastically increase the amount of resources required, while in practice it also allows training for a different object with few on-site images.

3.1 Approximating Levels

As seen previously, it is most important to approximate lower levels of the pyramid (*i.e.* larger image sizes), as they have the highest cost in terms of computation time. Throughout this work, we evaluate different downsampling strategies by applying them up to a certain level of the pyramid. Applying each scheme up to a level means that some or all intermediary levels are approximated from histograms from a lower level (source level depends on each scheme). After doing a hyperparameter (pyramid downscaling factor; C and risk gap of SVM; and box matching threshold) search (grid followed by random search) on the Dlib fHOG object detector [14] which we use as a baseline, we chose to use a $4/5$ downscale factor which in our implementation leads to a pyramid with 10 levels (0–9). In the following, we report the precision and execution time of each approach by applying the scheme (approximating histograms) up to a given level exclusively, while higher levels are computed in the same way as the baseline; this way, results at level 1 are equivalent for all schemes and baseline since the scheme is not applied to any level. Starting from level 2 and up to 9, the charts illustrate the effect of the chosen scheme on performance and speed. We also include level 10, which means that the entire pyramid (levels 0–9) is approximated, using the chosen scheme.

3.2 Image Downsampling

Constructing smaller scale images is originally performed in the Dlib baseline at each pyramid level using bilinear interpolation, while histograms and features are also extracted at each level. The power law based feature approximation approach [6] proposes to only subsample images at each octave (ratio of $1/2$) while approximating the intermediary final features (as opposed to histograms in our work). To test how approximation influences performance, regardless of octaves, we approximate histograms up to a given level while downsampling images for the higher levels in two ways: in the first scheme (dubbed *slow* method in the following) we continue to downsample images from each previous level, even if these intermediary images are not used for feature extraction, and in the second scheme (dubbed *fast* method) we do not keep intermediary images, but downsample from the original image directly to the level up to which the scheme is applied (see Fig. 1 for clarification). While significantly faster, one may argue that this method can lead to important information loss for higher levels (smaller image scales), due to the fact that entire pixels are ignored in the downsampled image.

We put this intuition to the test, and evaluate quality when skipping several levels when downsampling image. We note that using the *fast* approach, we obtain “pixelated” parts of the image, where gradient information is lost, but this only becomes clear when the gap between downsampled images increases, in our case, further than 4 or 5 levels. We therefore expect that having at least some intermediary images, such as in the case of the power law approach [6] which subsamples images at each octave, should improve results.

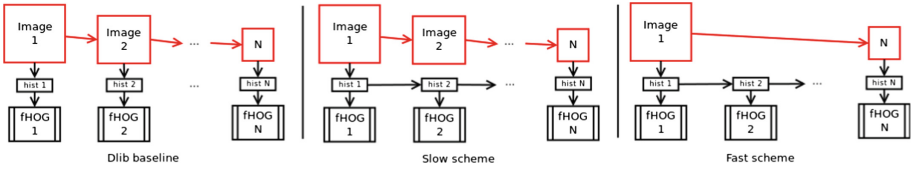


Fig. 1. Image downsampling approaches of Dlib baseline (left), *slow* (center) and *fast* (right) schemes applied up to level N .

To study the impact of the two image downsampling strategies, we measure speed and performance at each level in the pyramid. To visualize the progression in function of level, we plot the average execution time (Fig. 2 left) obtained by a single image pyramid, measured for 10 configurations. Each configuration i consists in an fHOG detector that performs feature downscaling using the respective image downsampling strategy (*slow/fast*) until level $i - 1$ and then, from level i onwards it performs the default feature extraction (which corresponds to downsampling the image and extracting features at each step).

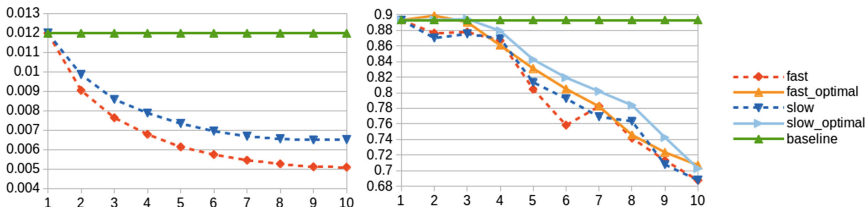


Fig. 2. Average execution time (in seconds) on modern computer (left) and average precision (right) of *slow* vs. *fast* image downsampling when applying each scheme up to each level of the pyramid (yielding multiple configurations of each scheme). Dotted lines show the performance of the detector in each case using the same hyperparameters initially found for the baseline. Because features differ when approximation is used, we retrained each configuration on the same dataset to obtain the true average precision for each level (indicated as “optimal” in figure), being equivalent to having a different model per configuration.

As expected, the *fast* approach is more desirable in terms of execution time. However, we find that the performance of extracted fHOG features depends on the quality of the image at higher levels, but remain robust to drastic downsampling at lower levels. Figure 2 (right) shows that, with hyperparameter tuning for each configuration, *slow* downsampling outperforms the *fast* method overall. However, this also implies a significant loss in execution speed. Nevertheless, we note that for the first few levels, the performance difference is not as pronounced.

Therefore, if the approximated gap between downsampled images is small enough, the *fast* downsampling strategy should retain enough information to minimize performance loss while offering good execution speed gains.

3.3 Histogram Downsampling

In this work, we refer to “histogram” as the frequency of gradients binned into each of the 18 orientations described by [8] that is computed *before* calculating the final 31-dimensional fHOG features, while [6] describe feature downsampling on *final* features. Downsampling final features would seem much faster, because recomputing and normalizing them is directly avoided, however it turns out that the time lost on this process is regained in our approach because the downsampling is done on 18 dimensions instead of 31. This leads to very similar runtime for both approaches, but we observe higher performance loss in scaling final features. This loss could be alleviated by smoothing the feature maps as [6] propose, but this would inevitably lead to slower runtime only to reach detection performance similar to our approach. In all experiments we use downsampling on 18-dimensional histograms and then compute and normalize the 31-dimensional fHOG features.

As with images, histograms can be downscaled using bilinear interpolation of bins between adjacent cells, either by always starting from the first level and obtaining the rest (which we dub *direct* method), or by successively obtaining level $i + 1$ from level i (*progressive* method), see Fig. 3 for clarification. Because the algorithm requires histograms for all levels, the *progressive* method yields faster overall computation time as the source histogram is smaller, but leads to additional blur that, contrary to the case of images, decreases detection accuracy.

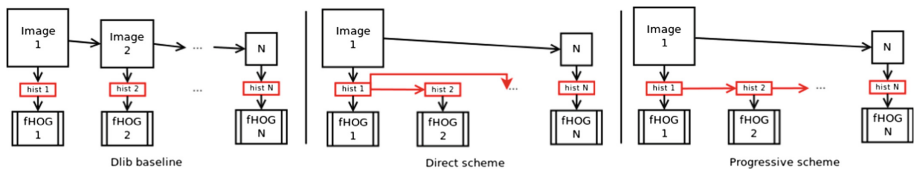


Fig. 3. Dlib baseline (left) and histogram downsampling schemes – *slow* (center) and *fast* (right) – applied up to level N .

In the following we evaluate execution speed and performance of each method, by successively applying it up to a given level ranging from 2 to 10, where 10 is actually a completely approximated pyramid, the entire pyramid has 10 levels (0–9) as in the previous results. From Fig. 4 we observe that the speed of both *progressive* and *direct* schemes are very similar, with little loss at higher levels for the *direct* histogram downsampling.

As for the *slow* and *fast* schemes, we compute the average precision of the detector using *progressive* and *direct* histogram downsampling with hyperparameters of the baseline and with best scores after a parameter search for each level. Results in Fig. 4 show that *direct* downsampling outperforms *progressive* by a small but real margin. We must note however that this advantage only appears after a few levels, where the blur introduced by the *progressive* method accumulates.

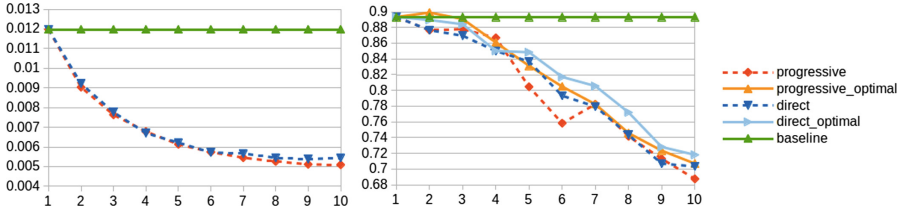


Fig. 4. Average execution time (in seconds) on modern computer (left) and average precision (right) of *progressive* vs. *direct* histogram downsampling. Dotted lines show the performance of the detector in each case using the same hyperparameters initially found for the baseline.

4 Adaptive Feature Pyramid Construction

It is clear that a trade-off exists between detection performance and the frame-rate at which the algorithm can run. While it is ideal to obtain accurate detections, in real setups such as the RoboCup competition the robot must also spend computational resources on other tasks, such as maintaining balance while walking. In fact, resource consumption varies throughout the game, depending on the situation. Therefore, it is desirable to have an adaptive control of the trade-off between accuracy and speed, while maximizing detection precision (*i.e.*, minimizing false positives).

In the previous section, we evaluated the drop in performance that comes with “skipping” feature computations up to each level of the pyramid. Meanwhile, the power law approach [6] provides a good trade-off: approximately 4% loss in average precision (in our setup, on images of robots) for almost doubling the speed of feature extraction. Here we evaluate a hybrid¹ between the skipping approach described previously and the power law based method.

We begin by skipping feature extraction up to level N exclusively, while retaining it at levels that coincide with a $1/2$ downscale of the image (octave). This way, we obtain a method that is bounded in speed and average precision by the original baseline (upper bound) and the power law based results (lower bound). The setup presented herein uses $4/5$ downscale from one level to the next, therefore octaves correspond roughly to levels 3, 6 and 9. The proposed scheme skips levels excepting those corresponding to octaves and applying the scheme up to (but excluding) level N , thus we have equivalence between levels 3–4, 6–7 and 9–10 as the same conditions are met.

We note that the approximation of levels following an octave is done using the result that was obtained from a downsampled image, therefore the quality of the histogram is superior to the case where the approximation had continued from the first level, as is the case in the previously described results.

In Fig. 5 we compare our approach with the original algorithm, power law method and the previously described level skipping strategies. As with the optimal versions of previous strategies, we performed a hyperparameter search and

¹ Full code and dataset available at: <https://github.com/polceanum/fast.fhog>.

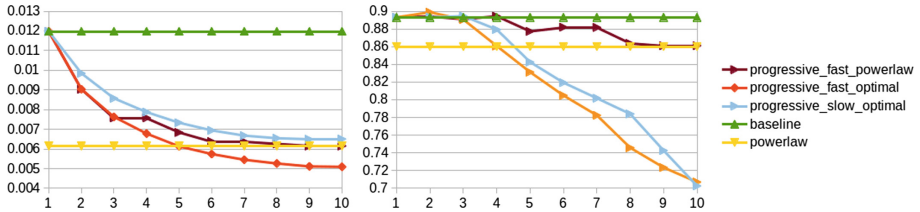


Fig. 5. Average execution time in seconds (left) and average precision (right) of studied feature approximation methods. Baseline and power law based approaches [6] shown as a straight lines, due to no level parameter.

retrained each configuration on the same dataset (equivalent to having a different model for each level to which the scheme is applied).

We observe the importance of retaining image downsampling at octave intervals as described by [6]. Applying our scheme up to level 4 does not sacrifice average precision, even though the most computationally expensive levels are approximated. At higher levels, average precision gradually decreases until it matches its lower bound, the power law baseline.

As the previous experiments have shown, gains in execution speed are significant especially for the first few levels of the pyramid. Figure 5 illustrates how execution time drops with each level, on the modern computer. At level 4, which had no average precision loss, the scheme offers 1.57x speed increase relative to the original algorithm. Increasing the level up to which the scheme is applied to 7 gives a 1.87x speed increase with only $\sim 1\%$ decrease in average precision.

We note that the *slow* image downsampling strategy could give slightly higher average precision results, but the loss in execution speed would be much higher.

5 Results

The processor equipped on the NAO v4 robot platform is, according to our estimates, approximately 25 times slower than the modern CPU on which we ran the evaluation. This is due to several factors such as lower frequency, less processor cache and other aspects which are outside the scope of this paper. These differences impose a hard standard on what algorithms can be run on this model of robots.

We evaluated the scheme on the NAO robot, obtaining gains in execution speed similar to the PC version (see Table 1). In fact, the speed increase is 1.68x without average precision loss relative to the dlib baseline, which is higher than the PC version, due to optimizations that are not available on the robot. At level 7, we obtain 1.95x speed increase with only 1.14% loss in average precision, while at level 10 (which is equivalent to the power law approach) the speed increase is 2.05x but the loss rises to 3.21%.

Table 1. Summary of proposed scheme performance. Average precision (AP), feature extraction execution time in milliseconds on modern computer (TMC) and on robot (TR), false positives per image (FPPI) and miss rate (MR) are shown for each level up to which the scheme is applied. For total detection time that includes SVM classification, add ~ 7 ms to TMC and ~ 117 ms to TR. Small FPPI variations most probably due to different SVM hyperparameter C values for each level.

Level	1	2	3–4	5	6–7	8	9–10
AP	89.3%	89.4%	89.4%	87.7%	88.2%	87.0%	86.1%
TMC	11.9	9.0	7.5	6.8	6.3	6.2	6.1
TR	298.2	227.1	177.0	162.3	152.4	148.9	145.3
FPPI	5.4×10^{-2}	8.6×10^{-2}	3.7×10^{-2}	2.1×10^{-2}	3.2×10^{-2}	4.3×10^{-2}	3.7×10^{-2}
MR	11.3%	8.6%	10.2%	11.8%	11.3%	12.9%	13.5%

We note that the time needed to compute the feature pyramid on the robot is still elevated, and thus more optimizations will be required. However, the ~ 150 ms drop with minimal loss in average precision is an important improvement in this case. To retain smooth motion and cognition, the algorithm can be broken down into multiple steps, and tracking can be performed in between. The important aspect is that the number of false positives per image is low, while some such cases are actually correct hits which were not annotated in the dataset (see Fig. 6).

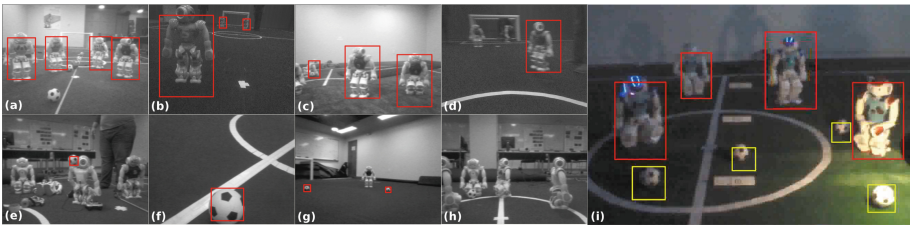


Fig. 6. Robot (a–d) and ball (e–h) detection examples and extreme lighting conditions (i). Note: true positives in (a, b); false negative in (c) (fallen robot was not annotated in the dataset; false negatives in (d, h) due to excessive blur and similar background; detection in cluttered image (e); near (f) and far (g) detections. Slight box mismatches due to sliding window size; over time, detected bounding boxes vary slightly, especially visible in the extreme lighting scenario.

Results on ball detection are also satisfactory with the ability to detect soccer balls that are close and far away. We do note however that when the background has very similar color and texture, detection does not perform well. The algorithm is relatively robust to lighting conditions, as well as to a reasonable amount of motion blur. We notice that a limitation is represented by situations in which the objects “blend in” with the background. Outside the scope of the RoboCup

competition, we expect that our approach will offer a better, adjustable trade-off between average precision and execution speed. For more difficult object detection problems, if hardware resources are more readily available, finer sampled image pyramids may improve results, and the gain from approximating intermediate levels becomes much more pronounced.

6 Conclusions and Future Work

In this work we provided a detailed evaluation of the trade-off between feature extraction speed and detector average precision, at each level of the feature pyramid. In our experiments, we used histogram downsampling instead of final feature downsampling used in related work. Results showed that this trade-off is not linear and that average precision is not lost by skipping the first few levels of the pyramid, which in fact account for a major part of the total computation time. We compared these results with the dlib library and with a state-of-the-art method based on image downsampling power law as baselines. Based on this analysis, we developed a hybrid method which is upper bounded by dlib and lower bounded by the power law approach in both execution time and average precision. We significantly improved the execution time compared to the dlib library and obtained a better trade-off than proposed by [6]. In practice, the proposed method can be adapted, by changing the level up to which it is applied, to favor average precision or execution speed. This way, on a modern computer, we obtain 1.57x increase in pyramid construction speed without any loss in average precision, $\sim 1\%$ average precision loss with 1.87x speed increase, and finally the same results as power-law approach when reaching the lower bound.

Execution speed gains are retained on the robot implementation, where we obtain 1.68x speed increase compared to the baseline with no loss and 1.95x increase with $\sim 1\%$ average precision loss, compared with 2.05x obtained with the power law baseline that presents $\sim 3\%$ average precision loss.

Following from the observation that the first few levels of the pyramid account for the majority of execution time, and that in our approach we compute the first level (level 0), extra time should be saved by upscaling level 0 from higher levels. Future work will include performance evaluation of this idea, as well as vectorizing histogram downsampling to the extent possible, including adopting complementary optimization techniques from related work. While this work improved the execution speed of the algorithm, the resulting framerate is still low on the NAO robot and requires further optimization (including vectorization which is not fully taken advantage of in this work), however it is possible to divide the feature pyramid extraction and object detection algorithm into steps that can be executed over multiple cognition cycles and couple the detection process with computationally cheaper tracking.

Acknowledgements. We thank the RoboCanes team of the University of Miami for providing their RoboCup software platform for our research. We also thank François

Lasson for help with hyperparameter search, and anonymous reviewers for their valuable feedback. The work in this paper was partially funded by the ANR project SOMBRERO (ANR-14-CE27-0014).

References

1. Albani, D., Youssef, A., Suriani, V., Nardi, D., Bloisi, D.D.: A deep learning approach for object recognition with NAO soccer robots. In: Behnke, S., Sheh, R., Sarel, S., Lee, D.D. (eds.) RoboCup 2016. LNCS (LNAI), vol. 9776, pp. 392–403. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68792-6_33
2. Brandão, S., Veloso, M., Costeira, J.P.: Fast object detection by regression in robot soccer. In: Röfer, T., Mayer, N.M., Savage, J., Saranlı, U. (eds.) RoboCup 2011. LNCS (LNAI), vol. 7416, pp. 550–561. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32060-6_47
3. Budden, D., Fenn, S., Walker, J., Mendes, A.: A novel approach to ball detection for humanoid robot soccer. In: Thielscher, M., Zhang, D. (eds.) AI 2012. LNCS (LNAI), vol. 7691, pp. 827–838. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35101-3_70
4. Cao, T.P., Deng, G.: Real-time vision-based stop sign detection system on FPGA. In: 2008 Digital Image Computing: Techniques and Applications, DICTA 2008, pp. 465–471. IEEE (2008)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893. IEEE (2005)
6. Dollár, P., Appel, R., Belongie, S., Perona, P.: Fast feature pyramids for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(8), 1532–1545 (2014)
7. Dollár, P., Belongie, S.J., Perona, P.: The fastest pedestrian detector in the west. In: British Machine Vision Conference, vol. 2, p. 7 (2010)
8. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010)
9. Genter, K., et al.: UT Austin Villa: project-driven research in AI and robotics. *IEEE Intell. Syst.* **31**(2), 94–101 (2016)
10. Gudi, A., de Kok, P., Methenitis, G.K., Steenbergen, N.: Feature detection and localization for the RoboCup Soccer SPL. Project report, Universiteit van Amsterdam, February 2013
11. Hemmati, M., Niar, S., Biglari-Abhari, M., Berber, S.: Real-time multi-scale pedestrian detection for driver assistance systems. In: 2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 1–6. IEEE (2017)
12. Iandola, F.N., Moskewicz, M.W., Keutzer, K.: libHOG: energy-efficient histogram of oriented gradient computation. In: International Conference on Intelligent Transportation Systems (ITSC), pp. 1248–1254. IEEE (2015)
13. Khandelwal, P., Hausknecht, M., Lee, J., Tian, A., Stone, P.: Vision calibration and processing on a humanoid soccer robot. In: 2010 The Fifth Workshop on Humanoid Soccer Robots at Humanoids (2010)
14. King, D.E.: Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* **10**(Jul), 1755–1758 (2009)
15. King, D.E.: Max-margin object detection. arXiv preprint [arXiv:1502.00046](https://arxiv.org/abs/1502.00046) (2015)

16. Menashe, J., et al.: Fast and precise black and white ball detection for robocup soccer. In: Akiyama, H., Obst, O., Sammut, C., Tonidandel, F. (eds.) RoboCup 2017. LNCS (LNAI), vol. 11175, pp. 45–58. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00308-1_4
17. Metzler, S., Nieuwenhuisen, M., Behnke, S.: Learning visual obstacle detection using color histogram features. In: Röfer, T., Mayer, N.M., Savage, J., Saranlı, U. (eds.) RoboCup 2011. LNCS (LNAI), vol. 7416, pp. 149–161. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32060-6_13
18. Mizuno, K., Terachi, Y., Takagi, K., Izumi, S., Kawaguchi, H., Yoshimoto, M.: Architectural study of HOG feature extraction processor for real-time object detection. In: 2012 IEEE Workshop on Signal Processing Systems (SiPS), pp. 197–202. IEEE (2012)
19. Sze, V., Chen, Y.H., Einer, J., Suleiman, A., Zhang, Z.: Hardware for machine learning: challenges and opportunities. In: 2017 IEEE Custom Integrated Circuits Conference (CICC), pp. 1–8. IEEE (2017)