Reduced Search Space for Rapid Bicycle Detection

M. Nilsson[1], H. Ardö[1], A. Laureshyn[2] and A. Persson[2]
[1]*Centre for Mathematical Sciences, Lund University, Lund, Sweden*
[2]*Traffic and Roads, Department of Technology and Society, Lund University, Lund, Sweden*

Keywords: Bicycle Detection, Search Space, RANSAC, SMQT, Split up SNoW.

Abstract: This paper describes a solution to the application of rapid detection of bicycles in low resolution video. In particular, the application addressed is from video recorded in a live environment. The future aim from the results in this paper is to investigate a full year of video data. Hence, processing speed is of great concern. The proposed solution involves the use of an object detector and a search space reduction method based on prior knowledge regarding the application at hand. The method using prior knowledge utilizes random sample consensus, and additional statistical analysis on detection outputs, in order to define a reduced search space. It is experimentally shown that, in the application addressed, it is possible to reduce the full search space by 62% with the proposed methodology. This approach, which employs a full detector in combination with the design of a simple and fast model that can capture prior knowledge for a specific application, leads to a reduced search space and thereby a significantly improved processing speed.

1 INTRODUCTION

Observing road-users is of great interest in Intelligent Transportation Systems (ITS) and an important tool in city planning. Detecting and counting bicyclists on roads are one of the many important aspects in this context. Cycling is often seen as an important part of a sustainable transport system, because cyclists do not pollute, are quite, take little space and cycling have very positive health effects (Pucher et al., 2010). Bicycle counting has usually been a difficult, expensive and labor-intensive task for road authorities. Hence, reliable and cost effective daily report of cyclist flow and frequencies at locations in a city, enables well argued decision making by municipalities to initiate new, or improve existing, infrastructure.

Several sensor technologies exist to measure traffic in general, for example, sensors based on infrared beams, passive infra-red, laser scanners, inductive loop detectors, hoses measuring air pressure and cameras using computer vision (Klein et al., 2006). This paper focuses on the latter. In particular, it focuses on designing a rapid bicycle detection framework for a specific application. The application addressed involves low resolution video, produced from a camera placed at the side of a road. The extension of the application is to investigate how wind, and potential windshileds, affect daily bicycling flow. Ongoing work is to collect video data over a year, and the aim is to investigate bicycle flows automatically with the approach described in this paper. It should be emphasized that processing speed is of great concern with one year of video data in mind.

There are a limit amount of computer vision papers focused specifically on bicycle detection and tracking (Ardeshiri et al., 2011). However, there are techniques proposed for pedestrian detection which can be tailored towards the task of bicycle detection. Some proposed methods for bicycle detection and tracking involves motion detection (Heikkila and Silven, 1999), top-view mounted stereo cameras (Belbachir et al., 2010), pedal motion (Takahashi et al., 2010), wheel extraction (Rogers and Papanikolopoulos, 2000; Ardeshiri et al., 2011) and part based learning methods (Felzenszwalb et al., 2010; Cho et al., 2011). Which all have their own merits and flaws regarding performance in detection and processing speed.

Approaches to improve detection speed in object detection, in general, focus on making improvments in the categories *features*, *classifier*, *prior knowledge*, *cascades*, *parallel (GPU) implementation* and/or *search strategy* (Benenson et al., 2012). In this paper an object detector previously successful in face detection is utilized (Nilsson et al., 2007). That object detector addresses *features*, *classifier* and *cas-*

Nilsson M., Ardö H., Laureshyn A. and Persson A. (2013).
Reduced Search Space for Rapid Bicycle Detection.
In *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods*, pages 453-458
DOI: 10.5220/0004264804530458

cades. In order to improve detection speed, this paper proposes to take advantage of additional *prior knowledge*, steaming from the particular application of bicycle detection addressed in the paper, as well as the existing detector, in order to design a practical and more efficient *search strategy*.

The paper is organized as follows. The next section discusses the bicycle detection application and the low resolution video data used. Section 2 describes the object detection framework used for bicycle detection. Section 3 presents the proposed search space reduction. Section 4 presents experimental results. Finally, conclusions are presented.

2 BICYCLES IN LOW RESOLUTION VIDEO

Due to the privacy concerns as well as sensor cost, low-resolution recordings are preferred and considered in this work. The video used is collected using an AXIS 211 surveillance camera collecting a MPG video of resolution 320×240. The camera placement is at the side of the road and bicycles to be detected are in, or close to, profile view. This camera setup yields video where bicyclists are about 40×40 pixels in size, see Fig. 1.

The low resolution introduces several restrictions to methods suitable for the task. For example basing the detection on wheels (Ardeshiri et al., 2011) will not result in a reliable detector since in many cases the circular or elliptic shapes are simply not found in the low resolution image, see for example top left patch in the right part of Fig. 1. Similarly, basing the detection on pedal motion (Takahashi et al., 2010) is not reliable since, if visible at all, the pedal motion will be too few pixels and thereby makes it indistinguishable from noise. Furthermore, in some cases the cyclists are in fact gliding by the camera due to gained momentum before entering field of view. Hence, there are no pedal motions to be detected. Basing the detection on a part based system (Felzenszwalb et al., 2010; Cho et al., 2011) might be a possible way to design the detector. However, utilizing the default high resolution settings will fail in detecting bicycles in low resolution video. For example, our tests using a part based detector (Felzenszwalb et al., 2010) trained on PASCAL VOC (Everingham et al., 2007) bicycle data applied on the low resolution videos addressed here, resulted in basically no detections. Furthermore, considering the amount of available information which can be used for detection, a part based method can be considered unnecessary complex in the scenario addressed in this paper. Hence, in this work the focus is on rapid single patch detection with the aim for real-time operation on embedded systems.

Figure 1: Example of image from low resolution (320×240) video with a bicycle to detect and examples of bicycle patches.

3 BICYCLE DETECTION USING CLASSIFIER CASCADE

The main bottleneck in the framework is indeed the detection part. Hence, rapid detection is the main concern. Using features that can be computed fast and have desirable properties with regard to illumination changes, such as Local Binary Patterns (LBP) (Ojala et al., 1994) or local Successive Mean Quantization Transform (SMQT) (Nilsson et al., 2005; Nilsson et al., 2007), are therefore of great interest. Both consist of binary patterns formed by comparing pixels within 3×3 patches. SMQT is expected to be more robust since it compares with the mean over the entire patch.

Combining features, such as LBP or SMQT variants, with a classifier cascade, allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions (Viola and Jones, 2001). An efficient classifier cascade based on the split up SNoW (Nilsson et al., 2007), which produces a cascade, is employed in this paper. It should be noted that the following search space discussion could be equally valid if another object detector, using similar a scanning windows approach, is used (Viola and Jones, 2001; Dalal and Triggs, 2005; Zhang et al., 2007; Felzenszwalb et al., 2010).

The classifier takes a patch of size M_p row pixels and N_p column pixels as input. The default search space for the object detector is typically created by resizing the original image with M rows and N columns. Hence, scanning the image with a sliding window in the original size results in $(M-M_p+1)(N-N_p+1)$ calls to the classifier. To search for various sizes of the object the original image is resize to various sizes. Consider a vector of K resize factors in relation to the original size

$$\mathbf{s} = [s_1, s_2, \ldots, K]^T, \quad (1)$$

then the image size at scale k can be found as

$$\begin{aligned} M_k &= \lfloor M \cdot s_k + 0.5 \rfloor \\ N_k &= \lfloor N \cdot s_k + 0.5 \rfloor. \end{aligned} \quad (2)$$

Note that due to rounding the true scales can now differ from s_k and also be different in M and N. The true scales factors used are now M_k/M and N_k/N. The number of calls, C_k, at scale k will be

$$C_k = (M_k - M_p + 1)(N_k - N_p + 1) \quad (3)$$

and the total number of classifier calls C with all K scales will be

$$C = \sum_{k=1}^{K} C_k. \quad (4)$$

The results after applying the detector are predicted positions of bicycles. A single detection vector d consists of a rectangle box, defined by the top left point (x_1, y_1) and the bottom right point (x_2, y_2), concatenated with the corresponding classifier value v as

$$\mathbf{d} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ v \end{bmatrix}. \quad (5)$$

The resulting detections that are overlapping are merged by a greedy non-maximum suppression.

4 REDUCED SEARCH SPACE USING PRIOR KNOWLEDGE

Prior knowledge regarding positions and scales where bicycles reasonably can occur are of particular interest in this bicycle application. Hence, information about the lane (Aly, 2008; Bao et al., 2011) or geometric layout (López et al., 2010) are of interest. With the application in mind, it is known that the camera is placed on the side of the road and it is assumed that it is observing a fairly straight piece of a path/road. However, there is no information about the position of the lane, distance between the camera and the lane, or the focal length used. The prior knowledge that the bicycle is traveling on a lane and that it is geometrically constrained is employed, but no attempt to extract information about the lane or geometric layout is performed. Rather, the intention is to utilize the (full search space) detector initially and use those detections in an analysis to reduce the search space and further utilize this information to reduce the search space for the detector. The following analysis utilizes the full detector results before performing non maximum suppression. This since non maximum suppression will reduce the number of inliers (several correct detections on a bicycle would be merged), which is undesired in the following analysis.

The way these detections, see Eq. (5), are used are as follows. The top center coordinates, $x_{tc} = (x_1 + x_2)/2$ and $y_{tc} = y_1$, as well as the bottom center coordinates, $x_{bc} = (x_1 + x_2)/2$ and $y_{bc} = y_2$, are calculated from each detection. The reason to use both top and bottom centrum points instead of the center points of the detection is that information about the scale can be captured. Additionally, note that the width of a detection, $w = x_2 - x_1$, is directly related to the scale index k.

Utilizing line models, on the detected top and bottom positions, is the first step to reduce the search space. Due to false positives from the object detector, which can be seen as outliers in position, there is a need to perform some kind of robust estimation. For example, robust estimators might be the Iterative Reweighed Least Square (IRLS) (Lawson, 1961; Burrus et al., 1994), Quantile Regression (QR) (Koenker and Bassett, 1978; Koenker, 2005) or RANdom SAmple Consensus (RANSAC) (Fischler and Bolles, 1981). In this paper RANSAC, with a threshold value θ to identify if a points fits well, is utilized for line fitting using Total Least Square (TLS) (Golub and Van Loan, 1980) error. The standard way of calculating the number of iterations for RANSAC in this case is

$$\frac{\log(1-p)}{\log(1-f^2)} \quad (6)$$

where p is the probability of finding an uncontaminated sample and f is the fraction of inliers. This estimate is overly optimistic due to the amount of inlier noise present in the application at hand. Therefore, an additional parameter, g, the probability that the inliers chosen produce a good model, is introduced yielding more realistic estimate of the number of iterations

$$\frac{\log(1-p)}{\log(1-f^2 g)}. \quad (7)$$

The result from the RANSAC fit will be two lines in the centrum of the inlier top and bottom points. In order to find the upper and lower boundaries of the inliers, additional investigation is required. For the sake of the argument lets consider the top line only. It is desired to move the top line upward to the upper boundary of the inlier points. To perform this action all points above the line are considered and their orthogonal distance to the line is collected in a set. Utilizing this set and finding the τth quantile q_τ, for a properly chosen τ, is used to move the inception of the line by adding the quantile multiplied with an additional margin factor b, that is $b \cdot q_\tau$, to the inception. Similarly, the bottom line is moved downwards by investigating the orthogonal distances of the points un-

der the bottom line. Thus, these operations will yield two lines with the purpose to capture the area in which the bicycles occur.

While the two lines capture an area in which the bicycles occur, it is also desired to investigate the scales of the detections inside this area. The different scales found within the area is captured in a histogram h_k. A threshold γ_h on the normalized histogram $h_k / \sum_{k=1}^{K} h_k$ is introduced in order to remove potential false positives, with scales not representative of bicycle size, within the area of concern. In this way, scales can potentially be removed from the search space. The remaining scales are further investigated with respect to the positions within the area and their minimum and maximum columns values are used as an additional boundary. Hence, each scale considered is now enclosed in position by four lines. Thus, the found reduced search space have potential to reduce significant amount of classifier calls which is desired from a processing speed perspective. Furthermore, the number of false positives within a frame can be lowered with this reduced search space.

5 EXPERIMENTS

MPG video recorded during daytime (8 hours) of resolution 320×240 is used for training. Bicycles are marked and used as positive samples. Negative samples are extracted from the background and increased in a bootstrapping manner. An example of a video frame can be found in Fig. 2. The patch

Figure 2: Example of image from video.

size used is $M_p = 32$ and $N_p = 32$. The full search space for the classifier cascade is built up by repeatedly downsizing with a scale factor of 1.2, that is $\mathbf{s} = \left[1, \quad \frac{1}{1.2}, \quad \frac{1}{1.2^2}, \quad \ldots\right]$ until $M_k < M_p$ or $N_k < N_p$. This approach leads to $K = 12$ scales and 167281 classifier cascade calls in the full search. An example of the full detector results from a 30min video are visualized in Fig. 3. Note that a vast amount of

Figure 3: Bicycle detections as red boxes from 30min video.

detections are correct on the lane, but there are also false positives below and above. The mid top and bottom points from detections used in the analysis can be found in Fig. 4

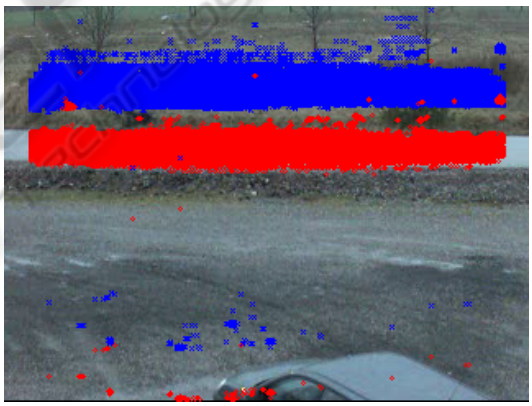

Figure 4: Top (blue) and bottom (red) points from object detector.

In the first search space reduction step, the RANSAC line fitting with parameters $p = 0.99999$, $f = 0.5$, $g = 0.1$ and $\theta = 50$ are performed on the top and bottom points. This results in lines following the center of the inliers, see Fig. 5. Following the estimate from RANSAC is the statistical analysis with quantiles described in the previous section. The parameters used are $\tau = 0.85$ and $b = 2.2$ in order to move the intercept, see Fig. 6. Utilizing the two lines capturing an area of bicycle occurrence, the scale of the detections inside the area are investigated and the histogram h_k and threshold $\gamma_h = 0.001$, described in the previous section, is utilized, see Fig. 7. The result from the histogram analysis is that, in this scenario,

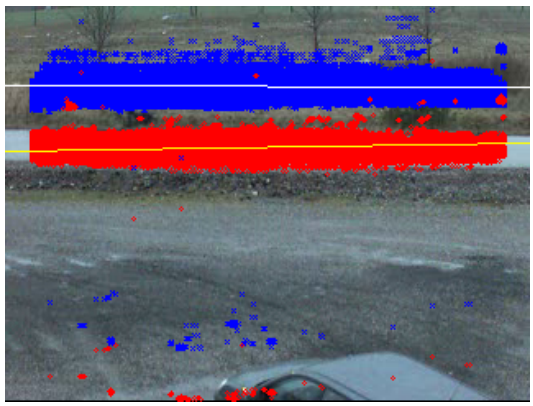

Figure 5: Top line (white) and bottom line (yellow) after RANSAC fit.

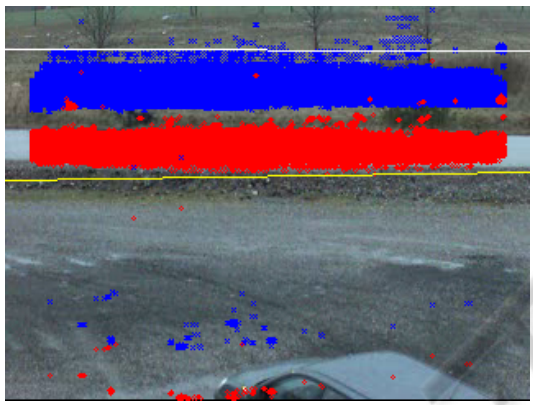

Figure 6: Top line (white) and bottom line (yellow) after moving to outer boundary.

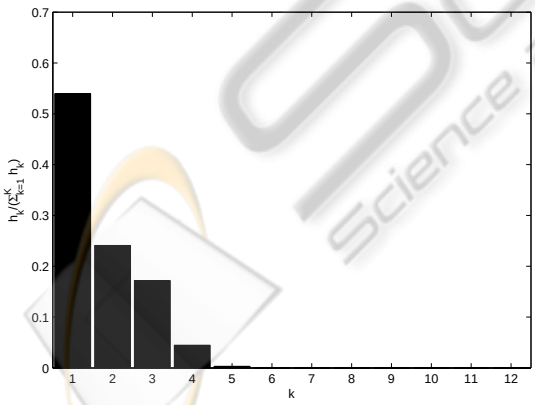

Figure 7: Normalized histogram of scales from points within the area defined by two lines found in Fig. 6.

the scales $k = 6, 7, \ldots, 12$ can be omitted. The remaining scales enclose possible positions by the two lines found from RANSAC and the two vertical lines as described in the previous section, see an example for one scale in Fig. 8.

Figure 8: Example of one scale and corresponding four lines enclosing potential positions for detection.

In the scenario addressed, the reduced search space results in 65014 classifier calls compared to the full search of 167281. Hence, the speach space be reduced by 62% by employing the proposed method. The processing speed is directly related, since a position check with four lines could be considered negligible in comparison to the classifier, to the number of classifier calls. This will lead to a similar expected processing time reduction. This will have a great practical impact when considering of a year of video data.

6 CONCLUSIONS

The paper addresses the application of bicycle detection in a specific scenario, aiming at investigating a year of video data. The paper utilizes an object detector in combination with a search space reduction method based on prior knowledge about the application. The object detector, based on local SMQT features and split up SNoW Classifier, is trained on bicycles and non-bicycles. The results from the detector is further analyzed, using RANSAC and statistical analysis, and a search space reduction is presented. Given the application addressed, is was possible to show that only around 38% of the full search space needs to be investigated and thereby a significant improvement in processing speed can be achieved.

REFERENCES

Aly, M. (2008). Real time detection of lane markers in urban streets. In *Intelligent Vehicles Symposium, 2008 IEEE*, pages 7 –12.

Ardeshiri, T., Larsson, F., Gustafsson, F., Schon, T., and Felsberg, M. (2011). Bicycle tracking using ellipse

extraction. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1 –8.

Bao, S. Y., Sun, M., and Savarese, S. (2011). Toward coherent object detection and scene layout understanding. *Image and Vision Computing*, 29(9):569 – 579.

Belbachir, A., Schraml, S., and Brandle, N. (2010). Real-time classification of pedestrians and cyclists for intelligent counting of non-motorized traffic. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 45 –50.

Benenson, R., Mathias, M., Timofte, R., and Gool, L. J. V. (2012). Pedestrian detection at 100 frames per second. In *CVPR*, pages 2903–2910. IEEE.

Burrus, C. S., Barreto, J. A., and Selesnick, I. W. (1994). Iterative reweighted least-squares design of fir filters. *IEEE Transactions on Signal Processing*, 42(11):2926–2936.

Cho, H., Rybski, P., and Zhang, W. (2011). Vision-based 3d bicycle tracking using deformable part model and interacting multiple model filter. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4391 –4398.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2007). The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

Felzenszwalb, P., Girshick, R., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627 –1645.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Golub, G. H. and Van Loan, C. (1980). An analysis of the total least squares problem. Technical report, Ithaca, NY, USA.

Heikkila, J. and Silven, O. (1999). A real-time system for monitoring of cyclists and pedestrians. In *Visual Surveillance, 1999. Second IEEE Workshop on, (VS'99)*, pages 74 –81.

Klein, L. A., Mills, M. K., Gibson, and P., D. R. (2006). *Traffic Detector Handbook: Volume I and II*. U.S. Dept. of Transportation, Federal Highway Administration, Research, Development, and Technology, Turner-Fairbank Highway Research Center, McLean, VA, 3rd edition edition.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. ISBN 0-521-60827-9.

Koenker, R. and Bassett, Gilbert, J. (1978). Regression quantiles. *Econometrica*, 46(1):pp. 33–50.

Lawson, C. L. (1961). *Contribution to the Theory of Linear Least Maximum Approximations*. PhD thesis, University of California, Los Angeles, Calif.

López, A., Serrat, J., Cañero, C., Lumbreras, F., and Graf, T. (2010). Robust lane markings detection and road geometry computation. *International Journal of Automotive Technology*, 11:395–407.

Nilsson, M., Dahl, M., and Claesson, I. (2005). The successive mean quantization transform. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 429–432.

Nilsson, M., Nordberg, J., and Claesson, I. (2007). Face detection using local smqt features and split up snow classifier. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.

Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition - Conference A: Computer Vision Image Processing*, volume 1, pages 582–585 vol.1.

Pucher, J., Dill, J., and Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine*, 50:S106–S125.

Rogers, S. and Papanikolopoulos, N. (2000). Counting bicycles using computer vision. In *Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE*, pages 33 –38.

Takahashi, K., Kuriya, Y., and Morie, T. (2010). Bicycle detection using pedaling movement by spatiotemporal gabor filtering. In *TENCON, IEEE Region 10 Conference*, pages 918 –922.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518.

Zhang, L., Chu, R., Xiang, S., Liao, S., and Li, S. Z. (2007). Face detection based on multi-block lbp representation. In Lee, S.-W. and Li, S. Z., editors, *ICB*, volume 4642 of *Lecture Notes in Computer Science*, pages 11–18. Springer.