A New Pedestrian Dataset for Supervised Learning

Gary Overett, Lars Petersson, Nathan Brewer, Lars Andersson and Niklas Pettersson
National ICT Australia
Locked Bag 8001, Canberra, Australia
{gary.overett,lars.petersson,nathan.brewer,lars.andersson,niklas.pettersson}@nicta.com.au

***Abstract*— This paper presents a comparative analysis of different pedestrian dataset characteristics. The main goal of the research is to determine what characteristics are desirable for improved training and validation of pedestrian detectors and classifiers. The work focuses on those aspects of the dataset which affect classification success using the most common boosting methods.**

Dataset characteristics such as image size, aspect ratio, geometric variance and the relative scale of positive class instances (pedestrians) within the training window form an integral part of classification success. This paper will examine the effects of varying these dataset characteristics with a view to determining the recommended attributes of a high quality and challenging dataset. While the primary focus is on characteristics of the positive training dataset, some discussion of desirable attributes for the negative dataset is important and is therefore included.

This paper also serves to publish our current pedestrian dataset in various forms for non-commercial use by the scientific community. We believe the published dataset to be one of the largest, most flexible, and representative datasets available for pedestrian/person detection tasks.

I. Introduction

The majority of current visual detection systems are based on supervised or semi-supervised learning using marked up training datasets. While much research is committed to the task of improving learning on a given dataset the actual nature and suitability of the training set is seldom, if ever, examined. This is probably because the collection, markup and extraction of datasets is often set aside as a necessary but unimportant part of the research task, with researchers rushing to progress to the 'actual' research.

Consequently, little research exists to help the scientific community formulate their training datasets, and many important questions remain unanswered. For example, pedestrian shape and pose variations necessitate that the bounding rectangle (of a training image patch) include a certain amount of background/non-pedestrian pixels. Yet no literature exists to help us choose a suitable aspect ratio for the bounding rectangle or an appropriate amount of background to include. Indeed, it is possible that patterns exist in the background pixels. For example, pedestrians tend to be found on smooth planar surfaces, therefore it is possible that including non-pedestrian image data around and below the feet of pedestrians is worthwhile.

National ICT Australia is funded by the Australian Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence Programs.

There are very few pedestrian datasets available for research benchmarking or for use as off-the-shelf training sets. An early pedestrian dataset, which has been available for use, is the MIT training database [1], [2] which contains 509 training and 200 test images for validation. Apart from being very few in number the set expresses a relatively small range of poses with only front or back views of pedestrians. Such small training sets quickly overtrain and comparative results are difficult to produce once detectors reach a certain level where the training set is almost completely solved. This was found to be the case by Dalal & Triggs [3], who's classifier exhausted the MIT dataset, thus necessitating a larger more difficult replacement set.

The new INRIA dataset, produced in [3], is freely available in a format which can be used with relative ease in most training tasks. It consists of 1239 positive training images which, with their left-right reflections, make a set of 2478 images. Validation data consists of 566×2 pedestrian images. The set is substantially more difficult because it explores the full range of upright pedestrian poses, from walking to stationary, and from arbitrary directions. The dataset comes with 1218 negative (pedestrian free) images which are large enough to allow for resampling to arbitrarily large negative training sets. The INRIA dataset is currently a popular benchmarking dataset having been used in numerous comparative pedestrian detection studies in CVPR 2007 [4], [5].

Other small, purpose-specific training sets can be found, such as USC pedestrian sets A and B [6] with 313 and 217 pedestrians respectively. Set A provides partially occluded humans while Set B provides a dataset of inter-human occlusions. A third set, USC pedestrian set C [7], provides another 232 unoccluded persons.

The work in this paper has been conducted as part of a smart cars project [8] at the Australian National University/National ICT Australia. One of the aims of this project is greater vehicle perception of target classes such as road signs and pedestrians.

Within the project we have developed a flexible training data extraction tool called TDB (Truth DataBase). This tool allows for rapid configured extraction of tailor-made training data based on previously marked up input images. See Figure 1 in Section II. Easy markup and extraction has allowed the creation of a much larger pedestrian dataset. Consisting of 25551 individual pedestrian images. With mirroring of the pedestrian images this allows for a dataset of 37344 positive training images up to a scale of 80 pixels in height[1], with 6879 validation images free of mirroring[2]).

In [9], it is shown that the nature of common weak learners for boosting means that a lack of training data can have very adverse affects on the overall classification results. A large amount of validation data is also important in drawing robust comparisons between different methods (benchmarking) and in measuring success for very powerful classifiers.

II. TDB - The Truth Database

TDB is designed to facilitate a flexible dataset extraction process. Its key features to date are:

- Ability to extract positive class instances at multiple scales
- Numerous options for adjusting the position of pedestrians within the extracted rectangle and adjusting the amount of background pixels surrounding the pedestrians.
- Ability to extract pedestrians with random values inserted for some parameters, e.g. randomly rotate images according to some distribution or randomly shift/translate pedestrians within the positive detection window.
- Ability to extract consistent data sets at different scales. A problem arises when extracting data at different sizes. It is possible that the database carries 25982 potential pedestrians at least 40 pixels in height but only 25551 pedestrians of at least 80 pixels in height. If we are to compare these two scales it is desirable to have TDB export the same 25551 pedestrians for both scales. This is possible by extracting the 40 pixel high pedestrians with a required source height of 80 pixels.
- Ability to extract pedestrian data into separate folds/sets, with the added constraint that all pedestrians from a single source image in the TDB database are placed into the same fold. This guards against pollution across folds. Thus any fold can be used as validation for K-folds validation and the like.

Unfortunately, at this stage, it is not possible to publish TDB source and due to privacy reasons we cannot release the full resolution marked up database. However, we are able to publish numerous datasets with different attributes for general research use. TDB's ease of use also means that we are able to take requests from the scientific community for made-to-order datasets below a given resolution. An example screen shot from the TDB interface is shown in Figure 1.

As research continues and as our pedestrian dataset continues to grow TDB is improved to include the latest information for extracting the best possible datasets for pedestrians.

[1]The number of pedestrians available for extraction depends on the size of the desired extracted pedestrians. Thus, if pedestrians of a smaller size are extracted, TDB may find more than 25551 instances.

[2]Validation pedestrians are not mirrored in order to offer the purest validation sets possible.

Fig. 1. A typical TDB view of marked up data showing speed signs and pedestrians.

III. Method

Currently, the most popular supervised learning method for this kind of classification task is boosting. While many boosting algorithms have been created, the real-valued variant of Adaboost [10] (RealBoost [11]) is currently the most widely accepted. Its formulation of confidence-rated predictions is also a core starting point of many other popular boosting variants, such as, LogitBoost [12] and WaldBoost [13].

As such, performing our experiments on RealBoost classifiers is a natural choice, which gives our results the most relevance to other researchers. See Algorithm 1.

$trainClassifiers(X, Y, F)$:
$X = \{x_1, x_2, ..., x_N\}$, the set of example windows
$Y = \{y_1, y_2, ..., y_N\}, y_i \in -1, 1$, the corresponding labels
$D_1(i) = 1/N$, the set of training weights
For $t = 1, ..., T$ (or until the desired rate is met)

1) Let $F_t = \{f_1, f_2, ..., f_M\}$ be the new set of features for this round.
2) Train classifiers h_j using distribution D_t. The classifier takes on two possible values: $h_+ = \frac{1}{2} ln\left(\frac{W_{++}}{W_{+-}}\right)$ and $h_- = \frac{1}{2} ln\left(\frac{W_{-+}}{W_{--}}\right)$ for positive and negative examples respectively. W_{pq} is the weight of the examples given the label p which have true label q.
3) Select the classifier h_t which minimises

$$Z_t = \sum_{i=1}^{N} D_t(i) exp(-y_i h_t(x_i)) \quad (1)$$

4) Update distribution $D_{t+1}(i) = \frac{D_t(i) exp(-y_i h_t(x_i))}{Z_t}$

The final strong classifier (cascade stage) is

$$H(x) = sign\left(\sum_{t=1}^{T} h_t(x)\right) \quad (2)$$

Alg. 1: RealBoost

Producing comparable results in RealBoost can be difficult. RealBoost is a 'greedy' algorithm and is therefore prone to a certain degree of randomness. Small selection differences at early rounds can significantly change the final learnt classifier. This effect is minimized if one uses only a single stage classifier rather than attempting to build a cascaded classifier [14]. Using a single stage classifier leads to less robust, and slower, classification results. However, a principle aim of this work is to yield *comparative* results to

TABLE I

DATASETS USED FOR TRAINING EXPERIMENTS

Dataset Group	Values
Sizes	$(8 \times 20), (16 \times 40), (32 \times 80)$
Wide Aspect Ratio Sizes	$(16 \times 20), (32 \times 40),$ (64×80)
Relative Scale ($\frac{\text{patch\_height}}{\text{person\_height}}$)	$\frac{95}{100}, \frac{100}{100}, \frac{110}{100}, \frac{120}{100}, \frac{160}{100}, \frac{200}{100}$
Random Translations [a] **(%)**	0-20% in 2% intervals
Random Rotations (degrees)	0-16° in 2° intervals

[a]Translations are as a percentage of the patch dimensions, e.g. 10% extracts images which are horizontally shifted by 10% of the patch width (32 pixels) and vertically shifted by 10% of the patch height (80 pixels).

test different datasets, not to generate classifiers to surpass the current benchmark results.

Revealing consistent differences between subtly different datasets can require a fairly large training stage. Yet, training time increases linearly with the number of rounds added to a RealBoost stage. Given this trade-off, we have selected to perform all experiments up to 100 RealBoost rounds. This allows enough rounds to tease out most training differences in reasonable training time. Some differences are apparent with very few rounds and investigation of these behaviours is often useful, especially if conclusions are to be drawn for use in shorter cascade classifiers in the future. For this reason we also examine the ROC performance curves at rounds 1, 5, 20, 50 and 100. Table I shows the different training datasets used.

To avoid an explosion in the number of required experiments, we do not create datasets using combinations of different group attributes, e.g. no translated and rotated dataset is extracted.

For each dataset, we produce a training and validation subset. TDB ensures that there is no overlap between source images extracted for training and validation. Additionally, the source images for the training data in one dataset group and another are guaranteed alike. This allows us to swap validation sets between groups, e.g. we may take the classifier trained on random translations of 6% and validate this classifier on the 2% random translation validation data without fear of using a polluted training set.

This allows 3 kinds of analysis:

1) **Like Validation:** For a classifier trained on a given training set, validation is performed using the like validation set, e.g. if the training data was comprised of random rotations of $c°$ the validation also consists of the same $c°$ random rotations. For randomly perturbed datasets (rotated or translated) this analysis provides information about the importance of accuracy in data markup and collection, e.g. is it important to markup pedestrians locations with high pixel accuracy and without camera rotation? For dataset groups where the size and aspect ratio of the training set are explored we can only use *like validation*.
2) **Pure Validation:** For a classifier trained on a given dataset (with random perturbation) we swap in the unchanged/pure dataset, e.g. consider a classifier trained on randomly translated pedestrians which is validated on the pure untranslated data. This reveals if there are any classification benefits to randomly shifting the training data, i.e. does the randomly perturbed dataset generalise better so as to improve results on the pure validation task?
3) **Pure Training with Perturbed Validation:** For a classifier trained on pure unperturbed data, we swap in a randomly perturbed validation set to examine the robustness of the classifier to input noise. In this way we are able to determine how quickly performance drops off if randomly perturbed data is given as input.

TABLE II

EXPERIMENTAL PARAMETERS

Parameter	Experimental Value
Booster	RealBoost
Rounds	up to 100
Class Type	Pedestrians
Size	32×80
Relative Scale	$\frac{120}{100}$
Random Translations	none
Random Rotations	none
Weak Classifier Learner	SRB Method [9]
# Haar-like Features	100K (monochrome)
# Train Positives	37344
# Train Negatives	37344
# ROC Validation Positives	6879
# ROC Validation Negatives	6879

This results in far more ROC analysis curves than can be shown in this paper. The most interesting curves will be included and important insights from the full range of experiments will be described.

Table II shows the default parameters of all experiments. These are the parameters used if no change is specified by Table I. The use of 100K randomly generated monochrome Haar-like features is deliberately very large compared to typical boosting experiments. It is important that the pool of potential features is oversampled to ensure that full space of the features over training patches is explored regardless of size. For example, if 5K features were produced for 8×20 and 64×80 image patches, the larger image patches would be relatively undersampled. Note that the set of features F is renewed for each round t of RealBoost, see Algorithm 1.

Random translations and rotations are drawn from a uniform distribution within the target range, e.g. a TDB extraction can specify random rotations within the range $[-2°, 2°]$. Initially a Gaussian distribution was used for extraction. However, RealBoost's weighting scheme quickly applies the most weight to the outliers of the distribution (e.g. the 5% most rotated training images). This is undesirable for a training set. Instead the uniform distribution (with hard edges) allows us to observe the effect if images are distorted up to some limit. This information is useful to markup operators, who can be aware of the required accuracy limits in their data markup to achieve good results.

IV. Experiments

A. Image Sizes and Aspect Ratios

The *size* of extracted pedestrian image patches determines the level of person detail which can be learnt by a classifier. At very small scales there will be no opportunity for the classifier to learn fine person details such as facial features. In general, we expect classifiers to achieve better results on larger images. However, it is highly desirable to use the smallest training size which gives good results. This gives two important advantages. Firstly, most learning algorithms will have speed and memory benefits from smaller training images. More importantly, however, the final detector will be able to detect pedestrians at a lower resolution. This can be used to lower the resolution of input detection video sequences or allow detection of pedestrians shrunk by perspective. The resulting ROC curves at 100 training rounds are shown in Figure 2.

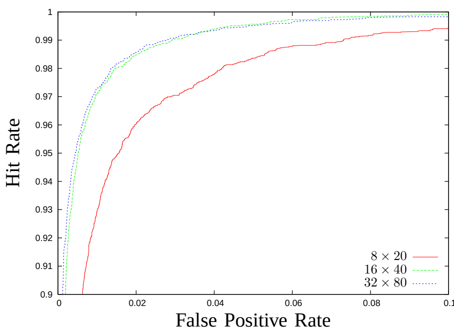

Fig. 2. ROC comparison curves at round 100 for different dataset size. As expected, larger sizes perform best. However, this improvement is less significant beyond a patch height of about 40 pixels. This trend is very clear through all stage sizes, i.e. it is already apparent from the first round where the classifier consists of just a single feature.

The *aspect ratio* of the training set is also found to be quite important in the early rounds of a classifier (i.e. few features) with lower resolution images. We believe this is largely because of the simple shape of the basic Haar-like feature types, see Figure 3. In order for the learning algorithm to observe pedestrian patterns against the non-pedestrian background it is important that there is enough space to fit the Haar-feature well to the pedestrian and its background space. This requires a non-minimal amount of background, see Figure 4.

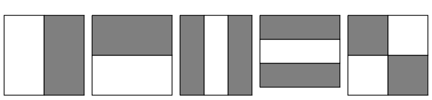

Fig. 3. Basic Haar-like Feature types

To show the effect of different aspect ratios we trained classifiers using a variety of wider and taller aspect ratios. In Figure 5 we see that *wider aspect ratios* are beneficial only for lower resolutions and only for small classifiers. It is likely that at higher resolutions other features which do not cover the space around the pedestrian are favoured by RealBoost.

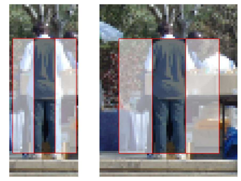

Fig. 4. The importance of non-minimal backgrounds for Haar-like feature type recognition. Clearly the wider image patch allows for a stronger fit of the Haar-feature to the pedestrians shape. Figure 5 shows the effect this has in the early rounds of low resolution images.

Having found that at higher resolutions widening the aspect ratio had little effect it could be expected that similar results might be revealed for *taller aspect ratios*. However, improved ROC performance was found. This was most significant if the aspect ratio height was increased at the bottom of the source images. I.e. to include image data below the feet of the pedestrian. An easy explanation might be that pedestrians are often found on relatively smooth planar surfaces and that RealBoost may learn this as a cue to further improve performance. However, this warrants more investigation as the TDB extraction process can enlarge a source patch to beyond the edge of the source image resulting in black 'background' pixels. As the feet of pedestrians are often at the base of images this could be a source of bias in the training sets. See Figure 5.

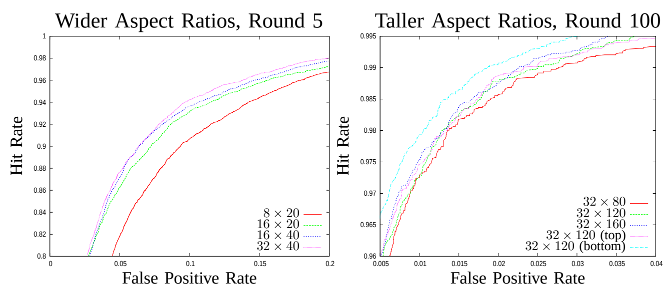

Fig. 5. ROC comparison curves for wider and taller aspect ratios. For lower resolution images, it is also clear that the inclusion of non-pedestrian background space, through a wider aspect ratio, significantly improves the recognition. However, it was noted in further experiments that this effect quickly drops off and makes little difference beyond 20 rounds. For taller aspect ratios it appears that inclusion of background space at the bottom of the image patch may be more important than at the top.

Of course, it is also possible to extend the Haar-like feature set in Figure 3 to include feature types which fit pedestrians with less background space required. Though, introducing multiple less-simple feature types introduces other complications, such as, a dilution of the spatial density explored by the randomly generated feature set.

B. Relative Scale

Like the aspect ratio, the relative size of pedestrians within the training patch alters the amount of background space and changes the level of pedestrian detail. For this experiment set we use a number of training datasets with varying relative scale. We also include the unity scale, where pedestrians are the same height as their bounding patch, and an undersized patch, where the pedestrian is slightly taller than the patch height. See Figure 6.

Fig. 6. Pedestrians extracted at a range of relative scales. Relative scales are labelled in the form patch_height/person_height.

Figure 7 shows the results for these classifiers using *like validation*. From these results we can see that an optimal choice of relative size will depend on the size of the classifier in terms of features and, if cascade methods are used, the size of the individual stages are important. The most apparent effect is that including no background space around the pedestrian results in consistently worse classification. We are hesitant to conclude that relatives scales with much larger patch sizes are favourable as this result may include the same bias as seen in Section IV-A, where source patches around the pedestrian include black pixels because the source patch reaches beyond the source image border.

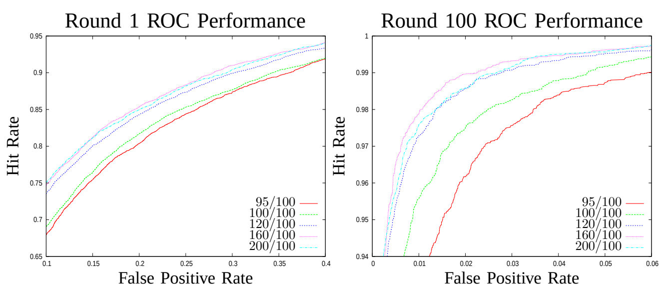

Fig. 7. ROC comparison curves shown at the 'knee' for RealBoost rounds 1 & 100 using different relative sizes. In the early RealBoost rounds the greatest classification success is found when selected patch sizes are much larger than the pedestrian, i.e. lots of background information is included. As the training rounds progress, the results favor a slightly more moderate scaling of pedestrians within the patch. If patch height is equal or less than pedestrian height the results are consistently far worse across all rounds.

ROC curve performance was also analysed using *pure training on perturbed validation* analysis. The main aim of this was to determine the rate at which a trained classifiers performance drops off as pedestrian size is allowed to shift. This is important in predicting the quality of pedestrian detection when classifying data using a scale-space pyramid to detect pedestrians at different sizes in the image. There is not room in this paper for a full analysis of these results. However, it was found that changes in pedestrian size of just 5% leads to a substantial and rapid drop off in detection rates. This suggests that classifiers, for detecting multi-scale pedestrians using a scale-space pyramid architecture, should use a scale change of less than 5% between levels.

C. Random Translations

The position of pedestrians within the training and validation patches will affect Haar-like feature weak learning. It is expected that with greater variance in the pedestrians position that correct classification rates will drop.

Classification rates on *like validation* for translations up to 4% are not significantly poorer. This indicates that the classification success is only moderately affected by random translations to validation and training data. However, if the classifier is only trained on pure training data the validation results quickly drop off with perturbed validation. This suggests that the classifier gains little from training on unperturbed data but loses robustness on perturbed validation tasks. This means a classifier can be trained/created on randomly perturbed data of less than 4% and then used to classify pedestrians using a larger step size, e.g. not searching at every pixel step across the image.

Figure 8 shows the different classifiers being tested on *pure validation* data. This shows that the dataset with random translations of 2% slightly outperforms its counterparts suggesting that its learning has been more general while the classifier trained on pure data is slightly too specific.

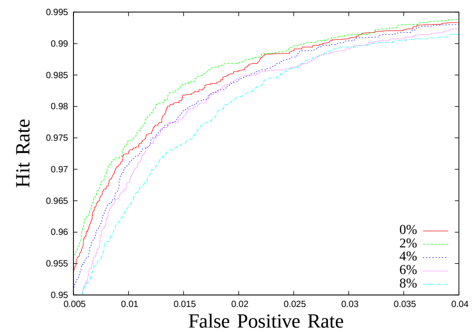

Fig. 8. Pure validation ROC comparison curves shown at the 'knee' of a 100 feature classifier. The ROC curve for the classifier trained on the 2% random translation dataset weakly dominates all other classifiers. This suggests that a randomly translated training set can help RealBoost to learn a more general classification rule.

D. Random Rotations

The orientation of a pedestrian in an image will affect the relative location of pedestrian pixels to each other and to background pixels within a patch. For this experiment we will use a number of datasets built with different degrees of random rotation, ranging from $0°$ to $16°$.

Figure 9 shows the *pure training on perturbed validation* analysis. Rotations of 4 degrees or less in the validation set are not significant to performance. If larger rotations are applied the performance begins to drop off. This result also suggests that detectors will perform best if input cameras are kept fairly level.

Pedestrians generally exhibit a high degree of lean and tilt as they move and walk. Thus, even the 'non-rotated' datasets probably contain a fair amount of natural rotation. This may act to ensure that the pure training set is already populated with a fair variety of random pedestrian rotational angles. Despite the natural rotation in the pure data set we still find that applying random rotations to the training sets can help the classifier toward greater generalised performance. For example, using *pure validation* analysis we found that the classifier trained on data perturbed by up to 4 degrees had a very similar ROC performance to the classifier trained on unperturbed data. This suggests that perturbing by small amounts has little effect on performance.

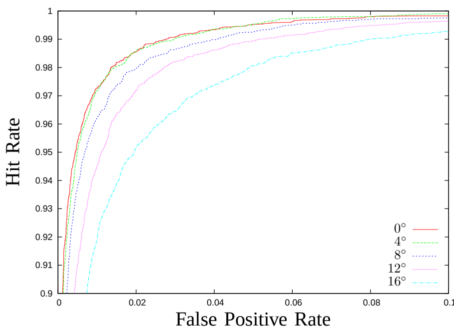

Fig. 9. ROC comparison curves shown at the ‘knee’ of a 100 feature classifier. The classifier is trained on non-rotated data with randomly rotated validation sets. Small rotations do not appear to adversely affect classification.

V. Dataset Availability

The NICTA[3] Dataset is made freely available to the scientific community via the organisation’s website at `www.nicta.com.au/computer_vision_datasets`. Any publications resulting from the use of this dataset, and derivative forms, should include the appropriate acknowledgement. For further information, please visit the website.

The final dataset contains 25551 unique pedestrians, allowing for a dataset of over 50K images with mirroring. Additionally, TDB allows the generation of several permutations per source image in order to further bolster the training set. Large negative datasets will also be provided, although researchers training cascaded classifiers may require their own bootstrapped negative datasets. Apart from the datasets described in this paper the authors are willing, within reason, to produce further pedestrian datasets for the scientific community.

Figure 10 shows a selection of pedestrians from the dataset. Most images were captured using normal digital camera hardware in normal urban environments, in multiple cities and in different countries.

Fig. 10. A selection of 32x80 color images from the NICTA dataset.

The negative set is drawn from a set of 5207 high resolution pedestrian free images in varied environments. Both the negative and positive sets are divided into unique folds for tasks such as validation.

VI. Conclusion

The effect of various pedestrian dataset attributes has been examined. By performing a large number of training and validation experiments on various forms of the same dataset, we have determined a number of desirable dataset characteristics.

It has been found that *larger patch sizes* (40+ pixels in height) yield better results. For Haar-like feature types it is found that *wider aspect ratio patch sizes* outperform narrower aspect ratio patches for low resolution images with small classifiers. The *relative size* of pedestrians within the pedestrian patches also determines classification success. Patches should not be the same height or smaller than the pedestrians they contain. That is, they should include a significant amount of background image around the pedestrian. RealBoost classifiers with few features (less than 10), or cascade stages made up of few features, may benefit from datasets with pedestrians at a relatively small scales compared to their bounding patch.

Applying random translations to the location of pedestrians within their bounding patch does not significantly affect detection until the random translations exceed 4% of the patch width and height. There is some evidence that training generalisation improves if random translations are inserted into the training set. If training is performed on unperturbed training data, classification performance can be expected to drop off more quickly if input patches to a given classification task contain such translations.

Pedestrian rotations of less than 6° in training and validation are not particularly significant.

The NICTA pedestrian dataset has been made available for non-commercial use in various forms for researchers requiring a large, challenging and flexible dataset for supervised learning applications. It is hoped that this dataset is used in providing quality classifier training and classifier benchmarking data in future research.

[3]National ICT Australia

References

[1] A. Mohan, C. Papageorgiou, and T. Poggio, “Example-based object detection in images by components,” *PAMI*, vol. 24, no. 4, pp. 349–361, April 2001.

[2] C. Papageorgiou and T. Poggio, “A trainable system for object detection,” *IJCV*, vol. 38, pp. 15–33, 2000.

[3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Computer Vision and Pattern Recognition*, vol. 01, pp. 886–893, 2005.

[4] O. Tuzel, F. Porikli, and P. Meer, “Human detection via classification on riemannian manifolds.” in *CVPR*. IEEE Computer Society, 2007.

[5] P. Sabzmeydani and G. Mori, “Detecting pedestrians by learning shapelet features,” in *CVPR*, 2007.

[6] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors,” in *Proceedings of the IEEE International Conference on Computer Vision*, vol. I, oct 2005, pp. 90 – 97.

[7] ——, “Cluster boosted tree classifier for multi-view, multi-pose object detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2007.

[8] L. Petersson, L. Fletcher, A. Zelinsky, N. Barnes, and F. Arnell, “Towards Safer Roads by Integration of Road Scene Monitoring and Vehicle Control,” *The International Journal of Robotics Research*, vol. 25, no. 1, pp. 53–72, 2006. [Online]. Available: http://ijr.sagepub.com/cgi/content/abstract/25/1/53

[9] G. Overett and L. Petersson, “Improved response modelling on weak classifiers for boosting,” *IEEE International Conference on Robotics and Automation*, 2007.

[10] Y. Freund and R. E. Schapire, “A short introduction to boosting,” *Journal of Japanese Society for Artificial Intelligence*, vol. 14, pp. 771–780, sept 1999.

[11] R. E. Schapire and Y. Singer, “Improved boosting using confidence-rated predictions,” *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999. [Online]. Available: citeseer.ist.psu.edu/article/singer99improved.html

[12] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: a statistical view of boosting,” *The Annals of Statistics*, vol. 38, no. 2, pp. 337–374, 2000.

[13] J. Sochman and J. Matas, “Waldboost - learning for time constrained sequential detection,” *Computer Vision and Pattern Recognition*, vol. 2, pp. 150– 156, 2005.

[14] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Computer Vision and Pattern Recognition*, vol. 01, p. 511, 2001.