Learning Pre-attentive Driving Behaviour from Holistic Visual Features

Nicolas Pugeault* and Richard Bowden

Centre for Vision, Speech and Signal Processing,
University of Surrey, UK
{n.pugeault,r.bowden}@surrey.ac.uk
http://www.ee.surrey.ac.uk/CVSSP/

Abstract. The aim of this paper is to learn driving behaviour by associating the actions recorded from a human driver with pre-attentive visual input, implemented using holistic image features (GIST). All images are labelled according to a number of driving–relevant contextual classes (eg, road type, junction) and the driver's actions (eg, braking, accelerating, steering) are recorded. The association between visual context and the driving data is learnt by Boosting decision stumps, that serve as input dimension selectors. Moreover, we propose a novel formulation of GIST features that lead to an improved performance for action prediction. The areas of the visual scenes that contribute to activation or inhibition of the predictors is shown by drawing activation maps for all learnt actions. We show good performance not only for detecting driving–relevant contextual labels, but also for predicting the driver's actions. The classifier's false positives and the associated activation maps can be used to focus attention and further learning on the uncommon and difficult situations.

1 Introduction

The objective of this manuscript is to learn the relationship between behaviour and visual stimulus in the context of driving. This is an extremely complex task due to variability in both the visual domain as well as the actions performed by the driver. Such actions are arguably dependant upon high level reasoning and context. However, we demonstrate that pre-attentive vision based upon simple holistic descriptors can account for the majority ($\sim 80\%$) of a driver's actions using minimal training ($< 1\%$).

The act of driving require little active attention for an experienced driver, allowing extended driving periods of several hours while at the same time having a conversation, thinking about an itinerary, etc. Indeed, this fact is a source of hazard, as an inattentive driver is less likely to react to unexpected emergencies. This article studies how pre-attentive visual perception can be used to learn aspects of driving behaviour by observing a human driver, releasing attention for other tasks such as tracking, traffics sign recognition, planning, etc. The

* Corresponding author.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part VI, LNCS 6316, pp. 154–167, 2010.

learning is performed by recording the driver's actions (eg, braking, steering) at each frame together with a coarse labelling of each frame according to a set of driving contextual categories (eg, motorway, junction, pedestrian crossing). We choose to use holistic image features (so-called GIST) as a functional equivalent to pre-attentive vision in humans. GIST are a class of visual descriptors that encode a global representation of a visual scene's content, as opposed to local image features. This holistic aspect, together with the low resolution it requires, is consistent with the visual signal processed by the periphery of the retina in the absence of (relevant) gaze fixation. This is in stark contrast with feature–based methods that rely on high resolution extraction of sparse descriptors, and therefore belong to attentive vision.

Holistic representations of visual scenes have received a lot of attention during the last decade [1,2,3,4]. The rationale behind the use of holistic image descriptors for visual context description is that they are insensitive to the small variations that abound in complex scenes and hamper classification based on local features. This is especially critical in urban scenes, where the amount of visual information and variability is enormous. The original version of the GIST was proposed by Oliva & Torralba, who compared two descriptors based on the Fourier transform of image intensity [1]. The first one was based on the Fourier transform computed on the whole image (DST); the second is based on a windowed Fourier transform (WDST), localised on a coarse 8 × 8 grid. The latter was shown to contain more information than the first, and was used to define a set of perceptual properties (roughness, ruggedness, etc.) that allow for scene classification. In later publications by the same authors, the Fourier transform was replaced with steerable [2,5], or Gabor wavelets [3], computed over varying scale and orientation and averaged over grids of varying sizes. The dimension of the feature vector was in some case reduced using PCA [6,3]. Renninger & Malik studied how human subjects could identify visual scenes even after very brief exposures (< 70ms), and proposed a GIST–like model as an explanation of those results [6]. Douze et al. compared GIST descriptors with bag-of-words approaches for image search, using the INRIA 'Holidays' and 'Copydays' datasets, and found that GIST descriptors yield lower performances than state of the art bag-of-word approaches, yet with a considerably lower computational and memory cost [4]. Siagan & Itti, used similar descriptors for the identification of indoor and outdoor scenes in a mobile robotics context [3,7]. Their implementation differs insofar as they use different filter banks, including centre-surround colour sensitive filters, and the resulting feature vectors were post-processed using PCA and ICA. Ackerman & Itti used spectral image information for steering a robotic platform on a path following scenario on two simple tracks [8]; in contrast, we consider a large database of real urban scenes. Kastner et al. [9] use a GIST variant for road type context detection, limited to the three categories 'highway', 'country road' and 'inner city'; their main contribution was the hierarchical principal component classification (HPCC).

In contrast, in this article we attempt to detect 13 contextual labels of varying difficulty pertaining to scene environment, road type, junction type along with

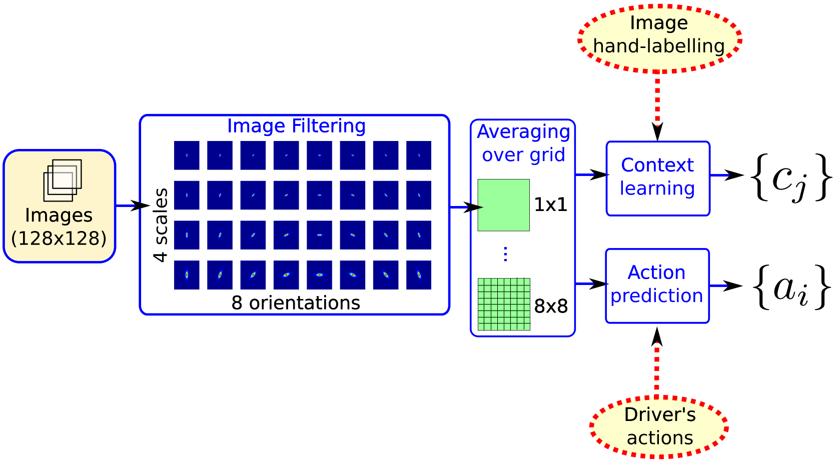


Fig. 1. Overview of the pre-attentive driving behaviour learning framework

some other attributes. Moreover, we learn relations between the visual context and five of the driver's actions: the activation of each of the three pedals, plus steering. We then show how these classifiers can be reversed to provide activation maps that determine the salient visual information that influences each action. The framework we propose is illustrated in Fig. 1: images are first resized and the contrast is normalized, then they are convolved with a filter bank, and the response is averaged over a grid; this forms the GIST descriptor. Then, two experts are learnt from these descriptors: the first one learns to detect contextual categories using hand labelled training samples; the second learns to predict the driver's actions. In this graph, the red dotted arrows represent information that is only provided at the training stage.

2 Methods

In this section we describe the learning framework illustrated in Fig 1: first, in section 2.1 we describe the GIST descriptor used, and propose a novel formulation of the descriptor; second, in section 2.2 we briefly discuss the learning algorithm.

2.1 Holistic Image Descriptors (GIST)

GIST are holistic image descriptor that encode a whole visual scene in one feature vector, generated by a coarse scale local filtering of a low resolution version of the image. The exact implementation varies in the literature, and the exact type of filters used does not seem to bear a major effect on the performance for context detection. In this work, we start by downscaling the images to 128×128 and normalizing the contrast, before filtering the resulting image with a bank of

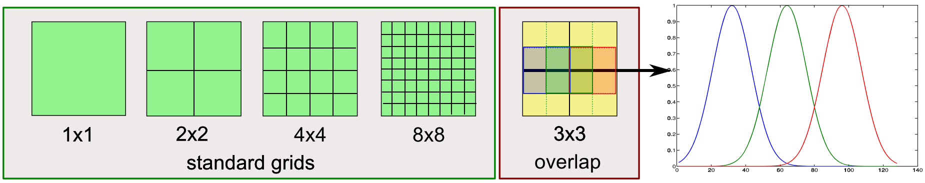


Fig. 2. Illustration of the grid averaging process. The left hand side shows the standard GIST grids, for sizes ranging from 1×1 to 8×8. The middle shows the effective cells for 2×2 grid with overlap: the green, red and blue square represent three overlapping squares. On the right, the graph shows an horizontal slice of this last grid, with overlapping Gaussians.

Gabor filters tuned to 8 different orientations and 4 scales; this results in $p = 32$ jets. The data size is then reduced by averaging the jets over a coarse grid laid over the image. Here again, the size of the grid used vary in the literature (we investigate the effect of this parameter in section 3.3); Oliva and Torralba reported a better performance of 4×4 versus 1×1 grids for context detection [1]. In this article we consider grids of size 1×1, 2×2, 4×4 and 8×8, separately and in combination (see Fig. 2).

One issue with this classical implementation is that the GIST vector can be very sensitive to small shifts of the features that lie close to the grid's boundaries. We propose an alternative sampling procedure based on overlapping smoothed cells. In this approach, adjacent rows of cells are overlapping by 50%, leading to an effective number of 144 cells for a 8×8 grid (see Fig. 2). Each cell's data vector $\mathbf{H} = (h_1, \cdots, h_p)$ is computed by averaging each jet $F_k, k \in \{1, \cdots, p\}$ according to a Gaussian kernel of variance one quarter of the grid cell's width:

$$h_k(x_0, y_0, s) = Q \sum_{x,y} F_k(x, y) \exp\left[-\left(\frac{x - x_0}{s/4}\right)^2 + \left(\frac{y - y_0}{s/4}\right)^2\right], \qquad (1)$$

where (x_0, y_0) is the centre of the grid cell, s is the cell width in pixels and Q is a normalization constant. The overlapping grid cells and the Gaussian smoothing are used to reduce the GIST vector sensitivity to small displacements at the grid's boundaries, and is shown to significantly improve performance on action prediction.

We will dispense with the additional PCA and/or ICA post-processing that is commonplace in the GIST literature (eg, [3]). Although reducing the feature dimension can be useful for some processes, we will rely on the boosted classifier to reduce dimensionality selectively through feature selection for each target category.

2.2 Classification

We use Boosting for learning both contextual labels and actions, as it has been shown to be successful for input selection and recognition [10,11]. We use a variant called *GentleBoost*, that has been shown to be more robust to noisy datasets [12]. Boosting is based on combining the weighted responses of a population of simple classifiers (called 'weak learners') into one robust classifier. The weak learners $l_i = (d_i, \tau_i, s_i)$ we used are simple decision stumps, each one applying a threshold τ on one of the feature vector's dimension d

$$R(l, \mathbf{v}) = \begin{cases} +s & \text{if} \quad v_d > \tau \\ -s & \text{otherwise} \end{cases}, \tag{2}$$

where, $s = \{-1, +1\}$ encodes the sign of the threshold that is applied. For each round of boosting i, the input dimension that best separates positive and negative examples is chosen, and the weights are updated. The classifier is therefore described by $\mathbf{L} = \{(l_1, w_1), \ldots (l_i, w_i), \ldots, (l_N, w_N)\}$, and the response is given by:

$$R(\mathbf{L}, \mathbf{v}) = \sum_{i=1}^{N} w_i \cdot R(l_i, \mathbf{v}). \tag{3}$$

As the number of weak learners is lower than the number of input dimensions, the learning process is effectively performing feature selection from the high dimensional input, and the weight of each weak learner provides a cue of the relative importance of each input towards the decision. In the following, and unless stated otherwise, the classifier was always trained using 1,000 samples from the dataset (0.7%), with half of the training set containing positive examples, and half negative examples. This positive/negative ratio was enforced to ensure that a sufficient number of positive examples were shown to the classifier, even for infrequent categories. Unless otherwise stated, the classifiers are evaluated on the rest of the dataset (ie, $> 99\%$ of the data).

2.3 Activation

In order to focus attention and direct higher level processes to relevant areas of the image, we need to evaluate which parts of the visual scene the predictors are tuned to, and whether they contribute to the activation or the inhibition of the action. We experimented with different ways to formalise what the predictors are responding to, and settled on reprojecting the Gaussian smoothing kernel in section 2.1 for each weak learner, weighted by this learner's weight. Thus the activation map is given by the mixture of Gaussians:

$$A(\mathbf{v}) = \sum_{i}^{|\mathbf{L}|} \left(w_i \cdot R(l_i, \mathbf{v}) \cdot G(l_i) \right), \tag{4}$$

for all weak learners l_i. In this equation $G(l)$ is the Gaussian kernel centred at the GIST grid cell l_i is associated with, with a variance of one fourth of the cell's width. The resulting map provides, for all images, an illustration of which image areas activate or inhibit each action.

Figure 9 shows the activation maps for each action for several example scenes, where the image is overlaid by green for excitation and red for inhibition.

3 Results

We evaluated the learning on a sequence taken from an instrumented car. The sequence contains 158,668 images for a total of about 3 hours of data, encompassing a variety of driving situations and settings. The dataset is illustrated in Fig. 3. The driver's actions were recorded from the car for each frame in the sequence.

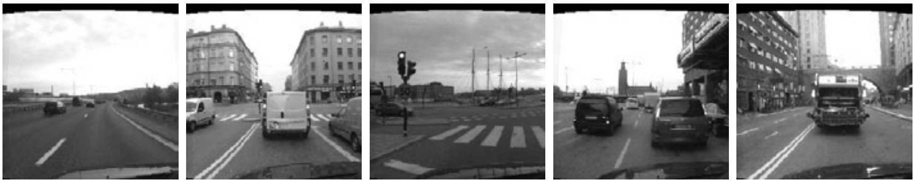


Fig. 3. Some example images taken from the 158,668 in the sequence

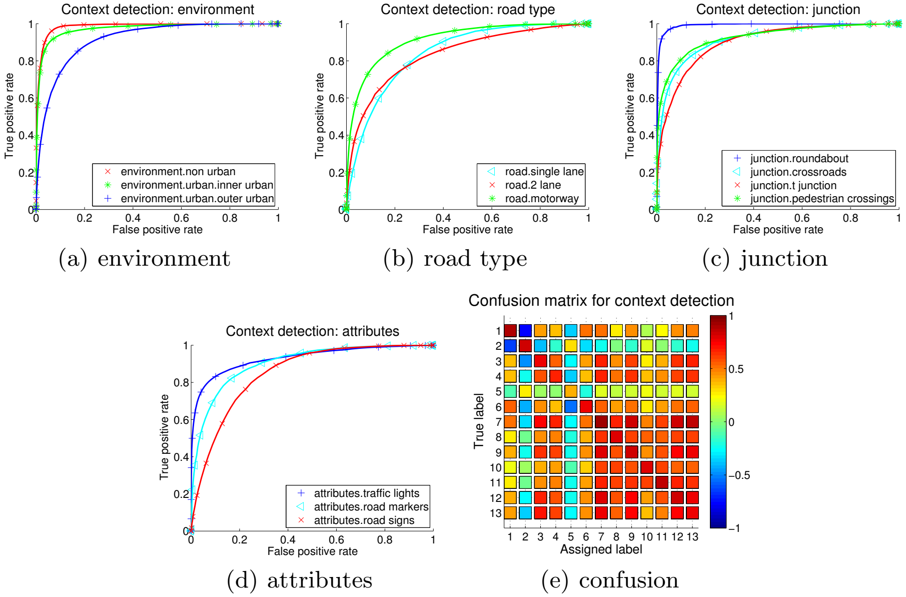
3.1 Learning Context Classes

Context information was provided in the form of a coarse labelling of each frame in the sequence pertaining to 13 classes. The number of frames labelled for each class is recorded in Table 1. The context classes are separated in four categories: *environment*, *road*, *junction* and *attributes*.

Table 1. Context labels associated to all images in the sequence (total: 158,668 frames)

Index	Category	Label	Count
1	environment	non-urban	47,923
2	environment	inner-urban	82,424
3	environment	outer-urban	28,321
4	road	single lane	31,269
5	road	two lanes	86,879
6	road	motorway	38,880
7	junction	roundabout	2,007
8	junction	crossroads	17,366
9	junction	T-junction	7,895
10	junction	pedestrian crossings	29,865
11	attributes	traffic lights	21,799
12	attributes	road markers	6,462
13	attributes	road signs	3,387

We trained an ensemble of Boosted decision stumps for each context class, using 100 rounds of Boosting on 1,000 frames chosen randomly; the performance was then evaluated on the rest of the dataset (more than 150,000 frames). Fig. 4 shows receiver operating characteristic (ROC) curves for all context classes, grouped by category. The confusion matrix is drawn in Fig. 4(e).



(a) environment (b) road type (c) junction

(d) attributes (e) confusion

Fig. 4. (a–d) ROC curves for the detection of different type of contextual information (see Table 1); (e) confusion matrix. All plots are for a combination of all overlapping grids for 100 rounds of Boosting, averaged over 10 runs.

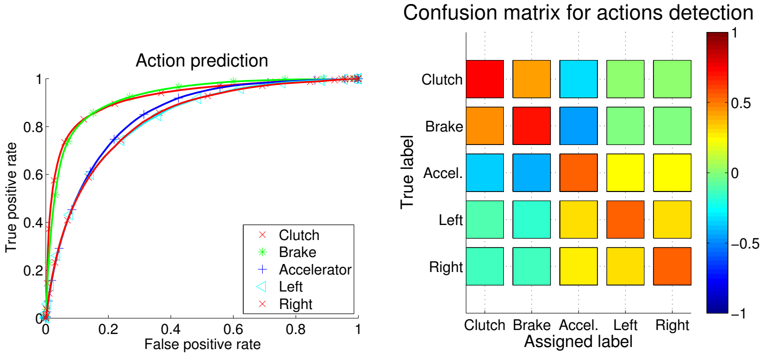
All classes are detected with good performance (note that all detectors are processed independently, without enforcing mutual exclusivity). The detection of the *environment* classes performs especially well, and the best performance is reached for distinction between 'inner urban' and 'non urban'. The lower detection performance for 'outer urban' is likely to be due to the somewhat fuzzier definition of the class; this is confirmed by the higher confusion value between 'non urban' and 'outer urban'. This high performance is consistent with published results in the literature. These categories are obviously global context categories and high performance validates other researchers' findings that GIST–type descriptors perform well for context recognition.

However, the performance is surprisingly high for other (more difficult) categories which make less use of global context. For the *road* category, confusion values are high between the 'single lane' and 'inner urban' classes, and the 'motorway' and 'non urban' classes, which are naturally consistent with expectations. The detectors for *junction* and *attributes* show a good performance for all classes (the very high performance on the 'roundabout' class may be due to the relatively low number of examples in the database). The confusion matrix shows a large confusion between all *junction* and *attributes* classes, and the 'inner city' class. This is consistent with the reality of traffic settings, and it should be noted that traffic lights (for example) are fundamentally *local* visual events, and therefore what is detected in this case is the visual context in which they are *likely* to occur, which is indeed a town centre intersection.

3.2 Learning Driving Actions

In a second experiment, we learnt to predict driver's actions from the gist features. The actions we considered are the pressing of one of the three pedals (Accelerator, Brake and Clutch) and the action of steering left or right. The actions were discretised, and therefore the amplitude of each action was disregarded for this experiment. Note that observation of the data revealed that the actions of pressing the clutch or the brake were binary actions anyway.

The classifier used was GentleBoost with decision stumps as weak learners; it was trained for 100 rounds with 1,000 randomly selected data points (less than 1% of the dataset).



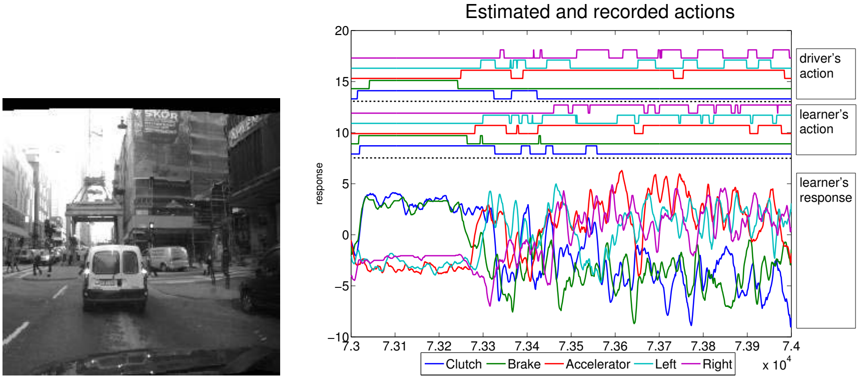
(a) ROC curves

(b) Confusion matrix

Fig. 5. Performance of the action prediction: (a) ROC analysis, and (b) confusion matrix. The results are for 100 rounds of boosting on the combined grids GIST descriptors; the training is done with 1,000 random frames, and tested with the rest of the dataset.

The action prediction performance is recorded in Fig. 5: the 'clutch' and 'brake' actions are predicted well (with 80% true positives for 10% false negatives); the two predictions also share a strong confusion value. This effect is driven by the large number of cases where the driver brings the car to a stop, pressing concurrently both brake and clutch. The performance when predicting the accelerator pedal and steering left or right is lower (80% false positives for 30% false negatives) but still good considering the large variability in the data. There is positive confusion values between steering and acceleration, which is consistent with good driving technique. The positive confusion between left and right steering is likely to come from the intersection situations, where steering left or right is equally plausible from visual information only.

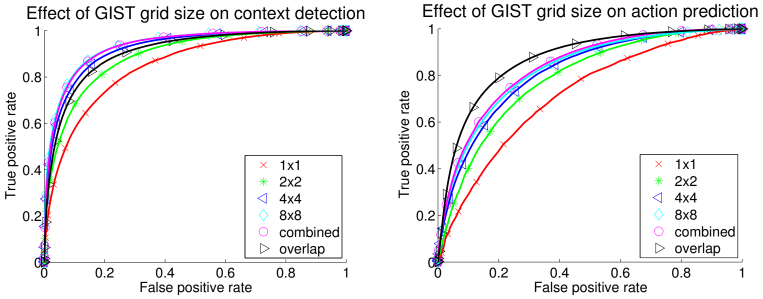
Fig 6 illustrates the quality of the action prediction on a short subsequence: the graph show curves for each action, for the driver and for the learnt response potential and final decision, respectively from top to bottom. The classifier was trained for 100 rounds on 1,000 frames taken randomly out of the 158,668. The classifier's response was smoothed using a 5–points moving average to remove the



(a) one image

(b) actions

Fig. 6. Illustration of the driver's and the system's elicited actions, on a short subsequence (1,000 frames). (a) first image in the sequence; (b) from top to bottom: the driver's action, the system's elicited actions, and the system's raw response. The predictor's response was smoothed using a 5–points moving average.



(a) GIST grid — context

(b) GIST grid — actions

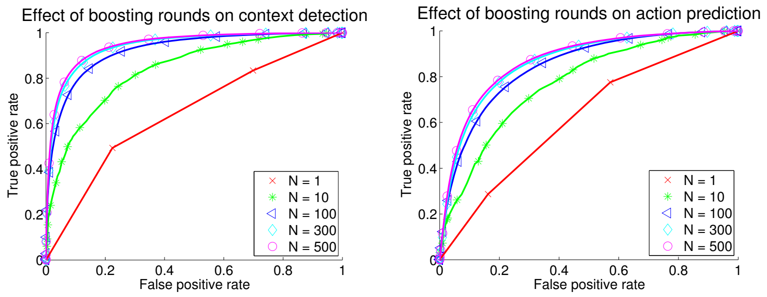
Fig. 7. Analysis of the effect of the GIST grid size on performance for (a) context detection and (b) action prediction

isolated outliers. The prediction for 'Brake', 'Clutch' and 'Accelerator' (accelerator pedal) is of very good quality for the whole sub-sequence. The prediction for steering 'Left' or 'Right' is not as reliable, but follows nonetheless the same patterns as the driver's.

3.3 Evaluation of the System's Parameters

We evaluated the influence of the GIST grid size and of the number of Boosting rounds on the detectors' performance, the results are displayed as ROC curves in Fig. 7 and 8. These ROC curves show the average performance over all classes and over 10 successive trainings of the detectors, each time with 1,000 randomly selected samples, and evaluated on the rest of the dataset. Figs. 7 show the performance for different GIST grids: 1×1, 2×2, 4×4, 8×8, and combinations of them all with and without overlapping. Each curve was obtained for 100 rounds of boosting. The best performance was obtained for using 8×8 grid and no additional performance was gained when using jointly a combination of all grids. The performance remained very good when using a 4×4 grid but dropped when using coarser histograms. When using overlapping smoothed grids, the performance for the context detection task was not improved compared to the 8×8 grid (Fig 7(a)); on the other hand, the performance for action prediction was significantly improved (Fig. 7(b)). This is likely to be due to less reliance upon global context and localised higher variability in the aspects of visual scenes relevant for predicting actions; eg, the position and the shape of the vehicle being followed can change to large extent. The non-overlapping grid used in classical GIST implementations make the feature vector sensitive to changes at the grid's boundaries, whereas an overlapping grid is less affected.

Fig. 8 shows the performance obtained for varying the number of rounds of Boosting, using an overlapping smoothed grid. No significant improvement was obtained by rising from 300 to 500 rounds, and 100 rounds yielded good performance. The performance for a single round of boosting was given as a baseline for a single decision stump's performance. Similar results were obtained when using other grids.



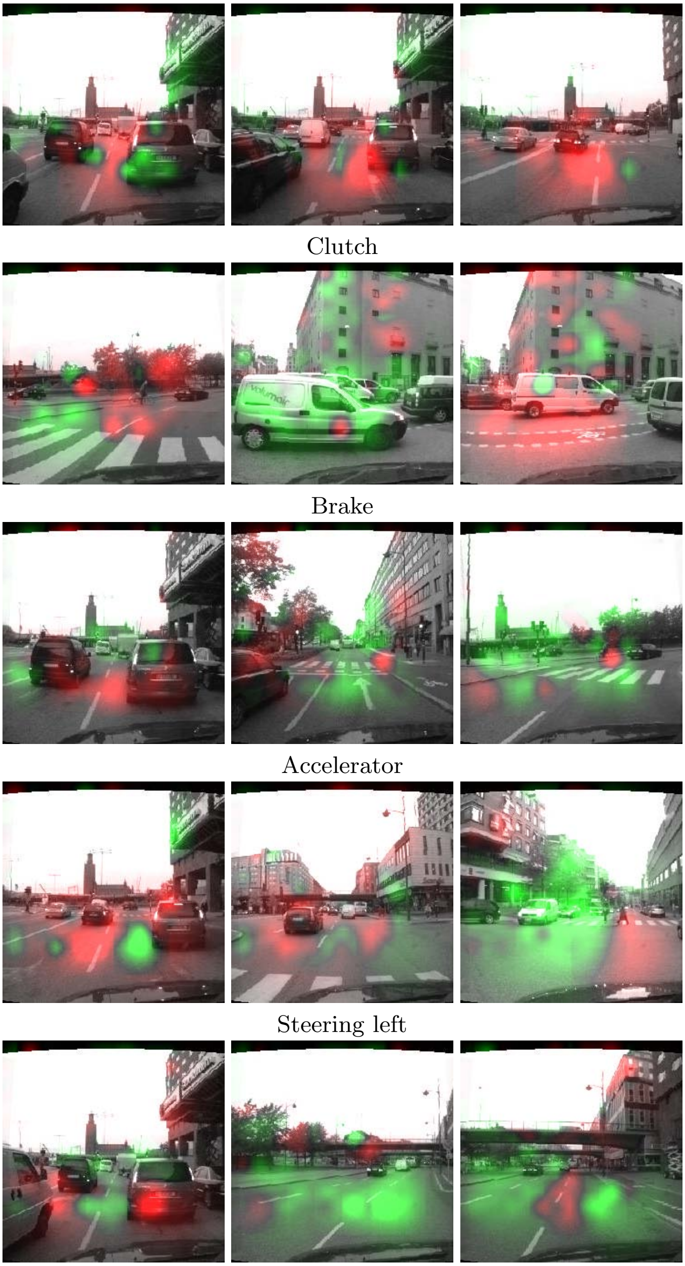
(a) Boosting rounds — context

(b) Boosting rounds — actions

Fig. 8. Analysis of the effect of the number of rounds of Boosting on performance for (a) context detection and (b) action prediction

3.4 Predictors' Activation

In order to get a better insight in what rules the system learns from the driver, we use the classifier inversion described in section 2.3 to identify what parts of the visual scenes activate the different action predictors. In Fig 9, the activation maps for three different situations are shown, for all actions. On those maps, the original image is overlaid with green on the excitatory areas and red on the inhibitory areas.



Clutch

Brake

Accelerator

Steering left

Steering right

Fig. 9. Activation maps on selected frames for each action predicted by the system; green shows activation and red inhibition. Areas of empty road activate acceleration and steering towards them, while inhibiting braking and pressing the clutch. Conversely, other vehicles on the road inhibit acceleration and steering towards them, while exciting braking, pressing the clutch and steering away from them—see text.

We expect the clutch pedal to be depressed when reducing speed to a minimum or when the car stops. On the left image, we can see that the clutch is activated by the presence of another car immediately in front. On the middle image the car is further away, and the same image area, now empty of cars, inhibits the 'clutch' predictor; a similar inhibition pattern is visible in the right image where the road ahead is free. The second row shows consistent activation patterns for the 'brake' action: on the left, the empty road in front inhibits the predictor whereas the pedestrian crossing area activates it. On the middle, the presence of a car immediately in front leads to a strong excitation, whereas on the right an empty space yields a strong inhibition. As expected, the 'accelerator' activation is the opposite of the 'brake': activated by empty spaces and inhibited by other vehicles in front. The activation maps for steering actions are somewhat more difficult to interpret, as expected from the lower prediction performance. The 'left' and 'right' actions appear to be activated by obstacles and to promote veering away from them (see left images). They also seem to react to the vehicle's position in its lane, as evidenced by the sharp inhibition of steering generated on the central white line (see the bottom–right image).

4 Discussion

In this article we attempt to model driving behaviour by learning the relationship between a human driver's actions and holistic image descriptors. Supervision comes in two forms: first, a coarse labelling of the images in terms of a variety of driving–relevant contextual categories; second, a frame per frame record of the driver's actions when faced with this situation. We use GIST features as an equivalent to human pre-attentive vision, for encoding the visual input, and attempt to learn, for all images, both the associated labels and the driver's actions.

The GIST descriptor is a generic approach for holistic image features, and has several free parameters. Experimenting with different type of grids for the GIST descriptor, we found that the best performance was obtained for a 8×8 grid. Moreover, the small difference in performance between 8×8 and 4×4 grids make in unlikely for finer grids to increase performance notably, for a high computational cost. Instead, we proposed an overlapping grid smoothed using Gaussian functions, that lead to a significant performance improvement for action prediction (see Fig. 7(b)).

We found that the optimal performance is reached with a relatively low number of rounds of Boosting for both context detection and action prediction (100 rounds); this is a large dimension reduction compared to the original feature vector (6,496 for the combined overlapping grids). Therefore, the relatively high dimensionality of the original feature vector is not an issue after the training stage as each classifier only uses a small carefully selected proportion of it. Those dimensions and their respective contribution to the classifier's response can be reprojected in the image domain as discussed in section 2.3, and produces the activation maps shown in Fig. 9.

The high performance of pre-attentive vision for detecting the environment class (‘non urban’, ‘outer urban’, or ‘inner urban’) is consistent with previous results in the literature. Very good performance was also obtained when detecting more complex aspects of the driving context such as T–junctions, pedestrian crossings, or even traffic lights (see Fig. 4).This shows that holistic features do carry a large amount of visual information relevant for interpreting driving scenarios. Moreover, the success in detecting what are essentially *local* events (eg, traffic lights) shows the high contextual prior that permeates most driving visual scenes: the presence of an intersection in an urban setting, for example, is a strong predictor for the presence of a traffic light, or road markings.

The performance with which the driver’s actions can be predicted from holistic image features, is a more unexpected result (see Figs. 5 and 6). Indeed, the system does not have insight into the driver’s intentions and lacks any formal knowledge of the highway code. The fact that the driver’s actions can be predicted at all, only from transient holistic image features, illustrates the intuition that most of a driver’s actions are completely determined by the context in which he is, and only a small fraction is determined by intention, attentive vision and high–level reasoning. These cases are of special importance for learning an attentional model of the driver’s behaviour: we expect the false positives to be the instances in the dataset where the driver’s pre-attentive actions were inhibited by higher–level considerations. If we consider the case of crossing traffic at an intersection, pre-attentive vision may learn to slow down before the intersection, but the driver will then need to actively assess whether the way is free or if he needs to stop. The activation maps shown in Fig. 9 provides us with a useful indication of which parts of the scene are relevant for taking a decision; together with the driver’s gaze, they provide a way to focus the attention of a higher level feature–based learning on the most promising parts of the visual scene. Therefore, the learning of a more complex model can be bootstrapped by the activation maps at false positives and the driver’s gaze can be combined to learn the attentive components of driving. In this context, the pre-attentive model serves as a filter to focus attentional learning towards the rare instances where it is required, and the aspects of the scenes that may be of importance.

5 Conclusion

Holistic image descriptors have received a lot of attention in the recent year, both from the computer vision and the psychology communities, as a good model for fast, pre-attentive vision, and a good feature for scene identification. We used such GIST features for learning driving behaviour from a human driver, and obtained very good results both for the detection of visual context labels and for the prediction of the driver’s actions. The fairly high performance of the action prediction illustrates the fact that only a small proportion of the driving actions require formal understanding of the driver’s intentions or the highway code. This is a vivid illustration of the strong priors at work during normal driving behaviour, and of how much information pre-attentive perception can carry, as 80% of a driver's actions can be predicted. Such a performance allows to focus attention on learning the more complex rules that underlie the 10–20% of problematic cases.

Acknowledgments. This research was supported by the EU project DIPLECS (FP 7 ICT project no. 215078).

References

1. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175 (2001)
2. Torralba, A.: Contextual priming for object detection. International Journal of Computer Vision 53, 169–191 (2003)
3. Siagian, C., Itti, L.: Rapid biologically-inspired scene classification using features shared with visual attention. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 300–312 (2007)
4. Douze, M., Jégou, H., Sandhwalia, H., Amsaleg, L., Schmid, C.: Evaluation of gist descriptors for web-scale image search. In: CIVR 2009: Proceedings of the ACM International Conference on Image and Video Retrieval (2009)
5. Torralba, A., Oliva, A., Castelhano, M., Henderson, J.: Contextual guidance of attention in natural scenes: The role of global features on object search. Psychological Review 113, 766–786 (2006)
6. Renninger, L., Malik, J.: When is scene identification just texture recognition? Vision Research 44, 2301–2311 (2004)
7. Siagian, C., Itti, L.: Biologically inspired mobile robot vision localization. IEEE Transactions on Robotics 25, 861–873 (2009)
8. Ackerman, C., Itti, L.: Robot steering with spectral image information. IEEE Transactions in Robotics 21, 247–251 (2005)
9. Kastner, R., Schneider, F., Michalke, T., Fritsch, J., Goerick, C.: Image–based classification of driving scenes by a hierarchical principal component classification (HPCC). In: IEEE Intelligent Vehicles Symposium, pp. 341–346 (2009)
10. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55, 119–139 (1997)
11. Viola, P., Jones, M.: Robust real–time object detection. International Journal of Computer Vision 57, 137–154 (2001)
12. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. The Annals of Statistics 28, 337–407 (2000)