

# Multi-modal Speed Limit Assistants: Combining Camera and GPS Maps

Claus Bahlmann<sup>1</sup>, Martin Pellkofer<sup>2</sup>, Jan Giebel<sup>3</sup>, and Gregory Baratoff<sup>3</sup>

<sup>1</sup>Siemens Corporate Research, Inc.  
755 College Road East  
Princeton, NJ 08540, USA  
claus.bahlmann@siemens.com

<sup>2</sup>VDO Automotive AG  
Wernerwerkstr. 2  
93049 Regensburg, Germany  
{martin.pellkofer,jan.giebel,gregory.baratoff}@continental-

<sup>3</sup>VDO Automotive AG  
Peter-Dornier-Str. 10  
88131 Lindau, Germany  
corporation.com

## Abstract

*We propose a method for fusing two modalities of information for speed limit assistants: (i) camera based speed sign recognition and (ii) digitized speed limit maps combined with a GPS sensor. The fusion is based on a Bayesian framework. Here, we rely on two modeling assumptions: (i) the speed sign recognizer's score being probabilistic and (ii) a model describing speed limit sign probabilities conditioned on the map information. Speed limit assistants incorporating the proposed fusion can particularly benefit over uni-modal solutions in situations, where a solution based on a single modality is ill-posed, that is, adverse lighting or weather conditions in case of camera based speed sign recognition, and dynamic traffic guidance systems, construction zones, or incomplete maps in case of GPS maps. We give exemplary evidence of the proposed solution's effectiveness.*

## 1. Introduction

In recent years, the technology of *speed limit assistants* (SLAs) has emerged as an important application for automotive advanced driver assistant systems (ADAS). The aim of such systems is to inform or warn the driver about effective speed limits, or, in a more active setting, supplement an adaptive cruise control (ACC) system. Current solutions for SLAs are usually based on one out of two different technologies: (i) camera based speed sign recognition and (ii) digitized maps with a GPS interface, as reviewed in the following.

Camera based speed sign recognition aims at detecting and classifying speed limit signs in image sequences. Camera sensors are becoming increasingly cheap and ubiquitous in automobiles. Since cameras are extremely versatile, this solution is particularly attractive when combined with additional front facing camera

ADAS applications, such as those involving the detection of lane markers, vehicles, or pedestrians.

A number of highly accurate speed sign recognition systems has been developed in recent years [3, 2, 1, 6, 4], both in the context of gray-scale and color videos. For a detailed literature review it is referred to those references. As a general remark, state-of-the-art systems seem to converge to a two stage approach, which addresses both accuracy and speed: Firstly, image frames are scanned for possible sign occurrences using a particularly fast detection algorithm (based on geometric or color information), and speed sign hypotheses are generated for regions with sufficient sign evidence. Secondly, these hypotheses are verified and categorized using a statistical classifier. Most often, the classifier is also able to eliminate false alarms from the first stage, based on the classifier's confidence value, for instance. State-of-the-art systems achieve remarkable performance nowadays, however, 100% accuracy cannot be guaranteed. In particular, adverse lighting (frontal or backward sunlight) and weather (rain, snow) conditions or night situations can impede recognition.

The second class of SLA technologies relies on a digitized map of speed limit zones, which, in combination with a GPS sensor, can provide the speed limit for the current vehicle position. The cost of such solutions is becoming less significant, as an increasing number of vehicles is already equipped with navigation systems, and hence, a GPS sensor. This technology does not require complex processing. Its accuracy mainly depends on the accuracy of the provided map data.

However, here lies a major problem of this approach. Firstly, available maps are static representations. They fail if speed limits change over time. For instance, modern traffic guidance systems make use of variable message signs, which dynamically adapt speed limits dependent on weather and/or traffic conditions. Also, construction zones require authorities to temporarily lower the speed limits for the affected road sec-







Limitations of uni-modal SLA technologies	
Camera based	Map/GPS based
	
Adverse lighting conditions	Dynamic traffic guidance systems
	
Adverse weather conditions	Temporary construction zones
	
Night situation	Non-annotated rural roads

Figure 1: Each of the SLA technologies has specific limitations. For the camera based speed sign recognition, adverse lighting, weather, and night conditions can impede recognition. The map/GPS technology fails in dynamic environments, such as, variable speed limit zones and construction zones, as well as on rural roads, which have not yet been annotated.

tion. Secondly, speed limit zones are still not completely mapped to date; in particular, coverage is rather sparse for rural roads. Thirdly, map data only represents a snapshot at a certain time and may be outdated after a few years. Figure 1 summarizes the mentioned limitations for both approaches.

As explained above, none of today’s solutions is perfect on its own. The presented contribution aims at a scenario where both camera and GPS sensors are available. It presents a solution to combine information of both sensors, achieving increased accuracy.

## 2. Bayesian fusion of camera and GPS

We employ a Bayesian fusion framework. It is based on the speed sign recognizer’s confidence and a formulation that describes speed limit sign probabilities conditioned on the map information. Figure 2 summarizes the main idea.

Before explaining the proposed framework, we

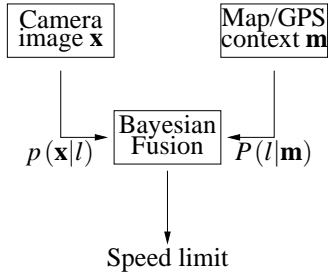


Figure 2: Proposed fusion framework for camera and map/GPS modalities: Both modalities provide a probabilistic output, which are fused within a Bayesian framework.

briefly introduce some nomenclature w.r.t. most camera based sign recognition systems and map/GPS representations.

### 2.1. Camera based SLA

The goal of camera based speed sign recognition is to detect an instance of a speed sign and assign it to one of a set of class labels. A typical German setting, for instance, would cover 14 speed limit signs, their corresponding suspension signs, and a general suspension sign, that is,

$$l \in \mathbb{L} = \{5, 10, 20, \dots, 130, 5\text{-end}, 10\text{-end}, 20\text{-end}, \dots, 130\text{-end}, \text{any-end}\} \quad (1)$$

Given a *speed sign class*  $l$ , it is straightforward for most situations to infer the corresponding *speed limit class*

$$k \in \mathbb{K} = \{5, 10, 20, \dots, 130, \text{no-limit}\}. \quad (2)$$

As reviewed above, the majority of camera based speed sign recognition algorithms employ a statistical classifier as final element in the processing chain. Many classifiers provide some sort of score proportional to the likelihood

$$u_l(\mathbf{x}) \propto p(\mathbf{x}|l), \quad l \in \{1, \dots, L\}, \quad (3)$$

given the feature vector  $\mathbf{x}$  of the sign hypothesis. Note that  $u_l(\mathbf{x})$  being the posterior probability  $p(l|\mathbf{x})$  would also be covered by Eq. (3), since it can be related to  $p(\mathbf{x}|l)$  via Bayes’ rule  $p(l|\mathbf{x}) \propto P(l)p(\mathbf{x}|l)$ .

Eq. (3) is the only constraint we will impose on the classifier in the context of the proposed fusion. This covers many of today’s popular solutions. For instance, our own systems [1, 4] employs maximum-a-posteriori (MAP) classification, other possible choices include polynomial, AdaBoost [8], multi-layer perceptron classifiers [5], etc. Support vector machines are a bit more involved, in the sense that they do not directly provide a probabilistic output. However, techniques exist, which

estimate probabilities in a post-processing using logistic regression [7].

## 2.2. Map based SLA

Typical digital traffic maps represent speed limit zones by dividing a network of roads into road segments. For each road segment a context tuple of variables, say  $\mathbf{m}$ , is available. The most useful information for our purposes are the *street type*

$$\begin{aligned} m_{ST} &\in \mathbb{M}_{ST} \\ &= \{\text{NONE, MOTORWAY, HIGHWAY, RURALROAD,} \\ &\quad \text{URBANROAD, TRAFFICCALMINGZONE}\} \end{aligned} \quad (4)$$

and the speed limit

$$\begin{aligned} m_{SL} &\in \mathbb{M}_{SL} \\ &= \{\text{NONE, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100,} \\ &\quad \text{110, 120, 130, no-limit}\}. \end{aligned} \quad (5)$$

For both cases, a value of 'NONE' indicates that no information is available, for instance, if streets have not been mapped yet.

Map-only SLA can simply output  $m_{SL}$  for the given road segment.

For the fusion framework explained below, variables  $m_{ST}$  and  $m_{SL}$  will form the context tuple

$$\mathbf{m} = (m_{ST}, m_{SL}).$$

## 2.3. Bayesian fusion framework

For the formulation of our multi-modal fusion we take a Bayesian perspective. In this respect, given a speed sign camera image  $\mathbf{x}$  and the context tuple  $\mathbf{m}$ , the maximum-a-posteriori classification paradigm assigns  $\mathbf{x}$  to the class  $\hat{l}$  with the highest a-posteriori probability, that is,

$$\hat{l} = \underset{l}{\operatorname{argmax}} \{P(l|\mathbf{x}, \mathbf{m})\}. \quad (6)$$

In Eq. (6), variables  $\mathbf{x}$  and  $\mathbf{m}$  represent the two modalities of sensor information. Typically,  $P(l|\mathbf{x}, \mathbf{m})$  cannot be measured directly. However, we can use Bayes' rule to translate  $P(l|\mathbf{x}, \mathbf{m})$  into functions of  $p(\mathbf{x}|l)$  and  $P(l|\mathbf{m})$ <sup>1</sup>

$$\begin{aligned} P(l|\mathbf{x}, \mathbf{m}) &\propto p(\mathbf{x}|l, \mathbf{m}) P(l|\mathbf{m}) \\ &= p(\mathbf{x}|l) P(l|\mathbf{m}). \end{aligned} \quad (7)$$

<sup>1</sup>Here and in the remainder, we follow the convention of using lowercase  $p$  for probability densities of continuous variables and uppercase  $P$  for probabilities of discrete variables.

The last equality holds, because knowledge about the road segment's street type or speed limit information does not impact the sign image appearance for a particular class  $l$ .

Eq. (7) shows that the posterior decomposes nicely into two terms:

1.  $p(\mathbf{x}|l)$  describes the speed sign appearance as provided by the camera sensor, and can usually be obtained from the statistical pattern classifier, as discussed above.
2.  $P(l|\mathbf{m}) = P(l|(m_{ST}, m_{SL}))$  enables a probabilistic modeling of the map/GPS context. We will elaborate the specifics in the next section.

It is worth noting that in above equations  $\mathbf{m}$  refers to the context tuple of the street segment commencing beyond the location of the speed sign, while the camera image  $\mathbf{x}$  is taken when the vehicle is located in the road segment preceding this location. In order to synchronize  $p(\mathbf{x}|l)$  and  $P(l|\mathbf{m})$ , two solutions are viable: (i) perform the fusion with a delay (of a few seconds), until the vehicle has entered the new road segment and the new  $\mathbf{m}$  is known; or (ii) project the context  $\mathbf{m}$  of the next available road segment. If it is ambiguous, e.g., due to road forks, another layer of probabilistic modeling, similarly to the techniques described above, is advised.

## 2.4. Modeling the map probabilities

Since  $p(\mathbf{x}|l)$  can simply be obtained from the statistical classifier, the remaining task is to model the map probabilities  $P(l|\mathbf{m})$ , as indicated in point 2 above. One may be tempted to estimate  $P(l|\mathbf{m})$  from a set of observations  $\left\{ \left( l^{(1)}, \mathbf{m}^{(1)} \right), \dots, \left( l^{(M)}, \mathbf{m}^{(M)} \right) \right\}$ . However, for our settings (see Eq. (1), (4), and (5)),  $P(l|\mathbf{m})$  corresponds to a table of  $L \cdot |\mathbb{M}_{ST}| \cdot |\mathbb{M}_{SL}| = 29 \cdot 6 \cdot 16 = 2784$  discrete probabilities. For reliable estimates, hundreds of thousands of labeled road segments are required.

Unfortunately, this amount of data is not available, hence, we resort to using expert knowledge and modeling assumptions, in order assign values to  $P(l|\mathbf{m})$ . This knowledge describes which speed limits  $k$ , or consequently, speed limit signs  $l$ , are how likely to expect in which context  $\mathbf{m}$ . Instead of assigning values directly for each matrix element  $P(l|\mathbf{m})$ , we pursue the strategy of defining a model based on a set of canonic traffic regulation rules and probabilistic assumptions, and derive  $P(l|\mathbf{m})$  from this model.

It is apparent that this model includes country specific regulations, in general. In the following, we exemplary state a model for German traffic regulations. Note, however, that it would be straightforward for respective experts to set up similar rules in other countries.

1. The range of permitted speed limits for different street types, according to general German

traffic regulations, are:  $k \in \{60, \dots, \text{no-limit}\}$   
for  $m_{\text{ST}} = \text{MOTORWAY}$ ,  $k \in \{5, \dots, 130\}$   
for  $m_{\text{ST}} = \text{HIGHWAY}$ ,  $k \in \{5, \dots, 100\}$  for  
 $m_{\text{ST}} = \text{RURALROAD}$ ,  $k \in \{5, \dots, 60\}$  for  
 $m_{\text{ST}} = \text{URBANROAD}$ ,  $k \in \{5, \dots, 30\}$  for  
 $m_{\text{ST}} = \text{TRAFFICCALMINGZONE}$ .

2. Any suspension sign resets the speed limits to  $k = \text{no-limit}$  for  $m_{\text{ST}} = \text{MOTORWAY}$ , to  $k = 130$  for  $m_{\text{ST}} = \text{HIGHWAY}$ , to  $k = 100$  for  $m_{\text{ST}} = \text{RURALROAD}$ , to  $k = 50$  for  $m_{\text{ST}} = \text{URBANROAD}$ , and to  $k = 30$  for  $m_{\text{ST}} = \text{TRAFFICCALMINGZONE}$ .
3. Assumptions about relation of true speed limit and map speed limit:
  - (a) The true speed limit  $k$  is never higher than the map speed limit  $m_{\text{SL}}$ . This is based on the assumption that  $k$  and  $m_{\text{SL}}$  only differ in construction zones (where cautious driving is advised), in zones with variable message signs (where  $m_{\text{SL}} = \text{NONE}$ ), and on streets without map information (where  $m_{\text{SL}} = \text{NONE}$ ).
  - (b) The class  $l \in \mathbb{L}$ , which corresponds to a given map speed limit  $m_{\text{SL}}$  is assigned the highest probability, say,  $P(l|\mathbf{m}) = a$ .
  - (c) All other speed limit signs  $l$  that correspond to  $k \neq m_{\text{SL}}$  and are permitted by Rule 1, are assigned a lower probability  $P(l|\mathbf{m}) = b$ ,  $b < a$ .

Above stated rules are adequate to derive the complete table  $P(l|\mathbf{m})$ . Remember that  $P(l|\mathbf{m})$  is a 3D matrix of dimension  $L \cdot |\mathbb{M}_{\text{ST}}| \cdot |\mathbb{M}_{\text{SL}}|$ . Table 1 shows the respective parts for  $m_{\text{ST}} = \text{MOTORWAY}$  and  $m_{\text{ST}} = \text{RURALROAD}$ , the remaining parts (for  $m_{\text{ST}} = \text{HIGHWAY}$ ,  $m_{\text{ST}} = \text{URBANROAD}$ ,  $m_{\text{ST}} = \text{TRAFFICCALMINGZONE}$ ,  $m_{\text{ST}} = \text{NONE}$ ) can be derived similarly.

The following remarks explain Table 1 more specifically: entries in the first rows ( $l = 5, \dots, 50$ ) for  $m_{\text{ST}} = \text{MOTORWAY}$  encode the existence of a minimum speed on German motorways (Rule 1). The next set of rows ( $l = 60, \dots, 130$ ) incorporate the logic of Rule 3, the rows  $l = 60\text{-end}, \dots, \text{any-end}$  that of Rule 2.

Interestingly, a large portion of entries has a value of 0, corresponding to configurations excluded by the proposed model. In this respect, a number of potential mis-classifications — which could arise from the camera based recognition (e.g., taking 40- for 90-speed limits in a city environment) — can be prevented, by taking into account the GPS map information.

The constraint  $a > b$  favors speed limit signs that are in accordance with the map information, if the statistical classifier is in doubt. It becomes clear that the performance of the fusion framework depends on carefully chosen parameters  $a$  and  $b$ , in particular the ratio  $a/b$ . Inappropriate settings can bias the SLA too much

towards one of the modalities. In the context of our maximum-a-posteriori classifier, a ratio of  $a/b = 1/0.7$  has been proven from experimental evaluation to perform well.

### 3. Experiments

It is clear that the fusion framework benefits over a GPS-only system in situations where no map information is available. In that case, it simply relies on the output of the camera system, which is more informative than random guessing.

In the following, we want to compare the effectiveness of the fusion framework with that of our plain camera based speed sign recognition system [1] on a qualitative level. Figure 3 shows the system behavior for the recognition of a speed limit sign “80”.

As can be seen from the blue box in the image frame, the detection module correctly identifies the sign. Input to the statistical classifier (which is a uni-modal Gaussian MAP classifier) is a pre-processed (i.e., geometrically normalized, masked, and histogram equalized) image, as shown by the enlarged area. The second column illustrates the output of the camera-only classifier, the third the fusion of camera and map/GPS. More precisely, the two speed limit icons indicate the Top-2 classes, as obtained from our maximum-a-posteriori classifier w.r.t.  $P(l|\mathbf{x}, \mathbf{m})$ . The three real numbers below correspond to the probabilistic values “ $P(l|\mathbf{x}, \mathbf{m}) / p(\mathbf{x}|l) / p(\mathbf{x}|l) \cdot P(l|\mathbf{m})$ ”, as defined in Section 2.3.

The camera-only system — without the knowledge of any map/GPS input — considers uniformly  $P(l|\mathbf{m}) = 1$  (note that we assume an implicit normalization of  $P(l|\mathbf{m})$  over  $l$ , such that  $P(l|\mathbf{m})$  is a probability). It can be seen that this results in a misclassification in this example, as  $p(\mathbf{x}|l = 60) = 0.50 > p(\mathbf{x}|l = 80) = 0.45$ . Most likely, this is caused by the digits printed unusually off-center in the image patch, and our classifier relying on accurately aligned speed limit numbers.<sup>2</sup>

The situation changes when map/GPS input is available (column 3), particularly the map probabilities  $P(80|\mathbf{m} = (\text{MOTORWAY}, 80)) = a = 1$  and  $P(60|\mathbf{m} = (\text{MOTORWAY}, 80)) = b = 0.7$  (taken from Table 1). The higher map probability  $P(l = 80|\mathbf{m})$  compensates for lower values in  $p(\mathbf{x}|l = 80)$ , resulting in a correct classification.

### 4. Conclusion

We have proposed a novel framework for information fusion in the context of speed limit assistants. It fuses two modalities of a camera based speed sign

<sup>2</sup>It can be argued that more complex segmentation or classification methods are less sensitive to such off-center digits. However, note that, whatever camera based systems may be used, it faces limitations, as have been summarized in Figure 1. In the context of this paper, we want to study the impact of sensor fusion in those difficult conditions.

$m_{ST} = \text{MOTORWAY}$																
$l \setminus m_{SL}$	5	10	20	30	40	50	60	70	80	90	100	110	120	130	no-limit	NONE
5																
10																
20																
30			0								0				0	0
40																
50																
60							$a$									
70								$a$								
80									$a$				$b$			
90										$a$						
100			0								$a$				$b$	$a$
110								0				$a$				
120													$a$			
130														$a$		
5-end																
10-end																
20-end																
30-end			0								0				0	0
40-end																
50-end																
60-end																
70-end																
80-end																
90-end																
100-end			0								0				$a$	$a$
110-end																
120-end																
130-end																
any-end																

$m_{ST} = \text{RURALROAD}$																
$l \setminus m_{SL}$	5	10	20	30	40	50	60	70	80	90	100	110	120	130	no-limit	NONE
5	$a$															
10		$a$														
20			$a$													
30				$a$												
40					$a$						$b$		0		0	$a$
50						$a$										
60							$a$									
70			0					$a$								
80									$a$							
90										$a$						
100																
110					0						0		0		0	0
120																
130																
5-end																
10-end																
20-end																
30-end																
40-end					0						$a$		0		0	$a$
50-end																
60-end																
70-end																
80-end																
90-end																
100-end																
110-end					0						0		0		0	0
120-end																
130-end																
any-end						0					$a$		0		0	$a$

Table 1: Map probabilities  $P(l | (m_{ST} \in \{\text{MOTORWAY}, \text{RURALROAD}\}, m_{SL}))$ : Entries are probabilities for a speed sign class  $l \in \{5, \dots, \text{any-end}\}$ , given the map context  $\mathbf{m} = (m_{ST}, m_{SL})$ . A value 0 corresponds to an impossible event,  $a$  to an event with high probability,  $b$  to an event with lower probability, i.e.,  $b < a$ . In our system, we use  $a = 1$  and  $b = 0.7$ .

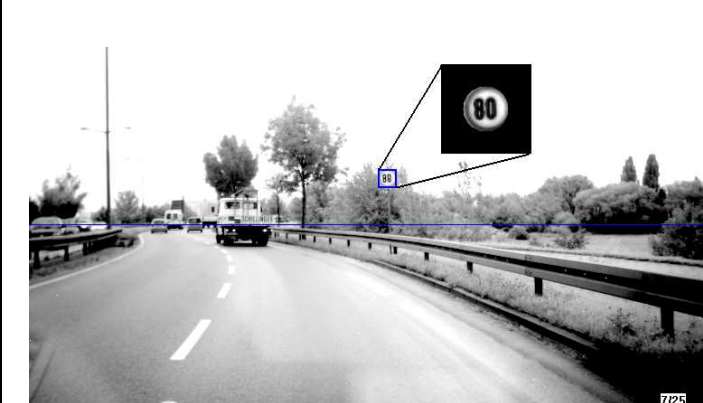


Detector output	Classifier output	
	Camera-only	Fusion Camera-GPS
		

Figure 3: An example showing the effectiveness of the proposed sensor fusion. Column 1 shows the detector output, column 2 the Top-2 of the camera-only classification, and column 3 of the fused classification. See Section 3 for further explanations.

recognition and digitized speed limit maps combined with a GPS sensor. The fusion is performed in a Bayesian setting. In this respect, we rely on a statistical classifier giving a probabilistic decision score and a probabilistic model of speed sign probabilities conditioned on the map information. We have shown how such model can be obtained from a set of simple traffic regulation rules. Although our example is derived from German traffic regulations, it can be straightforwardly extended to different countries. We have shown on a qualitative level that such framework can be advantageously used.

One aspect of future work is a quantitative system performance evaluation, since we here focus on a qualitative analysis.

As explained above, a reasonable choice of the parameter  $a/b$  is important for the system performance. In the present prototype, parameter tuning is performed rather adhoc. In general, however, there are many factors that influence an optimal setting, e.g., characteristics of the classifier, accuracy of the speed zone map, etc. A more systematic and automatic strategy, which takes into account those sources, could improve system performance and portability to different environments.

Further, the road segment information from speed zone maps not only can improve classification, but also detection of signs. In this respect, speed limit signs are expected at the boundary of those road segments. Promising future work lies in incorporating such knowledge into the detector.

## Acknowledgment

The authors would like to thank Jan Ernst from Siemens Corporate Research for valuable discussions and comments.

## References

- [1] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, 2005.
- [2] N. Barnes and A. Zelinsky. Real-time radial symmetry for speed sign detection. In *IEEE Intelligent Vehicles Symposium (IV)*, Parma, Italy, 2004.
- [3] D. M. Gavrila. Traffic sign recognition revisited. In *Mustererkennung (DAGM)*, Bonn, Germany, 1999. Springer Verlag.
- [4] C. G. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff. Real-time recognition of US speed signs. In *IEEE Intelligent Vehicles Symposium (IV)*, Eindhoven, Netherlands, 2008.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2325, Nov. 1998.
- [6] F. Moutarde, A. Bargeton, A. Herbin, and L. Chanussot. Robust on-vehicle real-time visual detection of american and european speed limit signs, with a modular traffic signs recognition system. In *IEEE Intelligent Vehicles Symposium (IV)*, Istanbul, Turkey, June 2007.
- [7] J. C. Platt. Probabilities for SV machines. In *Advances in Large Margin Classifiers*, pages 61–73. MIT Press, 2000.
- [8] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.