Autonomous Audio-Supported Learning of Visual Classifiers for Traffic Monitoring

Horst Bischof, Martin Godec, and Christian Leistner, *Graz University of Technology*

Andreas Starzacher and Bernhard Rinner, *Klagenfurt University*

Using acoustic detection and classification of vehicles, the proposed autonomous self-learning framework generates scene adaptive vehicle classifiers without the need to hand label any video data.

The steady increase of automobiles in operation impacts our lives in several ways. Road congestion causes severe economic consequences because of delays and energy waste; estimations on the total cost of traffic congestion range up to 1 percent of a gross domestic product. According to the European

Transport Safety Council (www.etsc.eu), approximately 39,000 people were killed in road collisions in 2008 in Europe.

Automated traffic monitoring plays an important role in increasing safety and throughput on the existing road infrastructure. Numerous road sensors capture traffic data that is analyzed to assess the current situation. This assessment can then trigger various counter actions such as warning drivers, reducing speed limits, or rerouting traffic. Given the traffic system's scale and complexity, it is invaluable to automate traffic monitoring and control as much as possible. Most current traffic-monitoring systems, however, only capture data from traffic sensors, so assessment requires continuous human supervision (see the "Intelligent Traffic Monitoring" sidebar for more details). Additionally, there is an increasing demand for mobile or portable systems that can monitor temporary events such as construction sites.

Powerful visual and acoustic classifiers exist, but to obtain high accuracy, these algorithms require a huge amount of hand-labeled data. Collecting this data is a tedious and cost-intensive task. The classifiers are usually trained in the lab and later applied (without adaptation) to many possible scenarios and might thus become unnecessarily complex. Additionally, typical appearance-based classifiers are sensitive to an object's orientation,[1] making it difficult to obtain well-performing general detectors.

In contrast, specialized detectors for specific scenes promise to perform better in terms of both accuracy and efficiency. Because specialized detectors reduce the task's complexity, they also drastically reduce the required amount of labeled training samples. In practical applications, these specialized detectors must fulfill several requirements: First, their training must be as autonomous as possible to avoid manual labeling for every site. Second, this autonomous learning

Intelligent Traffic Monitoring

In the near future, we will witness more than a billion automobiles in operation worldwide.[1] Automated traffic monitoring will therefore play an essential role in improving road throughput and safety. Current monitoring systems capture—usually vision-based—traffic data from a large sensory network, but they require continuous human supervision, which is extremely expensive. Future traffic-monitoring systems must become more intelligent to analyze and assess traffic situations in real time under virtually all weather conditions.

Robustness and adaptivity are key challenges for intelligent traffic monitoring. Numerous sensors are installed at various locations (such as on poles, on gantries, or even in the pavement) to capture the traffic and estimate different traffic parameters. This diverse setting typically requires tedious sensor calibration and adapting the analysis algorithms to the observed scenes. This calibration and adaptation should be done with as little human intervention as possible. On the other hand, robustness is a precondition for integrating traffic monitoring to various applications.

Research on intelligent traffic monitoring has been ongoing for many years. Because it is widely recognized that image-based systems are flexible and versatile for advanced traffic-monitoring applications, most research has focused on image and video analysis.[2–4] Various image-analysis methods are applied to the data from individual cameras to estimate traffic parameters. These parameters can be related to individual vehicles such as detection, classification, and tracking or to the traffic behavior over a given time period, such as lane occupancy or travel time.

Another stream of research focuses on improving robustness by exploiting data from multiple sensors. Sensor fusion techniques can exploit the different characteristics of homogeneous or heterogeneous sensors. Rama Chellappa and his colleagues introduced a Markov Chain Monte Carlo technique for joint audio-visual vehicle tracking.[5] Acoustic beamforming estimates the direction of arrival, which in turn guides the visual tracking. Andreas Klausner and his colleagues exploited acoustic and visual sensors for vehicle detection and classification by extracting discriminative features from the different sensors and performing sensor fusion at the feature or decision level, respectively.[6] Manish Kushwaha and his colleagues also exploited acoustic and visual information for vehicle tracking in urban environments.[7] They perform multimodal fusion on an embedded sensor network in an urban environment.

Recently, several traffic-monitoring systems have been deployed on a larger scale to evaluate automated traffic analysis under real-world conditions. Tomás Rodríguez and his colleagues described a vision-based traffic-monitoring system that can detect vehicles in real time.[8] The major objective is to tackle some of the challenges in real-world deployments such as shadows, occlusions, day-and-night transitions, and slow traffic, which prohibit existing monitoring systems from achieving stable accuracy.

Their proposed system works autonomously for a certain period of time without human intervention and can adapt automatically to several environmental conditions.

Similarly, an earlier work proposed an example-based algorithm to detect moving vehicles in a vision-based traffic-monitoring environment under changing conditions.[9] The algorithm was designed to learn from examples, so it does not need to incorporate any prior knowledge (such as a prior vehicle model). The algorithm was evaluated under several varying environmental conditions and has achieved a satisfying performance.

The Visatram real-time vision system for automatic traffic monitoring follows a 2D spatiotemporal image-based automatic traffic-monitoring approach.[10] It handles vehicle counting, vehicle velocity estimation, and classification using 3D measurements. Furthermore, Marco Rigolli and Michael Brady reinforced the need to improve road safety by investigating inferences about driver behavior and learning normal behavior driving modes.[11] They proposed an agent-based approach for analyzing the drivers' behaviors.

References

1. H. Gharavi, K. Prasad, and P. Ioannou, "Scanning Advanced Automobile Technology," *Proc. IEEE*, vol. 95, no. 2, 2007, pp. 328–333.
2. V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A Survey of Video Processing Techniques for Traffic Applications," *Image and Vision Computing*, 2003, vol. 21, no. 4, pp. 359–381.
3. K.-T. Song and J.-C. Tai, "Image-Based Traffic Monitoring With Shadow Suppression," *Proc. IEEE*, vol. 95, no. 2, 2007, pp. 413–424.
4. R. Cucchiara, M. Piccardi, and P. Mello, "Image Analysis and Rule-Based Reasoning for a Traffic Monitoring System," *IEEE Trans. Intelligent Transportation Systems*, vol. 1, no. 2, 2000, pp. 119–130.
5. R. Chellappa, G. Qian, and Q. Zheng, "Vehicle Detection and Tracking Using Acoustic and Video Sensors," *Proc. IEEE Int'l Conf. Acoustic, Speech and Signal Processing*, IEEE Press, 2004, pp. 793–796.
6. A. Klausner et al., "Vehicle Classification on Multi-Sensor Smart Cameras using Feature- and Decision Fusion," *Proc. ACM/IEEE Int'l Conf. Distributed Smart Cameras* (ICDSC 2007), IEEE Press, 2007, pp. 67–74.
7. M. Kushwaha et al., "Target Tracking in Heterogeneous Sensor Networks Using Audio and Video Sensor Fusion," *Proc. IEEE Int'l Conf. Multisensor Fusion and Integration for Intelligent Systems*, IEEE Press, 2008, pp. 14–19.
8. T. Rodríguez and N. García, "An Adaptive, Real-Time, Traffic Monitoring System," *Machine Vision and Applications*, Springer, 2009, pp. 781–794.
9. L.J. Zhou, D. Gao, and D. Zhang, "Moving Vehicle Detection for Automatic Traffic Monitoring," *IEEE Trans. Vehicular Technology*, vol. 56, no. 1, 2007, pp. 51–59.
10. Z. Zhu et al., "VISATRAM: A Real-Time Vision System for Automatic Traffic Monitoring," *Image and Vision Computing*, vol. 18, no. 10, 2000, pp. 781–794.
11. M. Rigolli and M. Brady, "Towards a Behavioural Traffic Monitoring System," *Proc. 4th Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, ACM Press, 2005, pp. 449–454.
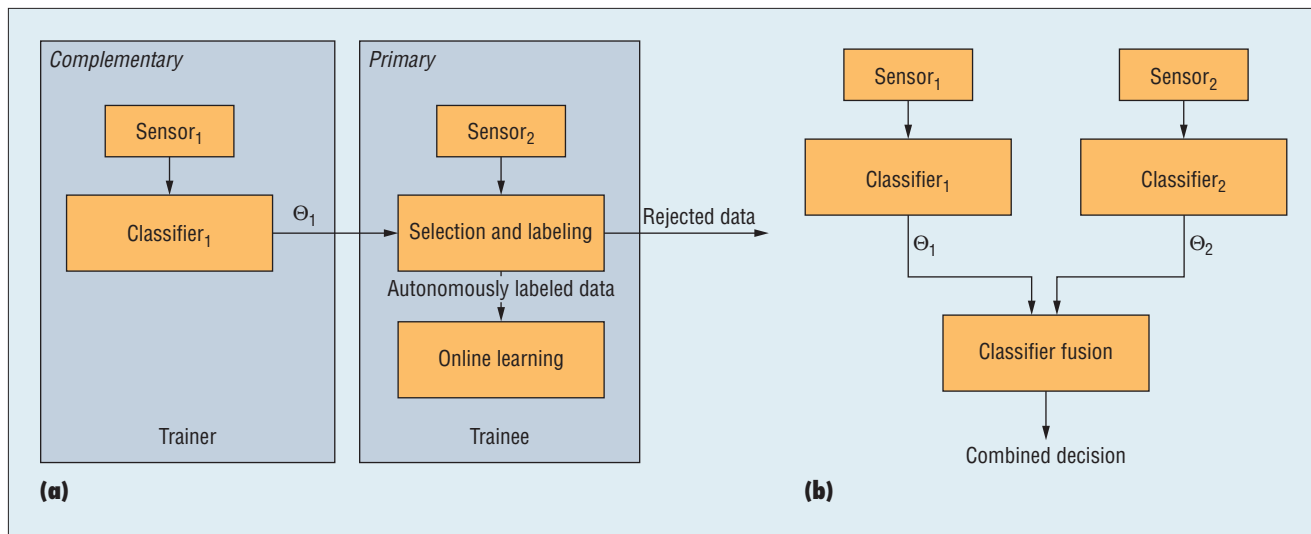
Figure 1. Self-training framework. (a) During the online training process, the audio sensor helps to select and label training data for the visual classifier. (b) The classification process combines the output from both sensors to improve performance.

must be performed continuously to allow for varying scenario conditions such as weather and illumination changes. Finally, these systems must be resource-effective to enable widespread use.

With these criteria in mind, in this article we focus on autonomous visual detection and classification of vehicles. We propose a self-learning framework with the goal of autonomously adapting multisensor classifiers to different sensor settings and scenes. Our system consists of a robust online boosting classifier that allows for continuous learning and concept drift. The learner is also less susceptible to class-label noise, which is hard to avoid in real-world self-learning applications. Furthermore, we incorporate an audio sensor as an additional, complementary source into the training process. This audio sensor source acts as teacher for self-learning of the primary visual classifier and helps to resolve ambiguities typically present in single-sensor settings.

To enable a mobile outdoor application, we implemented our system on an embedded platform and demonstrated it for vehicle classification on highways using audio and image

data. Our learning framework does not require any labeled visual data for online training and can significantly improve classification performance.

Self-Training Framework

Several traffic-monitoring systems exploit data from multiple and/or heterogeneous sensors.[2–4] Our system achieves robustness by first applying a robust online boosting classifier that also allows for continuous learning in order to train a visual appearance-based detector. Second, we incorporate an additional complementary sensor source (audio classification) into the learning process. The audio classifier acts as an autonomous supervisor. It is initially trained on a small set of labeled data and supports the visual online classifier in its continuous self-learning process. The audio classifier achieves sufficient accuracy with little training data and does not perform self-training, which ensures stability. This generic audio classifier is applicable to many scenarios—which justifies one-time human labeling—while the visual detector is trained autonomously for each individual scene. Furthermore, we abstain from complex microphone arrays and calibrations,

which are usually necessary for audio classification.

Our system uses a single consumer microphone acting as a teacher and complementary information source for the video classification, to reduce costs and allow for easy system deployment and maintenance. Another advantage of our approach is that we can use the audio classifier to resolve typical ambiguities between the vehicle classes (such as between cars and trucks), which are hard to resolve for visual classifiers but are easy for acoustic classifiers.

Figure 1 depicts the overall structure of our autonomous self-learning framework. Figure 1a describes the online training process using data from primary and complementary sensor sources, respectively. Figure 1b illustrates the collaborative classification process.

In the training process, both the audio and video sensors synchronously capture scene data. The complementary sensor acts as a trainer for the self-learning of the classifier of the primary sensor (trainee). The trainer's classifier is trained by using a small amount of hand-labeled audio data. For every detected object, the trainer performs a classification—using the a

priori trained classifier— and forwards a 2D parameter vector Θ_1 consisting of the decision (class label estimate) and its confidence value to the trainee. The trainee selects data only from objects with high classification confidence for its online training— that is, it refuses objects and its associated data when the trainer's confidence value is below a threshold. For selected objects, the trainee uses the trainer's classification result as a label.

After the audio-supported online training of the visual classifier, the trainer and trainee can operate as independent classifiers. To improve the overall performance, we combine the output of both classifiers based on their confidences (see Figure 1b).

Our proposed system is similar to previous systems based on cotraining,[5] where two classifiers are first trained independently on labeled data and then they train each other on unlabeled data. For instance, earlier works proposed cotraining a car detector[6] and using an audio-visual cotraining system for human gesture recognition.[7] However, our system differs from these approaches in three ways:

- we use continuous online learning,
- we do not need any human labeling effort for the visual classifier, and
- our audio classifier never performs self-updates, which ensures long-term system stability.

The latter point is supported by previous works that highlighted that cotraining's main assumption

(conditional independence[5]) is hard to fulfill in practice, and systems where an autonomous predictor (acting as a classifier) teaches another classifier have shown to perform better. For example, Peter Roth and his colleagues used a generative model to conservatively update an online classifier,[8] and Bo Wu and Ram Nevatia trained an online classifier using an "oracle" for pedestrian detection.[9] Our system differs from these two approaches in several aspects. First, we use an audio classifier as an autonomous teacher. Second, we use robust online boosting as a classifier and incorporate the teacher in the final classification process to resolve ambiguities among vehicle classes.

Acoustic Classification

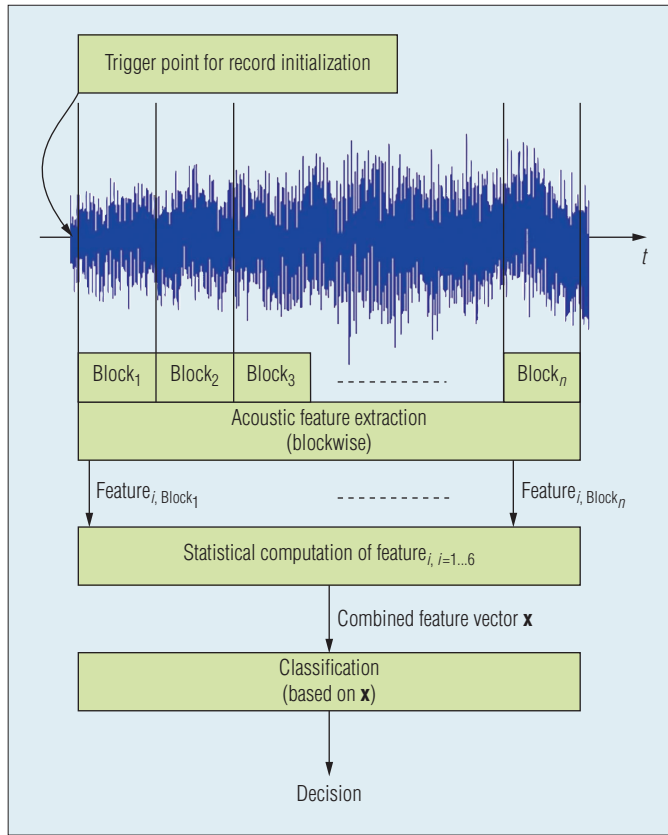Figure 2 depicts our acoustic classification system's basic structure.

Figure 2. Acoustic classification system. In our system, features are extracted from blocks of captured audio samples. These features serve then as input for the classifier.

A microphone captures the audio signal of passing vehicles. In the first processing step, we partition the audio samples into n blocks with a configurable block size. We then extract several acoustic features for each block individually. These block features are further abstracted into a single feature vector $\mathbf{x}$ by a statistical merging. The abstracted feature vector serves as input for the classifier.

The performance of the classification process strongly depends on the features' characteristics. Our goal is to select a set of highly discriminative features for the considered classes. In our case, we use six acoustic features, which are defined in the time, spectral, and cepstral domain, respectively.[10]

The short-time energy (ste) is a simple time-domain feature that is highly discriminative between cars and trucks and sensitive to noise. Spectral bandwidth (spbw), spectral roll-off point (spro), and two coefficients of band–energy ratio values (ber_6 and ber_7) exploit different characteristics of the vehicle's emitted acoustic spectrum. The spectral bandwidth measures the spread of frequencies around the spectral centroid. The spectral roll-off point indicates up to what frequency level a defined amount of percentage of the spectrum is accumulated. A higher roll-off value corresponds with more intense or higher frequencies. The band energy ratio values describe the ratio of energy in certain frequency subbands to the total signal energy. The ratio based on

the sixth and seventh subband yields more class-discriminative values than with the first five subbands. A cepstral analysis (cep) is performed as well.

We compute the combined value of feature i by statistically merging the $\text{Feature}_{i,\text{Block}k}$ for all blocks. This task is performed for all the different features. Thus, we combine the resulting features into a 6D feature vector $\mathbf{x}$:

$$\mathbf{x} = (\text{ste, spbw, spro, ber}_6, \text{ber}_7, \text{cep})^T. \quad (1)$$

We implemented and evaluated several classification algorithms such as k-nearest neighbor (KNN), linear and quadratic discriminant analysis (LDA, QDA), naive Bayes (NBC), a support vector machine (SVM), and an artificial neural network (ANN).[11] Each algorithm has its advantages and disadvantages depending on the data set. Therefore, the algorithm choice is based on the specific application domain. The classification algorithms return the estimated class labels with their confidence values as output.

Video Classification

A common choice in visual traffic analysis is simple background modeling (BGM), but this has several disadvantages. For instance, BGM is sensitive to shadows, cannot discriminate between different vehicle classes, and cannot detect vehicles in slow-motion scenarios such as traffic jams.

For visual classification, we therefore train an appearance-based model avoiding such problems. In particular, we follow previous work that showed that cascades of boosted classifiers and efficient image representation (integral images) lead to real-time appearance-based object detection systems.[1] However, our object detector differs in two aspects: First, we use online boosting for feature selection to allow for continuous learning without storing any training samples.[12]

Second, we use more robust loss functions for online boosting.[13] Using a robust learning algorithm is especially important in practice because label noise is an inherent problem in self-learning approaches.

In the training phase, we exploit the audio classifier to extract training data from scene-specific video streams captured by a noncalibrated consumer camera. To avoid hand labeling, we use a simple Gaussian background model to extract initial motion blobs.[14] We use the BGM to crop regions of interest for the boosting detector's training process. During operation mode, we only use the appearance-based detector. To extract proper training blobs, we apply different kinds of postprocessing such as size verification and positioning within the scene. Subsequently, we exploit the audio classifier, which can separate these samples into scenarios containing single vehicles of either class and scenarios containing multiple vehicles or no vehicle at all. We can also easily generate negative training examples from the scene with the audio classifier—that is, we crop random patches from the scene if neither the BGM nor the audio classifier are indicating that there are vehicles.

Because most traffic applications are concerned with both detecting vehicles and discerning different vehicle classes, we train two different detectors—one for trucks and one for cars. To resolve visual ambiguities among the different vehicle classes, we also incorporate the acoustic classifier to make the final classification between trucks and cars, which is an easier task for the audio classifier. We abstain from training a single detector for both cars and trucks because we must cover a high intra-class variance, which usually leads to higher model complexity and thus slower detectors.

Furthermore, we would lose the additional confidence provided by two visual detectors, which can be coupled with the audio classifier.

Collaborative Audio and Video Classification

During the classification phase, we unify the visual and audio cue by linearly combining the confidences of both classifier types. To classify a scene, we first generate a visual classifier by applying our two visual detectors for cars and trucks to identify several candidate regions where at least one of the two detectors provides a positive confidence. Then, we combine the visual classifiers' confidences with the confidences provided by the audio classifier. (All confidences are normalized to the range of [−1, +1] before fusion.)

To keep our approach simple, we use weighting parameters α and β for the combination of both confidences of the audio f_a and the visual f_v classifier. In particular, we use a simple arithmetic mean to weight the two confidences, both set to 0.5. (We can easily set α and β to more "reasonable" values—for instance, by using cross correlation on labeled samples or using more sophisticated weighting techniques.) Finally, by using a nonmaxima suppression, the highest vote is estimated by providing the according class for the candidate regions.

Experiments

Our experimental evaluation is based on real-world data sets of approximately 200 vehicles for each class (cars and trucks) from multilane freeway traffic. We partitioned the data sets into training and testing sets with 150 and 50 samples per class, respectively. Thus, we used 150 audio samples for each class to train the initial acoustic classifier. Figure 3 depicts our experimental setup. We directed
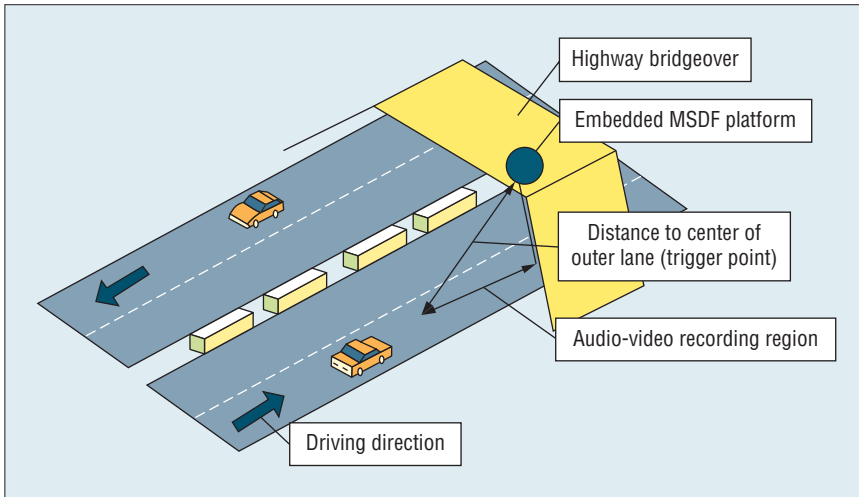
Figure 3. Experimental setup on a freeway with two lanes in both directions. The microphone and camera are connected to the embedded multisensor data fusion (MSDF) platform. The distance between the sensors and the outer lane is approximately 10 meters.
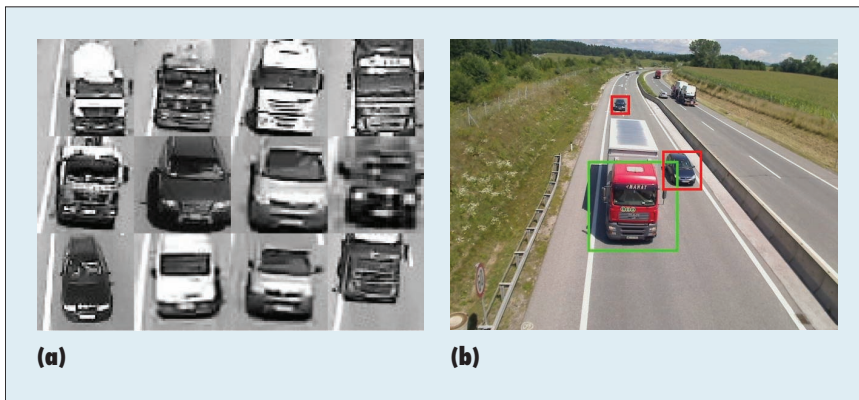

Figure 4. Visual vehicle detection and classification. (a) Some examples of automatically cropped sample patches and (b) the final detection output including the color-encoded classification result.

the microphone to the center of the outer lane. The audio data was recorded at 44.1 kHz in mono format with 16-bit resolution. The camera captured front-shot images at a frame rate of approximately 5 Hz.

Figure 4a shows some examples of cropped vehicle patches, and Figure 4b shows an example of the final detection and classification output. Video and audio recording were synchronized and started at a (virtual) trigger point. For each vehicle, the sensors captured up to four seconds of data—the actual recording

period depended on the vehicles' speed.

We performed the experiments on our MSEBX945 embedded computer board from Digital Logic with a SMX945-L7400 CPU module. This platform provides interfaces to several sensory devices such as audio, video, and laser sensors. We attached the microphone to a preamplifier from M-AUDIO, which was connected to the embedded platform via USB. The camera was directly interfaced with the platform via FireWire over MiniPCI.

Our experimental evaluation had two goals. We wanted to show that our autonomous framework enables online training of classifiers under real-world conditions without any hand labeling of the visual data. We also wanted to demonstrate that a collaborative classification of multiple sensors could gain significant performance improvements. For self learning, we used the audio sensor as a trainer, and we exploited classification for both cues.

In a previous work,[10] we showed that acoustic classifiers based on the feature vector given in Equation 1 achieve notable classification accuracies of up to 93.75 percent with quadratic discriminant analysis (QDA). The other algorithms we mentioned earlier achieved about 90 (for ANN, SVM, and LDA), 86.25 (for KNN) and 85 (for NBC) percent. We obtained all these results by five-fold cross validation. Thus, we use the QDA classifier as a trainer for our learning framework.

Autonomous Learning of Visual Classifiers

In our first experiment, we trained two vehicle detectors—one on car samples and the other on truck samples. For representation, we used simple Haar-like features[1] but abstained from training cascades because the classifiers can be kept simple due to their scene specificity. For all the experiments, we used 100 selectors each with 50 weak classifiers. For the online boosting, we applied a logistic loss function in the form of $\log(1 + e^{-yF(x)})$, which has proven more robust than the exponential loss usually applied in online boosting.[13] We set the starting shrinkage factor s_{start} to 1, but we decreased it with increasing number of selectors in the form of $s_t = s_{start}/(t + 1)$.

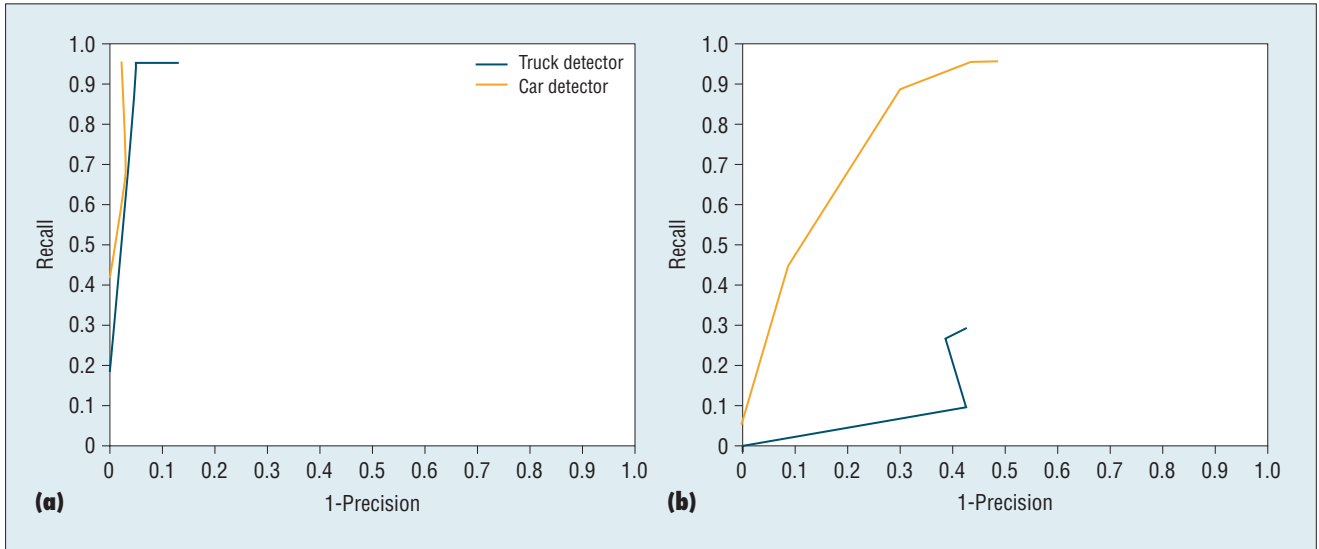As Figure 5a shows, our system can train well-performing car and

Figure 5. Automatically trained car and truck detectors. (a) The detectors achieve high classification performance when applied to test scenes only containing their training class. In this case, the detectors only discriminated the trained target class from the scene background. (b) When applied to scenes containing both vehicle classes, the performance degrades. The performance deteriorates dramatically, especially for the truck detections.

truck detectors without hand labeling any visual data. To demonstrate the practical relevance of our approach, we performed a second set of experiments where we degraded the performance of our teachers (audio classifiers). In particular, we varied the noise level from 0 percent (perfect teacher without any misclassification) to 25 percent (teacher with a 25 percent misclassification rate), which are ranges typically occurring in practice.

Tables 1 and 2 show that the recall rates hardly change with increasing noise level for both the car and truck detectors—that is, the number of false positives increases. The precision also remains constant for the truck detector, but the precision decreases with increasing noise for the car detector. However, we figured out in practice that if the recall rate stays high, a degraded precision can be corrected by applying smarter postprocessing in case of multiple detections. We did not apply postprocessing in this case and only applied classifiers to the class they have been trained on.

In the next two experiments, we tested the car detector only on

Table 1. Detector performance depending on different noise levels for cars.

Noise (%)	Recall	Precision	F-measure
0	0.95	0.78	0.86
5	0.95	0.48	0.64
10	0.98	0.40	0.57
25	0.98	0.34	0.50

Table 2. Detector performance depending on different noise levels for trucks.

Noise (%)	Recall	Precision	F-measure
0	0.98	0.17	0.29
5	0.98	0.16	0.28
10	1.00	0.15	0.27
25	1.00	0.15	0.26

sequences with cars and the truck detector only on sequences with trucks, respectively (Figure 5a). However, as Figure 5b shows, the performance degrades dramatically if the two detectors have to cope with instances of both classes at the same time. Training the car detector using some truck samples as negatives and vice versa leads to a decreased recall while the precision only slightly

increases. The main reason for this behavior is that the car detector cannot discriminate parts of a truck from real cars, leading to many false positives.

Collaborative Classification
In the third experiment, we used the same settings but focused on a collaborative classification of audio and video. The idea is that the visual
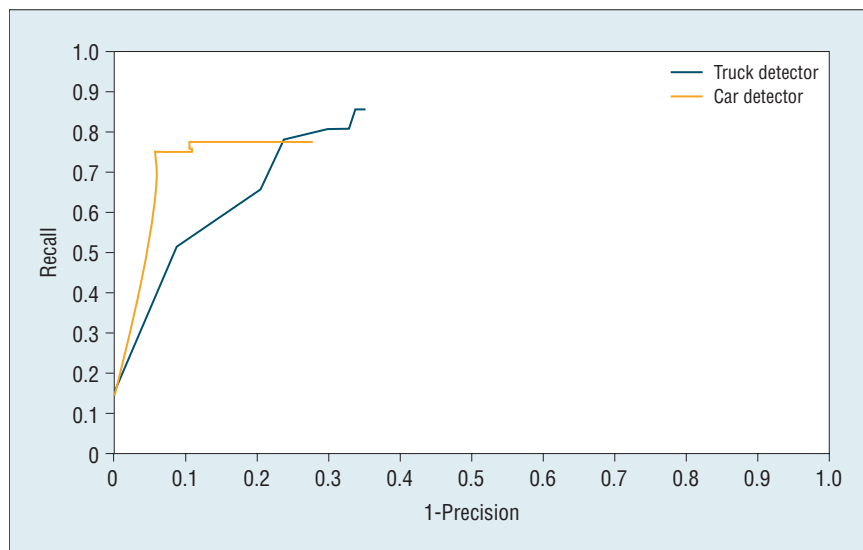
Figure 6. Final result of the collaborative classification using audio and video. The accuracy increases significantly if the audio classifier supports the visual detector in the final classification, especially for the truck detector.

Table 3. Classification performance using only a visual classifier.

Classifier	Recall	Precision	F-measure
Truck	0.29	0.57	0.39
Car	0.95	0.51	0.67

Table 4. Classification performance using visual and audio classifiers in combination.

Classifier	Recall	Precision	F-measure
Truck	0.85	0.71	0.78
Car	0.77	0.77	0.77

detector should be applied to locate the object in the video. Once an object has been detected, the audio classifier should support the visual detector to resolve ambiguities. In particular, after running both visual detectors over the video frame, we derived the final classification in a postprocessing step by computing a linear combination of the video and audio classifiers.

Figure 6 and Tables 3 and 4 depict the result of this collaborative classification, which leads to significantly improved detection results. By comparing the F-Measure, which gives an impression of the overall performance, we can see the improvement of the combined classification.

Because our approach does not need any calibration, it can be applied in mobile, flexible, low-cost traffic surveillance platforms. Potential future applications include traffic monitoring, free-flow toll collection, and law enforcement. Although we have demonstrated our multisensor method for vehicle classification, self-learning is a general concept with high potential for many applications. We are confident that it can serve as an important step toward versatile, autonomous, and intelligent traffic monitoring.∎

References

1. P. Viola and M. Jones, "Robust Real-Time Object Detection," *Int'l J. Computer Vision*, 2002, pp. 1–25.
2. A. Klausner et al., "An Audio-Visual Sensor Fusion Approach for Feature Based Vehicle Identification," *Proc. 2007 IEEE Int'l Conf. Advanced Video and Signal based Surveillance* (AVSS 2007), IEEE CS Press, 2007, pp. 1–6.
3. V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A Survey of Video Processing Techniques for Traffic Applications," *Image and Vision Computing*, vol. 21, no. 4, 2003, pp. 359–381.
4. A. Koutsia et al., "Intelligent Traffic Monitoring and Surveillance with Multiple Cameras," *Proc. 6th Int'l Workshop Content-Based Multimedia Indexing*, Hindawi Publishing, 2008, pp. 125–132.
5. A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-training," *Proc. 11th Ann. Conf. Computational Learning Theory*, Morgan Kaufmann, 1998, pp. 92–100.
6. A. Levin, P. Viola, and Y. Freund, "Unsupervised Improvement of Visual Detectors Using Co-training," *Proc. 9th IEEE Int'l Conf. Computer Vision*, vol. 2, IEEE Press, 2003, pp. 626–633.
7. C.M. Christoudias et al., "Co-training with Noisy Perceptual Observations," *Proc. IEEE Int'l Conf. on Vision and Pattern Recognition* (CVPR 2009), IEEE CS Press, 2009, pp. 1–8.
8. P.M. Roth et al., "Conservative Visual Learning for Object Detection with Minimal Hand Labeling Effort,"

THE AUTHORS

Horst Bischof is a professor at the Institute for Computer Graphics and Vision at the Graz University of Technology, Austria, a member of the scientific boards of the applied research centers ECV, VrVis and KNOW, and a board member of the Fraunhofer Institute für Graphische Datenverarbeitung (IGD). His research interests include object recognition, visual learning, motion and tracking, visual surveillance and biometrics, medical computer vision, and adaptive methods for computer vision. Bischof has a PhD in computer science from the Vienna University of Technology. Contact him at bischof@icg.tugraz.at.

Martin Godec is a research assistant at the Institute for Computer Graphics and Vision at the Graz University of Technology. His research focuses on efficient and robust learning of object representations for tracking and detection. Godec has an MSc in telematics (computer science and electrical engineering) from the Graz University of Technology. Contact him at godec@icg.tugraz.at.

Christian Leistner is a research and teaching assistant at the Institute for Computer Graphics and Vision at the Graz University of Technology and is working on his doctoral thesis on online and off-line SSL methods for object detection, tracking, and recognition. Leistner has an MSc in telematics (computer science and electrical engineering) from the Graz University of Technology. Contact him at leistner@icg.tugraz.at.

Andreas Starzacher is a research assistant at the Institute of Networked and Embedded Systems at Klagenfurt University and is working on his doctoral thesis on data fusion and embedded systems. Starzacher has an MSc in computer science from Klagenfurt University. Contact him at andreas.starzacher@uni-klu.ac.at.

Bernhard Rinner is a full professor and chair of pervasive computing at Klagenfurt University (Austria), where he is currently serving as Vice Dean of the Faculty of Technical Sciences. His research interests include embedded computing, embedded video and computer vision, sensor networks and pervasive computing. Rinner has a PhD in telematics from Graz University of Technology. He is a member of IEEE, the International Federation for Information Processing (IFIP), and Telematik Ingenieurverband (TIV). Contact him at bernhard.rinner@uni-klu.ac.at.

Proc. 27th Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) Symp., Springer Verlag, 2005, pp. 293–300.

9. B. Wu and R. Nevatia, "Improving Part Based Object Detection by Unsupervised, Online Boosting," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition* (CVPR 2007), IEEE CS Press, 2007, pp. 1–8.

10. A. Starzacher and B. Rinner, "Single Sensor Acoustic Feature Extraction for Embedded Realtime Vehicle Classification," *Proc. 2nd Int'l Workshop Sensor Networks and Ambient Intelligence*, IEEE Press, 2009, pp. 1–6.

11. A. Starzacher and B. Rinner, "Embedded Realtime Feature Fusion Based on ANN, SVM and NBC," *Proc. 12th Int'l Conf. Information Fusion* (Fusion 2009), IEEE Press, 2009, pp. 1–8.

12. H. Grabner and H. Bischof, "On-Line Boosting and Vision," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition* (CVPR 2006), vol. 1, IEEE CS Press, 2006, pp. 260–267.

13. C. Leistner et al., "On Robustness of Online Boosting: A Competitive Study," *Proc. 3rd IEEE Online Learning for Computer Vision Workshop* (ICCV 09), IEEE CS Press, 2009, pp. 1362–1369.

14. C. Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition* (CVPR 1999), vol. 1, IEEE CS Press, 1999, pp. 246–252.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.