Automatic Detection and Reading of Dangerous Goods Plates

Peter M. Roth, Martin Köstinger, Paul Wohlhart, and Horst Bischof
Graz University of Technology, Institute for Computer Graphics and Vision, Austria
{pmroth,kostinger,wohlhart,bischof}icg.tugraz.at

Josef A. Birchbauer
Siemens AG Österreich, Corporate Technology (CT T CEE), Austria
josef-alois.birchbauer@siemens.com

Abstract

In this paper, we present an efficient solution for automatic detection and reading of dangerous goods plates on trucks and trains. According to the ADR agreement dangerous goods transports are marked with an orange plate covering the hazard class and the identification number for the hazardous substances. Since under real-world conditions high resolution images (often at low quality) have to be processed an efficient and robust system is required. In particular, we propose a multi-stage system consisting of an acquisition step, a saliency region detector (to reduce the run-time), a plate detector, and a robust recognition step based on an Optical Character Recognition (OCR). To demonstrate the system, we show qualitative and quantitative localization/recognition results on two challenging data sets. In fact, building on proven robust and efficient methods, we show excellent detection and classification results under hard environmental conditions at low run-time.

1. Introduction

Dangerous goods transports such as for fuel or liquid nitrogen have to be marked by an orange plate covering the hazard class and the identification number of the hazardous substances. These plates are standardized according to the European Agreement concerning the International Carriage of Dangerous Goods by Road, commonly known as ADR[1]. In particular, these plates having a size of 40x30cm are either void (mixed transport) or contain two codes: the class/hazard-identification (*e.g.*, 3 or 33 for flammable liquids) and the UN number (*e.g.*, 1202 for diesel fuel or heating oil) [13]. For a more detailed discussion we refer to [13].

[1] Accord europen relatif au transport international des marchandises Dangereuses par Route. Published as document ECE/TRANS/202, Vol.I and II (ADR 2009) [13]

Knowing the current hazard potential of transports the safety in tunnels or marshaling-yards can be increased. Especially, in cases of accidents knowing the involved hazardous substances can help to settle the right steps. Due to the ever increasing number of surveillance cameras it becomes feasible to extract the provided information using automatic systems. The goal of this work is to automatically identify dangerous good transports by extracting the information from these dangerous goods plates, where the recognition of dangerous goods plates can be seen in the context of text reading in natural scenes (*e.g.*, [1, 2]) or in particular with automatic number plate detection (ANPR, *e.g.*, [4, 9, 12]). However, additionally also void plates have to be handled making the task more complicated.

Assuming a “soft”-calibrated acquisition setup providing horizontally aligned images, the main steps in such processing queues can be described as localization (of the plates) and reading the information by using an Optical Character Recognition (OCR). The most prominent way for the localization (detection) is to apply a sliding window technique (*e.g.*, [3, 15]). A previous learned model - typically a discriminative model estimated from positive samples (i.e., the plates) and from negative (i.e., all possible backgrounds) is applied on the image, and all locations that are consistent with the model are reported. Once the location is initialized the OCR is applied to extract/read the text.

However, such processing queues are not feasible for the given task for two reasons. First, the system should run in real-time, which is hindered by high resolution images and, second, in a real-world setup the system has to cope with changing conditions and images of low quality. Thus, in the following we introduce an efficient and robust system for detecting and reading dangerous goods plates. In particular, we propose a multi-stage system, including a saliency detection, a plate detector, and a recognition module. In the experimental results we show quantitative and qualitative detection and recognition results.

2. System Description

For our task, however, the processing queue described Section 1 is infeasible for two reasons. First, due to the large image sizes 100,000s of locations would have to be analyzed resulting in insufficient run-time. Second, since the dangerous goods plates can mainly be described as homogeneous regions it is quite hard to discriminate them from similar regions that can be found on dangerous goods transports (*e.g.*, trucks and trains) or even on the road, resulting in an unacceptable number of false detections.

Thus, in the following we propose a five-stage approach overcoming these problems. To reduce the detection time, we perform a *salient region detection* (based on a region-based segmentation) to identify possible candidate regions and run a *detector* on the identified regions. In an *enhancement* step the contrast in the detected plates is increased to improve the classification using an *Optical Character Recognition (OCR)*. Finally, these results are checked versus a given database in a *lookup* step. The whole processing queue is illustrated in Figure 1.

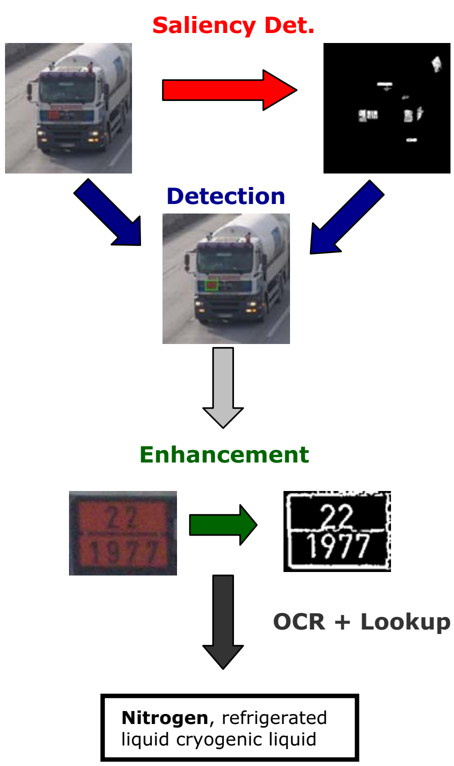
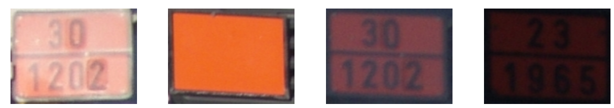


Figure 1: Dangerous goods plate detection/recognition system: (a) salient region detection, (b) detection, (c) enhancement, (d) OCR, and (e) lookup step.

2.1. Saliency Detection

Having in mind that all plates have to be orange, it would be obvious to use color information for segmentation, *i.e.*, to search for orange regions. However, due to the large variability in color appearance resulting, *e.g.*, from different illumination conditions, shadows, or dirt this is infeasible in practice. This is illustrated in Figure 2 for correctly identified plates: the color ranges from light orange over carmine to dark gray, which clearly shows that color is an insufficient information cue for our segmentation task.



(a) (b) (c) (d)

Figure 2: Examples of real-world dangerous goods plates showing that due to the high variability a simple color model is insufficient to get a stable classification.

Figure 2, however, also reveals that the plates are mainly characterized by homogeneous regions, which can perfectly be described by a region-based segmentation. In particular, in our system we build on the Maximally Stable Extremal Region (MSER) algorithm of Matas et al. [10], which has proven to be one of the best interest point detectors in computer vision (*i.e.*, it is invariant to affine transformations, allows for multi-scale detection, etc.).

The MSER method belongs to the family of watershedding algorithms [14], which generate a binary image B from an intensity image I by considering all possible thresholds θ:

$$\mathbf{B}_\theta = \begin{cases} 1 & \mathbf{I}(x) \geq \theta \\ 0 & \text{otherwise} . \end{cases} \tag{1}$$

A maximally stable extremal region (MSER) is then a connected region which is stable over a large number of thresholds. To estimate the MSERs, first all pixels are ordered by their intensity (*i.e.*, 0, ... ,255). Then, iteratively by increasing the threshold the corresponding pixels are added to the binary image and a list of connected components is returned.

The original formulation was limited to gray-scale images, however, for our application color would still provide valuable information. Donoser and Bischof [5] proposed to estimate a color-space transformation and applied the original MSER approach on the transformed images. In particular, they estimated a multivariate Gaussian distribution of the original RGB values and ordered the pixels by their Mahalanobis distance to this Gaussian distribution [5]. In contrast, Forssen [7] proposed to detect regions that are stable across a range of time-steps in an agglomerative clustering of image pixels based on proximity and similarity in color

[7]. The approach of Forssen is computationally too expensive and limited due to noisy images. However, for our purpose we can adapt the method of Donoser and Bischof. Since we cannot estimate a Gaussian distribution from the given data, we use Support Vector Regression [6] instead on a combined $[R, G, B, cos(H), sin(H), S, V]$ color-space to learn a robust measure for ”orangeness”. In this way, we obtain a likelihood image covering the most essential information for our task, which can be used as input by the standard MSER method. Please note, using the likelihood image improves the segmentation but does not fully solve the task (due to the high variability within the data).

2.2. Plate Detector

Once we have detected the salient regions, *i.e.*, the MSERs, within the image, as illustrated in Figure 1, we use this information to reduce the computational effort when running a sliding-window-based detector. Thus, the model is evaluated only for image locations that were identified during the segmentation process. In particular, to increase the detection performance we run three detectors in parallel: one for void plates, one for plates containing numbers, and one covering both cases.

In general, any classifier-based detector can be applied, but since, as can be seen from Figure 3(b), gradients provide an informative description of the data, in our system we apply the HOG-Detector of Dalal and Triggs [3]. Histograms of oriented gradients (HOGs) are locally normalized gradient histograms, which are estimated as follows. Given an image $\mathbf{I}$ the gradient components $g_x(x, y)$ and $g_y(x, y)$ for every position (x, y) the image is filtered by 1-dimensional masks $[-1, 0, 1]$ in x and y direction [3]. This is illustrated in Figure 3(a) and (b).

Then, to create the HOG descriptor, the image is divided into non-overlapping 10x10 cells. For each cell, the orientations are quantized into 9 bins and weighted by their magnitude. Groups of 2x2 cells are combined in overlapping blocks and the histogram of each cell is normalized using the L2-norm of the block, which is illustrated in Figure 3(c). Using the thus obtained descriptors we then learn a discriminative classifier using a Support Vector Machine. During the evaluation for all image locations pre-selected by the segmentation step a HOG descriptor is estimated and checked versus the previously learned model.

(a) (b) (c)

Figure 3: HOG descriptor: (a) input image, (b) gradient image, and (c) descriptor.

2.3. Contrast Enhancement

To obtain reasonable results of the OCR, we require images of high contrast. However, as can be seen from Figure 4(a), the detected regions cannot directly be used as input for the OCR step. Although, the structure of the plates assumes a bimodal nature, this cannot be captured by simple foreground-background segmentation methods. Due to the severe intra-class variance of the plates' background automatic thresholding methods such as Otsu's [11], which is used in most OCR systems, are prone to errors and artifacts. Thus, as a pre-processing step we binarize the potential plate regions by generating a color model in which the plate's orange background can be discriminated well from other colors. In particular, we apply the Support Vector Regression likelihood map introduced in Section 2.1. The required binarized image showing the letters and the frame of the plate are estimated by thresholding the convolution with a Laplacian of Gaussian filter. As shown in Figure 4(b), in this way excellent segmentations can be obtained.

(a) (b)

Figure 4: The original input images are insufficient as input for the OCR – a preprocessing step is required: (a) original and (b) pre-processed data.

2.4. Number Recognition

Now, having estimated the location of potential dangerous goods plates, we can extract the contained information, *i.e.*, the hazard identification number (HIN) and the UN number, using an OCR. However, the binary images obtained by the enhancement step as described in Section 2.3 can not directly be used as input for the OCR. The image still contains non-text artifacts such as the plates' border and minor segmentation inaccuracies, which would lead to wrong recognition results. Hence, to remove those lines, we apply a Hough transformation [8] to the binarized image. In Hough space the candidate angles are constrained and a non-maximum suppression next to the located maxima is applied. By fitting the Hough lines to three horizontal and two vertical lines, as illustrated in Figure 5(a) and (b), the border can be removed and the image patch can be subdivided into two separate parts which can be used as input for the OCR. This is illustrated in Figure 5(c).

In particular, we use a standard OCR software (*i.e.*, Tesseract OCR[2]). For feature extraction the OCR uses a polygonal approximation of the connected components of

[2]http://sourceforge.net/projects/tesseract-ocr

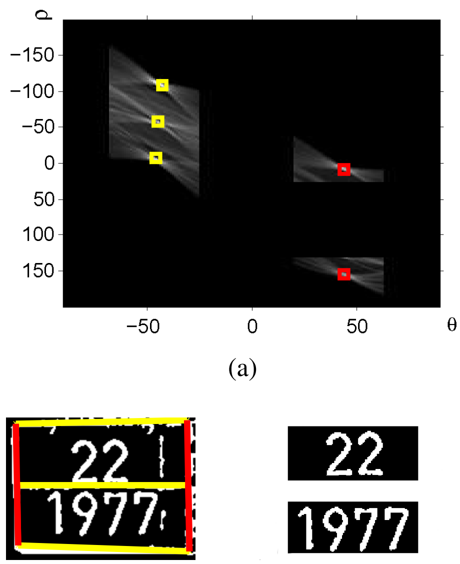


Figure 5: Number extraction for OCR: using a Hough transformation on the binarized images (a) the horizontal and vertical lines are identified (b) and removed yielding the input for the OCR (c).

the binary image. The extracted features are used to match to a set of allowed prototypes. To minimize the error rate and to speed up the text extraction, we reduced the allowed character set to numbers and the letter X, which indicates a dangerous reaction of the substance with water.

2.5. Lookup

Finally, after running the OCR the obtained results are checked versus the UN database containing all possible number combinations. Moreover, in this step errors of the OCR can be compensated due to restrictions given by the database entry. The unique 4-digit code allows direct inference on the hazard identification number (HIN). The HIN and UN are checked vice-versa for correspondence and this information can be used to detect and correct errors of the independent number recognition results. The errors are mostly based on mixed up digits (*e.g.*, 8 instead of 0) or clutter that resembles a digit in a similar geometric configuration. For instance, if the UN and the HIN were recognized as 1202 and 38, the HIN can be corrected to 30.

3. Experimental Results

To demonstrate our system, we generated two real-world data sets, *Trucks* and *Railway*. The *Trucks* data set illustrating different conditions during a whole day consists of 54 images showing trucks with and 177 images showing trucks without orange plates; all of them having a resolution of 4 mega-pixels. The *Railway* data set consists of 164 images showing wagons with and 380 images without plates. The resolution varies from 6 to 16 mega-pixels depending on the vehicle length. Hence, it is clear that an efficient system is required. In fact, on a customary computer system (*i.e.*, Intel Core2Duo, using only one core) the whole queue runs in less than 1 second for the *Trucks* and, depending on the image size, in 1-2 seconds on the *Railway* data set. If a less sophisticated processing queue (*e.g.*, skipping the saliency detection) is applied, the computation time per image can go up to 10 minutes! As can be seen from Figures 6 and 7 the data sets are challenging since they cover realistic problems such as changing illumination, shadows, cluttered backgrounds, polluted plates, and low contrasts.

Due to the constraints given by the specific application the two tasks mainly differ in the image acquisition setup and the resolution of the images. Since for trucks the orange plates have to be mounted in the front and the rear of the vehicle, existing high resolution image acquisition setups which are typically mounted on highways can be applied. In contrast for trains the plates are mounted at the side and due to the vehicle length a line-scan camera is required to capture the single images. The dangerous goods plates, however, follow the same conventions, *i.e.*, they are composed by a hazard-identification and an UN number, thus, the same processing steps can be carried out. To evaluate the system, we analyze both, the detection accuracy (localization) and the performance of the number recognition.

3.1. Trucks

First, we evaluate the plate detection on the *Trucks* data set, where we show the results obtained by the three classifiers discussed in Section 2.2. The obtained results are summarized in Table 1, where we show the recall (*i.e.*, the percentage of correctly detected plates) and as a measurement for the precision the false positives per image (FPPI).

	recall	FPPI
all plates	94.44%	0.12
plates with numbers	96.66%	0.01
void plates	79.16%	0.06

Table 1: Detection characteristics for all three detector types for the *Trucks* data set.

It can be seen that for the combined classifier for a recall of 94% only 1 false positive is reported per 10 frames. This is especially an excellent result since the void plates do not contain any informative structure and can easily be mixed up with other homogeneous regions in the images. Moreover, if reducing to plates containing numbers (which are the interesting ones), for a recall of 97% only 1 false positive per 100 images is reported. These results are illustrated in Figure 6. From Figure 6(a) it can be seen that even under difficult varying illumination conditions accurate detections can be obtained. In contrast, as shown in Figure 6(b), misses mainly result from highly polluted plates and low contrasts. In both cases the segmentation fails and no detections are reported. The false detections shown in Figure 6(c) are detected as void plates, which is plausible considering the appearance.

Next, we evaluate the recognition performance for both the HIN and the UN number. For that purpose all plates that were detected in the previous step are used as input for the OCR and the Lookup step. Table 2 shows these results, separated into plates with and without numbers. For the plates containing a dangerous goods information around 95% of all plates are classified correctly whereas for all void plates also a void response was returned, *i.e.*, none of these plates were wrongly classified! Hence, even if the detection step returns a small number of false detections (mainly void plates) this would not harm the overall system, since no (wrong) information is extracted. Thus, applying the combined classifier for this task is a considerable trade-off.

	plates with number	void plates
HIN	94.45%	100.00%
UN	96.36%	100.00%

Table 2: Recognition characteristics for HIN and UN number for the *Trucks* data set.

3.2. Railway

Next, we run the same experiments on the *Railway* data set. Since this data set does not contain images with void plates the detection task was run only using the model not representing void plates. The obtained detection and recognition results are summarized in Table 3 and Table 4. In addition, illustrative detection results are given in Figure 7.

It can be recognized that compared to the *Trucks* data set the detection rate is a little bit lower. This can be explained by the quality of the images (noise, low resolution, artifacts of the line-scan camera, etc.) and the partly distorted geometry of the plates which are mounted on the curved tank wagons. However, as can be seen from Figure 7(b) often even for a human it would be quite hard to recognize the orange plates. In contrast, Figure 7(a) shows that excellent detection results are obtained under varying environmental conditions, where the small number of false positives mainly result from homogeneous regions containing parts of text or logos. However, such detections return a void response during the OCR step and do not harm the recognition performance. In contrast, as shown in Table 4 for the correctly detected plates an excellent recognition performance is obtained.

	recall	FPPI
plates with numbers	93.75%	0.02

Table 3: Detection characteristics for the plate detector for the *Railway* data set.

	plates with number	void plates
HIN	96.67%	–
UN	95.33%	–

Table 4: Recognition characteristics for HIN and UN number for the *Railway* data set.

4. Conclusion

In this paper, we tackled the problem of automatic detection of dangerous goods plates, which can be of considerable interest for transportation safety. For instance, in case of accidents knowing which dangerous goods are involved in the accident could help to take the right necessary steps. In particular, we propose a five-stage method: salient regions detection (to reduce computational costs), detection of plates, contrast enhancement, text extraction, and a lookup in the database. For all of these steps we apply proven and widely used methods assuring the required stability. The qualitative and quantitative results, which were obtained on challenging data sets, show that the approach works quite robustly, even in realistic scenarios. Future work would include locating and recognizing the diamond-shaped dangerous goods symbol sign, reading the license plates by means of ANPR, and improving the image acquisition setup.

Acknowledgment

The work was supported by the FFG projects MDL (818800) and SECRET (821690) under the Austrian Security Research Programme KIRAS.

References

[1] U. Bhattacharya, S. K. Parui, and S. Mondal. Devanagari and bangla text extraction from natural scene images. In *Proc. Intl' Conf. on Document Analysis and Recognition*, pages 171–175.

[2] X. Chen and A. L. Yuille. Detecting and reading text in natural scenes. In *Proc. CVPR*, volume II, pages 366–373, 2004.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, volume I, pages 886–893, 2005.

[4] M. Donoser, C. Arth, and H. Bischof. Detecting, tracking and recognizing license plates. In *Proc. Asian Conf. on Computer Vision*, volume II, pages 447–456, 2007.

(a) (b) (c)

Figure 6: Illustrative detection results - *Trucks* data set: (a) correctly detected plates, (b) missed plates, and (c) false positives.

(a)

(b)

Figure 7: Illustrative detection results - *Train* data set: (a) correctly detected plates and (b) missed plates.

[5] M. Donoser and H. Bischof. Efficient maximally stable extremal region (MSER) tracking. In *Proc. CVPR*, volume I, pages 553–560, 2006.

[6] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In *Advances NIPS*, pages 155–161, 1997.

[7] P.-E. Forssén. Maximally stable colour regions for recognition and matching. In *Proc. CVPR*, 2007.

[8] P. V. C. Hough. Method and means for recognizing complex patterns. US patent nr. 3069654, 1962.

[9] W. Jia, H. Zhang, X. He, and M. Piccardi. Mean shift for accurate license plate localization. In *Proc. IEEE Conf. on Intelligent Transportation Systems*, volume I, pages 732–737, 2006.

[10] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. British Machine Vision Conf.*, pages 384–393, 2002.

[11] N. Otsu. A threshold selection method from grey level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[12] V. Shapiro, G. Gluhchev, and D. Dimov. Towards a multinational car license plate recognition system. *Machine Vision and Applications*, 17(3):173–183, 2006.

[13] United Nations. ADR: European agreement concerning the international carriage of dangerous goods by road. 2009.

[14] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. on PAMI*, 13(6):583–598, 1991.

[15] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, volume I, pages 511–518, 2001.