Route Segmentation into Speed Limit Categories by using Image Analysis

Philippe Foucher, Emmanuel Moebel and Pierre Charbonnier
Cerema/DTer Est/LR Strasbourg, ERA 27, 11 rue Jean Mentelin, BP9, 67035 Strasbourg, France

Keywords: Road Scene Classification, Speed Limits, Image Analysis, Route Segmentation, Evaluation.

Abstract: In this contribution, we address the problem of road sequence segmentation into speed limit categories, as perceived by the user. We propose an algorithm that is based on two processing steps. First, the images are classified independently using a standard random forest algorithm. Low-level and high-level approaches are proposed and compared. In the second phase, a sequential smoothing of the results using different filters is applied. An evaluation based on two databases of images with ground truth shows the pros and cons of the methods.

1 INTRODUCTION

Historically, most vision-based road scene analysis systems have been dedicated to detecting and recognizing components of the scene, such as pedestrians, cars, obstacles, signs, or road markings. In this paper, we address a more recent application of image analysis, which aims at classifying the road scene, as a whole, into a semantic category. Applications are related to the research about self-explaining roads, in which driving psychologists investigate about the relationships between visual characteristics of the road and its environment and road "readability" (Charman et al., 2010). In other words, the mental categorization of roads is known to have an impact on the driver's behavior. Hence, road legibility should be diagnosed and enhanced by appropriate treatments of the infrastructure, to improve road safety. For the moment, these works are based on the analysis of road image sequences by human operators, which represents a considerable work at the scale of a road network (in our application, image sequences are collected by an inspection vehicle along itineraries, typically one image every 5 meters).

In this work, we investigate the segmentation of routes into speed limit categories, as perceived by the user. Among the six speed limits defined in the French regulations, we here consider 4 categories : 50 km/h ; 70 km/h ; 90 km/h and 110 km/h. Note that, in France, speeds are limited to 50 km/h in built-up areas, 110 km/h on dual carriageway roads, that are separated by a central island or barrier and 90 km/h on rural two-way roads. In some dangerous situations, such as crossings, driving speeds are limited to 70 km/h (so this class is difficult to discriminate based on visual cues only, even for human operators). Sample images of the 4 categories are shown on fig. 1.

As a first contribution, we propose a two-step categorization method. The first step relies on a classification of individual images, which is performed using either a low-level approach or a high-level scheme based on an intermediate representation of the scene. Note that, in both cases, the classification into four categories is obtained using the random forest algorithm (Breiman, 2001). Of course, analyzing images individually is inappropriate, so the second step of the proposed method consists in a sequential smoothing of the classification results to obtain a consistent route segmentation. The second contribution of the paper is that we perform a systematic evaluation of the algorithm on large image databases.

The rest of the paper is organized as follows. We first propose a short review of related works in Sect. 2. Then, in Sect. 3, we describe the algorithm for classifying images into speed limit categories. In Sect. 4, the experimental setup is presented. In Sect. 5, we comment the experimental results. Sect. 6 is dedicated to conclusions and future works.

2 RELATED WORK

Different approaches have been proposed to classify natural scenes into semantic categories. We refer the

Foucher P., Moebel E. and Charbonnier P..
Route Segmentation into Speed Limit Categories by using Image Analysis.
DOI: 10.5220/0005312704160423
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 416-423
ISBN: 978-989-758-090-1

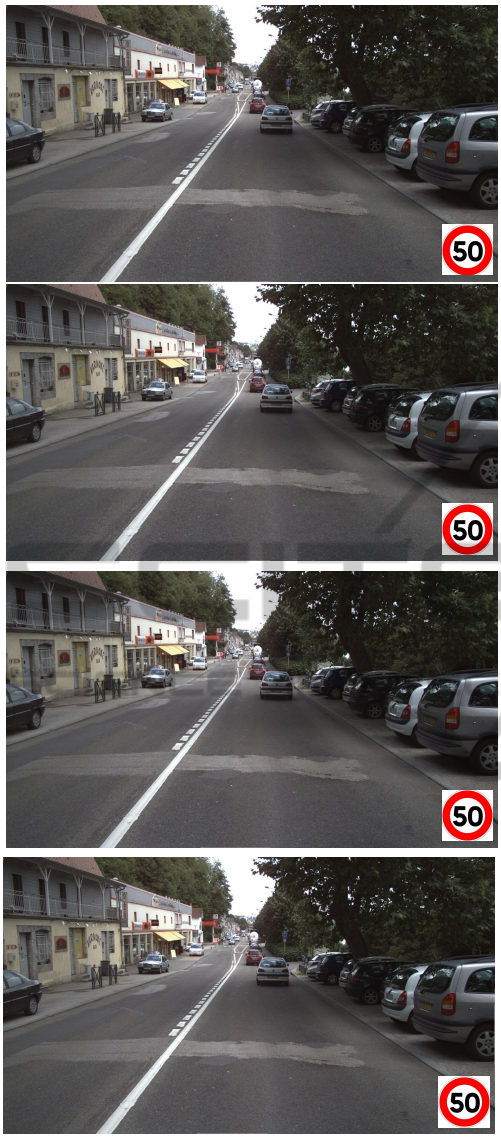


Figure 1: Examples of speed limit categories.

reader to (Bosch et al., 2007) for a global review of scene classification strategies. In a first approach, the classification is based on statistical learning using low-level visual features, such as color, edges or textures, i.e. images are classified without any semantic knowledge of the objects in the scene. In (Vailaya et al., 1998), the classification of outdoor images into "city" or "landscape" category is based on color and edge directions features. Experiments show that edge direction features are discriminative enough to be used alone in the algorithm. In (Wu and Rehg, 2011), the CENTRIST(*CENsus TRansform hISTogram*) descriptor was introduced and assessed on five graylevel image data sets, including scene recognition data. The use of CENTRIST improves the classification performances compared to other usual descriptors. It may be noticed that the mCentrist descriptor, a multi-channel version of CENTRIST, has been recently introduced (Xiao et al., 2014). On four image data sets (indoor and outdoor images), the authors have shown that mCentrist enhances the performances of CENTRIST in terms of scene classification.

In a second approach, an intermediate representation of the scene is built in order to bridge the semantic gap between low-level descriptors and high-level semantic information. The bag-of-words (*BOW*) approach is commonly used to obtain a semantic model of the image. In this framework, images are represented by histograms of visual words, which are generated from local image patches by some unsupervised classification algorithm using SIFT (Lowe, 1999) or histograms of oriented gradient (*HOG*) (Dalal and Triggs, 2005) as input features. These histograms of visual words are then considered as high-level attributes for scene classification. In (Bosch et al., 2007), a methodology has been proposed to compare the performances of the low-level and high-level strategies. The authors have shown that the use of an intermediate semantic modeling is more appropriate when the number of categories is high and in the presence of ambiguities between classes. In contrast, low-level features can be useful when the number of semantic categories is low and when the categories are easily distinguishable. The computation time is also lower.

We may note that, among the numerous existing scene classification works, only a few are dedicated to the particular case of road scenes. A scheme, based on low-level color and texture features, is proposed to classify road driving environment into four different categories (off-road, major road, motorway, urban road) in (Tang and Breckon, 2010; Mioulet et al., 2013). The aim of this research is to propose an autonomous sensor to adapt the vehicle dynamics (traction, braking, engine dynamics) to the driving environment and the descriptors are computed in three regions of interest, selected to extract relevant properties of the road and its surroundings. Road scene understanding is also investigated in (Ess et al., 2009) to identify road type. This research is focused on urban road scenes and the authors propose a two-step method. A supervised learning process is first used to segment images into semantic objects (walls, road, cars, grass...) and a classification algorithm is then applied to identify 8 road types and detect the presence of 3 kinds of objects (cars, pedestrians, pedestrian crossing).

In the literature, to our knowledge, there is no paper that concerns the classification of images into

speed limit categories. The closest work to our contribution is probably Ivan and Koren's paper (Ivan and Koren, 2014) in which the authors propose to automatically classify road scenes as built-up and non-built-up areas using a bag-of-words methodology. However, our approach differs in the number of categories and, moreover, we smooth the scene classification results along the sequence to obtain a relevant route segmentation. Note that, in (Ess et al., 2009), the effect of temporal smoothing using Markov Random Fields (*MRF*) in the segmentation process is evaluated, showing the benefits of using sequential information. However MRF's are not used in the scene classification step due to computational limitations.

3 PROPOSED METHOD

Our method aims at classifying images into four speed limit categories. It is a two-step algorithm with an image classification phase and a post-processing phase that smoothes the results by considering the arrangement of the images in a temporal sequence.

3.1 Image Classification Algorithms

We propose in fact two algorithms, based either on low-level features or on a high-level representation, as summarized in Fig. 2. In both cases, the scene is classified using the random forest meta-classifier (Breiman, 2001).

A random forest is an ensemble of decision trees that predict a class from a set of features. The prediction of a tree is obtained by using successive binary decision rules. Every tree is learned independently on a randomly selected sample of features of the training data set, with replacement (bootstrap process). The result of the random forest algorithm is a probability which is obtained by combining the outputs of the trees. A sample is assigned to the class that corresponds to the maximum probability. Note that random forests are intrinsically multi-class and can be easily parallelized. The input features of the random forest differ between our low-level and high-level approaches. In the low-level algorithm, mCentrist features are computed directly from the image. In the second approach, the features are extracted from an intermediate representation of the image, previously generated using a clustering algorithm.

3.1.1 Low-level Approach: Centrist and mCentrist Features

The Centrist descriptor (Wu and Rehg, 2011) is an

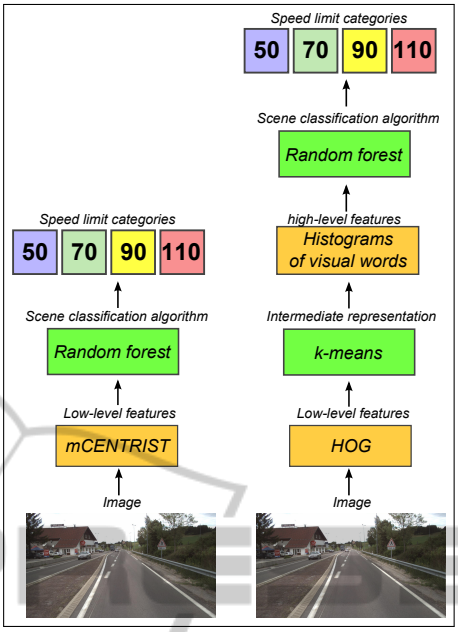


Figure 2: Scene classification methodology: low-level (left) and high-level (right) approaches.

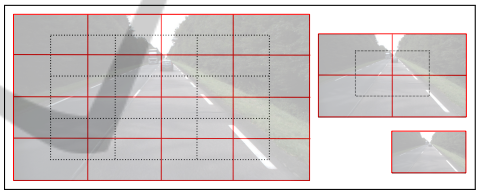


Figure 3: Spatial pyramid with levels 2, 1 and 0. For example, the level 2 representation (left image) is split into 25 blocks : 16 red blocks (continuous lines) and 9 black blocks (dash lines). There is an overlap of 25% between red blocks and blacks blocks.

histogram of the Census Transform (*CT*) values (Zabih and Woodfill, 1994). The CT value, which is akin to a Local Binary Pattern (LBP (Ojala et al., 1996)), corresponds to a byte that encodes the comparisons between the current pixel and its eight neighbours. More specifically, a bit is set to 1 if the pixel is higher than (or equal to) the central pixel, else it is set to 0. Neighbours are searched from left to right and from top to bottom:

$$\begin{array}{|c|c|c|}\hline 156 & 168 & 172 \\ \hline 156 & 156 & 157 \\ \hline 153 & 150 & 155 \\ \hline \end{array} \Rightarrow (10010111)_2 \Rightarrow CT = 151 \quad (1)$$

Spatial information is incorporated into the features by using a pyramid. In this scheme, multi-scale representations of the image are obtained by dividing it into several blocks, from which the Centrist descriptors are extracted. We use a 2-level pyramid with 31 blocks, as proposed in (Wu and Rehg, 2011). Note that the blocks are built with identical dimen-

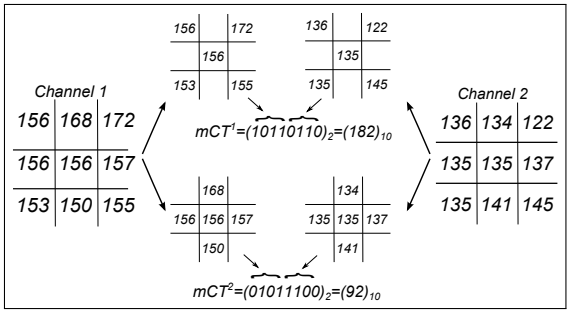


Figure 4: mCentrist computation by combining 2 channels.

sions and with an overlap of 25% between successive scales (see Fig 3).

Color information, that is ignored in the Centrist descriptor, can be dealt with using mCentrist, a multi-channel extension of Centrist, proposed in (Xiao et al., 2014), in which channels are combined pairwise as shown on Fig. 4. In our contribution, we consider a 4-channel color space (three channels in opponent color space and a Sobel contour image), as in (Xiao et al., 2014).

In Sect.5, we experimentally evaluate the influence of color and spatial information.

3.1.2 High-level Approach

In our bag-of-words approach, *HOG* (Dalal and Triggs, 2005) are first extracted from rectangular regions of interest (*ROI*) placed on a regular grid on the image. The number of bins n_{bins} of the HOG is usually a sub-multiple of 180°. The optimal value we found in our experiments is given in Sect. 5. It may be noticed that, since we use HOG features, color information is ignored at this step. The k-means algorithm is then used to generate a codebook by clustering the features into visual words. The random forest algorithm, with the histogram of words as input features, is finally applied to classify the scene into a speed limit category. The size of the patch, $n_c \times n_c$ pixels, and the number of words n_{words} in the visual vocabulary are empirically determined and the results are shown in Sect 5.

3.2 Route Segmentation

We are interested in segmenting routes into homogeneous road sections according to a speed limit criterion. The classification should be more robust and relevant by taking into account the image sequence than by using images individually. Considering the posterior probabilities over the sequence as a 1-D signal, we propose to smooth the outputs of the classifier by applying successively a mean filter and a morphological filter (Serra and Vincent, 1992) over the signal. The two filters aim at eliminating the small segments in the sequence. The number N_{filter} of neighboring images centered on the current image varies over the range $[0, 100]$. We remind the reader that 100 images correspond to 500 meters. The mean filter is first applied to the probabilities stemming from image classification. The four mean signals are then filtered by morphological closing and morphological opening operations and normalized. The structuring element of the morphological filter is a vector of dimension $N_{filter} \times 1$. Finally, the image is assigned to the class of maximum filtered probability.

4 EXPERIMENTAL SETUP

We consider real-world image sequences acquired by frontal cameras mounted on top of inspection vehicles. Images are taken every 5 meters during daytime under various, uncontrolled lighting conditions. The image size is 1920×1080. We will consider two kinds of data sets : *individual images data set* and *sequence data set*. Every image of the data sets is manually given a speed limit label in order to establish a ground truth (*GT*).

4.1 Individual Images Data Set

This data set is composed of 640 individual images (*i.e.* no sequential information is considered) equally split into the four categories. This data set is used for the training phase, for the determination of low-level features and for the empirical optimization of the BOW parameters n_{bins}, n_c and n_{words} in the high-level approach.

4.2 Sequence Data Set

The *sequence data set* comprises 11689 images with homogeneous sections of the different speed limit categories. Three different human operators have built three GT's (GT1, GT2 and GT3) according to their visual perception of the scenes, as illustrated in Fig. 5. Note that the number of images by category is highly imbalanced, with a majority (68%) of images in the 90 category. We quantify that 14.04% (respectively 11.16% and 9.06%) of the images have been categorized differently in GT1 and GT2 (respectively GT1-GT3 and GT2-GT3). The differences between ground truths are mainly located at the transitions between categories. However, we observe other ambiguities, often corresponding to sections that are categorized

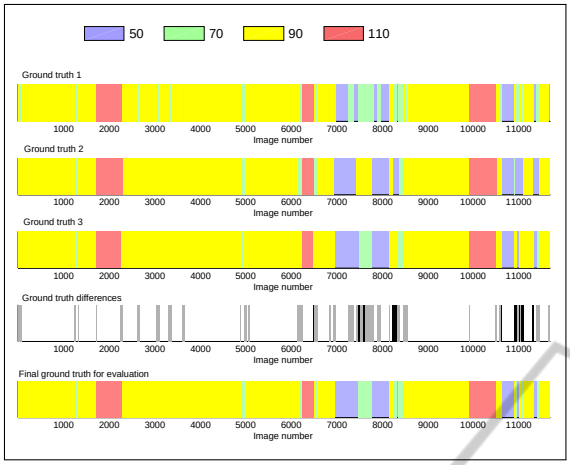


Figure 5: Ground truths for the sequence data set. The abscissa represents the image number. Colors correspond to labels (50 ; 70 ; 90 ; 110). The rows 1-3 are the 3 GT. The fourth row shows the ambiguities between GT's: white means no difference between GTs ; gray means that a GT differs from the two others; black means that the 3 GT's are different. The last row corresponds to GT_{eval}.

as 70 by one (or two) operator(s). To evaluate the performances of the algorithms, we build a single ground truth (GT_{eval}) by combining GT1, GT2 and GT3. The majority category in GT1, GT2 an GT3 is allocated to GT_{eval}. When we observe three different labels (1.82% of the images), the median label is assigned to GT_{eval}.

5 RESULTS AND DISCUSSION

In this section, the *individual images data set* is used to compare the performances with Centrist and mCentrist descriptors (§ 5.1) and to determine the best values n_{bins}, n_c and n_{words} for the BOW approach (§ 5.2). Note that the classification rate is calculated by cross-validation. The image classification algorithms are then tested on the *sequence data set* and the impact of sequential smoothing is analyzed (§ 5.4). Throughout the current section, the number of trees in the random forest is 200 for the low-level approach and 300 for the high-level approach.

5.1 Considering Centrist, mCentrist and Spatial Pyramid

In this experiment, we evaluate the classification on the *individual images data set* by using Centrist and mCentrist with or without spatial information in the low-level approach. The results are gathered in tab. 1.

Table 1: Scene classification into four speed limit categories by low-level approach. Results (true positive rate in %) on image data set by using Centrist, mCentrist and mCentrist + Spatial Pyramid (SP).

features	50	70	90	110	Overall
Centrist.	90.0	78.1	67.5	84.4	79.6
mCentrist	90.6	78.8	74.4	86.9	82.7
mCentrist+SP	95.0	79.4	83.8	96.2	88.7

Table 2: Confusion matrix (in %) for the low-level approach.

		Algorithm results			
		50	70	90	110
GT	50	95	5	0	0
	70	10	79.4	10	0
	90	1.9	9.4	83.8	5
	110	0	1.9	1.9	96.2

We distinguish the true positive rate (*TPR*) by category from the overall score, which is the mean of TPR for the four categories. By analyzing the results, it may be noticed that the use of color information has a strong impact for the classification into the category 90 (the TPR increases by 10.2% with mCentrist). The improvements of performances with mCentrist are less significant for the other categories. The TPR increases respectively by 0.7%, 0.9% and 2.9% for the categories 50, 70 and 110. It may be explained by the nature of the images in the category 90 which mainly contains rural images with typical color objects (vegetation, forests). The use of the spatial information has a great influence on the classification scores for all categories, except category 70 (for which the improvement is low). The TPR using mCentrist and SP are much higher than the TPR using mCentrist on the whole image. Hence, we retain the mCentrist descriptors computed on the blocks of spatial pyramid in the rest of our experiments. The confusion matrix of the low-level approach using mCentrist and SP is given in Tab 2. Note that the TPR of the categories 50 and 110 are much higher than the TPR of the intermediate categories 70 and 90. For these two intermediate categories, the misclassifications are due to confusions between the category 70 and 90. About 10% (resp. 9.4%) of images in the category 70 (resp. 90) in *GT* are classified in the category 90 (resp. 70).

5.2 Bag-of-Words Optimization

We focus on the determination of the values n_{bins}, n_c and n_{words} with a series of experiments using the *individual images data set*. The curves plotted on Fig. 6 show the overall TPR vs. the number of words (n_{words}) either for several values of n_c (top) or for three values of n_{bins} (bottom). In Fig. 6-top, we ob-

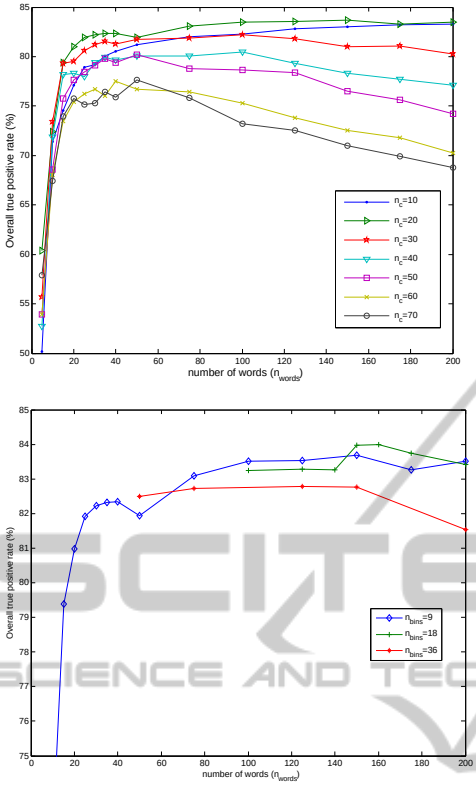


Figure 6: Overall TPR vs. the number of words in codebook (in abscissa) for several sizes of patches n_c (top) and for three for three values of n_{bins} in HOG (bottom).

serve that the TPR with $n_c = 20$ (and $n_{bins} = 9$) is systematically higher than the six other values whatever the number of words. We chose this value for the second experiment (Fig. 6-bottom) in which the number of bins is evaluated. We can see that the best performances are obtained with $n_{bins} = 18$ for $n_{words} = 160$. Hence, these values are chosen for the remaining of the paper.

Table 3: Confusion matrix (in %) for the BOW approach.

		Algorithm results			
		50	70	90	110
GT	50	90.6	8.1	0.6	0.6
	70	7.5	81.2	10	1.2
	90	1.2	5.6	88.8	4.4
	110	0.6	0.6	1.9	96.9

5.3 Comparison Between Low-level and High-level Approaches

The confusion matrix of BOW algorithm, applied to the individual images data set, is shown in Tab. 3. It may be noticed that, for the category 90 and 70, the high-level method (TPR = 88.8% and 81.2%) performs better than the low-level algorithm (TPR = 83.8% and 79.4%). On the other hand, the category 50 is better identified using mCentrist features than using bag-of-words model. In both cases, the lower performances are obtained for the category 70. In terms of computation time, the mCentrist extraction takes 4.2 seconds per image while the bag-of-words extraction is longer (10.1 seconds per image).

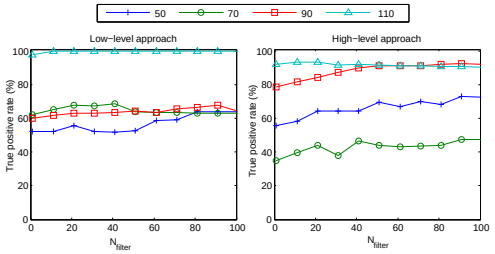


Figure 7: TPR vs. the number of images for the filters:N_{filter}.

5.4 Evaluation of Route Segmentation Algorithm

In this section, we evaluate the route segmentation method on the sequence data set. A smoothing, as described in Sect. 3, is applied on the results of the image classification. The results of route segmentation without filtering are shown on Fig. 8. By visually comparing the classification using either low-level approach (Fig 8-2^{nd}*row*) or high-level approach (Fig 8-4^{th}*row*) to the ground truth (Fig 8-1^{st}*row*), we note that the road sections are coarsely identified. However, we observe a lot of category changes between successive images. This confirms that it it necessary to smooth the individual classification results using neighboring results. The size of the filter N_{filter}, centered on the current image, can be empirically determined. The curves TPR by category vs. N_{filter}

Table 4: Confusion matrix (in %) for the route segmentation algorithm : (top) mCentrist approach ; (bottom) BOW approach. $N_{filter} = 21$.

		Algorithm results			
		50	70	90	110
GT	50	55.7	44.3	0	0
	70	5.1	67.6	27.1	0.1
	90	0.4	27.6	62.9	9
	110	0	0	0	100

		Algorithm results			
		50	70	90	110
GT	50	64.4	25.3	6.3	4
	70	4.3	44.1	51	0.1
	90	0.3	4	84.5	11.1
	110	0	0	6.5	93.5

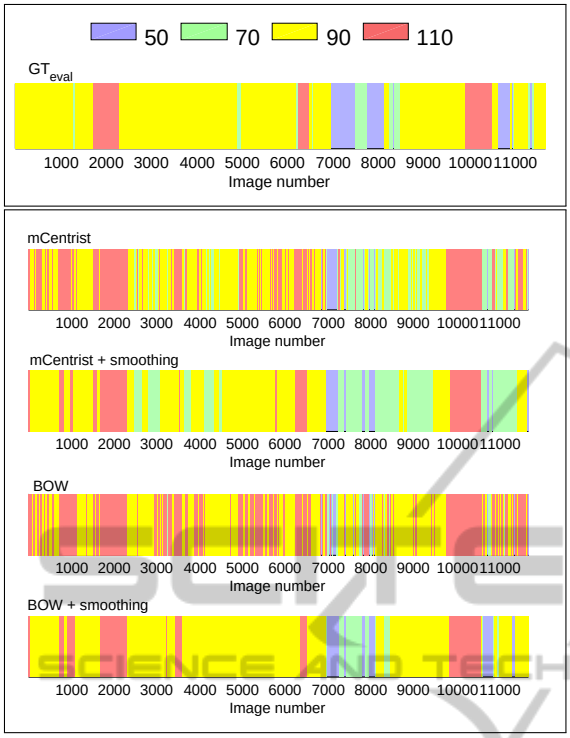
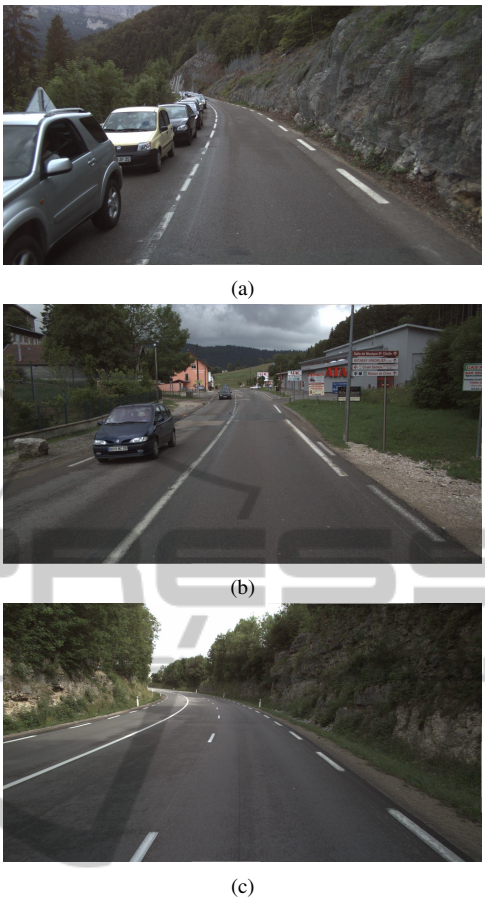


Figure 8: Route segmentation algorithm for the sequence data set. The abscissa represents the image number in the sequence. Colors correspond to the label (50 ; 70 ; 90 ; 110). The first row is the ground truth GT_{eval} as defined in Sect. 4. The next two rows are the segmentation using the low-level approach without (row 2) and with (row 3) smoothing. The last two rows are the segmentation using the high-level method without (row 4) and with smoothing (row 5). $N_{filter} = 21$.



(a) (b) (c)

Figure 9: Examples of false classifications : (a) image classified as 50 by the low-level method and as 90 by BOW, as in GT_{eval}; (b) ambiguous situation. The image is identified as 90 in GT; (c) image wrongly classified as 110.

are plotted on Fig 7. N_{filter} varies over the range $[0,100]$. The value 0 means that smoothing is disabled. A study of these curves shows that the improvement of performances with smoothing is significant for the two approaches whatever N_{filter}. The increase for the high-level approach is higher than for the mCentrist method. No value N_{filter} gives the optimum for all categories. We propose to use a rather low value $N_{filter} = 21$ to reduce the smoothing impact in the algorithm. Quantitative performances are gathered in Table 4. It may be noticed that the results for the *sequence data set* are lower than the TPR for the *individual images data set* (see tab. 2, 3). Both algorithms perform very well to classify images into the category 110. In the mCentrist approach, The TPR are quite homogenous for categories 50 (TPR=55.7%) , 70 (TPR=69.6%) and 90 (TPR=62.7%). We observe a lot of confusions between the three categories. The performances using BOW are quite better in the category 50 and much higher in the category 90 by comparing to the mCentrist approach. On the other hand, the TPR is very low in the category 70 (TPR=44.1%). It may be noticed that the categories 50, 70 and 90 represent variable situations (as shown in Fig 9-a and b, both images illustrate scene in category 90). The category 110 is more homogenous. This could explain that the performances are different for the category 110 with respect to the three others. Figure 8-(3^{rd} and 5^{th} row) shows the obtained route segmentation using filtering for both approaches. The road sections appear more homogenous and more similar to the ground truth. A careful examination of the results shows that wrong classifications are mainly located at the boundaries between sections. This can be observed on Fig. 8, about image numbers 2000, 7000, 8000 and 10000, where the width of the classified speed limit section is lower or higher than the corresponding one defined in GT_{eval}. A probable explanation is that the transitions are gradual and involve several images. However, other misclassified sections appear in Fig. 8. With the mCentrist approach, many

sections are wrongly identified, as category 70 (instead of 50 or 90), which confirms the quantitative results. This case is illustrated in Fig 9-b. Note that this situation is quite ambiguous, even for a human operator. With the BOW method, we observe a three-carriageway section that has been wrongly classified in the category 110 (90 in GT_{eval}) as shown in Fig 9-c.

6 CONCLUSION

In this paper, we addressed the problem of route segmentation into four speed limit categories using road scene analysis. We proposed a two-step algorithm that first classifies the images either by using a low-level approach or by using a high-level, semantic approach. In both cases, the second step is a sequential filtering to obtain a relevant route segmentation with homogenous road sections. The performances of the algorithm were evaluated on individual images and on a sequence data set. In the sequence data set, the true positive rate is satisfactory for the category 110. By using mCentrist, the TPR are homogeneous and vary over the range $[55.7, 69.6]$ for the categories 50, 70 and 90. In the BOW method, the performances are improved for category 50 and 90, but the TPR is low for the category 70. Wrong classifications correspond to situations that can be ambiguous, even for a human operator, e.g. transitions areas.

Future prospects include the use of robust algorithms, such as Markov chains, or semi-Markovian models in the sequential filtering. In a sequence, the number of images by category are highly imbalanced. This problematic shall be considered in the training phase. These improvements should be assessed on other sequence data sets to increase the true positive rates.

REFERENCES

Bosch, A., Munoz, X., and Marti, R. (2007). A review: Which is the best way to organize/classify images by content ? *Image and vision computing*, 25:778–791.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Charman, S., Grayson, G., Helman, S., Kennedy, J., de Smidt, O., Lawton, B., Nossek, G., Wiesauer, L., Furdos, A., Pelikan, V., Skladany, P., Pokorny, P., Matejka, M., and Tucka, P. (2010). Self-explaining road literature review and treatment information. Deliverable number 1, SPACE project.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 886–893, San Diego, USA.

Ess, A., Muller, T., Grabner, H., and Gool, L. V. (2009). Segmentation-based urban traffic scene understanding. In *Proc. British Machine Vision Conference 2009 (BMVC,2009)*, pages 84.1–84.11, London, UK.

Ivan, G. and Koren, C. (2014). Recognition of built-up and non-built-up areas from road scenes. In *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, Paris, France.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proc. International Conference on Computer vision(ICCV'99)*, volume 2, pages 1150–1157, Kerkyra, Greece.

Mioulet, L., Breckon, T., Mouton, A., Liang, H., and Morie, T. (2013). Gabor features for real-time road environment classification. In *Proc. IEEE International Conference on Industrial Technology (ICIT)*, pages 1117–1121, Cape Town, South Africa.

Ojala, T., Pietik ainen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59.

Serra, J. and Vincent, L. (1992). An overview of morphological filtering. *Circuits, Systems and Signal Processing*, 11(1):47–108.

Tang, I. and Breckon, T. (2010). Automatic road environment classification. *IEEE transactions on Intelligent Transportation systems*, 12(2):476–484.

Vailaya, A., Jain, A., and Zhang, H. J. (1998). On image classification : City images vs. landscapes. *Pattern Recognition*, 31(12):1921–1935.

Wu, J. and Rehg, J. (2011). CENTRIST: A visual descriptor for scene categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1489–1501.

Xiao, Y., Wu, J., and Yuan, Y. (2014). mCENTRIST: A multi-channel feature generation mechanism for scene categorization. *IEEE Transactions on Image Processing*, 23(2):823–836.

Zabih, R. and Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In *Proc. European Conference on Computer Vision (ECCV)*, pages 151–158, Stockholm, Sweden.