Thermal-Infrared based Drivable Region Detection

Jae Shin Yoon, Kibaek Park, Soonmin Hwang, Namil Kim,
Yukyung Choi, François Rameau and In so Kweon

***Abstract*— Drivable region detection is challenging since various types of road, occlusion or poor illumination condition have to be considered in a outdoor environment, particularly at night. In the past decade, Many efforts have been made to solve these problems, however, most of the already existing methods are designed for visible light cameras, which are inherently inefficient under low light conditions. In this paper, we present a drivable region detection algorithm designed for thermal-infrared cameras in order to overcome the aforementioned problems. The novelty of the proposed method lies in the utilization of on-line road initialization with a highly scene-adaptive sampling mask. Furthermore, our prior road information extraction is tailored to enforce temporal consistency among a series of images. In this paper, we also propose a large number of experiments in various scenarios (on-road, off-road and cluttered road). A total of about 6000 manually annotated images are made available in our website for the research community. Using this dataset, we compared our method against multiple state-of-the-art approaches including convolutional neural network (CNN) based methods to emphasize the robustness of our approach under challenging situations.**

I. Introduction

In recent years, Advanced Driver Assistance System (ADAS) has drawn both car manufacturers and researchers attention. Among various ADAS applications, drivable region detection has been considered as a fundamental step to solve various types of problem such as collision avoidance, path planning, and autonomous navigation. Different kinds of approaches have been developed to tackle these tasks using visible light cameras.

Earlier studies [2], [16], [21], [23] usually exploit the dominant orientation and the texture of scene to detect drivable road originating from a vanishing point. These approaches assume that the road is located in between the vanishing lines, however, the vanishing lines detection process is highly sensitive to noisy data and occlusions caused by objects on the road. Hoiem *et al.* [10], [11] propose to use a combination of features such as texture, line and pixel intensities in order to classify the scene segments, unfortunately, this method is not suitable to detect drivable regions.

Recently, many researches [6], [9], [20], [22], [25] presented multiple region-growing based approaches from initial samples in an unsupervised way. Dahlkamp *et al.* [6] proposed the use of the Gaussian Mixture Model (GMM) in an on-line learning manner to propagate the drivable region. In [20], Grow-Cut algorithm [24] is adopted to propagate the drivable region from road samples extracted by a semi-ellipse mask. These methods are effective for highly contrasted images, but tend to fail when the background and the foreground are not clearly distinguishable.

The authors are with the Electrical Engineering Department, Korea Advanced Institute of Technology (KAIST), 335 Gwahak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea. (e-mail:jsyoon,kbpark,smhwang,nikim,ykchoi,frameau@rcv.kaist.ac.kr, iskweon77@kaist.ac.kr)

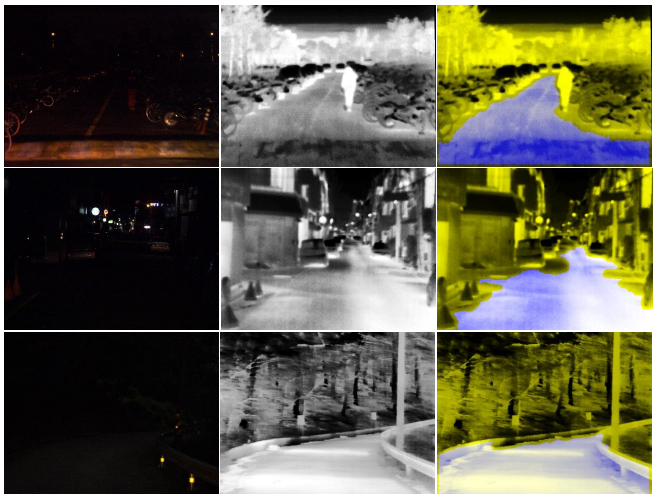


Fig. 1: The first and second columns are respectively the visible images and their thermal counterparts. Third column represents detection results from thermal images. Here, the blue color indicates the detected drivable region and yellow color represents the invalid drivable region.

More recently, many works have been based on the convolutional neural network (CNN). Brust *et al.* [3] proposed convolutional patch networks which are learned to classify different image patches so that their results can be used for pixel-wise labeling by simple thresholding. In [19], pixel-level semantic segmentation architecture was proposed, and their results can be used for drivable region detection as a 2-class segmentation problem. As our best knowledge, however, CNN based drivable region detection in thermal domain has never been tried because of the lack of training dataset corresponding to the various thermal condition.

Commonly, the methods developed for visible light cameras are particularly sensitive to low light conditions and specular reflections, especially in a outdoor environment. Nowadays, several works propose to utilize alternative kinds of sensors to overcome these limitations. One of these alternatives is the use of thermal-infrared cameras which are very robust to bad illumination conditions. Choi *et al.* [5] proposed a multi-spectral imaging system in which the visible and thermal images are aligned thanks to a specific optical setup. This system has been designed for autonomous driving purposes to ensure a sufficient visibility even during the night. Other types of applications take advantage of such

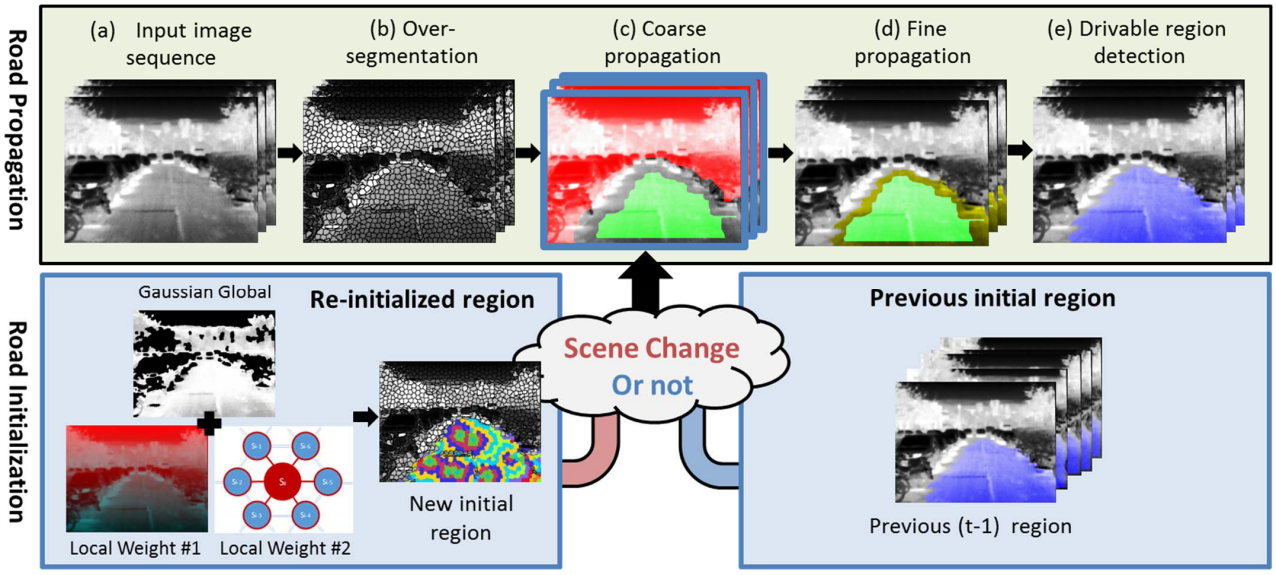


Fig. 2: Overview of the proposed algorithm. First row shows the road propagation step in video after obtaining initialized drivable region and second row shows the on-line road initialization procedure.

type of heterogenous sensor, for instance in [12] Hwang *et al.* presented a novel multi-spectral pedestrian detection technique and showed that the thermal image can be useful to improve the detection performances. While, Kolli *et al.* [15] and Cheng *et al.* [4] propose to use thermal images for accident prediction by recognizing the driver's emotion and activity.

In this paper, we propose a region-growing based road detection in the thermal domain to solve the problem originating from low light condition described in Fig. 1. More specifically, we focus our efforts on detecting safe drivable region for any type of road surface. To efficiently determine the drivable region, we first investigate initial sampling candidates from our scene-adaptive sampling method, and propagate this initialization using a label propagation approach. We further present a novel strategy to enforce the temporal consistency in a sequence of images with *coarse*-to-*fine* approach which gradually refines the optimal drivable region. We provide a large scale experiments to validate the robustness of the proposed method. To the best of our knowledge, this paper is the first to propose both, consistent experiments (with a considerably large number of images), and a freely available database (with ground-truth) for thermal camera road detection and navigation.

The rest of our paper is composed as follows: In Sec. II, we describe the overall system as well as the specific steps of the proposed algorithm. We then conduct comparative experiments in Sec.III. Finally, in Sec.IV, a short conclusion summarizes the contributions of this paper.

II. Proposed Method

In this paper, we propose a region-growing based road detection. Our approach consists in two stages: road initialization and road propagation for two successive images. In the road initialization step, we first roughly extract scene-adaptive road samples and find the optimal drivable area using a label propagation method. We then propose a strategy to efficiently estimate the drivable region in a video sequence by taking advantage of prior road information acquired over time. The overall system is shown in Fig. 2. The detail of each step is described in the following sections.

A. On-line road initialization

1) Road initial mask: In thermal domain, road has usually a homogeneous tendency (without sensor noise). We observed that this "texturelessness" can be useful to extract reliable road samples. Therefore, we utilize Gabor filter [14] to check the local homogeneity of the image. Thus, we extract all regions which are weakly responding to the Gabor wavelets [18] according to the Eq. (1)

$$M_i = \begin{cases} I_{i\in\mathfrak{G}}, & \text{if } \mathbb{T}_{\mathfrak{G}} < I_{i\in\mathfrak{G}} \\ 0, & \text{otherwise} \end{cases}, \qquad (1)$$

where $I_{i\in\mathfrak{G}}$ is the intensity of the i^{th} pixel response to the Gabor wavelets filter $\mathfrak{G}$ (this filter has eight orientations and a rather large radius), while $\mathbb{T}$ is an arbitrary threshold to separate the weak and strong homogeneous regions. We then distinguish the potential road candidate parts M_i as shown in Fig. 3-(b) (red area) by selecting the region close to the safe road point located at the middle bottom of the image. This initial step is very generic and can be applied on most of the road surfaces in order to compute a reasonable road initial mask.

2) Label propagation: The road sampling mask is not accurate enough to finely detect the drivable region. However, it can be used to initialize the precise drivable region detection. For this purpose, we first over-segment the scene with a conventional superpixel algorithm [1] to improve the

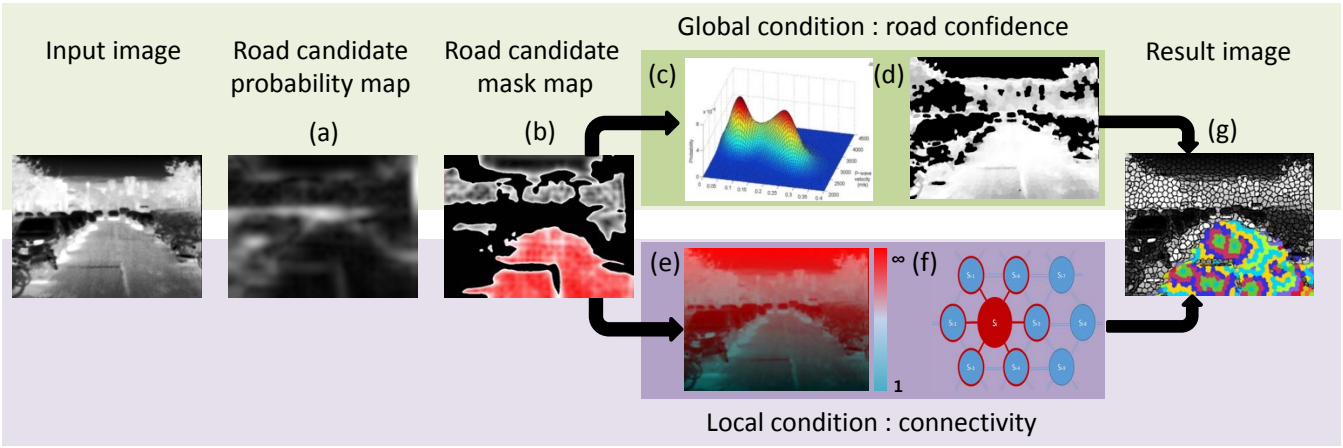


Fig. 3: Overall process of road initialization. (a) Original thermal-infrared image is firstly convoluted with Gabor wavelets. (b) Low response parts are extracted and initial road is completed by selecting the area that is closest to the safe road point (located at the middle bottom of the image). (c), (d) Gaussian mixture models ($k = 2$) are employed for a global condition with soft threshold. (e), (f) Two elements are considered to get local condition. First one is the ambiguity of thermal value according to the distance from camera, next one is dissimilarity between connected regions. (g) Finally, the label propagated image is obtained like this.

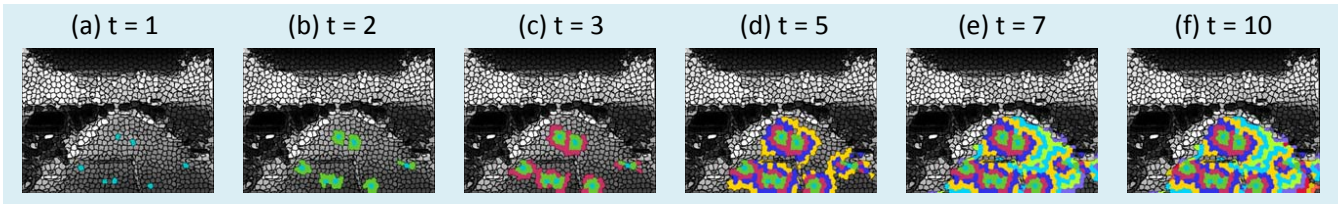


Fig. 4: Label propagation process with global and weighted local conditions. Color denotes the result of propagation for every step.

computational efficiency and robustness to the sensor noise with not the pixel-level processing but the cluster-level. Thus, we calculate the clustered mean of intensity of all the pixels within each superpixel region, this value is attributed to all the pixels in the superpixel.

To precisely segment the drivable region, the detected road has to satisfy both local (connectivity between superpixels) and global conditions. For the global condition, we build a Gaussian probability model with the sampled thermal intensities obtained from the initial mask. Since the dimension in the thermal domain is one (monochromatic image), we select the number of Gaussian model $k = 2$ to separate the drivable region and the background. We further run a Expectation and Maximization (EM) algorithm until convergence. Lastly, we complete the global condition map by processing the superpixels that contain a high probability value. Fig. 3-(d) is the global condition map, where the white area is satisfying global conditions, while the black parts does not. However, the Gaussian stochastic model cannot reliably separate the thermal image into drivable region and background as shown in Fig. 3-(d). In order to compensate this problem, we further formulate a weighted local condition considering the connectivity of neighborhood superpixels under the two following assumptions:

1) The mean of intensity differences with surrounding areas calculated from sampled labels is similar on the road.

2) The thermal ambiguity increases with respect to the distance of the camera to the surface. This phenomenon is caused by the attenuation of the heat over distance.

The weighted local condition LC_i is modeled by the Eq. (2):

$$LC_i = D_1 + \frac{(D_2 - D_1)}{M} \cdot (M - L_i) \ , \tag{2}$$

where M is the height of the image, L_i indicates the distance of the i^{th} superpixel S_i to the bottom of the image (as described in Fig. 3-(e)), D_2 is the ambiguity weight term from Eq. (4). D_1 is the average of the intensity differences between the reference S_i -within initial road mask- and its neighbors (red circle in Fig. 3-(f)). This relation is formulated in Eq. (3):

$$D_1 = \frac{1}{N_s} \sum_{i=1}^{N_s} \Big(\sum_{k=1}^{N_{variable}} \frac{1}{N_{variable}} |I_i - \hat{I}_{i,k}| \Big) \tag{3}$$

and

$$D_2 = \frac{1}{4} \frac{1}{N_s} \frac{1}{N_s - 1} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} |I_i - I_j| \ , \tag{4}$$

Algorithm 1 Random Start Label Selection in Initial Road

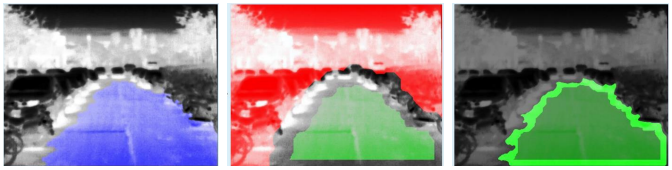
procedure RANDINITROAD
  visited label $L \leftarrow \emptyset$
  road label $R \leftarrow \emptyset$
  for i $\leftarrow$ 1 **to** Nr **do** ▷ for all random initial labels
    $NextLabel \leftarrow randi()$ mod Ns
      ▷ Ns the number of superpixels
  end for
  $L \leftarrow NextLabel$
  $R \leftarrow NextLabel$
end procedure

Algorithm 2 Label Propagation

procedure LABELPROP
  global condition $GC \leftarrow GetGlobalCondition()$
  local condition $LC \leftarrow GetLocalCondition()$
  while $NextLabel \notin \emptyset$ **do**
    $cnt \leftarrow 0$
    $stack \leftarrow 0$
    for all $i \in NextLabel$ **do**
      $val_r \leftarrow Superpixel(i)$
      $I \leftarrow SearchNeighborLabels(i)$
      for all $ind \in I$ **do**
        if $ind \notin L$ **then** ▷ Not visited yet
          $val_s \leftarrow Superpixel(ind)$
          if $val_s \in GC$ & $|val_r - val_s| \in LC$ **then**
            $cnt \leftarrow cnt + 1$ ▷ Satisfy conditions
            $stack(cnt) \leftarrow ind$
            $R \leftarrow R \cup \{ind\}$
          end if
          $L \leftarrow L \cup \{ind\}$
        end if
      end for
    end for
    $NextLabel \leftarrow stack$
  end while
end procedure

where $N_{variable}$ is the number of neighboring superpixels of S_i (which is variable from one superpixel to another) and $\hat{I}$ is the intensity of of each neighboring superpixels. While, N_s denotes the number of S_i within the initial road mask and I is the intensity of each S_i. Label propagation begins from randomly selected label inside the initial road mask as explained in algorithm 1. This initial stage is depicted in Fig. 4-(a). Then, road labels are transferred to the superpixels surrounding S_i if it satisfies the global and the weighted local conditions simultaneously. This process is iterated until there is no remaining label to be propagated. The details of this process are described in algorithm 2, this process is also illustrated in Fig. 4.

Due to sensor noise and irregularities stem from motion blur, the propagation results can contain some artifacts, like branches and holes. Therefore, we refine these artifacts with Morphological Image Analysis (MIA) [13] filters and



(a) Previous result (b) Coarse estimation (c) Fine propagation

Fig. 5: The process of *coarse*-to-*fine* drivable region detection in sequential frames. (a) blue: previous drivable region. (b) red: restriction of adjustable area (outside of dilation), green: coarsely estimated initial road (outside of erosion). (c) green: finely propagated drivable region from Grow-Cut.

weighted median filter [26]. The erosion filter is utilized first to remove the branch artifacts. We then restore the amount of eroded drivable region with the same size dilatation filter. Finally, we fill the holes and render the result by applying a weighted median filter with five iterations.

B. Road Propagation in Sequential frames

In this section, we present a strategy to efficiently deal with consecutive images. Unlike the previous road or drivable region detection methods which only consider a single image [2], [16], [21], [22], our algorithm is designed to handle a video data through a *coarse*-to-*fine* approach by incorporating temporal information. Our method continuously check the amount of scene variation to determine if the detection needs to be tracked or reinitialized. If the scene has changed too much, it automatically reinitializes the road mask using the single image detection described in Sec. II-A.2. Note that in this section the presented algorithm still relies on the use of superpixels and clustered mean.

1) Coarse-to-fine drivable region estimation: In order to detect the drivable region sequentially, our algorithm first utilizes the previously extracted drivable region to roughly initialize the road detection in the current frame from this assumption: the road properties between two consecutive frames are analogous. For this initialization step, thirty percent of the previous drivable area are removed using erosion filter (green in Fig. 5-(b)) and enlarged with dilation filter (outside of the red in Fig. 5-(b)). The remaining area between these two regions is the uncertainty region where the refinement step has to be operated. Finally, the Grow-Cut [24] algorithm is applied in the uncertainty region to precisely segment the current drivable region.

The Grow-Cut algorithm is a strategy which ensures an accurate segmentation from roughly labeled image. In order to utilize this method, three kinds of labels are assigned to three different areas, a label (+1) to the superpixel belonging to initial road detection (green in Fig. 5-(b)), a label (-1) to the background area (red in Fig. 5-(b)). While a label (0) is enforced in the uncertainty area (gray in Fig. 5-(b)). Restricting the size of uncertainty area in which Grow-cut algorithm is applied reduces the background influence as analyzed in Sec. III. It results in more accurate segmentation. The initial

superpixels labeled with ((+1), (-1)) l_a are propagated to the uncertainty area labeled with (0) l_d, if the following condition is respected:

$$(1-\frac{\parallel I_a - I_d \parallel_2}{max \parallel I \parallel_2}) \cdot \Delta_a > \Delta_d \ , \quad (5)$$

where I_a and I_d are mean value of the superpixels, while Δ_d and Δ_a are the strength of a superpixel respectively labeled l_a and l_d. These strengths are initialized to 1 for the superpixels l_a and to 0 to the superpixels l_d (the range of variation of the strengths is [0 1]). If l_d is propagated to l_a, its strength is updated in the following fashion:

$$\Delta_d = (1-\frac{\parallel I_a - I_d \parallel_2}{max \parallel I \parallel_2}) \cdot \Delta_a \ . \quad (6)$$

This process is reiterated until convergence as shown in the Fig. 4. For more specific explanation, please refer [24]. As a post-processing step, the propagated results are filtered by a weighted median filter [26].

2) Scene change: The appearance of the drivable area can be affected by multiple factors such as moving objects or speed bump. These occurrences can lead to a misdetection of the road. Hence, we propose a simple, but efficient approach to detect these failure cases in order to reinitialize the tracking. This failure case of detection is based on the use of normalized correlation indicator which can be expressed as follows:

$$\rho_{PH} = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(H_i - \bar{H})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(H_i - \bar{H})^2}}, \quad (7)$$

where H is the histogram of the pixel-intensity within the drivable region mask in the current image, P is the histogram computed ten frames before, and n is equal to 255 since the histogram range is limited to [0 255]. If this correlation indicator decrease rapidly, the system is initialized. With this reinitialization strategy, our algorithm is robust to any appearance change of the road, regardless of on-road or off-road situation.

III. Experiments

Dataset We performed experiments on three kinds of videos to show the effectiveness of our algorithm on thermal images. One of the video utilized is the *Campus* sequence from KAIST All-Day Benchmark [1]. To cover diverse scenarios, we also captured two new sequences entitled *Mountain* and *City*. These new sequences have been acquired using a FLIR A655sc thermal camera which have a better spatial resolution (640×480). Since thermal infrared wavelengths cannot penetrate the glass, our thermal camera is mounted on the roof of vehicle for obvious reason. The first sequence, *Campus*, contains 3546 frames mostly consisting of on-road scene with clear lane marking. *City* (1313 frames) and *Mountain* (1087 frames) sequences are representative of daily life scene with many moving objects and off-road scene respectively. We manually annotated the ground-truth of drivable regions for all sequences (5946 frames). The sequence databases and ground-truth can be found in our website.[2] The comparative evaluation results on each sequence are summarized in Table I with mean results. In addition, the qualitative results on a large sample of images are shown in Fig. 7 (these results are also available in our website).

TABLE I: Performance evaluation. Here, first row, second row, and third row are associated with ***FPR; False Positive Rate***, ***FNR; False Negative Rate***, and ***ErrorRate*** respectively. **Red**: Best, **Blue**: Second best.

	Campus	Mountain	City	Overall
Classifier	48.21	205.59	81.36	89.02
	3.07	0.54	1.24	2.18
	17.63	29.87	19.74	20.72
GMM	51.54	97.96	65.31	64.31
	9.98	5.69	8.72	8.80
	22.98	21.44	21.90	22.44
VP	34.36	274.59	44.60	89.28
	5.67	2.68	10.75	5.94
	15.13	43.96	16.64	21.77
GrowCut	52.78	167.47	82.77	83.59
	4.68	2.83	2.47	3.87
	19.88	26.11	21.36	21.53
AlexNet	42.78	221.24	67.89	80.95
	44.06	10.39	17.49	32.05
	42.10	41.11	29.15	39.07
CN24	46.94	93.39	55.84	57.40
	19.60	7.69	17.07	16.86
	36.4	21.10	28.06	36.77
FCN	12.01	13.42	5.79	10.89
	11.96	6.95	11.71	10.99
	10.29	6.34	9.77	**9.45**
Proposed	18.38	74.64	28.00	32.56
	7.03	6.44	8.18	7.11
	10.70	14.00	12.58	**11.77**

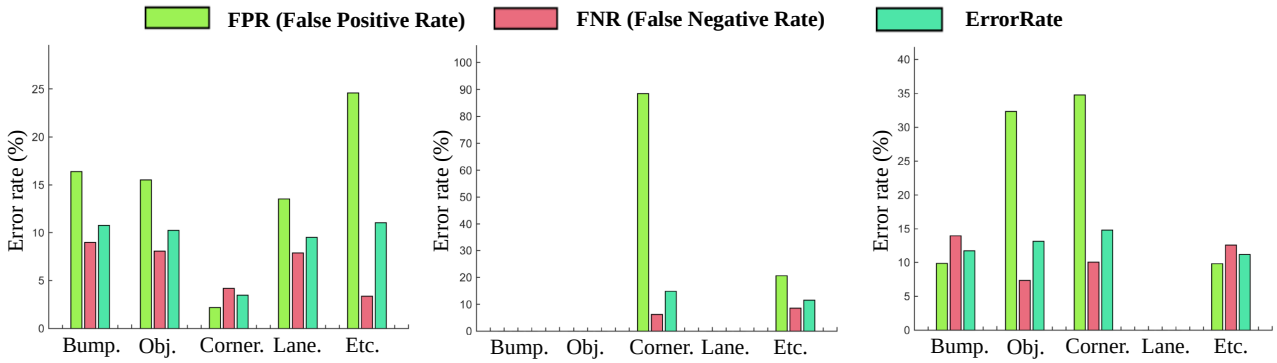
Evaluation criterion In order to evaluate the performance of our algorithm on each frame, we adopt three quantitative pixel-wise metrics [7], [8], *i.e.* *False Positive Rate* (FPR), *False Negative Rate* (FNR), and *ErrorRate* as following :

$$FPR(\%) = \frac{N_{FP}}{N_P} \times 100$$
$$FNR(\%) = \frac{N_{FN}}{N_N} \times 100$$
$$ErrorRate(\%) = \frac{(N_{FP}+N_{FN})}{(N_P+N_N)} \times 100$$

where N_{FP}, N_{FN}, N_P, N_N are the number of pixels which respectively correspond to wrongly detected drivable region, wrongly detected non-drivable region, ground-truth of drivable and non-drivable region.

[1] https://sites.google.com/site/alldaydataset/

[2] https://sites.google.com/site/drivableRegion/



(a) Campus – 3546 frames (b) Mountain – 1087 frames (c) City – 1313 frames

Fig. 6: Class accuracy with ErrorRate of the proposed method. Bump is when the camera-mounted vehicle pass through speed bump. Obj, Corner, and Lane are when the thermal camera is faced with moving objects, corner shaped road, and traffic lane respectively. Etc (miscellaneous situations) is when the state of camera does not belong to any case.

Baseline To verify the effectiveness of the proposed algorithm, we compared our method with several state-of-the-art [3]. One of the recent approaches is a growing-region based method which propagates binary road labels from few initial samples. Dahlkamp *et al.* introduced Gaussian Mixture Model (*GMM*) based method [6], while Lu *et al.* [20] propagated an initial semi-elliptic shape using the Grow-Cut algorithm (*GrowCut*). Another approach [16] takes advantage of geometric information such as vanishing point and lines (*VP*) recovered from texture information of the road. Hoiem *et al.* [11] proposed another approach utilizing comprehensive information (*i.e.*textures, pixel intensity, ground location and so on) to parse the road scene (*Classifier*). The ground part of their result is used to evaluate the drivable region detection performance.

For further comparison, we compare three kinds of state-of-the-art deep learning based approaches. We denote a first baseline as *AlexNet*, because we modified *AlexNet* [17] to handle the road patch classification. For the task, we convert fully-connected layers to the same dimensional convolutional layers to control the receptive field in 64×64 patches. With the architecture of *CN24* [3], we could train the 28×28 patch which contains the spatial location information. As last method, we trained the drivable region with the architecture designed by *FCN* [19]. Here, we adopt the 32 times upsampled prediction. To evaluate the performance, cross validation is processed. Namely, two sequences are trained, and remained sequence is validated with trained weight. Lastly, threshold which makes best performance is applied to *AlexNet* and *CN24* respectively to obtain the binary drivable region map.

A. Analysis of the proposed method

Overall the comparative results (Table I) show that our method outperforms other approaches except *FCN* [19], especially in the *Mountain* sequence. The proposed method can handle curved road scene thanks to our scene adaptive sampling mask (Sec. II-A). The proposed label propagation algorithm in Sec. II-A.2, and the strategy for successive images (presented in Sec. II-B) are very effective in the sequence *City* which contains many obstacles such as pedestrians and vehicles. Fig. 7 illustrates the stability of our method. For further analysis, we categorized each video sequence into five classes: Bump (speed bump), Obj (moving objects), Corner (right angle turn), Lane, Etc (miscellaneous situations) and analyzed the performance of our method for each class.

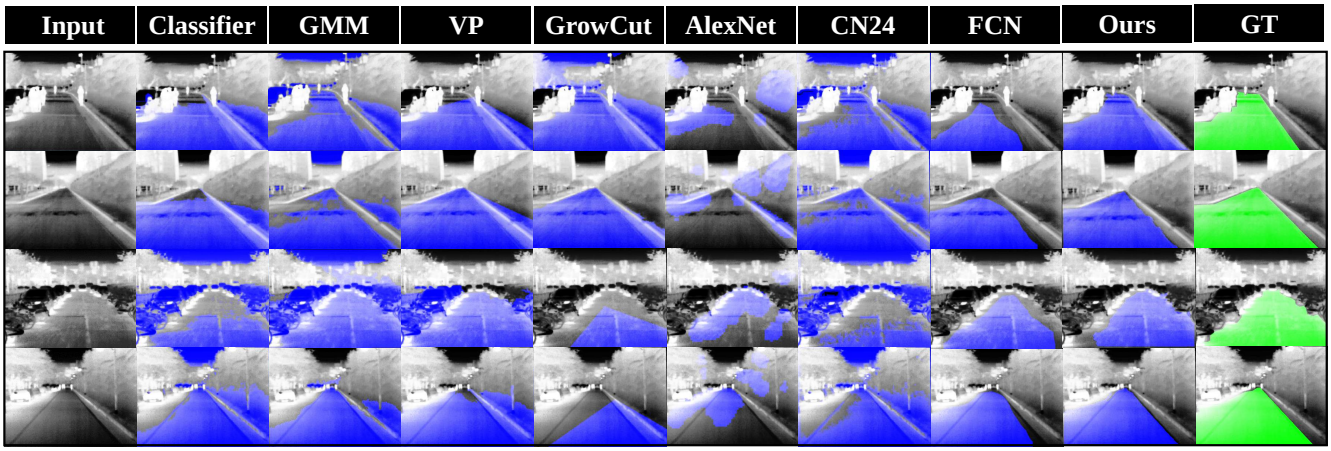
As shown in Fig. 6, moving objects and speed bump in each sequence are comparatively unfavorable to our algorithm. When the thermal camera mounted on the top of vehicle is facing moving objects, the shape of drivable region drastically changes between two consecutive frames. These rapid changes lead to a higher *ErrorRate*. However, even in these challenging situations, our algorithm never diverges thank to our reinitialization strategy. Meanwhile, corner in *Campus* sequence makes the drivable region well distinct from the background which is very favorable to detect drivable region (see Fig. 6-(a)).

B. Comparative evaluation

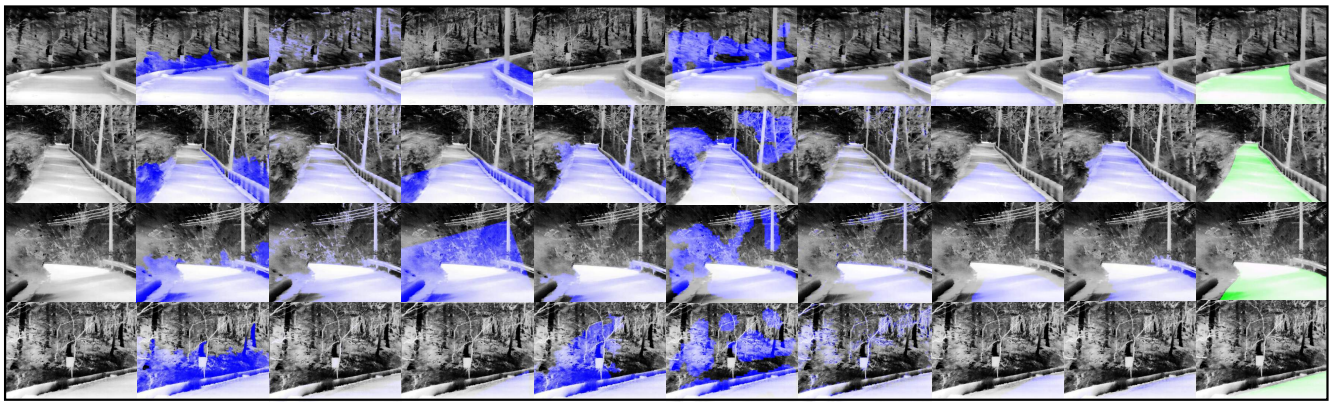
1) Comparison with hand-crafted feature based approach: Overall, the proposed method shows better *ErrorRate*. The *GMM* based method [6] searches background areas that have similar probability than the initial road samples. Therefore, it is prone to fail to detect clear drivable region even if exact initial road samples are extracted as shown in Fig. 7.

The general *GrowCut* method is only suitable for a limited range of images in the database when the road area is well distinctive from background. In other words, The Grow-Cut algorithm is very sensitive to cluttered background as shown in Fig. 7. On the other hand, our method shows quite stable *ErrorRate*, this can be explained by the use of sequential information which limit the segmentation to a smaller portion of the image, which drastically reduce the influence of the background. Furthermore, the sampling

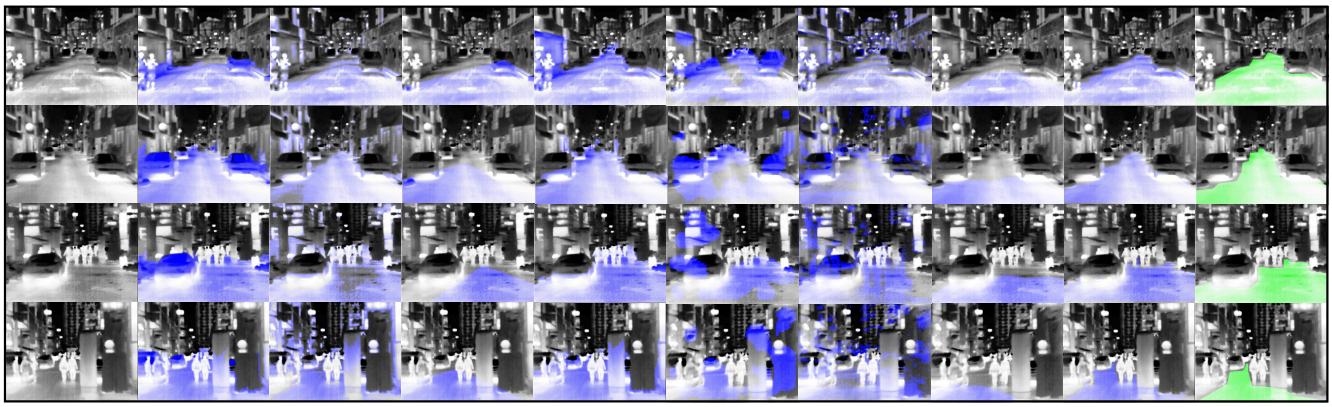
[3]Some state-of-the-art road detection algorithms are designed RGB images only. To perform experiments, thermal images are processed onto three channels for these RGB based algorithms.



(a) Campus sequence: 3546 frames



(b) Mountain seqeunce: 1087 frames



(c) City sequence: 1313 frames

Fig. 7: Qualitative result samples.

method in [20] is based on a very rough approximation using a simple semi-ellipse on the bottom part of the image. In practice this assumption is not always valid and this drawback is particularly problematic when the vehicle faces a curvy road, therefore, the performance variance of this algorithm is very high in the mountain sequence.

Although the vanishing point seems relatively easy to compute in the *Campus* sequence, the *VP* does not perform well due to the noise and motion blur in the thermal images. In the *City* sequence, many occlusions (objects on the road) complicate the computation of the vanishing point, it explains a relatively high *ErrorRate* in this sequence. The proposed method also shows better performances than the *VP* approach in the *Mountain* sequence which contains a lot of curved road frames.

The *Classifier* method shows comparable performances, although it is designed for scene understanding to separate three parts. However, as the *Classifier* just concerns approx-

imated ground parts in the scene, it is hard to estimate the exact drivable regions in non-trivial road scenes such as the *Mountain* sequence which contains only a few road pixels in some frames. Furthermore, it cannot handle moving objects as shown in Fig. 7.

As the proposed method is carefully designed for various situations, our approach outperforms many state-of-the-art techniques in terms of *ErrorRate*. Furthermore, in many robotics and autonomous vehicle applications, conservative choices might prefer to detect not just road area but the drivable region. For this reason, focusing on *FPR* rather than *FNR* is reasonable for the drivable region detection algorithms.

2) Comparison with deep learning based approach: As shown in Table I and Fig. 7, patch classification based approaches, *AlexNet* and *CN24*, do not suit for thermal-infrared based drivable region detection. This is attributed to the fact that thermal image is monochrome containing high similarity between background and drivable region. *CN24*, however, shows better results than *AlexNet* thanks to the spatial patch location prior.

The usage of combined local and global features in FCN helps to capture context within image, and it could boost the performance, especially in *Mountain* sequence. Therefore, *FCN* could show best performance. Our method is comparable to the method *FCN* in overall ErrorRate from Table I without supervised training. This is due to the robust spatial prior (such like global condition, local condition, and temporal reasoning) proposed in Sec. II. In terms of the applicability to various thermal environment in an unsupervised way, therefore, our method still effective for thermal infrared based drivable region detection.

IV. Conclusion

In this paper, we proposed an algorithm to robustly detect a drivable region with a thermal infrared camera. We present a novel scene-adaptive sampling technique and label propagation method. Also, our algorithm is suitable for video sequence by taking into account the temporal information between consecutive frames. We evaluated our method through a large number of experiments on three kinds of videos to demonstrate the robustness of our algorithm compared to other approaches. A detailed analysis proved that our algorithm is suitable for diverse shapes of road scene under various conditions.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2010-0028680).

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.

[2] J. M. Álvarez, A. M. López, T. Gevers, and F. Lumbreras. Combining priors, appearance, and context for road detection. *Transactions on Intelligent Transportation Systems (ITS)*, 15(3):1168–1178, 2014.

[3] C.-A. Brust, S. Sickert, M. Simon, E. Rodner, and J. Denzler. Convolutional patch networks with spatial prior for road detection and urban scene understanding. *arXiv preprint arXiv:1502.06344*, 2015.

[4] S. Y. Cheng, S. Park, and M. M. Trivedi. Multiperspective thermal ir and video arrays for 3d body tracking and driver activity analysis. In *Computer Vision and Pattern Recognition-Workshops (CVPRW)*, 2005.

[5] Y. Choi, N. Kim, K. Park, S. Hwang, J. S. Yoon, and I. S. Kweon. All-day visual place recognition: Benchmark dataset and baseline. In *Computer Vision and Pattern Recognition-Workshop (CVPRW)*, 2015.

[6] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski. Self-supervised monocular road detection in desert terrain. In *Robotics: science and systems (RSS)*, 2006.

[7] J. Fritsch, T. Kuhnl, and A. Geiger. A new performance measure and evaluation benchmark for road detection algorithms. In *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[8] C. Guo, S. Mita, and D. McAllester. Mrf-based road detection with unsupervised learning for autonomous driving in changing environments. In *Intelligent Vehicles Symposium (IV)*, 2010.

[9] C. Guo, S. Mita, and D. McAllester. Adaptive non-planar road detection and tracking in challenging environments using segmentation-based markov random field. In *International Conference on Robotics and Automation (ICRA)*, 2011.

[10] D. Hoiem, A. Efros, M. Hebert, et al. Geometric context from a single image. In *International Conference on Computer Vision (ICCV)*, 2005.

[11] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision (IJCV)*, 75(1):151–172, 2007.

[12] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[13] P. T. Jackway and M. Deriche. Scale-space properties of the multiscale morphological dilation-erosion. *Pattern Analysis and Machine Intelligence (PAMI)*, 18(1):38–51, 1996.

[14] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1233–1258, 1987.

[15] A. Kolli, A. Fasih, F. Al Machot, and K. Kyamakya. Non-intrusive car driver's emotion recognition using thermal camera. In *Nonlinear Dynamics and Synchronization (INDS) & 16th Int'l Symposium on Theoretical Electrical Engineering (ISTET)*, 2011.

[16] H. Kong, J.-Y. Audibert, and J. Ponce. Vanishing point detection for road detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2009.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[18] T. S. Lee. Image representation using 2d gabor wavelets. *Pattern Analysis and Machine Intelligence (PAMI)*, 18(10):959–971, 1996.

[19] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[20] K. Lu, J. Li, X. An, and H. He. A hierarchical approach for road detection. In *International Conference on Robotics and Automation (ICRA)*, 2014.

[21] O. Miksik. Rapid vanishing point estimation for general road detection. In *International Conference on Robotics and Automation (ICRA)*, 2012.

[22] P. Moghadam and W. S. Wijesoma. Online, self-supervised vision-based terrain classification in unstructured environments. In *International Conference on Systems, Man and Cybernetics (ICSMC)*, 2009.

[23] C. Rasmussen and T. Korah. On-vehicle and aerial texture analysis for vision-based desert road following. In *Computer Vision and Pattern Recognition-Workshops (CVPRW)*, 2005.

[24] V. Vezhnevets and V. Konouchine. Growcut: Interactive multi-label nd image segmentation by cellular automata. In *Graphicon*, 2005.

[25] Q. Zhang, Y. Liu, Y. Liao, and Y. Wang. Traversable region detection with a learning framework. In *International Conference on Robotics and Automation (ICRA)*, 2015.

[26] Q. Zhang, L. Xu, and J. Jia. 100+ times faster weighted median filter (wmf). In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2830–2837. IEEE, 2014.