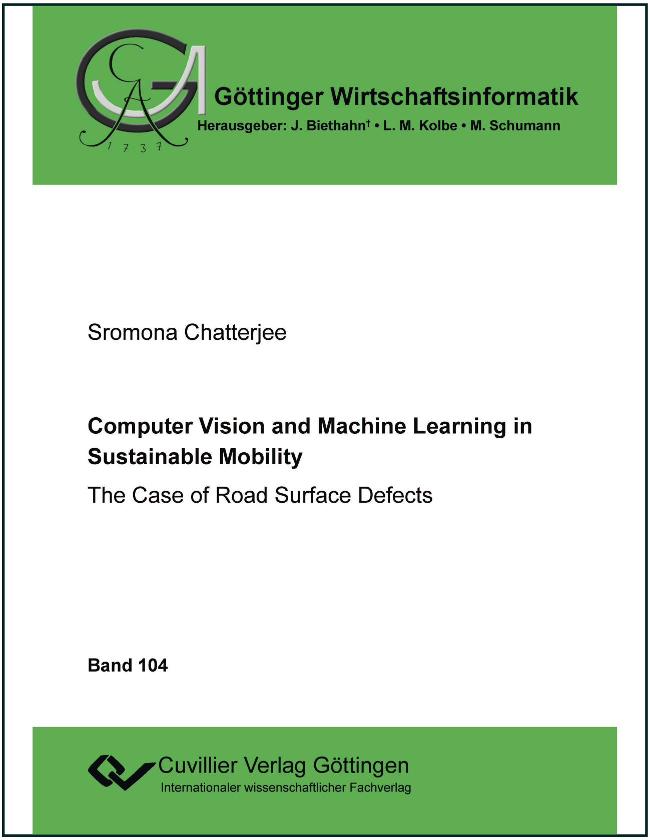
Sromona Chatterjee (Autor)

Computer Vision and Machine Learning in Sustainable Mobility: The Case of Road Surface Defects



Göttinger Wirtschaftsinformatik

Herausgeber: J. Biethahn† • L. M. Kolbe • M. Schumann

Sromona Chatterjee

Computer Vision and Machine Learning in Sustainable Mobility

The Case of Road Surface Defects

Band 104

Cuvillier Verlag Göttingen

Internationaler wissenschaftlicher Fachverlag

Chapter 1: Introduction

1.1 Motivation

Research in the area of pavement and infrastructure evaluation with automation techniques has been an active topic since the last several years (Chambon and Moliard 2011; Koch et al. 2015; Eisenbach et al. 2017). It constitutes of detection, evaluation and assessment of different pavement conditions such as, cracks, potholes, spalling, joint rupture, skid resistance, smoothness (BASt 2008; Miller et al. 1993; PMIS 2011; Zhang 2009). Pavement Management System (PMS) plays an important role in the maintenance of pavement networks and scheduling periodical maintenance to avoid incurring long term costs and road accidents (PMIS 2011; Hoeller 2012). This work uses advanced computer vision and machine learning (ML) techniques to analyze easily acquirable 2-D natural front-view scene images for crack related defect detection on roads, thereby helping in road surface evaluation.

In today's time, a huge increase in the urban population has created a manyfold demand on public infrastructures such as roads, tunnels, bridges. By 2050, urban population in the world will be around 66% (Chatterjee et al. 2018b; UN-DESA 2016). This creates a massive strain on infrastructures due to heavy traffic, movement of people, and the need for more connectivity. As an example, Koch and Brilakis (2014) noted that in Germany municipal roads cover 65% (400,000 km) of the entire road network, and their condition has deteriorated over the last years. In 2014 a comprehensive policy report released by the European Union noted significant backlog in road maintenance, requiring large investments for road rehabilitation and repairs (Gleave 2014). The report also pointed out that bad road conditions and soaring maintenance costs lead to a reduction in trade, an increase in accidents, high vehicle operating costs, an increase in CO_2 emission and pollution. Pollution could increase not only due to more fuel consumption because of bad roads, but also from a longer period of construction works. Consequently, delayed monitoring and maintenance give rise to a higher cost of rehabilitation. The impact of road maintenance expenditure can be seen in Figure 1.1. Therefore, it can be seen that timely and cost-effective monitoring of pavements is highly necessary to increase sustainability across all dimensions- social, economic, and environment (Chatterjee 2018b; Kahn 1995).

Traditionally, road monitoring and surface distress detection are predominantly done manually or by specialized vehicles fitted with costly sensors (Chatterjee et al. 2018a; Eisenbach et al. 2017; FHWA 2016; PMIS 2011; Radopoulou et al. 2016). Municipalities and certified

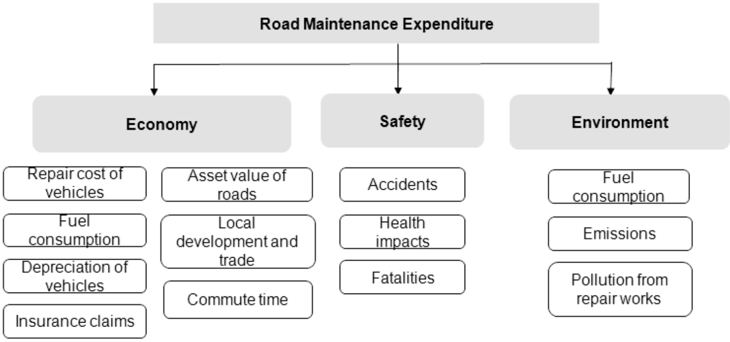


Figure 1.1: Impacts of road maintenance expenditure (taken and modified from Gleave (2014)).

service agencies inspect roads visually, or they take images to inspect them manually or in semi-automatic ways afterward. However, manual approaches are laborious and suffer from non-uniform, subjective, and delayed results (Chatterjee et al. 2018b). On the other hand, specialized vehicles are equipped with line scanner cameras or downward-facing cameras to take images of road surfaces, costly sensors, lasers, multiple cameras, artificial lighting, etc. These make such vehicles larger and unfit for narrower roads, lanes, or bike-paths, as well as highly expensive. The cost of such vehicles could be around $800,000 (Radopoulou et al. 2016). This makes them unaffordable and infrequently used by most municipalities. In some of the cases, aerial vehicles with imaging capabilities are also used (Dobson et al. 2014; MnDOT, 2018; Zakeri et al. 2016).

The ubiquitous nature of digital devices around us has given rise to unprecedented data traces (Galliers et al. 2015; Yoo 2010), thus enabling us to infer the world like never before (Yoo et al. 2012). Nowadays, Big Data enabled changes can be felt in all walks of our life catering to social interactions, economy, personalized services, to name a few. This has shown how causal relationships can be derived from data (Agarwal and Dhar 2014). This explosion of data, along with powerful analytics and computing facilities, has lately given rise to enhanced capabilities for management and decision-making (Agarwal and Dhar 2014; Kitchin 2014a; Rabari and Storper 2014).

Usage of everyday devices like smartphones or cars could thus be employed for road maintenance to offset the high cost of specialized vehicles and image acquisition. Lately, few efforts are seen to acquire front-view images using simple commodity cameras, like cameras mounted on cars, service vehicles, or smartphone cameras for road surface analysis (Chatterjee et al.

2018a; Mertz 2011; Tedesschi and Benedetto 2016; Varadharajan et al. 2014). However, to date most crack detection techniques for road surfaces use downward-view images acquired with special setups (Eisenback et al. 2017; Varadharajan et al. 2014). Figure 1.3 (A) shows a setup for front-view image data acquisition that is used in this work. Such kinds of acquisitions are much cheaper, easier, flexible, and faster. Data from different sources could be acquired and integrated. However, such capabilities are not yet fully harnessed (Chatterjee et al. 2018b). To get tangible real-world understanding from the massive data around us, we need to intelligently process the data for obtaining valuable insights (Goes 2014; Guenther et al. 2017). However, with more generalized data, data processing techniques to extract meaning from it become also highly challenging.

Regarding the algorithmic approaches that are portrayed in the literature, most crack and such defect detection approaches are highly image-processing based and use techniques related to thresholding, edge detection, segmentation, morphological operations, histograms, or statistical properties like, standard deviation (Koch et al. 2015; Mohan and Poobal 2017; Sinha and Fieguth 2006). A segmentation and fuzzy c-means clustering based approach is seen in Noh et al. (2017). Additionally, several image features in combination with Support Vector Machine (SVM) are seen also in many works (Eisenbach et al. 2017; Varadharajan et al. 2014). Lately, deep learning approaches are applied for pavement surfaces (Eisenbach et al. 2017; Gopalkrishnan 2018). In Zhang et al. (2016) a deep Convolutional Neural Network (CNN) is used for crack detection on asphalt roads. Pauly et al. (2017) also use CNN for crack detection using image patches. In Silvia and Licena (2018), a CNN, along with transfer learning using VGG-16 network, have been used for crack detection on concrete surfaces. Several techniques could also be found in defect detection for other infrastructures like sewer pipes (Su and Yang 2014) or quality inspection in steel parts (Soukup and Huber-Mörk 2014).

In general, the last years have seen the affirmation of either image-processing or of the use of ML requiring a lot of training data for pavement surface monitoring (Chatterjee et al. 2018b; Eisenbach et al. 2017; Koch et al. 2015; Li et al. 2014; Mohan and Poobal 2017). Moreover, some employ image-processing at the initial steps prior to the usage of ML techniques like, Artificial Neural Network (ANN), SVM (Li et al. 2014; Wu et al. 2016). As an example, in Wu et al. (2016) binarization and a novel morphology based operation are performed prior to the use of ANN for final crack detection. Further, most of the works use downward-view images. Lately, efforts are seen to use complete front-view scene images, similar to the type of images used in this work, for the purpose of crack and defect detection on road surfaces (Maeda et al. 2018; Radopoulou et al. 2016; Varadharajan et al. 2014). This work uses front-view scene images collected under normal daylight, as shown in Figure 1.3. Front-view images suffer from various challenges, such as lack of focused view of road surfaces, presence of external scene elements (e.g. building, vehicles, road signs, manholes), illumination scenarios and shadows, varied surface conditions and materials, lighter cracks, and different crack types (Chatterjee et al. 2018a; Varadharajan et al. 2014). Nevertheless, front-view images give more coverage area, flexibility, and scalability.

Intelligent Decision Support Systems (DSSs) that incorporate advanced data-driven approaches (Abella et al. 2017) and simple 2-D front-view scene images for automatically detecting and evaluating pavement surface defects (such as cracks) could enhance overall automation and effectiveness of a PMS (Chatterjee et al. 2018b). A combination of data analytics and cheaper measures to acquire data could help to increase the safety and timely maintenance of pavements. From these perspectives, the motivation of this dissertation is to develop computer vision and ML-based techniques to analyze 2-D front-view scene images for crack and related defect detection on pavement surfaces. This work handles different kinds of cracks such as single cracks, distributed multiple cracks, or network cracks. Cracks are one of the first defects to occur, and if not attended on time can lead to more severe defects (BASt 2008). This work further integrates the findings to propose a DSS for effective detection of road surface damages like cracks.

Thus, this research is motivated by both algorithmic development and practical aspects. Considering the algorithmic viewpoint, this work proposes two perspectives for crack and related defect detection on varied road surfaces: an ML-based classifier and a keypoint matching mechanism. Inspired by image matching applications, a keypoint matching approach has been developed in this work to detect cracks and defective pixels. Furthermore, it has been proposed in this work that Gamma mixture based clustering, fuzzified image feature descriptors, and texture features with ensemble ML approaches could better model the cracks. This stems from the observation that cracks are not easily distinguishable from the background and are often sparse. This work further provides an unsupervised learning approach using hierarchical clustering for road area detection and segmentation from the scene images. The images belong to various scene settings, such as rural and urban. Crack and defect detection algorithms are applied to the segmented road area. The developed crack detection algorithmic approaches in this work take into consideration the kind of image data available, challenges of such front-view image data, and how the data can be processed using classifier and keypoint matching techniques. The developed techniques do not depend on noisy image-processing based approaches and use ML, systematic image feature selection process, Gamma mixture model, and fuzzy-theoretic based analytical models to detect defective areas and pixels in images. The approaches have been developed in order to achieve more flexibility and adaptability.

On the other hand, from the viewpoint of practitioners and sustainability in infrastructure management, this work proposes ML-based techniques for crack and defect detection using simple and easily acquirable front-view scene images. Thus, human interventions could be greatly reduced for effective road monitoring using the proposed automated approaches in this dissertation. Usage of easily acquirable scene images helps in making road monitoring cost-effective and less time consuming. This, in turn, would increase road safety and road lifecycle, as well as reduce the occurrence of accidents and pollution. Finally, as part of this work, a prototype of computer vision and ML-based intelligent DSS for road surface evaluation using front-view images has been developed. The DSS has been designed using Design Science Research (DSR) approach in a modular manner to integrate the findings of the developed algorithmic approaches and viewpoints of different stakeholders. Such a DSS could thus help to achieve more automated, inexpensive, flexible, scalable, and timely road surface inspection practices.

1.2 Research Questions

In this section, the research questions that are dealt with in this dissertation are detailed. A block diagram, as shown in Figure 1.2, is presented showing how the research questions are connected. As discussed in the above section, image-processing based techniques are highly noisy. Manual effort is required to annotate training images for classifiers, which can also be subjective and time consuming. It is also seen that various crack detection approaches suffer when applied to different locations for the training and testing phases (Gavilan et al. 2011). Therefore, it is important to know how to characterize these cracks in flexible ways to achieve scalable and adaptable road maintenance practices. At the same time, alternative approaches to image-processing, feature-based classification, or an immense amount of labeled training data are also required for effective pavement surface evaluation. Additionally, the design of an intelligent and suitable DSS for road surface monitoring with interpretable outputs and construction is also of high necessity.

In general, it has been observed that cracks have special properties that make them challenging to be detected. Firstly, cracks are not as easily distinguishable as other objects like buildings, vehicles. Lighter and distributed cracks also increase the difficulty. Secondly, the number of defective or crack pixels is always much lower than non-defective ones in images, pointing towards a heavy-tailed distribution of gradient values. Crack pixels are generally darker than the non-crack ones. Figure 1.3 (D) shows a distribution of pixel intensity values in gray-scale colorspace. Thirdly, cracks or such defective pixels could be detected at a patch level (i.e. region level or superpixel) or at a pixel level, as shown in Figure 1.3 (F - G). On top of these, front-view

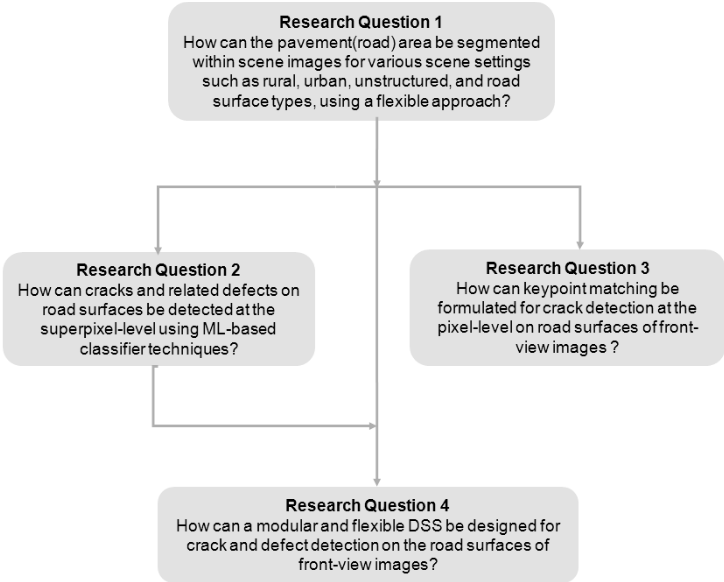


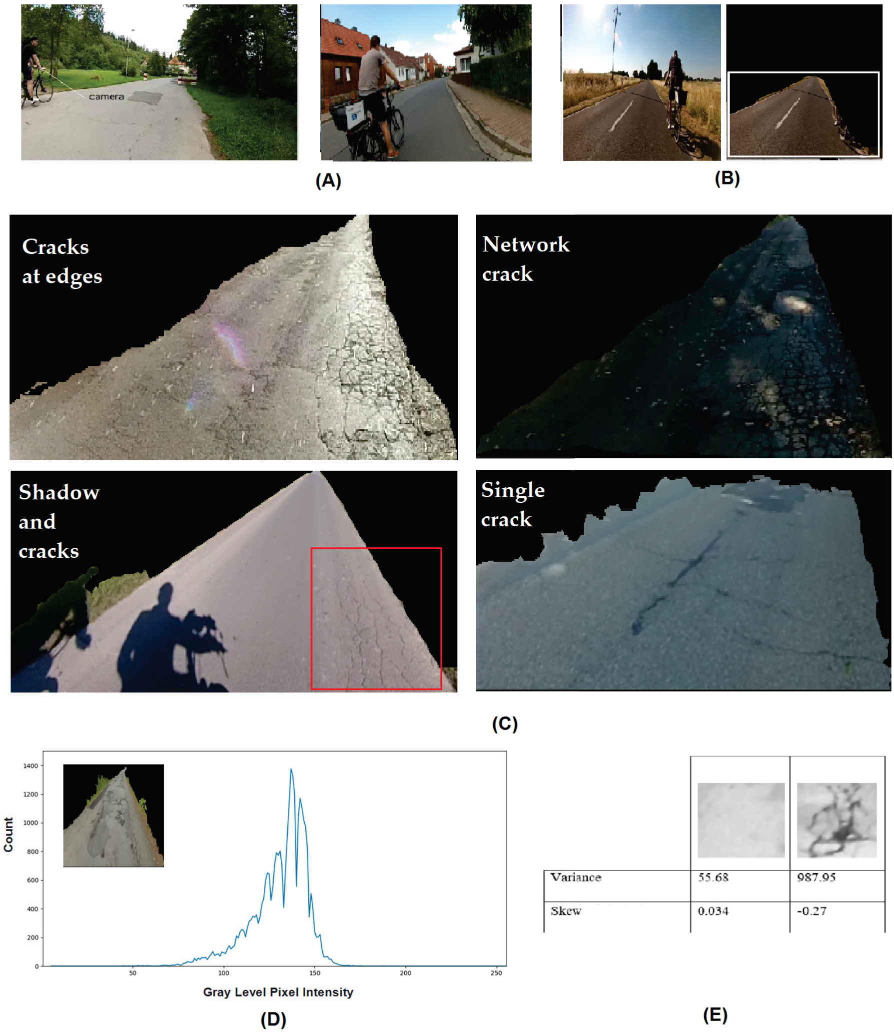
Figure 1.2: Relationship between different research questions.

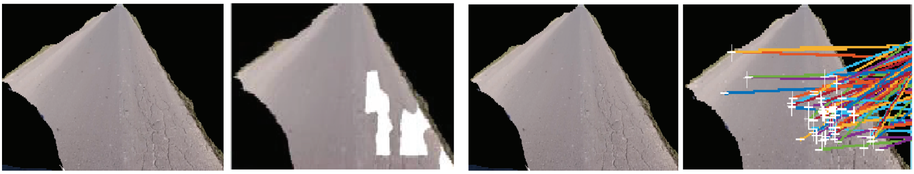
scene images taken under normal daylight also lack a closer view of road surfaces and may suffer from various illumination and surface textures, different crack and defect types, as in Figure 1.3. Such observations motivate the formulation of the following research questions.

1.2.1 Road Area Segmentation and Data Preparation

Figure 1.3 (A) shows how data is collected for this work. The 2-D scene images are collected under normal daylight from the driver's viewpoint. Simple high definition cameras are mounted on pedelecs/ e-bikes for the data collection. Thus, in order to prepare the data for crack detection on road surfaces, the road surface must firstly be detected in the scene images.

Different road segmentation algorithms are found in the literature based on techniques related to seed points (Kong et al. 2010; Zhou and Iagnemma 2010), deep learning and semantic segmentation (Alvarez et al. 2012; Brust et al. 2015) and camera based intrinsic features (Hoiem et al. 2005). Vanishing point estimation is another commonly used approach that is followed (Moghadam et al. 2012). However, vanishing point related techniques many times include lane


camera
(A)
(B)
Cracks at edges
Network crack
Shadow and cracks
Single crack
(C)
Count
1400
1200
1000
800
600
400
200
0
50
100
150
200
250
Gray Level Pixel Intensity
(D)
Variance 55.68 987.95
Skew 0.034 -0.27
(E)



(F) (G)

Figure 1.3: (A) data collection setup using a mounted camera on a pedelec. (B) Front-view scene image and the corresponding road surface image. Road surface images are obtained from scene images after applying the road segmentation approach. Bounding box shows the region-of-interest. (C) Example road surfaces with defects and crack-types that are handled in this work. (D) Characteristics of images containing defects: (left) Histogram of the gray-level intensity pixel values for an image containing cracks and defects. The intensity image in RGB colorspace is shown in the inset. (right) Statistical proprieties of defective and non-defective image patches. (F) Superpixel level crack detection. (G) Pixel level crack detection. Crack areas or pixels are marked in white.

markings or camera features; thereby making it less flexible or not applicable to rural roads where lane markings are not present. On the other hand, supervised learning based approaches mostly require a lot of manually labeled training data. Approaches related to 3-D segmentation is also seen in the publications (Hillel et al. 2014). In this work, the aim is to segment the road for 2-D scene images that include various scene settings (rural, urban, and unstructured) and road surface types (paved, unpaved); hence an adaptable approach is required that can handle different settings. So, the first research question that is answered in this work is the following:

> $ResearchQuestion1$: *How can the pavement (road) area be segmented within scene images for various scene settings such as rural, urban, unstructured, and road surface types, using a flexible approach?*

$ResearchQuestion1$ helps to remove the scene elements, such as buildings, vehicles, sky, people; thereby detecting only the road surface. This helps to handle front-view images more effectively for crack detection, as external scene elements would create more uncertainties if not removed.

Accordingly, an unsupervised learning approach based on hierarchical clustering has been developed in this work. Additionally, shadows (such as from buildings, trees) could be identified as part of the algorithm, so that they do not interfere at the later stages of crack detection. Here, external information such as camera characteristics, lane markings, road shape or boundaries are not considered. Furthermore, the number of clusters is also not required to be given. Hence, $ResearchQuestion1$ helps to come up with an adaptable and flexible approach of road area segmentation for various scene settings and road surface types. Once the road has been segmented, cracks and defects could be detected on the road surfaces.

1.2.2 Classifier Based Defect Detection on Road Surfaces at the Superpixel Level

To identify cracks and related defects on the road surfaces of front-view images using state-of-the-art ML classifiers, it has been observed in the literature that a systematic study is required to ascertain what are the most relevant image features for such tasks (Chatterjee et al. 2018a). Image features for understanding objects or regions in images could be based on shape, color, texture, histogram-based features, geometric properties, to name a few. As cracks are not always easily distinguishable due to various surface conditions, texture, illumination, as seen in Figure 1.3 (B - C, F), it is required to understand suitable feature types and frameworks. Furthermore, suitability of the kind of ML approaches, such as ensemble learning, Artificial Neural Network (ANN), for the task also needs to be investigated. This motivates the second research question as follows:

> $ResearchQuestion2$: *How can cracks and related defects on the road surfaces of front-view images be detected at the superpixel level using ML-based classifier techniques?*

The sub-research questions that are handled within $ResearchQuestion2$ are the following:

> $ResearchQuestion2.1$: *How state-of-the-art ML classifiers can be applied for crack and related defect detection?*
> $ResearchQuestion2.2$: *What are the most relevant image features for crack detection at the superpixel level, so as to make the cracks more distinguishable?*

Here, $ResearchQuestion2$ helps to answer the research goal of using superpixel and ML classifiers for crack detection on the road surfaces. Superpixel is a group of pixels and has unique characteristics (Achanta et al. 2012). For example, a superpixel containing cracks has more variance (variance could be calculated using the gray-level pixel intensities within the supeprixel) than the one without crack, as shown in Figure 1.3 (E). Crack pixels are generally more darker than non-crack ones, thus a considerable difference in intensity values are there. Hence, the research question aims to find the most relevant feature types suitable for the task at hand. Furthermore, the research question helps to ascertain suitability of different state-of-the-art ML classifiers and does not use noisy image-processing techniques for the task of crack detection at the superpixel level.

1.2.3 Keypoint Matching Based Defect Detection on Road Surfaces at the Pixel Level

Apart from recognizing cracks at the superpixel level, pixel level recognition is a step ahead. For recognizing different crack shapes, types (e.g. single cracks or network) and properties at later stages, it is highly useful to detect cracks and such defects at the pixel level.

Furthermore, supervised learning approaches require a lot of labeled training data, and can face challenges to handle lighter cracks and varied road texture types (e.g. when the surface contains different road texture types at different locations). Labeling could suffer from non-uniformity and human bias. Moreover, it is seen that the density of crack pixels in an image is much less than that of non-crack pixels. Figure 1.3 (D) shows the distribution of pixel intensity values (in gray-scale) of an image containing defects. It is also seen that the distribution of gradient values (as cracks can be inferred as edges) in such images is highly heavy-tailed and this characteristic can be used to build an alternative approach for crack detection. Thus, the third research question in this work is summarized as follows:

> $ResearchQuestion3$*: How can keypoint matching be formulated for crack detection at the pixel level on road surfaces of front-view images?*

The sub-research questions that are handled within $ResearchQuestion3$ are the following:

> $ResearchQuestion3.1$*: How can keypoints be generated in images for crack detection, considering density of defective crack pixels are much less in any image?*
> $ResearchQuestion3.2$*: How can keypoints be encoded with descriptors for crack detection, so as to make the descriptors more discriminative?*
> $ResearchQuestion3.3$*: What is the matching criteria to ascertain matched keypoints across images for crack detection?*

Here, noisy techniques like image-processing are not used for crack and defect detection at the pixel level, unlike most followed methods. This research question aims to use an analytical model consisting of fuzzy image feature descriptors and a keypoint matching approach for the first time in order to detect cracks on road surfaces. Keypoints in images are interesting points or regions, such as edges or corners, and local image features could be defined at such keypoints (Lowe 2014). In this way, objects or regions in images could be represented as a collection of keypoints. This approach is seen in image matching problems, where objects are matched across images based on the properties of the detected keypoints (Kumar et al. 2016, 2019). So, keypoints encoded with descriptors can be matched between two images to see how similar they are.