Improving Pothole Recognition through Vision Tracking for Automated Pavement Assessment

C. Koch\textsuperscript{1}, I. Brilakis\textsuperscript{2}
\textsuperscript{1}Ruhr-Universität Bochum, Germany; \textsuperscript{2}Georgia Institute of Technology, USA
koch@inf.bi.rub.de

Abstract. Pavement condition assessment is essential when developing road network maintenance programs. In practice, pavement sensing is to a large extent automated when regarding highway networks. Municipal roads, however, are predominantly surveyed manually due to the limited amount of expensive inspection vehicles. As part of a research project that proposes an omnipresent passenger vehicle network for comprehensive and cheap condition surveying of municipal road networks this paper deals with pothole recognition. Existing methods either rely on expensive and high-maintenance range sensors, or make use of acceleration data, which can only provide preliminary and rough condition surveys. In our previous work we created a pothole detection method for pavement images. In this paper we present an improved recognition method for pavement videos that incrementally updates the texture signature for intact pavement regions and uses vision tracking to track detected potholes. The method is tested and results demonstrate its reasonable efficiency.

1 Introduction

Existing road networks range in age, condition and performance. Several maintenance programs have been established in order to monitor the ongoing road performance, to predict future pavement conditions and assess long term need, to support investment planning and decision making, to assess the road safety and to identify rehabilitation and maintenance treatment. As part of these programs, pavement surface condition assessment is identified as a key component that includes reliable measurements on distresses like cracks, potholes, rutting, etc. (Vos et al. 2008).

In current practice, visual pavement data is automatically collected by sophisticated digital inspection vehicles, and then reviewed by technicians in order to manually detect and assess surface distress based on manuals (e.g. FHWA 2003, ZTV ZEB-StB 2006). Due to the high costs the number of inspection vehicles is very limited, which results in survey cycles of one year (MnDOT 2009) to four years (BASt 2008). In addition, the automated data collection is performed for highways only. Municipal road networks, which represent a very large part of a country’s road network, are predominantly surveyed manually. In Germany, for example, municipal roads cover 65 % (400,000 km) of the entire road network, and their condition has rapidly decreased over the last years (Steinauer and Kemper 2005). In order to ensure long-term optimal road conditions at a minimal economic investment, comprehensive and cheap condition surveying and assessment of municipal road networks is essential (Steinauer and Kemper 2005). Facing the amount of roads each agency (state, city, municipality, etc.) has to maintain, the solution of using dedicated vehicles to monitor all pavement assets on a regular basis is not feasible.

The overall objective of this research is to test whether an omnipresent, distributed and mobile network of passenger vehicles equipped with a high-speed wide-angle backup camera can sense and comprehensively assess pavement conditions over large areas at high fidelity. Distress findings from the network are transmitted to a central server and visualized online for
the public and all interested agencies. As one part of distress detection and assessment this paper focuses on pothole recognition in pavement videos.

In general, manually reviewing and assessing visual pavement data is a time-consuming and expensive procedure and, in addition, the results are highly influenced by the subjectivity and experience of the human raters (Bianchini et al. 2010). To improve the accuracy and reliability of surveys, a lot of research aims to automate the detection and assessment of pavement surface condition. Concerning potholes, available detection methods are either based on 3D surface reconstruction employing range sensors or stereo vision, or on vibration data using accelerometers. While the former generally suffer from high equipment costs and high computational effort, the latter lack accuracy and reliability.

Complementing our previous method (Koch and Brilakis 2011), this paper presents an enhanced pothole recognition method that takes advantage of a sequence of images from pavement videos. First, texture signatures for non-distress pavement regions are incrementally updated using a number of preceding frames instead of a single frame. Thereby, the global pavement surface appearance is taken into account resulting in more robust detection. Second, detected pothole regions are tracked in subsequent frames utilizing vision tracking algorithms. This enables convenient pothole counting in a pavement videos avoiding inefficient re-detection and matching.

The proposed method is implemented and tested using pavement videos captured by a remote-controlled robot vehicle. In order to evaluate the method’s performance, common metrics are measured comparing test results with manually identified ground truth. The resulting performance indicates that potholes can be recognized in pavement videos with reasonable efficiency.

2 Background

2.1 3D reconstruction based pothole detection

Current research efforts in automating the detection of potholes can be divided into 3D reconstruction-based and vibration-based methods. 3D surface reconstruction methods rely on 3D point clouds provided by range sensors or by stereo-vision algorithms using a pair of video cameras. Li et al. (2010) have presented a 3D laser scanning system for pavement rutting and pothole detection. Although laser scanning systems collect data in real-time, they suffer from high initial costs, and the need for significant power and frequent maintenance, which makes this technology unfeasible for a passenger vehicle network. A stereo-vision based surface model for comprehensive pavement conditioning has been proposed by Hou et al. (2007). With the availability of a 3D point cloud, Chang et al. (2005) have presented a clustering approach to quantify the severity and coverage of potholes. Jiaqiu et al. (2009) have created a method to identify and measure sag deformations (potholes and depression). The drawbacks of stereo-vision based approaches are that they require a complete 3D reconstruction of the pavement surface resulting in a high computational effort due to the very repetitive pavement texture.

2.2 Vibration based pothole detection

Besides vision and laser scanning based methods, a vibration-based system for the preliminary pavement condition survey has been proposed (Yu and Yu 2006). In analogy to cameras “seeing” a pavement surface, a vibration-based sensor (e.g. accelerometer) “feels”
the ground conditions based on the vehicle’s mechanical responses. Yu and Yu (2006) have identified the advantages of a vibration-based system as requiring small storage, being cost-effective and amenable for real-time processing. The projects BusNet (De Zoysa et al. 2007) and Pothole Patrol (Eriksson et al. 2008) share the same basic idea of combining vibration sensors with GPS and using mobile nodes to sense road conditions in terms of pothole detection and positioning. Rode et al. (2009) have proposed an integrated pothole detection and warning system that distributes collected defect data to the participating vehicles assisting drivers in avoiding potholes. However, Erikson et al. (2008) found that vibration-based approaches could provide wrong results, e.g. in case of bridge expansion joints that have been detected as potholes by mistake. Thus, complementing vibration systems with visual sensors could help to create more accurate and more reliable recognition models for various types of distress.

2.3 Previous work in visual pothole detection

In order to overcome the aforementioned limitations, the authors created a vision-based method for automated pothole detection in asphalt pavement images (Koch and Brilakis 2011). For each pothole, it was found that (1) it includes one or more shadows that are darker than the surrounding region, (2) its shape is approximately elliptical due to a perspective view, and (3) its surface texture is much coarser and grainer than the texture of the surrounding intact pavement. Based on these three distinctive visual characteristics a detection method was proposed (Fig. 1). Pavement images are first segmented into defect and non-defect regions using histogram shape-based thresholding. Then the geometric properties of a defect region are used to approximate the potential pothole shape through morphological thinning and elliptic regression. Subsequently, the texture inside a potential defect region is extracted and compared with the texture of non-defect pavement regions in order to determine if the region of interest represents an actual pothole.

Figure 1: Previous method and enhancements (highlighted)

Although the experimental results proved that our method works, it is still limited to single images and has to be applied subsequently to every single frame when processing pavement videos. This approach is computationally inefficient since the same pothole has to be re-detected and matched again and again, and with regard to automated assessment from videos, it is difficult to count individual potholes over a sequence of frames. Moreover, in the texture comparison procedure the representative texture for non-distress pavement is determined in the same image. In fact, the area of clean pavement is small and quite scattered in an image that contains a pothole due to associated cracks. Since texture is a 2D phenomenon, it is concluded that a small and scattered region is not the best representative of the intact pavement surface regarding a number of neighboring video frames.
3 Method enhancements

In order to detect and count potholes in pavement videos, the method presented in this paper enhances the previous image-based method in terms of two aspects (Fig. 1). On the one hand, texture signatures for non-distress pavement regions are incrementally updated, and on the other hand, detected pothole regions are tracked in subsequent frames.

3.1 Texture extraction and comparison

Based on our third observation, the texture of the pothole region is to be compared with the texture of an intact pavement surface. In contrast to our previous method, the texture of non-distress pavement is described as the average texture of \( n \) regions \( R_{o,1}, R_{o,2}, \ldots, R_{o,n} \) subsequently cropped from the center of preceding video frames, for which the upstream image segmentation method could not detect any severe distress, such as big cracks, patching, etc. (Fig. 2). To achieve that, four spot filters are applied on each region \( R_{o,i} \) separately, and the filter responses are used to set up the corresponding feature vectors \( f_{o,i} \in \mathbb{R}^5 \). The first component of each vector is the standard deviation (STD) of the region’s gray intensities, the other four components correspond to the STD of the respective filter response’s intensities (Koch and Brilakis 2011). The average feature vector \( f_o \) is then used to represent the texture of the intact pavement. While processing successive video frames, this texture signature \( f_o \) is incrementally updated and, if a pothole candidate is detected, compared to the feature vector \( f_i \) representing the inside pothole texture.

![Figure 2: Improved texture extraction and comparison procedure](image)

The advantage of this procedure over the previous method is twofold. First, computing filter responses involves the calculation of a two-dimensional convolution with a performance depending on both the kernel size (31x31, (Koch and Brilakis 2011)) and the number of image pixels. Instead of performing the convolution on a complete video frame (640x480 pixels), only small center regions (\( n \times 100 \times 100 \) pixels) and the pothole candidate region are used, which reduces computation time significantly. The second advantage lies in the capability of taking the global pavement appearance into account. Instead of using a scattered region of non-distress texture from a single frame, an averaged and more global texture appearance is modelled, which makes the texture comparison procedure more robust and less sensitive to abrupt surface appearance changes.

3.2 Pothole tracking

Since there is a huge amount of vision-based tracking methods available, an appropriated one has to be identified first. Yilmaz et al. 2006 distinguish three basic categories of visual object
tracking: point-based tracking, kernel-based tracking and silhouette-based tracking. In point-based tracking the object of interest is represented by multiple feature points that need to be detected and matched in consecutive frames. Since the pavement texture is characterized by a very repetitive pattern, identifying and matching feature points is quite difficult. In silhouette-based tracking the focus is on the object’s exact contour that is modelled and used for tracking. Under our method, the shape of detected potholes is approximated as an ellipse and, so far, does not represent the accurate pothole contour. Besides this and its sensitivity to noise silhouette-based tracking is not suitable for our purpose. Kernel-based tracking considers both the shape and the appearance (texture) of an object in terms of its whole region, is less sensitive to noise and can cope with object scale and illumination changes (Yilmaz et al. 2006, Makhmalbaf et al. 2010). For these reasons, we propose kernel-based tracking to track detected potholes in pavement videos.

Ross et al. (2008) have proposed a robust kernel-based tracking method that has been successfully tested and validated in outdoor environments where target objects undergo large appearance changes due to varying lighting conditions and object distances. Moreover, the object’s appearance is incrementally learned throughout the tracking process without prior training. These features adequately meet the requirements of pothole tracking. In pavement videos the appearance of a pothole changes in scale, might change in illumination and is difficult to train due to the huge variety of potholes. So far we haven’t compared different kernel-based methods since our tracking results are very satisfying (see section 4.2).

Once a pothole is detected in a video frame, the corresponding region (as a result of the shape extraction procedure, Fig. 1) is tracked in the subsequent frames until it leaves the viewport. To achieve that, the detection algorithm is suspended and the detected pothole region is marked as the region of interest as the bounding box of the approximated ellipse (Fig. 3b). This rectangular area is then committed to the visual tracking algorithm (Fig. 3b) that tries to trace this region within the subsequent frames (Fig. 3c-d). Once the region of interest leaves the viewport the tracking algorithm is stopped (Fig. 3d), the detection algorithm is resumed and the number of detected potholes is incremented (Fig. 3e). The tracking algorithm has to be performed on every detected pothole independently, and in case of multiple potholes, it is executed in parallel. However, so far we make the assumption that only one pothole appears in the viewport.

![Figure 3: Frame sequence in pothole tracking](image)

Utilizing vision tracking for pothole recognition in videos reduces the computational effort of the whole method, since expensive pothole re-detection and matching in several subsequent frames is avoided. This way, pothole tracking allows for convenient pothole counting in order to determine the magnitude of pothole distress in pavement videos.
4 Implementation and results

4.1 MATLAB prototype

In order to test its performance, the pothole recognition method presented in this paper has been implemented in MATLAB version 7.11.0 (R2010b) utilizing the embedded Image Processing Toolbox. Within the frame of our first experiments, it is decided to use three non-defect frames to determine the non-defect texture signature (Fig. 2). The implemented pothole tracking procedure is based on the kernel-based tracking method presented in (Ross et al. 2008). In order to adapt and test this method, several options need to be set appropriately. First, the expected object movement from one frame to next is to be specified in terms of translation, rotation, scaling and scaling direction. In accordance with our experimental setup (resolution, frame rate, vehicle speed, view angle etc.) we assume that a detected pothole moves about 3px in x, 20px in y direction and changes in scale about 2% (neither rotation nor skewing). Second, it needs to be specified how many frames the algorithm should remember when performing the incremental model update. Since a pothole has to be tracked over a few (10 to 30) frames only, all frames should be remembered, the so-called forgetting factor is set to 1. There are a few other options that were initialized with standard values described in (Ross et al. 2008).

4.2 Experiments and results

A database of 39 pavement videos was collected in order to test the effectiveness of the pothole recognition method. These video files were captured around the Georgia Tech campus using a remote-controlled robot vehicle that is equipped with a HP Elite Autofocus Webcam and a Viliv S5 Ultra Mobile PC (Fig. 4, left). According to our future vision, the camera is installed at an altitude of about 60 cm above the pavement surface and its view is directed down and backwards (45°). The video resolution is fixed at 640x480 at 30 frames per second. Figure 4 shows a selection of pothole recognition results column by column.

![Pavement sensing robot (left) and selected pothole recognition results (right)](image)

In order to validate the proposed method, two different metrics are defined for pothole tracking and overall recognition (detection & tracking). The tracking procedure is validated for correctly detected potholes. Since our objective is to count potholes, it is not important how accurate the position of a pothole region is tracked. It is rather essential that the tracker does not lose a pothole so that (1) the tracking procedure stops and the same pothole is re-detected and counted twice, or (2) the tracking procedure continues although a pothole has left the viewport. The latter might result in missing subsequent potholes since the detection
procedure is suspended. For these reasons, the number of frames in which the pothole regions is correctly tracked, that is, the pothole region intersects the tracking rectangle, is counted. The tracking performance is then measured as the percentage of correctly tracked frames over the number of frames in which the pothole is visible after detection.

To measure the overall recognition performance in videos, precision and recall are calculated based on the number of True Positive (TP, number of correct recognitions), False Positive (FP, number of incorrect recognitions), and False Negative (FN, number of actual potholes that are not recognized). For each video these values are counted and precision (TP / (TP+FP)) and recall (TP / (TP+FN)) are calculated. Table 1 presents an extract of the measured performances and the total recognition precision of 75% and recall of 84%.

Table 1: Extract of measured performances for pothole tracking and recognition

<table>
<thead>
<tr>
<th>Video</th>
<th># of Frames</th>
<th># of Potholes</th>
<th>Tracked Frames [%]</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP</td>
</tr>
<tr>
<td>#1</td>
<td>267</td>
<td>3</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>#2</td>
<td>604</td>
<td>3</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>#3</td>
<td>52</td>
<td>3</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>#4</td>
<td>140</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>#5</td>
<td>365</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>#6</td>
<td>127</td>
<td>1</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>#7</td>
<td>639</td>
<td>1</td>
<td>92</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>#37</td>
<td>315</td>
<td>3</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>#38</td>
<td>217</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>#39</td>
<td>182</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>10180</td>
<td>51</td>
<td>94</td>
<td>42</td>
</tr>
</tbody>
</table>

4 Conclusions

Pavement condition assessment is a key component when developing road network maintenance programs. In current practice, highway networks are sensed using a few dedicated and expensive inspection vehicles. However, municipal roads, which cover more than the half of a country’s road network, are predominantly surveyed manually. Facing the amount of roads, the solution of using dedicated vehicles to monitor all municipal pavement assets on a regular basis is not feasible. For this reason, we propose an omnipresent network of passenger vehicles to comprehensively sense and assess pavement conditions of large municipal road networks based. As one part of distress detection and assessment this paper focuses on pothole recognition in pavement videos. Existing research approaches either come along with high equipment and computation costs, or they can only facilitate preliminary and rough pavement condition surveys.

Complementing the previously proposed method, this paper presented an enhanced pothole recognition method. First, texture signatures for non-distress pavement regions are incrementally updated using a number of preceding frames instead of a single frame. Thereby, we take the global pavement surface appearance into account. Second, detected pothole regions are tracked in subsequent frames utilizing vision tracking algorithms. This enables convenient pothole counting in a pavement videos avoiding inefficient re-detection and matching. The method was implemented and experimental results show that pothole recognition precision and recall can reach 75% and 84%, respectively.
References

BASI (2008), Road Monitoring and Assessment (ZEB - Verfahrensbeschreibung), Federal Highway Research Institute (Bundesanstalt für Straßenwesen - BASIt), http://www.bast.de


ZTV ZEB-StB (2006), Zusätzliche Vertragsbedingungen und Richtlinien zur Zustandserfassung und –bewertung von Straßen (ZEB), German Road and Transportation Research Association (FGSV), No. 998.