Multi-resolution approach for fine structure extraction: Application and validation on road images
Nicolas Coudray, Argyro Karathanou, Sylvie Chambon

To cite this version:
Nicolas Coudray, Argyro Karathanou, Sylvie Chambon. Multi-resolution approach for fine structure extraction: Application and validation on road images. International Joint Conference on Computer Vision Theory and Applications, VISAPP, May 2010, France. hal-00614116

HAL Id: hal-00614116
https://hal.science/hal-00614116
Submitted on 9 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
MULTI-RESOLUTION APPROACH FOR FINE STRUCTURE EXTRACTION

Application and validation on road images

Nicolas Coudray, Argyro Karathanou
Laboratoire MIPS, Université de Haute-Alsace, Mulhouse, France
nicolas.coudray@uha.fr, e0602532@uha.fr

Sylvie Chambon
Laboratoire Central des Ponts et Chaussées, LCPC, Nantes, France
chambon@lcpc.fr

Keywords: Segmentation, multi-resolution, road images, cracks.

Abstract: In the context of fine structure extraction, this paper presents a new method based on multi-resolution segmentation applied for the detection of road cracks. A method already developed to detect low-contrasted biological membranes has been adapted to detect cracks on images: crack features are defined as heterogeneities rather than transitions of closed regions characterizing the membranes. This new methodology is quantitatively validated on reference segmentations and compared to an adapted filtering and Markovian modeling algorithm.

1 INTRODUCTION

Detecting fine structures is very helpful in a lot of domains: to extract ceramic damages (Elbehiery et al., 2005), to find cracks in underground pipes (Iyer and Sinha, 2005), to detect road network in satellite images (Geman and Jedynak, 1996), to follow vessels in medical images (Frangi et al., 1998). Since 1990, a lot of algorithms have been proposed in the domain of crack detection on road pavement surface. In fact, every country needs to evaluate, periodically, the quality of roads and most of this work is done manually, which is expensive, non-reproducible, dangerous and not very efficient. In consequence, a lot of efforts have been made in the field of research on automatic or semi-automatic procedures for detecting deterioration on the roads and, in particular detecting cracks. In 2003, the report of Schmidt (Schmidt, 2003) gives a good summary of the technologies developed in this field. The main difficulty of this task is that the default is not well contrasted compared to the texture of the road, and that it represents a very small part of the images (about 1.5% of the image).

The most recent algorithms proposed to tackle this problematic are based on multi-resolution approaches. Even if they are the most efficient, they have shown their limits by giving a sparse detection. In the field of detection of biological membranes, recent methods based on multi-resolution watershed segmentation provided great performances, even if the objects of interest are low-contrasted (Coudray et al., 2007). By analogy, the problem of biological membranes detection is similar to crack detection. The goal of this paper is to adapt this watershed method to road crack detection, and compare its performances with a Markovian modeling-based algorithm developed specifically for this purpose (Chambon et al., 2010).

Firstly, a brief state of the art of fine structure extraction is exposed. Secondly, the studied crack algorithms are presented: the Markovian modeling algorithm and the multi-resolution segmentation. Thirdly, the protocol used to analyze the performances is given. And finally, before concluding, results are summarized and discussed.

2 FINE STRUCTURE EXTRACTION

We distinguish four methods: those based on a THRESHOLD (Koutsopoulos and Downey, 1993), methods combining a threshold with mathematical morphology tools (MORPHOLOGY) (Tanaka and Uematsu, 1998), NEURAL NETWORK-based methods (Bray et al., 2006), and MULTI-SCALE filter-


Then, we also have to choose the number of directions for the adapted filtering and, considering the aspect of the road cracks, it is realistic to use these directions: $[0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}]$. Interested readers can find more details about this work in (Chambon et al., 2010). The goal of this paper is to compare this method with the new one presented below.

3.2 Multi-Resolution Detection of Heterogeneities (MRDH)

This method is inspired from the 3-step segmentation method developed for the segmentation of low and heterogeneously contrasted biological membranes (Coudray et al., 2007). The aim was to partition the image by detecting the closed contours of membranes. In this initial algorithm, the edges are firstly thresholded at different scales, and, secondly, the resulting binary images are combined in a Reconstructed Gradient-Like (RGL) image. Finally, the watershed algorithm is applied on the RGL image to obtain the searched partition. For the cracks, the third step has been adapted and a fourth step has been added to discard spurious detections. The steps are detailed below.

A multi-resolution transform is first applied to the initial image $I(x, y, 1)$ using a pyramidal transform: images $I(x, y, s)$ are obtained at scale $s$ using:

$$
I(x, y, s) = \sum_{m=-k}^{+k} \sum_{n=-k}^{+k} G(m, n, s) I(sx + m, sy + n, 1),
$$

where $G(m, n, s)$ is the smoothing average filter. Dyadic pyramids reduce the size of the image by a factor of 2 between each scale. Here, non-dyadic pyramids are used to better represent and identify the features, with $s \in V = \{1, 2, \ldots, 10\}$ (at scales coarser than 10, the images become too small to be analyzed). Heterogeneities are detected and thresholded at each scale, leading to binary images $B(x, y, s)$. In this study, features to be identified, i.e. the cracks, are better described as local heterogeneities than local transitions (e.g. Figure 1), making a local standard deviation filter more appropriate than a gradient filter like Sobel, for instance.

![Initial image example](image)

Figure 1: Initial image example ($I(x, y, 1)$) – The crack corresponds to a local low-contrast heterogeneity.
In the next step, binary images are combined to form the RGL image $I_{RGL}(x,y)$:

$$I_{RGL}(x,y) = \max_{s \in V_s} \left( (11 - s) \times B(x,y,s) \right).$$  \hspace{1cm} (2)

With the $(11 - s)$ weighting, the gray-level on the RGL image corresponds to the finer scale at which the feature has been identified. Details identified with a better precision have therefore a higher value (e.g. Figure 2).

The next step aims to thin the RGL image features to obtain 1-pixel wide objects, and has been adapted from our initial work to this road application. It needed to be modified since, contrary to membrane edges, cracks do not form closed regions: therefore, we can either apply the watershed algorithm after considering the background of the RGL image as seeds (i.e. pixels which value is 0 and which touch non-0 value pixels), or by applying a gray-scale thinning algorithm (Redding, 1996). For its simplicity and lower computational time, the watershed approach has been used in this study. The resulting image $I_{pc}$ (e.g. Figure 3) points to the potential crack pixels (PCP).

A last step has been added to validate $I_{pc}$ and reduce the false detections: a PCP is validated if its contrast is above a given threshold $C_{Tr}$. To make the measure more robust to the noise, instead of analyzing each PCP independently, the contrast is averaged on segments (a segment is a set of contiguous pixels having at most two neighboring PCP, see Figure 4 for an illustration of this vocabulary). The average contrast $C_{seg}$ of a segment is the difference between the average gray-level of the segment ($GL_{seg}$), and the average gray-level of the neighboring background ($GL_{Bkg}$):

$$C_{seg} = GL_{Bkg} - GL_{seg}. \hspace{1cm} (3)$$

A segment is considered as targeting a crack if $C_{seg} > C_{Tr}$ (e.g. Figure 6). Measures have been realized to set the main parameters of the algorithms, i.e. the kernel size of the standard deviation filter and the threshold $C_{Tr}$; we identified that the best results were obtained with a $5 \times 5$ pixels kernel and $C_{Tr} = 50$. We will describe the analysis methodology, and we will restrain the presentation of the influence of $C_{Tr}$ on the results.

4 ASSESSMENT METHODOLOGY

To evaluate and compare the crack detection methods, we had to choose: a/ the tested images, b/ how to determine the "ground truth" segmentation or reference segmentation and c/ the criteria in order to quantitatively evaluate the results.
4.1 Images

The algorithms were evaluated on 3 sets of test images: on 14 synthetic crack images, on 10 real images, and on 10 pre-processed real images. For synthetic images, it is easy and reliable to give a ground truth segmentation. For real images, estimate a "ground truth" or a reference segmentation is more complicated but the images are more realistic than the synthetic images.

The 14 synthetic images were built using different kinds of backgrounds: cracks were added on 8 images built with a random dot texture (sample $S_1$), and on 6 images of road containing no cracks, 2 of them acquired with a static camera (sample $S_2$), and 4 of them dynamically acquired with a camera embedded on a vehicle (sample $S_3$). For the last four ones, controlled lights were added. For all the 14 images, the cracks were randomly added, with a random shape and a random gray-level (Figure 7).

On the 10 real images, 4 were acquired using a static camera and 6 using the dynamic system (Figure 7).

In the third set of test images, the 10 real images were pre-processed using:

1. **Threshold** – In order to reduce the light halo in some images (the last six ones presented in Figure 9), each pixel over a given threshold is replaced by the local average gray levels.
2. **Smoothing** – A mean filter of size $3 \times 3$ is applied.
3. **Erosion** – An erosion with a square structuring element of size $3 \times 3$ is applied.
4. **Restoration** – This last pre-processing tries to combine the advantages of all the previous methods in three steps: histogram equalization, thresholding (like Threshold), and erosion (like Erosion).

4.2 Reference Segmentation

For real images, we briefly explain how the manual segmentation is validated. Four experts manually segmented the images with the same tools and in the same conditions. Then, the four segmentations were merged, following these rules:

1. A pixel labeled as crack by almost two experts was kept;
2. Every pixel near to a pixel kept by step 1 was also kept.

The second rule is iterative and stops when no pixel is added.

Then, the result is dilated with a structuring element of size $3 \times 3$. Results of the manual segmentations are presented in Figure 9 (for more details, the study of the reliability of these reference segmentations can be found in (Chambon et al., 2010)).

4.3 Quantitative Assessment

In Figure 8, the evaluation criteria are presented. They are the most used in the literature, except "accepted" pixels. In consequence, we have included accepted pixels in the computation of the similarity coefficient. For estimating accepted pixels, a threshold is applied on the distance between the detected pixel and the reference segmentation. This threshold is estimated using the mean distance between each of the four segmentations used for establishing the reference.

The DICE values lie on the $[0; 1]$ interval, but to simplify the visualization of the results, we display $\text{DICE} \times 100$ in all figures and tables of section 5.

5 RESULTS AND DISCUSSION

5.1 Synthetic Images

Figure 10 compares results for the different methods with synthetic images. It shows that the multi-resolution with $C_T = 50$ is the best to identify most of
Corrects = True positives (TP)
Positives (P)

False
Positives (FP)

Accepted

False
Negatives (FN)

Figure 8: Evaluation criteria – Representation of the recovery between an estimated segmentation (positives, P) and a reference segmentation.

Table 1: Results on Synthetic Images – The 3 image categories are: $S_1$, fully synthetic images, $S_2$, background acquired with a camera and $S_3$, background acquired with the dynamic system. Results in bold indicate best values and it illustrates how the MRDH approach outperforms the AFMM, in particular, with the most difficult images (sample $S_3$) where DICE is correct (greater than 50) whereas it is very low with the other methods.

<table>
<thead>
<tr>
<th></th>
<th>AFMM</th>
<th>MRDH, $C_{Tr} = 30$</th>
<th>MRDH, $C_{Tr} = 40$</th>
<th>MRDH, $C_{Tr} = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP+Acc</td>
<td>87</td>
<td>46</td>
<td>15</td>
<td>68</td>
</tr>
<tr>
<td>FN</td>
<td>1.5</td>
<td>3.7</td>
<td>2.6</td>
<td>2.8</td>
</tr>
<tr>
<td>DICE</td>
<td>92</td>
<td>63</td>
<td>24</td>
<td>81</td>
</tr>
<tr>
<td>$S_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP+Acc</td>
<td>81</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>FN</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>DICE</td>
<td>92</td>
<td>63</td>
<td>24</td>
<td>81</td>
</tr>
<tr>
<td>$S_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP+Acc</td>
<td>91</td>
<td>66</td>
<td>59</td>
<td>91</td>
</tr>
<tr>
<td>FN</td>
<td>18</td>
<td>6.7</td>
<td>6.9</td>
<td>18</td>
</tr>
<tr>
<td>DICE</td>
<td>94</td>
<td>80</td>
<td>70</td>
<td>94</td>
</tr>
</tbody>
</table>

5.2 Real Images

Figure 11 displays the results obtained on the real images. We see that the amount of FN is higher than with synthetic images, but we note that these values are probably overestimated, the human-reference images tending to select only the major cracks. The AFMM algorithm identified almost half of the cracks selected by the human, and a bit more than half were identified with the multi-resolution method. Considering the DICE measure, the multi-resolution tests gave the best overall results; though, on camera acquired images, DICE values can be considered as equivalent for both methods.

Figure 11: Summary of the algorithm performances on real images.

5.3 Real Pre-Processed Images

Here, we compare the AFMM method with the MRDH method (only the one with the more efficient...
parameter, i.e., $C_r = 50$, is discussed here) when the real images are pre-processed. In particular, we evaluate the influence of the 4 pre-processing described in section 4.1.

Table 2: DICE of pre-treated images: the three numbers show the average values: with the 10 images (with the 4 images acquired in static way, with the 6 images acquired in a dynamic manner).

<table>
<thead>
<tr>
<th></th>
<th>AFMM</th>
<th>MRDH, $C_r = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>46 (49.44)</td>
<td>55 (52.57)</td>
</tr>
<tr>
<td>Threshold</td>
<td>64 (68.61)</td>
<td>55 (60.50)</td>
</tr>
<tr>
<td>Smoothing</td>
<td>66 (77.59)</td>
<td>52 (63.42)</td>
</tr>
<tr>
<td>Erosion</td>
<td>59 (58.60)</td>
<td>57 (64.51)</td>
</tr>
<tr>
<td>Restoration</td>
<td>52 (49.55)</td>
<td>57 (59.55)</td>
</tr>
</tbody>
</table>

Table 2 shows the DICE measures obtained. It can be seen that the four pre-processings enhance the overall performances.

When pre-processed with the threshold, smoothing, erosion and restoration, the AFMM’s algorithm is better, with the best results being obtained after a smoothing. Also, smoothing and threshold have an even more beneficial effect when static images are used.

According to the DICE measure, the performances of MRDH applied after pre-treatment are quite disparate. In fact, it illustrates how this new method is globally quite robust against the acquisition conditions. Images do not need to be pre-processed in order to increase the performances and this is an important superiority compared to the AFMM method which is highly dependent on the acquisition conditions. Analyzed in details, we notice that for static images, DICE values seem to be enhanced by each pre-processing, while DICE values seem to decrease for "dynamic" images.

6 CONCLUSIONS

A methodology for adapting a multi-resolution-based segmentation to the field of road crack detection has been introduced. The experimental results demonstrate the efficient results obtained on 14 synthetic images and 10 significant real images. Moreover, it shows how it outperforms a previous method based on adapted filtering, in particular with the most difficult images acquired dynamically and that present illumination defaults.

The next step of this work will be to validate this work on more data (32 more images will be available soon). It has been shown that MRDH improves the percentage of true detections (this percentage is higher than with AFMM), but the AFMM presents less false negatives. The next improvement will be to reduce this phenomenon by introducing more constraints on the final step, like for example, an active contour constraint instead of using only the constraint on the contrast.

REFERENCES


