Anomaly detection in urban drainage with stereovision

Meijer, D.W.J.; Luimes, R.; Knobbe, A.J.; Bäck, T.H.W.

Note: To cite this publication please use the final published version (if applicable).
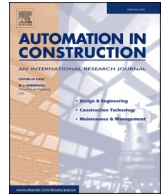
Anomaly detection in urban drainage with stereovision

Dirk Meijer [a,c,*], Rianne Luimes [b], Arno Knobbe [a], Thomas Bäck [a]

[a] *Leiden University, the Netherlands*
[b] *Eindhoven University of Technology, the Netherlands*
[c] *Centrum Wiskunde & Informatica (CWI), the Netherlands*

ABSTRACT

This work introduces RADIUS, a framework for anomaly detection in sewer pipes using stereovision. The framework employs three-dimensional geometry reconstruction from stereo vision, followed by statistical modeling of the geometry with a generic pipe model. The framework is designed to be compatible with existing workflows for sewer pipe defect detection, as well as to provide opportunities for machine learning implementations in the future. We test the framework on 48 image sets of 26 sewer pipes in different conditions collected in the lab. Of these 48 image sets, 5 could not be properly reconstructed in three dimensions due to insufficient stereo matching. The surface fitting and anomaly detection performed well: a human-graded defect severity score had a moderate, positive Pearson correlation of 0.65 with our calculated anomaly scores, making this a promising approach to automated defect detection in urban drainage.

1. Introduction

The SewerSense project aims to improve sewer condition assessment, in part through novel inspection techniques and data processing [1]. Many sophisticated technologies have been suggested to improve sewer quality assessment [2,3], but to our knowledge, none have yet been implemented on a large scale, while the preliminary results are promising. In this work, we consolidate several techniques into RADIUS (Robust Anomaly Detection In Urban drainage with Stereovision), a framework for anomaly detection in sewer pipes.

The RADIUS framework is designed to be compatible with the existing workflow of the trained operator, as well as be future-proof for a completely automated sewer inspection system, for which advances are being made rapidly. The framework also has a low up-front investment in terms of equipment, as it uses two cameras for data collection, and image processing steps that can be performed on a consumer-grade computer within a reasonable amount of time.

Our proposed method revolves around the technique of *computer stereovision*, which uses two or more well-calibrated cameras placed side by side, in order to create a sense of depth, similar to how the binocular vision in humans is used to capture the spatial configuration of one's surrounding. In that sense, the proposed method is a type of 3D ranging technique that promises to produce a fairly faithful 3D reconstruction of

the interior of a sewer pipe in the form of a 3D point cloud with associated color information. Since the raw output of our stereovision setup captures the pipe's surface in considerable detail (the extent of this is determined by the resolution of the cameras employed), it can theoretically be used to recognise various different categories of pipe defects that have a spatial nature. These include deposits, holes, cracks, intrusions and exposed granulates. Some of these may in fact be harder to correctly classify using traditional single CCTV setups, since without further spatial clues, they cannot be distinguished (for example, a crack (on the surface) vs. an intruding root (away from the surface)).

While in terms of types of defects to be recognised, our method is open-ended, we focus in this paper on the recognition of two specific types of defects, namely deposits and exposed granulates. For other defect types (such as misaligned joints) the larger part of our proposed pipeline remains unchanged, but in the final stages provisions need to be made to account for the different spatial phenomena at hand. When focusing on deposits and exposed granulates, we need to recognise areas in the pipe where the surface is further inward or outward than one would expect. In other words, there is an *expected* pipe geometry (the pipe model) and a *measured* pipe geometry, and any deviations between these will be expressed in an anomaly score. Although this form of anomaly detection sounds fairly straightforward, it relies on the availability of a pipe model, which is actually non-trivial. First of all, the

---

precise location and orientation of the camera pair inside the pipe is uncertain. We can only assume that the cameras are pointing roughly along the main axis of the pipe, not too far removed from the center of the pipe. Second, our method should be robust with respect to the shape of the pipe, such that different pipe topologies can be dealt with, without having to reconfigure the recognition system. This means we will take a data-driven approach that assumes that the larger part of the pipe is unaffected, such that a 'normal' geometry can be derived from that, and the outliers with respect to this geometry are the anomalies.

In broad strokes, our method works as follows. In the image acquisition stage, any radial lens distortion is removed from the images, and slight misalignment between the left and right image are corrected. An existing semi-global algorithm for stereo matching [4] then produces pairs of corresponding pixels in both images. The computed disparity between each pair can be translated into a distance for this pixel: closer points will appear further apart in the two images. The resulting point cloud already captures the pipe geometry, but needs to be further processed in order to automatically identify the defects of interest, in our case deposits and exposed granulate. The next stage of surface fitting combines a parameterised surface model, one that can adapt to a wide range of pipe types, with the robust regression algorithm RANSAC [5]. This algorithm assumes that the majority of the data fits a predefined model class (our parameterised pipe model), but also that a certain fraction of the data constitutes outliers. This allows us to use the parameterised pipe model without the fit being influenced by these outliers. The deviation of the measured geometry from the expected geometry as predicted by the RANSAC model is used as an anomaly score.

The main contributions of this paper are:

- We demonstrate how a faithful and high-resolution reconstruction of the pipe surface, including its defects, can be obtained with stereo cameras and a stereo matching algorithm.
- We propose a generic pipe surface model that is able to model the pipe geometry of a range of pipe shapes (including circular and egg-shaped), captured under various angles. This pipe surface model has the attractive property that it falls in the category of functions that can be statistically fit with the Ordinary Least Squares method [6], making it computationally efficient.
- We propose a method based on the RANSAC algorithm to fit point cloud data that is a mixture of regular pipe surface and anomalies.
- We define a global anomaly score that quantifies the amount of deviation from the pipe model per image pair.

The paper is outlined as follows. Section 2 outlines selected prior work. Section 3 contains an overview of prerequisite knowledge. The RADIUS framework is described in full detail in Section 4. Section 5 gives an overview of the data and the experiments, the results of which are summarised and discussed in Section 6. Section 7 discusses the limitations of the framework, its envisioned applications, and possible future work.

2. Prior work and motivation

In the field of 3D ranging techniques, which our approach belongs to, the use of laser scanners for sewer pipe inspections has been thoroughly researched, see for example [7–10]. The reasons we have opted to go with stereovision instead of laser scanning are threefold, i) the equipment cost of two cameras versus that of a laser scanner is significantly lower, making this approach more accessible, ii) a stereovision setup has no moving parts, which matters in real-world scenarios, where the environment of a sewer pipe can be very abrasive to moving parts in particular, and iii) the point cloud obtained from stereovision will be linked directly to images with a color component, whereas a laser scanner only provides the geometry.

2.1. Stereovision

While not much research has been done on the use of stereovision in the context of sewer condition assessment, the use of stereovision in the general context of sewer maintenance is not new. Most works restrict their approach to cylindrical pipes, as these our fairly common, but our work does not impose that limitation.

Ahrary et al. (2005) [11] propose an algorithm for navigation of an autonomous vehicle through a sewer network based on stereovision. Later, Ahrary et al. (2008) developed a computationally efficient stereo matching algorithm specifically for sewer pipes [12]. We do not use either of these algorithms, as processing power is not a limitation in our research, and navigation of an autonomous vehicle is outside our scope.

Tangentially related to this work, Koodtalang et al. [13] use stereovision to determine manufacturing defects in pipes prior to installation.

Gunatilake et al. [14] combine a stereovision setup with two laser profilers to map the images recorded by the cameras onto the potentially more accurate point cloud produced by the laser profilers. This produces a high-resolution RGB-D dataset for later inspection, either by a trained expert or another algorithm. The method is tested on a single, heavily corroded pipe, as well as an artificial pipe.

Most closely related to the work presented in this paper, Huyhn et al. have published two works [15,16] on anomaly detection in sewer pipes with stereovision. They demonstrate the visibility of artificial defects in point clouds generated from stereovision. A critical difference to their approach is that our approach requires no human operator to center the defect into the camera's field of view, but instead is able to highlight anomalies in the entire pipe from a set of images, and is thus more suitable for automated defect detection.

2.2. Anomaly detection

Anomaly detection has been used extensively in sewer condition assessment [17–22] as a stepping stone from "traditional" data that is gathered for manual classification, towards automation of the inspection process. Meijer et al. [17] performed principal component analysis on various feature descriptors of a labelled set of CCTV images and compared the partial reconstruction with the actual values for an unsupervised approach, and compared this with a convolutional autoencoder. Myrans et al. [18] performed anomaly detection on sewer CCTV images by training a random forest and a support vector machine on GIST-features. Myrans et al. [19] later expanded on this by exploring the use of a one-class support vector machine, a type of support vector machine designed specifically for anomaly detection. Moradi et al. [20] similarly used a one-class support vector machine to detect anomalies from SIFT features, and combined this approach with localization of the pipe through text recognition. Fang et al. [21] performed anomaly detection on sewer CCTV video footage by performing principal component analysis (PCA) on various local feature descriptors. Russo et al. [22] use a convolutional autoencoder to detect anomalies in CCTV images.

While numerous other works that use computer vision or image processing to detect defects in sewer pipe images exist, we have limited this section to only those that perform *unsupervised anomaly detection*, as they are most similar to this work. For a more broad perspective on recent advances in this field, please refer to Haurum and Moeslund [23].

3. Prerequisites

3.1. Sewer pipeline quality assessment

Sewer pipes have to be periodically inspected to ensure proper function. CCTV inspection is one of the most common approaches: a remotely operated vehicle is lowered into a manhole, equipped with a camera and possibly other sensors, to gather data for classification by a trained operator [24]. This process is labor-intensive, and often does not

lead to reliable results, as the classification and severity ratings thereof are highly subjective, differing not only between operators, but also for the same operator on repeated inspections [25]. It would be beneficial to automate (parts of) the process, which should lead to higher consistency of the assessments, as well as improvement of the overall quality of the assessments.

3.2. Three-dimensional point clouds

A point cloud is a collection of data points that include two- or three-dimensional spatial coordinates [26]. We use 'point cloud' to refer to a three-dimensional point cloud specifically. A point cloud may be the result of physical measurements, including but not limited to: laser scanning, tomography, or photogrammetry (including in this work). The data points that make up the point cloud also may contain additional information, such as velocity, color, etc., depending on the type of measurement that was performed.

Laser scanning may refer to one of two different techniques: either laser telemetry, or laser triangulation. Laser telemetry emits a narrow beam pulse, then measures the time of flight until the beam returns to determine the distance to the reflecting surface [27]. Laser triangulation emits a narrow beam and uses a camera adjacent to the laser emitter to determine the distance to the reflecting object, based on the location of the laser's dot in the camera's field of view [28]. To generate a point cloud with this technique, the emitter and sensor must move to 'scan' across a surface. The trade-off between telemetry and triangulation is mostly one of range and precision. Laser telemetry has a high range (in the order of kilometers), but a low precision (in the order of millimeters), whereas laser triangulation has a low range (in the order of meters), but a high precision (in the order of ten microns).

Photogrammetry refers to the process of inferring information about physical objects through analysis of photographic images [29]. More specifically, stereophotogrammetry is the process of inferring the three-dimensional shape of an object from multiple images, recorded from different positions. Similar to laser triangulation, this process uses triangulation between points that appear in multiple images to determine its location. The precision of photogrammetry is dependent on the camera sensors, lenses, lighting conditions, proximity to the measured object, and working resolution, but a precision in the order of tenths of millimeters is achievable. As such, there are no inherent limitations that make photogrammetry unsuitable for use in sewer pipes. A more detailed overview of the specific case of stereophotogrammetry that we employ, computer stereovision, is given in Section 3.3.

Tomography is beyond the scope of this work, but has been used for sewer pipe measurements in the past, see for example [30,31].

3.3. Computer stereovision and epipolar geometry

Computer stereovision, or simply 'stereovision', is a computer vision technique in which two side-by-side cameras simultaneously record an image. The correspondence between points that appear in both images gives us information on the distance from the cameras to that point, similar to how the correspondence between the left and right eye allows humans to perceive depth [32].

To illustrate the principle, we examine the *epipolar plane* [33] of two horizontally aligned cameras and an object that is visible to both cameras, as in Fig. 1. The problem can be significantly simplified by the camera axes being parallel, which is an achievable configuration for this application. C_1 and C_2 are the two cameras, and **P** the point of interest. Both cameras have identical physical properties, and we consider C_1 to be the *reference camera*. f is the focal distance of the cameras, b is the baseline distance between the cameras, two physical distances that we know precisely. I_1 and I_2 are the virtual image planes, one focal length distance in front of the cameras.

We first calculate X and Z, the physical location of **P** in the epipolar plane, from the perspective of our reference camera C_1. Consider d_1 and
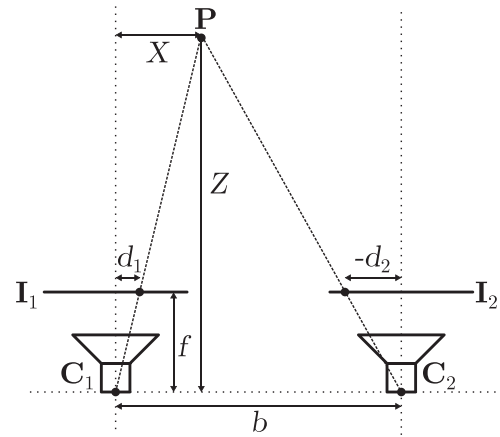

Fig. 1. Epipolar geometry with parallel camera axes.

d_2, the projected locations of **P** onto I_1 and I_2, relative to the centres of the image planes.[1] Similar triangle geometry allows us to solve for Z and X:

$$d_1/f = X/Z \tag{1}$$

$$-d_2/f = (b-X)/Z \tag{2}$$

$$(d_1 - d_2)/f = b/Z \tag{3}$$

$$Z = b \cdot f/(d_1 - d_2) \tag{4}$$

$$X = d_1 \cdot b/(d_1 - d_2) \tag{5}$$

Two important things should be noted at this point:

- The Y coordinate of **P**, which is orthogonal to the X-Z plane shown in Fig. 1, also has to be computed.

$$Y = d_y \cdot b/(d_1 - d_2) \tag{6}$$

where d_y is the vertical distance from the centre of the projection of **P** on I_1. Since the cameras are aligned in the horizontal plane, there is no need to take a vertical shift into consideration, as any point will be projected on both virtual image planes at equal height.

- For the calculated coordinates to be represented in physical units, we can either express d_1 and d_2 in physical units, or we can express f in pixels instead of physical units. Either conversion is done by finding the physical size of a pixel on the camera's sensor array. In this work, we will assume the focal length f is expressed in pixels.

Stereovision algorithms apply this principle to all pixels in an image: each pixel in the image produced by the reference camera is matched to a pixel in the second image, and the difference in horizontal positions of these pixels produces the *disparity* $(d_1 - d_2)$. More specifically, for each pixel in the reference image, a local neighbourhood around the pixel is compared to a patch of the same size as this neighborhood, in the same position in the other image, but shifted horizontally. The horizontal shift that minimises the difference between the two image patches is considered the best match.

This introduces multiple difficulties, as finding the correspondence between pixels is not feasible to do exhaustively in a short amount of time, making it a heuristic search process, expected to have multiple

---

[1] Note that we take d_2 to be a negative value in this case, as it is to the left of the centre of I_2. If **P** would be on the same side of both camera axes, d_1 and d_2 would have the same sign.

local optima. Images that have some periodicity in the horizontal direction may result in the correspondence being off by a multiple of the period. An even bigger challenge arises when the selected neighbourhood patch is entirely smooth: matching the exact location will become difficult, as small shifts lead to little difference in matching quality. Practically, an exact alignment of the cameras is also difficult to achieve, and any physical camera and lens are going to introduce distortion to the recorded images [34], both of which will have to be corrected before the matching process commences.

A more detailed overview of the specific stereovision algorithm used is given in Section 4.2.

3.4. Anomaly detection

Anomaly detection, sometimes referred to as outlier detection, is a machine learning problem aimed at finding instances in a dataset that deviate from the majority [35]. It has many applications, from fraud detection to noise removal. Anomaly detection can be supervised, semi-supervised, or unsupervised. This distinction indicates whether all training data is labelled as anomalous by some expert, only some of the data is labelled, or whether no such data is available, respectively. This work relies on unsupervised anomaly detection, meaning there is no knowledge of the anomalousness of any instance.

Unsupervised anomaly detection relies in most cases, including ours, on robust[2] regression. This means that we look for some model that explains the behaviour of *most* instances in our dataset. Any instances not explained by this model are considered to be anomalies or outliers.

An important quality of the model we fit on our data is that it has limited complexity. If the model's complexity is too high, it may fit the anomalies that we are trying to detect as well, meaning they become inliers and are no longer detected as anomalies. Still, a certain degree of complexity may be required in order to account for regular aspects of the pipe geometry. This implies a trade-off between how complex we allow patterns to be, and when complexities become anomalies. In this case, we let this trade-off be informed by the expected geometry of the undamaged sewer pipes.

4. Framework

We propose a framework for anomaly detection from stereovision measurements in sewer pipes, as shown in Fig. 2. The framework consists of five major steps: image acquisition, semi-global stereo matching, three-dimensional geometry reconstruction, robust pipe surface fitting, and anomaly detection and processing. These steps are designed to be executed in sequence, each step's output being the next step's input. The five steps composing the framework are each discussed in detail in Sections 4.1–4.5.

4.1. Image acquisition

The stereovision setup requires two cameras placed side-by-side, at equal height, pointed in the same direction. Perfect alignment of the cameras is near impossible, but correcting a slight misalignment is possible. The setup is then directed into the pipe, such that the camera axes are mostly parallel to the pipe axis. For in-situ inspection, this means that the setup has to be attached to the pipe inspection vehicle, while aimed directly into the pipe.

Any optical lens introduces some radial distortion to an image [34], meaning that points at different distances from the lens axis have different levels of magnification. As the lens axes of the two cameras are parallel but translated, this may introduce a difference in magnification of a point between the two cameras, and thereby a difference in vertical position. Depending on the severity of this distortion (or if the cameras
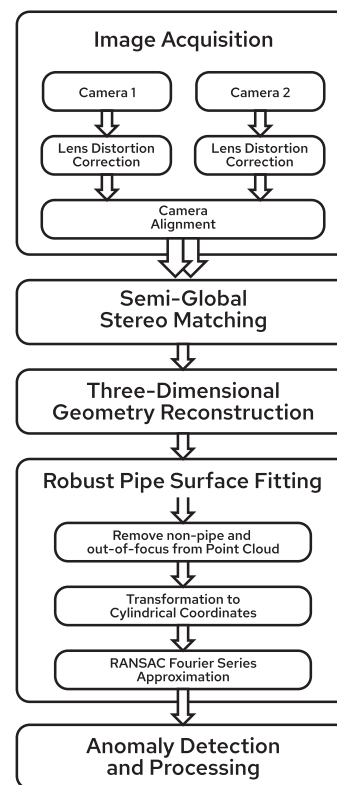
---

² *Robust* in this context refers to a reduced sensitivity to noise or outliers.

Fig. 2. Overview of the proposed framework.

and lenses are not of identical make), correction may be required for the images to be suitable for stereo matching.

By taking pictures of a chessboard pattern from different angles and distances, we can observe the effects of the radial distortion: without any distortion, the lines on a chessboard should be entirely straight, but slight curves may appear as a result of the radial distortion. Radial distortion can be reversed digitally by performing a second radial distortion to undo the first. The correct inverse distortion parameters can be estimated from the deviations in the images of the chessboard pattern, as outlined in more detail in [36].

Once images from both cameras are free of (extreme) lens distortion, an alignment between the cameras must also take place, in order to compensate for a vertical misalignment or rotation of one camera around its axis. From a set of images with several visible landmarks (the same images of the chessboard pattern may be used), we can estimate any vertical shift between the images, as well as a possible rotation along the camera axis, and correct this with a simple affine transformation [37]. If the camera axes themselves are not perfectly aligned, this may be visible as a horizontal shift of points that are very far away, as a point on the horizon should in theory have the same position in both images.

After camera alignment, the first step is complete and the images can serve as input for the second step: semi-global stereo matching.

4.2. Semi-global stereo matching

As described in Section 3.3, stereo matching relies on comparing projected locations of a three-dimensional point onto two-dimensional images. Stereo matching an entire image is generally done by comparing a position in the reference image to a horizontally shifted position in the second image, often to sub-pixel accuracy [38]. The shift that best matches the pixels in the reference image to the pixels in the second image is called the *disparity* for that pixel. The comparison may be done by minimising the difference in values of the pixels, but better results may be obtained by using a matching cost that relies less on

absolute values, such as cross-correlation, Hamming distance, or Birchfield-Tomasi dissimilarity [39].

Matching single pixels from the two images is going to lead to a substantial amount of incorrect matches. Suggested solutions for this include matching a window around each pixel to a window of the same size, enforcing some type of smoothness between disparity in neighboring pixels, and various combinations thereof [40].

Unique to our problem is the fact that the sewer pipe axis is parallel to the camera axes. The surface we are most interested in, the pipe wall, is perpendicular to the image plane. This results in the Z-distance and disparity to gradually change and not be constant anywhere inside the pipe. Window-based stereo matching methods are designed to perform best when large patches of the reference image have the same disparity, which is not the case in this scenario.

This gradual change requires the window around the pixel that is to be matched to be small (we suggest <10 pixels on either side). The larger the window is, the more difficult it will be to match it properly. The extremes of the window are expected to have a different disparity from the center pixel, so too large a window will be impossible to match correctly.

To enforce smoothness, we suggest using Hirschmüller's *semi-global matching* algorithm [4], which adds two regularisation parameters. These parameters, P_1 and P_2, penalise a pixel having a different disparity from its neighbouring pixels. The best match for each pixel in the reference image is chosen as the match that minimises the sum of the matching cost M and the regularisation cost R. This introduces a circular dependency, as the regularisation cost depends on the best match of neighbouring pixels, which itself depends on the regularisation cost. Because of this, the algorithm usually requires multiple passes to stabilise.

Each neighbouring pixel contributes 0, P_1, or P_2 to the regularisation cost, depending on whether the neighbour has the same disparity, an absolute difference in disparity of 1, or a larger difference in disparity with the current pixel, respectively. Specifically, each match is given the regularisation cost:

$$R(d_1, d_2) = \begin{cases} 0 & \text{if } d_1 = d_2 \\ P_1 & \text{if} |d_1 - d_2| \leq 1 \\ P_2 & \text{if} |d_1 - d_2| > 1 \end{cases} \quad (7)$$

Again taking into account the fact that we expect the disparity to change gradually, we suggest setting $P_1 \ll P_2$. A value of $P_1 = 0$ may prove successful, but could also lead to erratic results. With a small or no penalty on small differences in disparity, but a large penalty on larger disparities, we can enforce the type of smoothness that we expect in sewer pipe images. The best value of P_2 depends on the window size chosen for matching, the matching cost itself, and the number of neighbours that the semi-global matching algorithm considers (commonly 4 or 8).

To reduce false detections further, we also suggest using a *uniqueness ratio*, which requires that the best disparity must have a score that is at least u times as large as the next best disparity. This may lead to correct disparities also being discarded, but this happens most commonly in very smooth regions, where the exact disparity is difficult to pinpoint. For the purposes of anomaly detection, this is not an issue, as these regions are unlikely to contain anomalies.

With an accurate estimation of the disparity for each pixel, these disparities can be used to reconstruct a three-dimensional point cloud.

4.3. Three-dimensional geometry reconstruction

As outlined in Section 3.3, we can triangulate the three-dimensional location of a point visible in both cameras once we know the disparity, using eqs. (4), (5), and (6). Eqs. (5) and (6) may be rewritten as

$$X = (x - x_0) \cdot Z/f \quad (8)$$

$$Y = (y - y_0) \cdot Z/f \quad (9)$$

where $[x, y]$ is the pixel position of the triangulated point in the reference image, and $[x_0, y_0]$ is the pixel position of the center of the reference image. This means that the center of the image will be projected to some position on the Z-axis, both the X and Y coordinates of the point being zero. Note that no two points can have the exact same X and Y coordinates, as one would occlude the other in that case.

Doing this for every pixel in the image (that has a valid disparity) gives us a point cloud with one point for every pixel. It may be useful to keep the RGB values of the pixels attached to the points in the point cloud for easier inspection and later processing. But a few important caveats arise when we move away from the ideal purely mathematical situation as introduced in Section 3.3 though.

The value of b will have some non-zero error, which leads to a scaling of the entire point cloud by a factor of $\frac{b'}{b}$ where b' is the measured baseline and b the actual baseline. The more accurately b' is measured, the closer to 1 this scaling factor is. It should also be noted that if correct physical dimensions of the point cloud are not important to the application, this does not have to be taken into consideration.

If the camera axes are not entirely parallel in the epipolar plane, the calculated disparity will have a small error. This may lead to a perceived deviation in radius along the length of the pipe, meaning a cylindrical pipe may appear conical in the point cloud.

While these are both issues to be aware of, they do not hinder the pipe model we propose in this work.

4.4. Robust pipe surface fitting

At this point, we have a three-dimensional point cloud of a pipe which can be used to estimate the original pipe geometry as a mathematical model, excluding any anomalies (henceforth simply referred to as the 'geometry').

While the image will be perfectly in-focus at a specific distance, a region known as the *depth of field* around this distance is determined to be the range with *acceptable* levels of focus [41]. The distance range of the depth of field is a simple function of the focus distance, the focal length, and the aperture size, which is adjustable in most cases. A smaller aperture size will give a larger depth of field, at the cost of less light reaching the camera sensor, leading to more sensor noise at equal exposure times [41]. To inspect an entire pipe, we might move the pipe inspection vehicle through the pipe at small intervals, taking measurements at each interval. This means that a larger depth of field leads to fewer measurements needed to process a unit length of pipe, as a larger portion of the pipe can be captured in a single photograph. This results in a point cloud of a pipe, centered approximately around the Z-axis. Points with a Z value outside the depth of field can be discarded, as we are better off estimating the geometry of such points when the cameras are positioned at a different position along the pipe.

A transformation to cylindrical coordinates at this point allows for a more natural notation of the geometry. We define:

$$r = \sqrt{X^2 + Y^2} \quad (10)$$

$$\theta = arctan_2(Y, X) \quad (11)$$

where arctan$_2$ is the two-argument arctangent, which spans the interval $(-\pi, \pi)$. We can now without loss of generality express each point in (r, θ, Z).

A naive approach at this point might be to fit a cylinder model of

$$r = r_0 \quad (12)$$

to capture the geometry of the pipe, where r_0 is the radius of the pipe. There are a few reasons why this is a poor approach:

i. The *Z*-axis may not be the precise center of the pipe, depending on how accurately it was possible to align the reference camera's axis with the pipe axis.

ii. The pipe might not have a circular profile. In our experiments we use both cylindrical and egg-shaped pipes, but any pipe with a somewhat smooth profile should work with our approach.

iii. The radius and center of the pipe may appear slanted in the point cloud along the *Z*-axis as a result of a slight misalignment of the camera axes in the epipolar plane.

We can address each of these issues in order.

To address the first issue, we assume for now that the pipe is cylindrical along the *Z*-axis, but not perfectly centered. Using a polar coordinate representation of an off-centre circle, [42] we may express the geometry as

$$r = \sqrt{r_0^2 - d^2 sin^2(\theta - \phi)} + d cos(\theta - \phi) \tag{13}$$

where *d* is the distance between the axis of the cylinder and the *Z*-axis, and ϕ is the angle at which the distance to the *Z*-axis is maximal. It can be observed that if $d \ll r_0$ (the center of the pipe is close to the center of the image), we may simplify eq. (13) to

$$r \approx r_0 + d cos(\theta - \phi) (\text{for} \ \ d \ll r_0) \tag{14}$$

At this point, we take a small sidestep to rewrite eq. (14) using a trigonometric identity into

$$r = r_0 + a sin(\theta) + b cos(\theta) \tag{15}$$

It can be seen that these two forms are identical when $d = \sqrt{a^2 + b^2}$ and $\phi = \arctan_2(b, a)$. The reason for this rewrite is entirely practical: we still have two unknowns to solve for, but both unknowns are now parameters of a linear function, meaning that we can now solve *a* and *b* with Ordinary Least Squares regression [6], whereas that would not be possible for the parameter ϕ, because it is inside a cosine.

To address the second issue, the possibility of pipes with non-circular profiles, we need a more complex function to describe the radius *r* as a function of the angle *θ*. To prevent modeling possible defects into the geometry, making them impossible to detect as anomalies, we use a limited approximation of *r* in terms of functions of *θ*. Different approximations can be used, but we suggest the use of a Fourier series approximation, as the radius is inherently periodic as a function of the angle. We redefine the model as

$$r = r_0 + \sum_{k=1}^{K} (a_k sin(k\theta) + b_k cos(k\theta)) \tag{16}$$

where *K* dictates how many Fourier components are used to approximate the radius. It may be seen that eq. (15) is an instance of eq. (16), with *K* = 1. For 'egg'-shaped pipes, we find a value of *K* = 6 to be generally sufficient, but any pipe profile with corners or otherwise non-smooth sections may require a higher value of *K*. Fig. 3 illustrates how an 'egg'-shaped pipe may be expressed in these Fourier components.

To address the third and final issue, we allow the radial parameters to change along the *Z*-axis, that is, along the pipe axis. To account for both a translation and scaling of the profile along the *Z*-axis—corresponding to a misalignment of camera axis and pipe axis, and a measurement error in the baseline distance, respectively—we allow each of the previously introduced parameters to vary linearly along the *Z*-axis. For every term in eq. (16), we add another term with a different parameter, and multiply by *Z*, giving us:

$$r = (r_0 + \rho_0 Z) + \sum_{k=1}^{K} \begin{pmatrix} a_k sin(k\theta) + b_k cos(k\theta) + \\ \alpha_k Z sin(k\theta) + \beta_k Z cos(k\theta) \end{pmatrix} \tag{17}$$

Eq. (17) is the model we will use to fit the transformed point cloud data, but as we expect anomalies, we will have to employ a robust
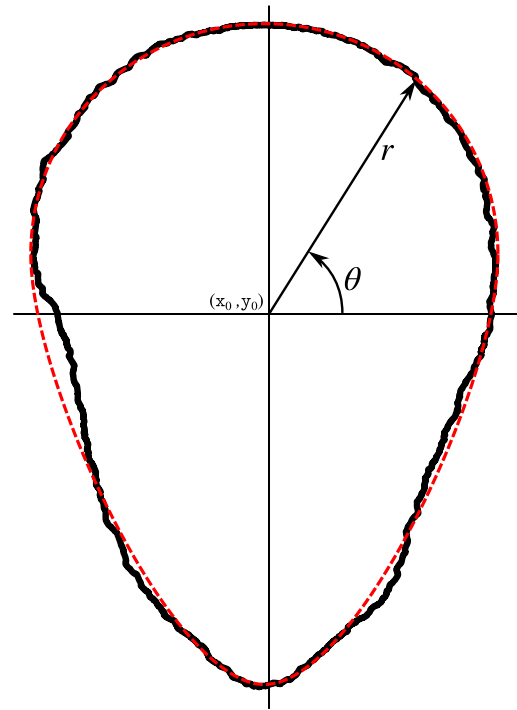

Fig. 3. Cross-sectional profile of a (fictional) deformed 'egg'-shaped pipe (black, solid), with a fit of Eq. (16) for *K* = 6 (red, dashed). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regression method.

The robust regression method we use is RANSAC, short for 'random sample consensus' [5]. RANSAC fits a model a large number of times on a 'minimal subset' of data, then selects a model fit that has both a large amounts of inliers, and a low error rate for those inliers. In this context, a minimal subset is the minimum number of points we need to fit the model. As our model has $2 + 4K$ parameters, we need as many points.

Then we determine the inliers, the points that are accurately described by this fit, according to some maximum difference between the actual value of *r* and the one predicted by the fit, known as the *inlier threshold*. If the number of inliers meets a set minimum, the model is fit a second time, but on all inliers this time. We store this new fit, along with the error rate on its inliers. This process is repeated a large number of times, after which we select the fit with the lowest error rate on its inliers.

The value of the inlier threshold will depend on the variance of the predicted variable. The minimum number of inliers required will usually be defined as a percentage of all data points, depending on the ratio of anomalous data points we expect to have. The chance that a minimal subset will result in a good fit of the data is low, but this first fit is only used to determine which points are the inliers that we want to perform the second fit on. Depending on how likely it is that the first fit reaches the minimum number of inliers under the chosen inlier threshold, the number of times the process should be repeated can differ by orders of magnitude: 10, 100, or 1000 could all be reasonable numbers.

If we choose the RANSAC algorithm parameters reasonably, this should give us a fit of the model described in eq. (17) that accurately describes a large portion of the data points, while not taking the actual anomalies into account.

4.5. Anomaly detection and processing

The final step in the framework, anomaly detection, is trivial at this stage: the difference between the actual value of *r* and the value

predicted by the best fit is an 'anomaly score'. We might threshold these scores to distinguish anomalies from non-anomalies, or consider the scores themselves as a continuous indicator.

Depending on the context in which the framework is employed, we have several suggestions for further processing of the anomaly scores:

- the ratio of anomalous pixels to regular pixels may be an aggregated indicator of anomalousness of a length of pipe,
- for further human inspection, the anomaly scores can be visualised in either an interactive point cloud or the original images, to indicate areas that might warrant attention,
- if pixel-wise classification is the goal, the anomaly scores can augment the RGB values of a pixel in a subsequent classifier.
- the size of continuous anomalous regions, as well as the absolute anomaly scores in such regions, might be used for defect identification or even severity.

In the experiments performed for this paper, we have calculated a global anomaly score per image set as follows:

$$A = \frac{1}{N} \sum_{i=1}^{N} \| min(r_i - \widehat{r}_i, 0) \| \tag{18}$$

where r_i is the radius of a point in the point cloud and $\widehat{r}_i$ is the predicted radius for that point. The point anomaly scores are clipped into the range $(-\infty, 0]$ and we calculate the average absolute value over all points in the point cloud. The point anomaly scores are clipped to only negative values, as otherwise the many points outside of the inner pipe wall present in our point clouds would skew the global anomaly score. This clipping would not be necessary in in-situ inspections, because no points outside the pipe should be visible, except if caused by defects.

5. Experimental setup

We evaluated the efficacy of the framework in a laboratory experiment. A total of 26 sewer pipe segments in various conditions were photographed with a stereocamera setup. Two Basler Ace2 A1920-160umBAS area scan cameras were fitted with lenses with a 16 mm focal length and attached side by side to a metal plate. The baseline was determined to be 29 mm, the lenses were focused at approximately 1.5 m distance, and the lens aperture was set to an *f*-number of 6, meaning the aperture diameter was equal to $16/6 \approx 2.667$ millimeters. A single pixel of an object at the in-focus distance corresponds to approximately 2.2 mm in real-world coordinates. In the ideal circumstances the algorithms can detect a shift of 1/16th of a pixel, so the maximum sensitivity we can expect to achieve is in the order of 1/10th of a millimeter.

The plate was attached to a rail, to allow for movement of the setup along the camera axes. The cameras were directed into the sewer pipe, which was covered at both ends with a piece of cloth. The pipe segments were illuminated with an LED light placed behind the cameras. The entire setup is depicted in Fig. 4.

22 of the 26 sewer pipes were photographed from both ends, the other 4 sewer pipes were photographed from one end only, giving us a total of 48 image sets.

5.1. Image data

Fig. 5 shows examples of stereo images sets of the sewer pipes as obtained with the experimental setup. Subfigure 5 shows a typical naturally aged cylindrical pipe, containing plenty of texture to use for the stereo matching. Subfigure 5b shows a typical naturally aged egg-shaped pipe, the reason we need a non-cylindrical model. Subfigure 5c shows a new cylindrical pipe, lacking sufficient texture to accurately stereo match.
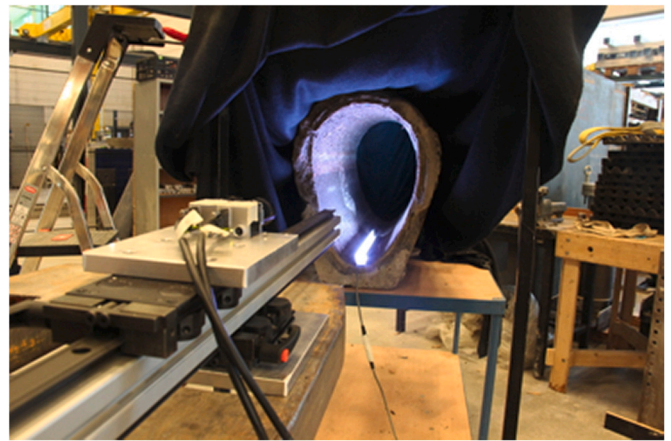

Fig. 4. Experimental setup (not the final lighting setup).

5.2. Implementation details and parameters

The framework was implemented in OpenCV 4.5 [43] and Python 3.6 [44]. Using the default implementation of semi-global stereo matching in OpenCV, we chose a search range between 20 and 220 pixels of disparity, used a block size of 7, $P_1 = 100$, $P_2 = 10,000$, and a uniqueness ratio of 10.

After stereo matching and geometry reconstruction, the red cloth background of the images is removed automatically with a flood fill and a valid Z-range can be selected by the user, or a default of $Z \in [1.5, 2.0]$ meters from the camera may be used.

After conversion to cylindrical coordinates, we fit the model in eq. (17) with $K = 6$. RANSAC is run for 10 iterations, the initial fit is calculated on 50 randomly selected datapoints, inliers are determined by a maximum absolute difference of 0.005, and the second fit is calculated on the inliers if those inliers make up at least 90% of the datapoints. The fit with the lowest error on its inliers is chosen, or if no initial fit had enough inliers, the entire process is repeated with 10 iterations.

The best fit from the RANSAC model is applied to all data points, including those outside the Z-range that the model was fit on, and the deviation from the fit is presented for visual inspection in both a point cloud and the reference image.

The code of our implementation is available to try as a demo. It can be found at:

`https://github.com/data-flux/StereoDemo`[3]

## 6. Results and discussion

6.1. Stereo matching and geometry reconstruction

Let us start our discussion of the results by considering the first half of our approach, the stereo matching and geometry reconstruction. Getting an objective, unequivocal assessment of the produced point cloud is challenging, since we do not have a golden standard measurement of the 3D pipe geometry to compare the point clouds with. Because of this, we will mostly have to rely on subjective validation of the results. We asked a human assessor to judge each point cloud on how accurate and consistent the point cloud reconstructed the original 3D geometry, on a scale from 1 (very poor) to 10 (very accurate). Over the 48 image sets, an average rating of 8.4 was assigned, indicating that the 3D reconstruction was quite good. In Fig. 6.1, four examples of different types of pipes are given where the reconstruction was successful
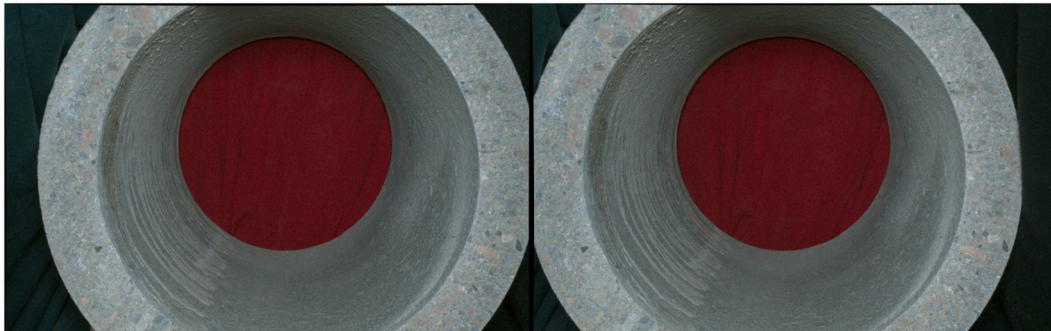
---

[3] Note to reviewers: This repository is set to private, pending approval of our funding committee, please find the demo among the supplemental materials

(a) A typical naturally aged cylindrical sewer pipe.
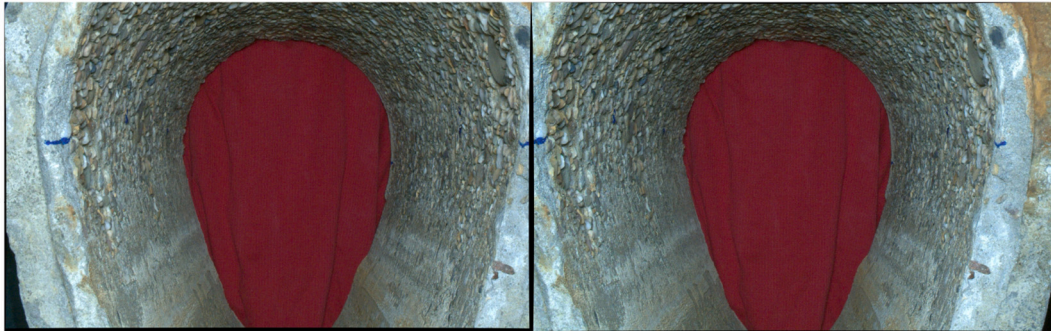
(b) A typical naturally aged egg-shaped pipe.

(c) A typical new cylindrical sewer pipe.

Fig. 5. Three examples of stereo image sets as obtained with the experimental setup. All images have lens distortion correction, the left images also have a vertical translation correction according to the process described in Section 4.1.

(average rating of 8.5). The left picture shows the left image of each stereo pair, and the right image shows the point cloud with points colored by the color from the original (left) image. The characteristics of the virtual lens of the generated image deviate from the physical lens somewhat, and the virtual camera was deliberately placed somewhat further ahead. This allows the viewer to somewhat appreciate the 3D nature of the point cloud (rather than reproducing the images on the left without knowing the depth). Specifically, the forward camera position allows observing any occluded areas by 'seeing around' the humps on the pipe surface. Especially in the point cloud image on the first row of Fig. 6, areas of occlusion behind the various deposits are clear. Viewed from the side, this point cloud has considerable gaps behind all deposits.

A minority of the pipes were not reconstructed perfectly. Five image sets (from four pipes) scored a value below 8 (average score 6), with the lowest being two scores of 5. What these five image sets have in common, compared to the remaining successful image sets, is that they involve pipes with areas of smooth and monochrome surface, as can be seen in Fig. 7. Such areas are especially common in relatively new pipes which do not show much deterioration. The surface of such pipes will be hard to stereo-match, since few surface features can be used to match

pixels in the stereo pair. The result is that entire patches of pixels have an undetermined depth. Further contributing factors to this poor matching are low lighting and lack of lens focus. With improved focus and lighting, the tiniest surface features will lead to proper matching again.

6.2. Surface fitting and anomaly detection

Next, we consider the quality of the surface fitting and subsequent anomaly detection steps of our framework on the 48 image sets. In Fig. 8, the four pipes of Fig. 6 are shown again, with the local anomaly score assigned to the derived point cloud. As can be seen, most of the clear deposits present are identified by the anomaly detection method. In the first row, all of the clear deposits on the left-hand side have been identified, and also some of the less obvious deposits on the right can be made out, for example on the low-right at half-depth. Additionally, some of the surface roughness is indicated at the top. Note also the large occlusion areas, which do not play a role in our detection algorithm, but further strengthen the identified anomalies. The pipe in the second row shows some deposits hanging from the top of the pipe, as well as considerable unevenness in the surface texture throughout the pipe that
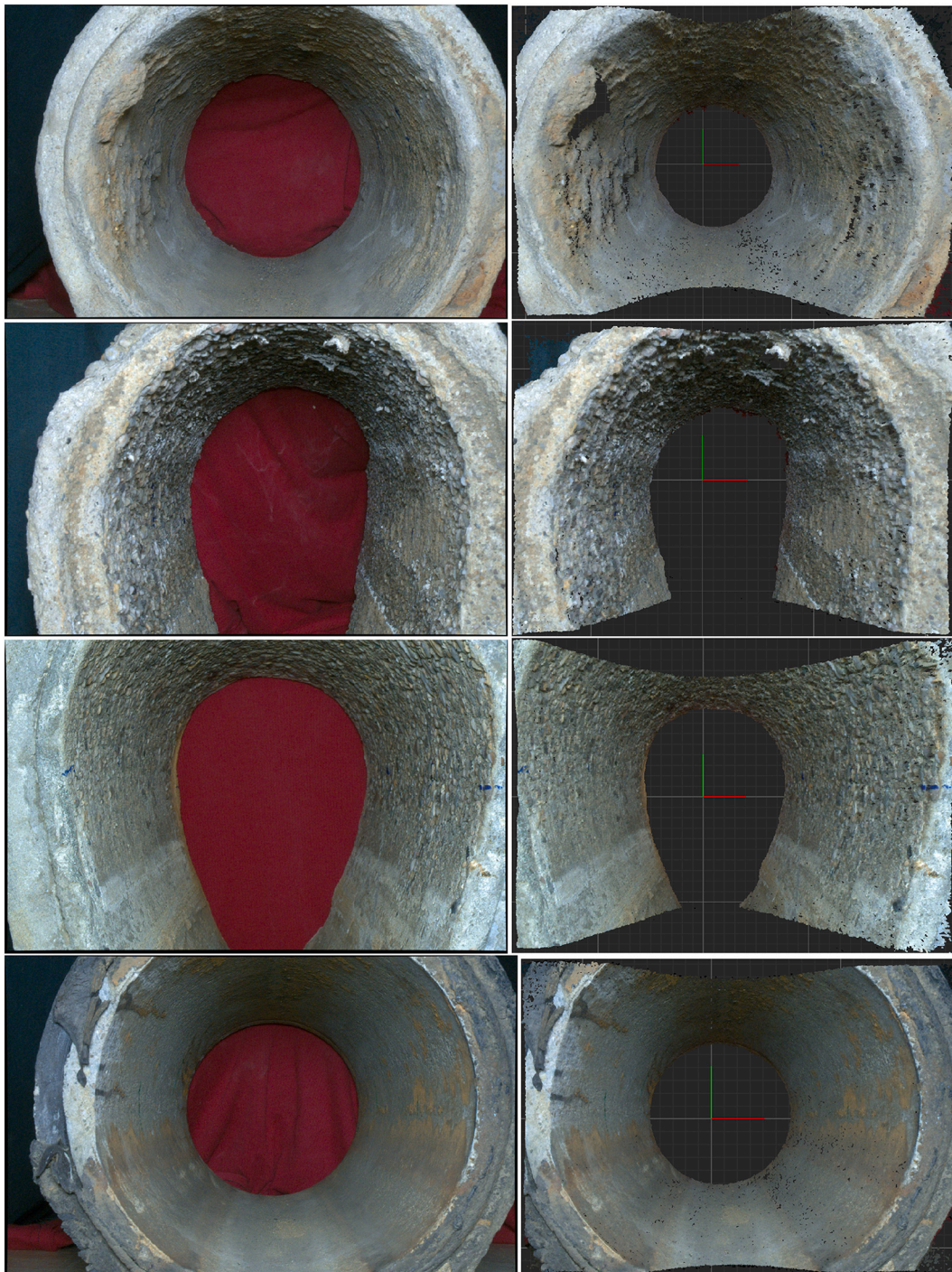
Fig. 6. Four examples of the stereo matching. Left shows the reference image, right shows the reconstructed point cloud.

is also captured by the framework as local anomalies. In this pipe, however, we also see evidence of some false positives in the detection, in the lower far corner of the pipe. It appears that here, the point cloud contains too little data in the lower regions (closer to the camera, the field of view does not contain the bottom of the pipe), such that the model is possibly inaccurate. In this particular area, there indeed appear to be some deposits around the flood line, but not to the extent indicated by the model. This phenomenon, that also plays a small role in row three, appears to be an artefact of the (in hindsight somewhat unfortunate) choice of lens that doesn't allow a full view of the pipe. The problem is easily remedied by removing the near and far end of the pipe and focussing on the middle band of the point cloud, which incidentally

also concerns the pixels with the best focus. Remember that in the field, the camera will be slowly inched forward through the pipe, allowing for a complete sweep of the pipe. Our framework can thus focus on the band of data where results are the most reliable.

Of the 48 image sets, a total of six images could not be fitted properly with our RANSAC method (not reaching the required fraction of inliers). Five of these concern the cases mentioned in the previous section as suffering from a poor stereo matching. The resulting point cloud has a non-negligible number of points that erroneously lie inside the pipe, reducing the number of inliers. The sixth pipe that could not be fitted properly had an entire section broken off, probably during extraction from the soil. The remaining 42 cases (87.5%) were properly fitted,
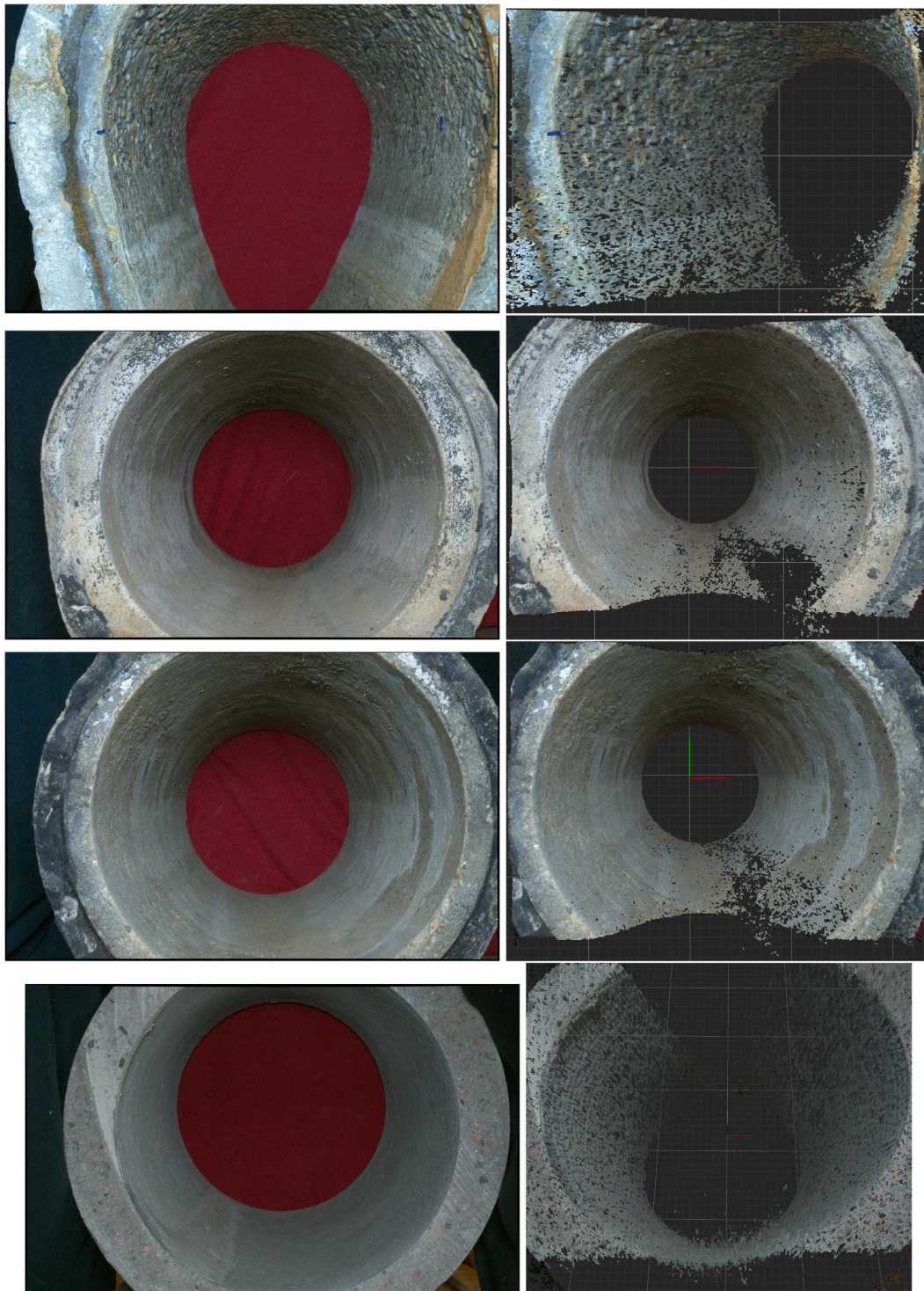
Fig. 7. Images of pipes that had problems with stereo matching, due to patches of pixels with mostly constant surface color. Especially in areas where along a horizontal line there is little color variation, this might lead to considerable confusion in the matching. For the top-left image, this confusion is limited to the lower right-hand corner near the flood line. For the other images, the problems occur in the lower half of the images, that are overly smooth.

producing the marked point clouds demonstrated in Fig. 8, as well as a single global anomaly score per image, as defined in eq. (18). Prior to validating the anomaly detection, the photographs of the pipes were also graded subjectively in terms of defect severity, to have an independent ground truth to compare the detected anomalies levels with. For the 42 cases where our framework produced a global anomaly score, a Pearson's correlation of $r = 0.65$ with the ground truth was obtained, which is, according to the customary interpretation, a moderate, positive

correlation.

6.3. Discussion

Although our framework demonstrates good results on the two defect types of exposed granulate and deposits, not all pipes are correctly assessed. The main weakness of our method appears to revolve around the images with smooth pipe surfaces. The problem with these pipes
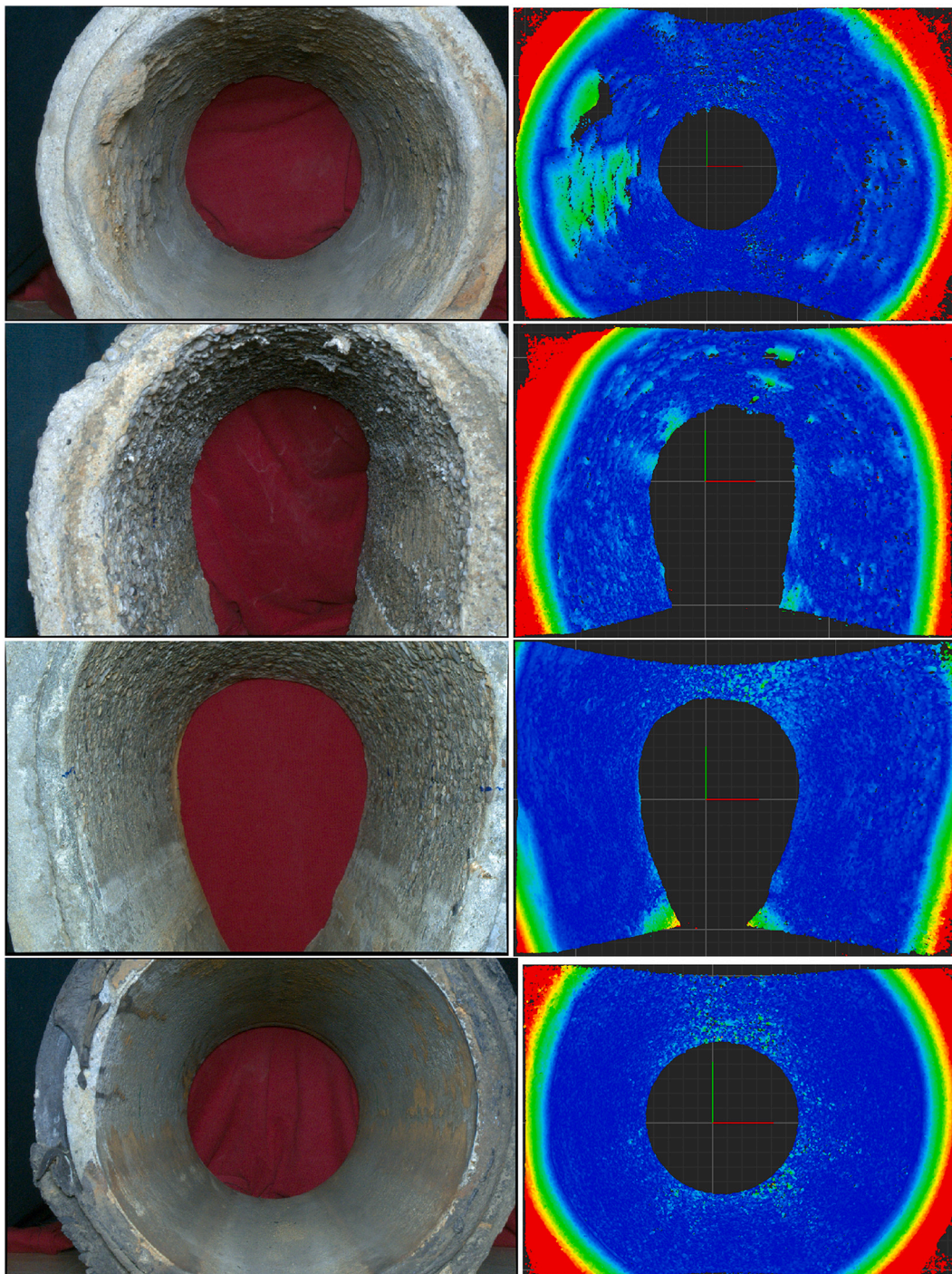
Fig. 8. Examples of four anomaly scores. Left shows the reference image, right shows the identified anomalies. Note that the anomalies on the edges and outside of the pipes are ignored.

occurs in the first part of our framework, the stereo matching, which indeed is known to be problematic in images with limited texture. The upside of this limitation is that pipes with smooth surfaces (often fairly new pipes) typically do not contain any defects. The main problem here is hence one of false alarms: the method sometimes erroneously identifies defect in smooth pipes because points are incorrectly placed in the 3D space. In future work, we will investigate whether stereo matching could also produce a confidence score, indicating the quality of the stereo matching in each region of the image. If successful, the method would only identify a defect if both the confidence is high (in other words, not a smooth surface) and the anomaly score is high.

Whenever the stereo matching produces at least a reasonable result, the surface fitting and anomaly detection correctly identifies the various defects present. It should be noted that the framework even identified defects that were overlooked in the initial subjective quality grading of single images (not stereo pairs). In that sense, other than merely automating parts of the inspection process, our framework also has the potential to outperform human inspectors in certain respects.

Our experiments were performed in the lab, which may have had some minor effects on the outcome, although both in a positive and a negative sense. On the positive side, our experiments were perhaps made more challenging than necessary due to some initial choices that

ended up being suboptimal. For example, the chosen lenses were not of sufficiently wide angle, such that in the region of focus not the entire pipe could be captured, especially for egg-shaped pipes. The effect of this on the results is that for some regions of the pipe, such as the corners of the point clouds corresponding to the edges of the image, not enough evidence of the regular geometry is available in order to reliably decide on deviations from that geometry. This effect, which may cause both false positives and false negatives locally, can be observed in the third row of Fig. 8. This problem can be easily corrected by taking a wider lens. Another unfortunate choice concerns the low lighting conditions, which could have been corrected with a longer exposure of the images. Low lighting has not played a significant role (as the mostly good results demonstrate), but may have contributed to the lack of matching in areas with smooth surfaces.

Another property of the lab set-up was that we search for anomalies in the entire pipe, which is not necessary in the field, and produces some challenges with limited focal depth and missing parts of the pipes (due to extraction). In a more practical setting, in an actual in-situ pipe, a single image pair would only be inspected for a somewhat narrow band, corresponding to the region of focus. After that, the pipe inspection vehicle would move forward by a small distance and the process would be repeated. Note that although our framework makes no such assumption, being in pipes of mostly the same geometry would allow one to assume a certain steady geometry, and deviations from this could be more easily recognised.

On the negative side, our lab set-up is less realistic as we only inspect individual pipe segments, not longer pipe systems. As a result, we have no images of pipe joints, which in most cases would produce a (false positive) deviation from our fitted surface. Although we expect the pipe joints to be easily matched by the stereo matching, and proper handling of these 'acceptable anomalies' would not be difficult, our current method does not include such facilities, nor have we been able to test this. As future work, it would be interesting to develop a method that uses 3D point clouds of joints in order to recognise joint-related defects such a misaligned joints or signs of leakage. A final difference between our setting and actual sewer systems is that the inspected pipes, after having been washed out, will be wet, whereas our lab pipes were recorded in a dry state. We do not expect this difference to have a significant effect on the results, but would need to set up new experiments to assess this.[4]

In future experiments it may be interesting to look at discontinuities in the points on the pipe surface as well, as these may be caused by occlusions. An occlusion of a portion of the pipe surface may be just as informative as the geometry of the visible parts of the pipe surface, but could be indicative of foreign objects present in the pipe, such as roots. However, because poor stereo matching may also result in such discontinuities, it may be advisable to require that such discontinuities exist only in the point cloud, and not in the disparity map.

7. Conclusion

7.1. Summary

In this work we have proposed RADIUS, an anomaly detection framework for sewer pipes based on computer stereovision. The framework consolidates several successful techniques into a sequential process, to allow for anomaly detection in an automated fashion from stereo photographs without intermediate user input. We performed experiments to demonstrate the efficacy of the framework and conclude that it is successful in detecting defects present in physical pipes as anomalies in the three-dimensional geometry, and moves the state of the art closer towards fully automated sewer asset management.

7.2. Limitations

The major limitation of this work is the varying quality of the data obtained in the lab, as a result of our limited experience with the hardware. Some of the images were made in poor lighting conditions and without proper camera calibration. Repetition of the experiments was not possible due to time and budget constraints, and because a part of the pipes have been subjected to destructive full-scale testing in other experiments. A secondary limitation related to this is the total number of experiments performed.

A more intrinsic limitation of the approach of stereovision is that smooth, undamaged concrete may not contain enough texture or markers to accurately match the reference image to the secondary image. While this is potentially an issue for a large portion of sewer pipes, we argue that such regions are unlikely to contain any anomalies. That said, since the absence of a match may also be an indicator of occlusion, we advise authors of future research to distinguish causes of a lack of a match: in the case of a too smooth pipe, the cause is likely a match that does not meet the uniqueness ratio required, while still having a relatively high matching score, as opposed to a patch in the reference image that does not appear in the secondary image due to occlusion.

In spite of these limitations, we feel the efficacy of the framework has been more than adequately demonstrated.

7.3. Recommendations

While anomaly detection may be a goal in itself, the authors hold the view that it is a stepping stone towards fully automated sewer condition assessment. To this end, we recommend future research to be performed into a follow-up step for the proposed framework: automated classification of the anomalies into defect classes. We feel that (the deviation from) the surface found through robust regression has a lot of potential for classification, as it will theoretically contain very little noise, as well as have a notion of "expected" behaviour.

It must be noted that while we have shown stereovision to be a viable tool for sewer pipe defect classification, the added value in practical settings has yet to be demonstrated. We have designed this framework to be compatible with current inspection practices: (monovision) CCTV inspection can still be performed while collecting data from two camera sources for stereovision experiments. This data may be used in parallel to both inch the industry towards automation of inspections, as well as to improve manual inspection techniques with an additional mode of data. State-of-the-art sewer defect detection solely based on CCTV data may suffer from a relatively large false positive rate [3], but the additional depth information provided by an additional camera could lower this significantly. While the ambition of automated inspection is currently en vogue (again), the added value of multi-sensor inspection for more reliable, precise, and complete detection of a range of observable sewer defects, is an important added value worth researching further.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

---

[4] note to reviewers: the experiments were performed prior and during the COVID-19 lock-down, and some of the pipes were subsequently part of destructive full-scale testing. As such, we cannot easily redo the experiments.

Scientific Research), the RIONED Foundation, STOWA (Foundation for Applied Water Research) and the Knowledge Program Urban Drainage (KPUD).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.autcon.2022.104285.

References

[1] TISCA programme funded by NWO-TTW, Sewersense – Multi-Sensor Condition Assessment for Sewer Asset Management, 2016–2020, URL: https://www.nwo.nl/en/projects/15343 (last accessed on 2022-01-07).

[2] F. Tscheikner-Gratl, N. Caradot, F. Cherqui, J.P. Leitão, M. Ahmadi, J.G. Langeveld, Y. Le Gat, L. Scholten, B. Roghani, J.P. Rodríguez, M. Lepot, B. Stegeman, A. Heinrichsen, I. Kropp, K. Kerres, M. do Céu Almeida, P.M. Bach, M. Moy de Vitry, A. Sá Marques, N.E. Simões, P. Rouault, N. Hernandez, A. Torres, C. Werey, B. Rulleau, F. Clemens, Sewer asset management–state of the art and research needs, Urban Water J. 16 (2019) 662–675, https://doi.org/10.1080/1573062X.2020.1713382.

[3] D. Meijer, L. Scholten, F. Clemens, A. Knobbe, A defect classification methodology for sewer image sets with convolutional neural networks, Autom. Constr. 104 (2019) 281–298, https://doi.org/10.1016/j.autcon.2019.04.013.

[4] H. Hirschmüller, Accurate and efficient stereo processing by semi-global matching and mutual information, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 2, IEEE, 2005, pp. 807–814, https://doi.org/10.1109/CVPR.2005.56.

[5] M.A. Fischler, R.C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography vol. 24, ACM New York, NY, USA, 1981, pp. 381–395, https://doi.org/10.1145/358669.358692.

[6] A.S. Goldberger, Econometric Theory, John Wiley & Sons, New York, 1964 (ISBN: 0-471-31101-4).

[7] M. Lepot, N. Stanić, F.H. Clemens, A technology for sewer pipe inspection (part 2): experimental assessment of a new laser profiler for sewer defect detection and quantification, Autom. Constr. 73 (2017) 1–11, https://doi.org/10.1016/j.autcon.2016.10.010.

[8] N. Stanić, F.H. Clemens, J.G. Langeveld, Estimation of hydraulic roughness of concrete sewer pipes by laser scanning, J. Hydraul. Eng. 143 (2017) 04016079, https://doi.org/10.1061/(ASCE)HY.1943-7900.0001223.

[9] O. Duran, K. Althoefer, L.D. Seneviratne, Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network, volume 4, IEEE (2007) 118–126, https://doi.org/10.1109/TASE.2006.873225.

[10] C.H. Bahnsen, A.S. Johansen, M.P. Philipsen, J.W. Henriksen, K. Nasrollahi, T.B. Moeslund, 3d sensors for sewer inspection: a quantitative review and analysis, Sensors 21 (2021) 2553, https://doi.org/10.3390/s21072553.

[11] A. Ahrary, L. Tian, S.-i. Kamata, M. Ishikawa, An autonomous sewer robots navigation based on stereo camera information: 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05), IEEE, 2005, pp. 628–633. doi: https://doi.org/10.5555/1105924.1106125.

[12] A. Ahrary, M. Ishikawa, A fast stereo matching algorithm for sewer inspection robots vol. 3, Wiley Online Library, 2008, pp. 441–448, https://doi.org/10.1002/tee.20297.

[13] W. Koodtalang, T. Sangsuwan, B. Noppakaow, A design of automated inspections of both shape and height simultaneously based on stereo vision and plc, in: 2018 18th International Conference on Control, Automation and Systems (ICCAS), IEEE, 2018, pp. 1290–1294. URL: https://ieeexplore.ieee.org/document/8571609 (last accesson on 2022-01-07).

[14] A. Gunatilake, L. Piyathilaka, S. Kodagoda, S. Barclay, D. Vitanage, Real-time 3d profiling with rgb-d mapping in pipelines using stereo camera vision and structured ir laser ring, in: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2019, pp. 916–921, https://doi.org/10.1109/ICIEA.2019.8834089.

[15] P. Huynh, R. Ross, A. Martchenko, J. Devlin, Anomaly inspection in sewer pipes using stereo vision, in: 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), IEEE, 2015, pp. 60–64, https://doi.org/10.1109/ICSIPA.2015.7412164.

[16] P. Huynh, R. Ross, A. Martchenko, J. Devlin, 3d anomaly inspection system for sewer pipes using stereo vision and novel image processing, in: 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2016, pp. 988–993, https://doi.org/10.1109/ICIEA.2016.7603726.

[17] D. Meijer, M. Kesteloo, A. Knobbe, Unsupervised anomaly detection in sewer images with a PCA-based framework, in: International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI), 2018, pp. 354–359. URL:

[18] J. Myrans, R. Everson, Z. Kapelan, Automated detection of faults in sewers using cctv image sequences, Autom. Constr. 95 (2018) 64–71, https://doi.org/10.1016/j.autcon.2018.08.005.

[19] J. Myrans, Z. Kapelan, R. Everson, Using automatic anomaly detection to identify faults in sewers, in: Water Distribution Systems Analysis (WDSA) / Computing and Control for the Water Industry (CCWI) Joint Conference Proceedings vol. 1, 2018. URL: https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12030 (last accessed 2022-01-07).

[20] S. Moradi, T. Zayed, F. Nasiri, F. Golkhoo, Automated anomaly detection and localization in sewer inspection videos using proportional data modeling and deep learning–based text recognition, J. Infrastruct. Syst. 26 (2020) 04020018, https://doi.org/10.1061/(ASCE)IS.1943-555X.0000553.

[21] X. Fang, W. Guo, Q. Li, J. Zhu, Z. Chen, J. Yu, B. Zhou, H. Yang, Sewer Pipeline Fault Identification using Anomaly Detection Algorithms on Video Sequences vol. 8, IEEE, 2020, pp. 39574–39586, https://doi.org/10.1109/ACCESS.2020.2975887.

[22] Stefania Russo, Andy Disch, Frank Blumensaat, Kris Villez, "Anomaly detection using deep autoencoders for in-situ wastewater systems monitoring data". Anomaly detection using deep autoencoders for in-situ wastewater systems monitoring data, Copenhagen, Denmark, 2019.

[23] J.B. Haurum, T.B. Moeslund, A survey on image-based automation of cctv and sset sewer inspections, Autom. Constr. 111 (2020), 103061, https://doi.org/10.1016/j.autcon.2019.103061.

[24] N. Stanić, J.G. Langeveld, F.H. Clemens, Hazard and operability (hazop) analysis for identification of information requirements for sewer asset management, Struct. Infrastruct. Eng. 10 (2014) 1345–1356, https://doi.org/10.1080/15732479.2013.807845.

[25] J. Dirksen, F. Clemens, H. Korving, F. Cherqui, P. Le Gauffre, T. Ertl, H. Plihal, K. Müller, C. Snaterse, The consistency of visual sewer inspection data, Struct. Infrastruct. Eng. 9 (2013) 214–228, https://doi.org/10.1080/15732479.2010.541265.

[26] M. Levoy, T. Whitted, The Use of Points as a Display Primitive, URL: https://graphics.stanford.edu/papers/points/, 1985.

[27] D.M. Finlayson, B. Sinclair, Advances in Lasers and Applications, CRC Press, 1999 (ISBN: 9780750306324).

[28] J.G.D. Frana, M.A. Gazziro, A.N. Ide, J.H. Saito, A 3d scanning system based on laser triangulation and variable field of view, in: IEEE International Conference on Image Processing 2005 vol. 1, IEEE, 2005, pp. I–425, https://doi.org/10.1109/ICIP.2005.1529778.

[29] W. Linder, Digital Photogrammetry vol. 1, Springer, 2009 (ISBN: 978-3-540-92725-9).

[30] S.-H. Ni, Y.-H. Huang, K.-F. Lo, D.-C. Lin, Buried pipe detection by ground penetrating radar using the discrete wavelet transform, Comput. Geotech. 37 (2010) 440–448, https://doi.org/10.1016/j.compgeo.2010.01.003.

[31] D.-H. Koo, S.T. Ariaratnam, Innovative method for assessment of underground sewer pipe condition, Autom. Constr. 15 (2006) 479–488, https://doi.org/10.1016/j.autcon.2005.06.007.

[32] I.P. Howard, B.J. Rogers, Perceiving in Depth, Volume 2: Stereoscopic Vision, Oxford University Press, 2012 (ISBN: 9780199764150).

[33] R. Szeliski, Computer Vision: Algorithms and Applications, Springer Science & Business Media, 2010 (ISBN: 9781848829350).

[34] E. Hecht, Optics vol. 5, 2002. Addison Wesley San Francisco. (ISBN: 0133977226).

[35] A. Zimek, E. Schubert, Outlier Detection, Springer, New York, New York, NY, 2017, pp. 1–5 (ISBN: 978-1-4899-7993-3).

[36] Y. Wu, S. Jiang, Z. Xu, S. Zhu, D. Cao, Lens distortion correction based on one chessboard pattern image, Front. Optoelectronics 8 (2015) 319–328, https://doi.org/10.1007/s12200-015-0453-7.

[37] L.G. Brown, A survey of image registration techniques, ACM computing surveys (CSUR) 24 (1992) 325–376, https://doi.org/10.1145/146370.146374.

[38] H. Hirschmüller, S. Gehrig, Stereo matching in the presence of sub-pixel calibration errors, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 437–444, https://doi.org/10.1109/CVPR.2009.5206493.

[39] S. Birchfield, C. Tomasi, A Pixel Dissimilarity Measure that is Insensitive to Image Sampling vol. 20, IEEE, 1998, pp. 401–406, https://doi.org/10.1109/34.677269.

[40] N. Lazaros, G.C. Sirakoulis, A. Gasteratos, Review of stereo vision algorithms: from software to hardware, Int. J. Optomechatronics 2 (2008) 435–462, https://doi.org/10.1080/15599610802438680.

[41] N. Salvaggio, Basic Photographic Materials and Processes, Taylor & Francis, 2009 (ISBN: 9781138744370).

[42] J. Stewart, Calculus, 7th ed., Brooks/Cole, Cengage Learning, Belmont, CA, 2012 (ISBN: 0538497815).

[43] G. Bradski, A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library, O'Reilly Media, Inc., 2008 (ISBN: 0596516134).

[44] G. Van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009 (ISBN: 1441412697).

https://users.encs.concordia.ca/icprai18/ICPRAI2018 Proceedings.pdf (last accessed on 2022-01-07).