SIViP (2013) 7:865–871
DOI 10.1007/s11760-011-0275-z

ORIGINAL PAPER

Photo-mosaicing of images of pipe inner surface

Yuri Rzhanov

Received: 1 September 2011 / Revised: 18 October 2011 / Accepted: 19 October 2011 / Published online: 2 November 2011

Abstract This paper describes the algorithm for the construction of continuous visually consistent images of the inner surface of a pipe from a sequence of images acquired by a wide-angle camera that traveled inside the pipe. The algorithm is designed to be a proof of concept and performs well on simulated data (rendered images) even when camera poses (attitude and location) have errors as much as 5%. Photo-mosaics are suitable for traditional (visual) inspection or automatic processing for the detection of manufacturing faults, corroded areas, and cracks. It is demonstrated that the quality of the resulting mosaic depends how the camera is oriented with respect to the pipe axis and that the traditional orientation with an almost collinear camera optical axis and the pipe axis is not the optimal choice. The proposed system is useful for inspection of pipelines that cannot accommodate traditional devices (e.g., pipeline inspection gauges or crawlers), for example, small-scale boilers and gas systems.

Keywords Mosaicing · Stereoscopic processing · Pipe inspection · Monocular stereoscopy

1 Introduction

There are millions of miles of pipeline in the USA. These include only large-diameter pipelines that transport water, oil, gas, etc. Counting small local pipes, for example, in boilers and in chemical factories, total length of pipes will probably double or triple. All pipelines have to be regularly inspected for corrosion, cracking, and manufacturing flaws. Devices called "pigs" (pipeline inspection gauges) have been developed for inspection and cleaning of large pipelines. Typical techniques for inspection are magnetic flux leakage and ultrasound. In some cases, remote visual devices are used that crawl along the pipe and either record video data on a local hard drive or send live video feed to the operators via cable. For example, the company Envirosight (http://www.envirosight.com) produces DigiSewer—a wheeled crawler that constantly acquires images in 5-cm-wide strips and then assembles a continuous flat scan image of pipe interior for visual inspection.

"Pigs" and crawlers work well for nearly straight pipes with diameter large enough to fit such a device. Structures with small diameter pipes and sharp turns (e.g., in many boiler systems) require a different approach. Typically, these small structures are inspected using a small-size wide-angle camera and a light source attached to a flexible cable that is pushed or pulled through the pipe. A video stream is sent via the cable to external monitors for recording and visual inspection. Typical devices for this sort of inspection cost a few thousand dollars and have cable lengths up to 100 m (see, for example, http://aitproducts.com/pipe-inspection-camera). Obviously, it is desirable to create a continuous picture of the pipe interior to give the operator a better perception of the conditions inside the pipe and provide significant savings in storage of recorded data. Typical recorded visual imagery can be found in a YouTube clip (http://www.youtube.com/watch?v=EmwqPvjDb0k). The downside of this approach is necessity to constantly watch a video stream (in real time or afterward), the inability to accurately measure the length traveled by the camera head, and often incomplete coverage of the pipe interior surface.

This paper describes an algorithm that generates a continuous image of a pipe interior surface from video or still images acquired by a cheap version of an inspection system

Y. Rzhanov (✉)
Center for Coastal and Ocean Mapping, University of New Hampshire, 24 Colovos Road, Durham, NH, 30824, USA
e-mail: yuri.rzhanov@unh.edu

when a "pig" or a crawler cannot fit in a pipe. It is assumed that the construction of the video head provides adequate lighting and nearly full coverage of the pipe interior surface (for example, using a flexible head with fish-eye lenses). It is also assumed that the acquired images are texture-rich. Camera lenses must be externally calibrated, so that corrections for lens-related distortion can be made, and the images submitted for processing can be considered acquired by a pin-hole camera. A typical calibration of such a camera with a field of view significantly wider than 180° is described in [1]. The acquisition frame rate is coordinated with the camera advance along the pipe to guarantee substantial overlap ($\geq$60%) between any sequential images.

Simulated images are used for input for two reasons. First, it was found to be impossible to find video data properly collected, i.e., specifically for mosaicing purposes. Second, simulated data allows for direct comparisons with the ground truth and estimations of associated errors.

2 Theoretical background

The present work makes use of a number of previously published results that will be only mentioned and referenced below for the sake of brevity. The first one is the work of Lowe [2] that proposed a technique for the detection of salient points in digital imagery with subpixel accuracy and descriptors robust with respect to rotation and scale changes. The second is a paper by Jokinen [3] who suggested an elegant solution for fitting point cloud to a cylindrical surface thus recovering cylinder's diameter, location and orientation. The last but not least is the fundamental book by Hartley and Zisserman [4] summarizing latest developments in multi-view geometry and computer vision.

3 Algorithm

The processing steps are as follows:

1. Frames are corrected for lens distortion. This step is unnecessary in the simulation. Instead, the frame acquisition is simulated in a setup shown in Fig. 1 by means of ray-tracing.
2. Salient point features (keypoints) are extracted from a pair of sequential images (acquired respectively by "first" and "second" cameras) using SIFT [2], SURF [5], or a simpler Harris operator-based [6] algorithm. Keypoints from different images are matched pair-wise. Matching procedures could be based on a comparison of SIFT-like [2] or triplet vector [7] descriptors, or on a maximization of a normalized cross-correlation to guarantee a low percentage of outliers even in situations when illumination conditions change from frame to frame. The experiments for determination of the best-performing features have not been conducted. All the above-mentioned techniques worked sufficiently well to achieve the final goal. SIFT algorithm usually provides more salient features than other approaches due to its robustness with respect to rotation and changes in scale. The reported results have been obtained by the two-step feature matching: at first, SIFT-like descriptors were used, and then, matches with the normalized cross-correlation score less than 0.7 were rejected.
3. Two frame images with a partial overlap display the same features, and hence, these two views (or two cameras, as they are typically referred to in the computer vision community) are related. Extracted set of matches allow for the recovery of the fundamental matrix [4]. No matching procedure can guarantee a complete absence of outliers (incorrect matches), so, for a determination of the fundamental matrix, the robust RANSAC-based procedure [8] is employed, with number of iterations dependent on estimated percentage of outliers.
4. The optimal fundamental matrix is decomposed in a rotation matrix that describes the change in camera attitude between the first and second views, and a translation vector—shift of the camera position. Decomposition can only be obtained with a 4-fold ambiguity; the procedure retrieves two rotation matrices, and the translation vector is determined up to a sign and an arbitrary scale. Any ambiguity related to the sign and the choice of a rotation matrix can be removed utilizing the cheirality constraint, the condition that all the 3D points that correspond to a projection on the retinal planes are in fact in front of both cameras [4]. The only free parameter that remains is the scaling factor for the inter-camera translation.
5. Pair-wise matches are triangulated in 3D space, and a recovered point cloud is fitted to a cylindrical surface [3]. The fitting procedure can be extended to a gradually bending cylinder with an arbitrary, but known, radius of bending. In the case of a straight cylinder, the regression procedure provides a vector that defines the orientation of the cylinder axis, point in 3D space through which the axis passes (in the system of coordinates of the first camera pose), and cylinder's diameter (in the same units as the translation vector).
6. Because the inner diameter of the inspected pipe D is known *a priori*, the scaling factor for the translation vector can be determined. Converting the system of coordinates to the pipe-centric coordinate system (main axis along Z-coordinate, X and Y perpendicular to the pipe surface, see Fig. 1) allows the determination of poses (position and orientation) of both cameras. The choice of the new system of coordinates leaves one free parameter (angle of rotation of the pipe about its axis), and,

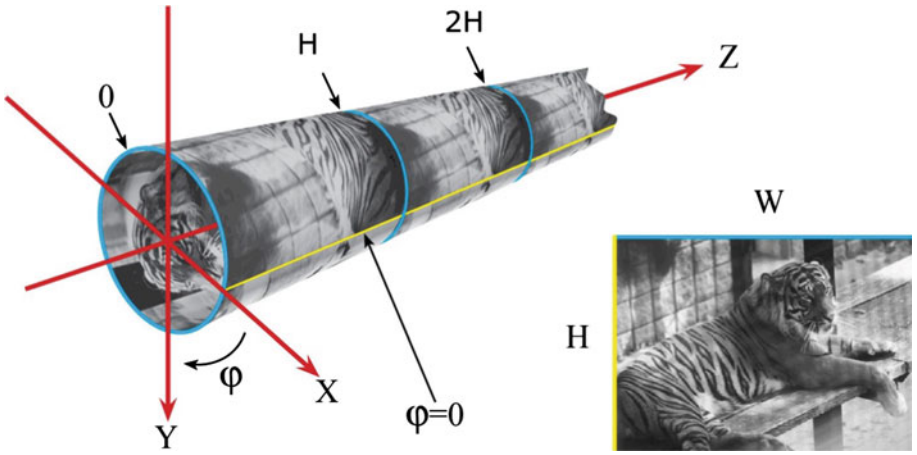

Fig. 1 Pipe-centric geometry. The *bottom-right* image is used as a wrapper for an inner surface of a pipe. Texture on a pipe surface is periodic along Z (repeats every H pixels) and has a seam at $\varphi = 0°$

for convenience, it is chosen such that the X-coordinate of the first camera is 0 (first camera is located directly below the pipe axis, $\varphi = 90°$.

7. Images acquired by the cameras (input images) are reprojected onto the pipe interior surface. The resolution of the reprojected images can be chosen arbitrarily. In the simulation, it is reasonable to choose the resolution such that it is close to that of the input images in the near field. This way, no information is lost in the process of reprojection. For example, if the typical dimension of the input image is 360 pixels, then it is sensible to assign one pixel to one degree angle in the pipe-centric cylindrical coordinates. The vertical dimension is oriented along Z axis (pipe axis), and the horizontal dimension follows the pipe surface, so that the reprojected image is "wrapped" around the pipe's inner surface. One pixel of reprojected image in the vertical dimension corresponds to $\pi * D/360$ in real units. To maintain square pixels in the reprojections, the pipe length W_z in Z direction should represent $W_z/(\pi * D/360)$ pixels.
8. "Wrapping" images from first and second cameras differ only by shifts in the horizontal (along the pipe) and vertical (rotation about the pipe axis) directions and a change of scale, which depends on the distance from the camera to the surface being imaged. There is no rotation between the images due to specifics of the reprojection procedure. The shifts and scale factor can be determined without repeating the matching process. Instead, matched pairs of point features extracted from the input images are reprojected onto the pipe inner surface and then converted to non-integer pixel locations on the "wrapping" images. The determination of the three parameters requires only two pair-wise matches. Usually, many more matches are available, thereby allowing formulation of an over-determined system of equations from which the parameters can be found in a least squares sense. The most accurate results are obtained when the contributions from the different matches are weighted, with the weight inversely proportional to the distance from the camera to the point projection.

Steps 1–8 are applied consecutively to all sequential pairs of input images, which lead to the construction of a continuous mosaic of inner surface of the inspected pipe.

4 Discussion of experimental setup

Position and orientation of a camera does not affect the mosaicing process. However, for more complete coverage of a pipe inner surface, the camera should take images being as close to the pipe wall as possible and look at the opposite wall. In the case of a wide-angle camera, this setup proved to deliver better results than that with a coaxial camera located near the pipe axis. The latter case leaves no blind zones, but the pipe surface can be seen in the near range only at the periphery of acquired images (see Fig. 4b), and the least distorted central part of the image is "wasted" on a low resolution poorly illuminated area. Low-resolution imagery leads to pixellation effects that introduce errors in localization of point features in the acquired images. This results in inaccurate determination of a pipe diameter (in camera translation units), which in turn leads to erroneous estimation of parameters relating consecutive "wrapping" images (shifts and scaling factors). The errors are less significant in the near range where pixellation effects are less pronounced.

The described algorithm essentially flattens the cylindrical surface, and thus, ghosting artifacts do not occur if the camera has been calibrated and a fundamental matrix has been determined without any error. Weak remaining artifacts are dealt with at the mosaic-building stage by using graph-cut optimization (as in, for example [9]).

5 Simulation

Two scenarios have been simulated. In both scenarios, two cameras were positioned at known locations and with known

poses with respect to the system of coordinates associated with the pipe of known diameter (pipe-centric system). The cameras had no lens-related distortion (pin-hole cameras). In the first scenario, a set of points was generated at random locations on the inner surface of the pipe. Only the points visible by both cameras were selected and projected onto the retinal planes (Fig. 2). At this stage, point locations were contaminated with some noise to simulate (a) pixellation (or rasterization) effects and (b) inaccurate localization of detected keypoints during the feature extraction procedure. Known pair-wise matches with no outliers were used to determine the fundamental matrix and, consequently, the relative pose change between the cameras up to a scaling factor.

A triangulated point cloud (see Fig. 3) allows the determination of the pipe pose and diameter [3] (or, inversely, relative pose of the second camera). The accuracy of the result depends on the origin of the introduced noise, its relative strength, and the size of the area that contains the matched keypoints [3]. Obviously, the higher the density of the matched keypoints and the larger the area that contains them, the more accurate the estimate of pipe parameters.

A robust determination of a pipe axis and diameter has been proposed in [10,11]. These algorithms account for any possible outliers and are relatively straightforward to implement. However, in this paper, the simplified approach described in [3] is used. Its steps are as follows:

- All triangulated points in the cloud are used to construct the triangulated irregular network (TIN) by Delauney triangulation.
- If points in a triangle are close enough, the normal to the triangle plane passes through or near the pipe axis.
- If two such normals are not collinear, then their cross-product is parallel (or almost parallel) to the axis. The final result for the axis vector is the average over all available cross-products.
- One of the coordinates of the point P through which the axis passes is arbitrarily chosen (for example, Y-coordinate P_y equal to that of the center of the point cloud).
- A least squares estimate is obtained for P_x and P_z using all the data points.
- An estimate for pipe radius is obtained by averaging the distances from all data points to the corresponding nearest point on the pipe axis.
- Given close estimates for 7 parameters that define the pipe, more accurate values are obtained by Levenberg–Marquardt iterative method.

There are four main differences with the approach described in [3]. Normals to the pipe surface are estimated only from the triples of points that are closer to each other than some empirically determined threshold. If closest neighbors are separated by too much, then the derived normal can only worsen the pipe-axis estimate even though the mean has more summands. Pairs of normals are used only if their dot-product is less than a certain threshold (usually 0.6) to guarantee that they are far from being collinear. The final optimization of the pipe parameters was not proposed in [3].

Points of the retinal planes can be reprojected onto the inner pipe surface, and their 3D locations can be compared with the locations obtained at the generation stage.

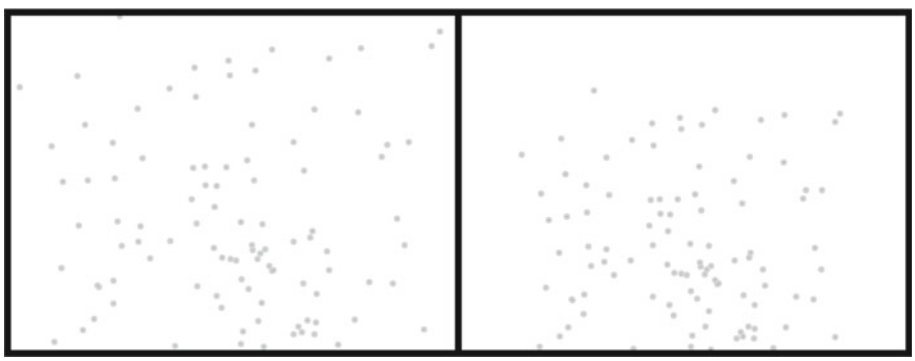

Fig. 2 Points on the pipe surface as seen by the first and second cameras. Images contain recognizable patterns allowing for pair-wise matching

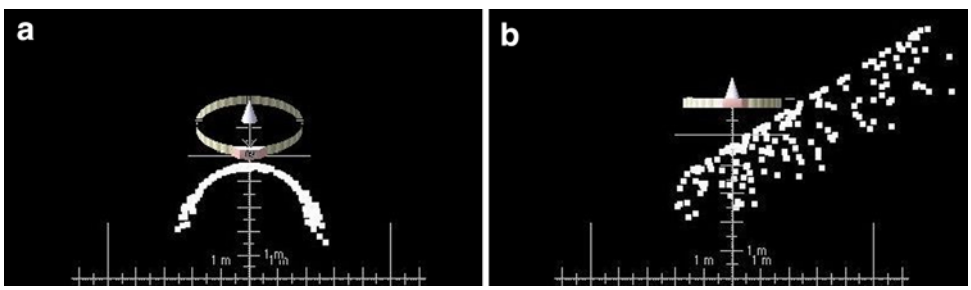

Fig. 3 Two views of triangulated point cloud (visualized using GeoZui4D, http://vislab-ccom.unh.edu/GeoZui4D). **a** Almost along the pipe axis. **b** Perpendicular to the pipe axis

The second scenario deals with the simulation of image acquisition and the reprojection of the rendered images. An image (Fig. 4a) with dimensions 500 × 360 pixels was used as a tile for the inner surface of the pipe. It was positioned on the pipe surface, so that the first column of the image was at $Y = 0$ and was neighboring the last column $(W - 1)$ of itself. In the Z-direction, the image was indefinitely repeated, so that the top row of image 0 was at $Z = 0$, the whole column occupies $H * \pi * D/360$ of the pipe length, and the top row of the Nth image was at $Z = N * H * \pi * D/360$.

Given the intrinsic parameters and pose of the camera, the rendered frame was formed by projecting the rays from the focal point onto the pipe surface. Figure 4b shows such a frame. The camera was shifted downwards from the axis and looks slightly up. The frame clearly shows a horizontal seam at $Y = 0$, multiple seams between consecutive wrapping images, and a vanishing point. If an image with toroidal periodicity is used, the seams do not appear on the rendered frame. It is obvious that such an inspection image is almost useless for mosaicing purposes. The resolution of the mosaic changes dramatically from the near field to the far field. Also, in a real situation, the far field would be illuminated worse than the near field. Better results can be obtained if the camera is shifted away from the pipe axis as far as possible and its optical axis is nearly perpendicular to the pipe axis. Figure 5 shows three frames rendered with such an arrangement, with slightly varying camera poses. The corresponding reprojections are shown below.

The camera poses are assumed to be known with sufficient accuracy, so that the discrepancies between the reprojections are only caused by the rasterization effects. However, even if these errors are significant, mosaicing of the reprojected images produces results visibly very similar to those without errors. For example, Fig. 6 shows three reprojected images with 5% error. The first image has a 5% error in Y position of the camera, the second image has a 5% error in camera roll angle, and the third image has a 5% error in both Y position and roll angle. Figure 7 shows mosaics created from nine image frames: Fig. 7a from the images without errors and Fig. 7b from the images reprojected with introduced errors. In both cases, the mosaics provide consistent and seamless views of the inner pipe surface, and the mosaics are visually almost indistinguishable from the ground truth image shown in Fig. 4a. This proves that the above procedure, even with an incomplete view of the pipe surface and hence inaccurate estimate of distance traveled by the camera and camera pose, can be successfully used for pipe inspection and a rough estimate of the locations where the pipe has areas of corrosion and visible defects.

Note that any effects of inhomogeneous illumination have not been modeled in the described procedure. In a real situation, the illumination inhomogeneity must be compensated

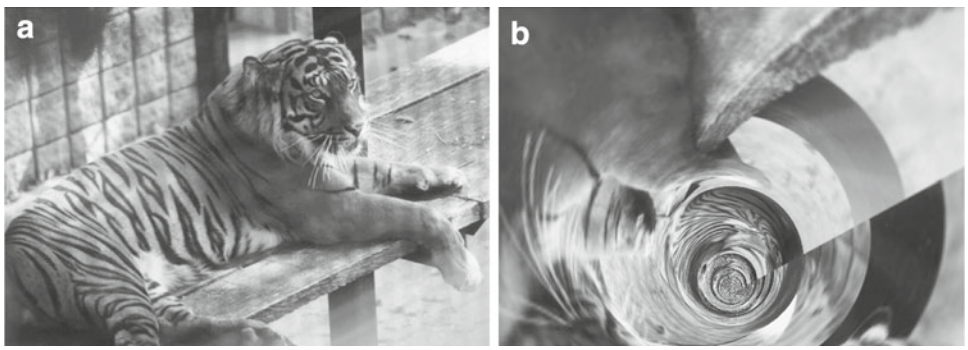

Fig. 4 **a** "Wrapping" image used in simulations. **b** Rendered frame with a camera looking almost along the pipe axis

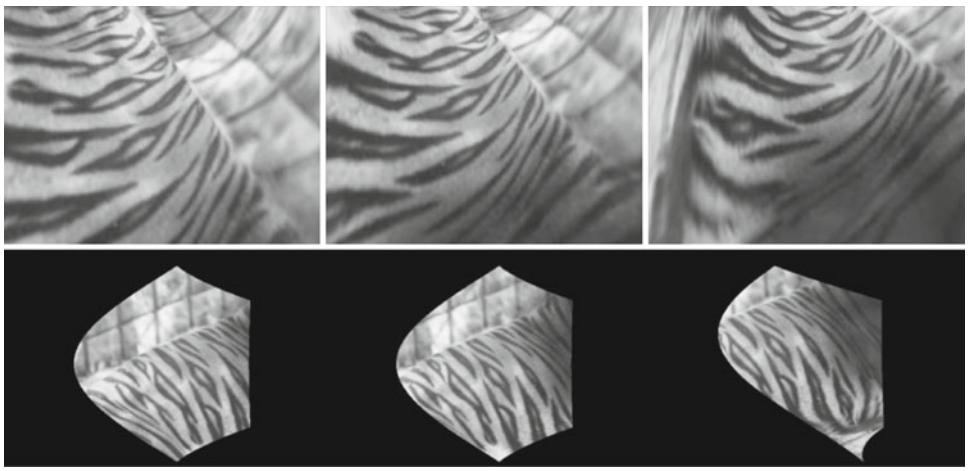

Fig. 5 Rendered frames and their reprojections. Explanations in the text

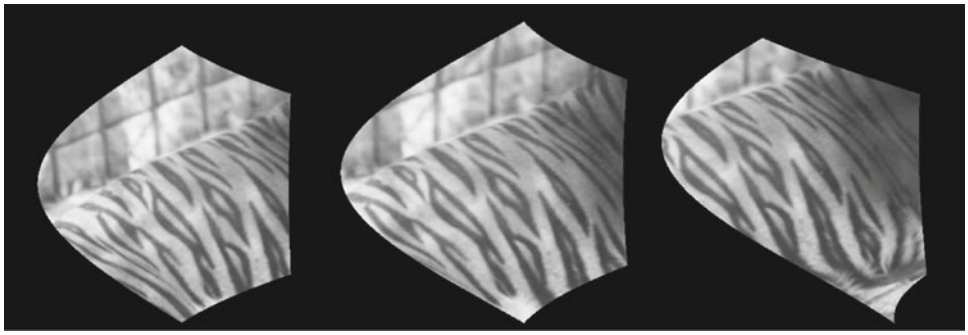

Fig. 6 Reprojected images with introduced errors

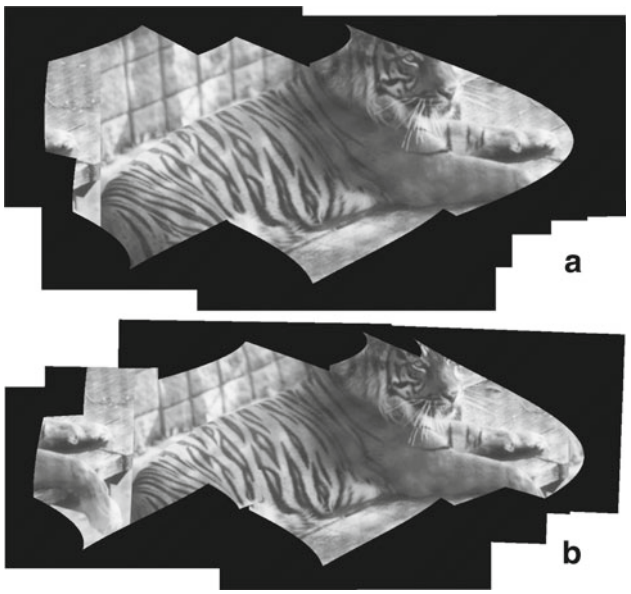

Fig. 7 **a** Mosaic of nine reprojected frames. **b** Mosaic created from the same frames reprojected with 5% error in either camera position or orientation

to obtain a mosaic without noticeable seams. Any brightness correction depends on the light source and its position with respect to the camera. However, a brightness correction does not pose much difficulty. First, the keypoint matching is robust with respect to variations in illumination [2]. Second, each pixel in the reprojected images has a known orientation (in the camera-centric system) and distance from the camera. Hence, a correction factor is easy to calculate from a simple calibration procedure and implement as a 3D lookup table (pitch-roll-distance). This sparse LUT can be interpolated to find any correction factor value for a specific pixel.

6 Conclusions

This paper proposes an algorithm that allows the processing of still or video images acquired inside a pipe and the construction of a continuous view inner surface of the pipe for manual visual inspection or automatic detection of corroded patches or cracks. The proposed approach requires certain conditions for image acquisition, specifically a wide-lens camera that travels near the pipe's wall and looks at the opposite wall, so that the acquired frames cover mostly the near-range areas. The described procedure serve as a feasibility study for a company involved in inspection of pipe systems that, because of the small diameter of the pipes or the complexity of the system, does not allow usage of a "pig" or a crawler. Further research includes building a prototype of an inspection system with a video head correctly oriented with respect to the cable, an attached source of illumination, and a pipe with a texture-rich pattern on the inner surface.

Acknowledgments The project was supported by NOAA grant NA05NOS4001153. The author would like to thank Doosan Babcock Energy Limited (UK) for asking the question whether this goal is achievable, and Prof. J. V. Gardner and C. Mitchell for help with the manuscript

References

1. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle and fish-eye lenses. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1335–1340 (2006)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoint. Int. J. Comput. Vis. **60**, 91–110 (2004)
3. Jokinen, O.: Reconstruction of quadric surfaces from disparity measurements. In: Tescher, A.G. (ed.) Applications of Digital Image Processing XVII, Proceedings of SPIE, vol. 2298, pp. 593–604. (1994)
4. Hartley, R., Zisserman, A.: Multiple view geometry. Second edn. Cambridge University Press, Cambridge. ISBN 0521 54051 8 (2003)
5. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. Computer Vision and Image Understanding (CVIU) **110**, 346–359 (2008)
6. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151. (1988)
7. Kanazawa, Y., Uemura, K.: Wide baseline matching using triplet vector descriptor. In: Proceedings of the 17th British Machine Vision Conference, pp. 267–276. (2006)
8. Torr, P.H.S., Murray, D.W.: The development and comparison of robust methods for estimating the fundamental matrix. Int. J. Comput. Vis. **24**, 271–300 (1997). doi:10.1023/A:1007927408552

9. Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M.: Interactive digital photomontage. In: Proceedings of the SIGGRAPH'04 (2004)
10. Chaperon, T., Goulette, F.: Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. In: Proceedings of the Vision, Modelling and Visualisation, pp. 35–42 (2001)
11. Rabbani, T., van den Heuvel, F.: Efficient Hough transform for automatic detection of cylinders in point clouds. In: ISPRS, Workshop "Laser scanning 2005" (2005)