ISPRS Journal of Photogrammetry and Remote Sensing}, 64 (2009) 6, 529-540, Elsevier 2009

Monocular 3D scene reconstruction at absolute scale

Christian Wöhler, Pablo d'Angelo, Lars Krüger

Daimler Group Research, Environment Perception
P. O. Box 2360, D-89013 Ulm, Germany

Annika Kuhl

Curtin University of Technology, Department of Computing
GPO Box U1987, Perth, Western Australia 6845

Horst-Michael Groß

Technische Universität Ilmenau, Faculty of Computer Science and Automation
P. O. Box 100565, D-98684 Ilmenau, Germany

Abstract

In this article we propose a method for combining geometric and real-aperture methods for monocular 3D reconstruction of static scenes at absolute scale. Our algorithm relies on a sequence of images of the object acquired by a monocular camera of fixed focal setting from different viewpoints. Object features are tracked over a range of distances from the camera with a small depth of field, leading to a varying degree of defocus for each feature. Information on absolute depth is obtained based on a Depth-from-Defocus approach. The parameters of the point spread functions estimated by Depth-from-Defocus are used as a regularisation term for Structure-from-Motion. The reprojection error obtained from bundle adjustment and the absolute depth error obtained from Depth-from-Defocus are simultaneously minimised for all tracked object features. The proposed method yields absolutely scaled 3D coordinates of the scene points without any prior knowledge about scene structure and camera motion. We describe the implementation of the proposed method as an offline and as an online algorithm. Evaluating the algorithm on real-world data, we demonstrate that it yields typical relative scale errors of a few percent. We examine the influence of random effects, i.e. the noise of the pixel greyvalues, and systematic effects, caused by thermal expansion of the optical system or by inclusion of strongly blurred images, on the accuracy of the 3D reconstruction result. Possible applications of our approach are in the field of industrial quality inspection; especially, it is preferable to stereo cameras in industrial vision systems with space limitations or where strong vibrations occur.

Key words: Close-range photogrammetry, Reconstruction, Metrology, Bundle adjustment, Industry

1 Introduction

The knowledge of three-dimensional structure plays an important role in many fields such as navigation, mapping, obstacle avoidance, and object detection. Depth-from-Stereo (Scharstein et al., 2002) was one of the first methods for recovering depth information as it is inspired by human vision. The known geometry of the cameras is used to triangulate the spatial position of corresponding points from two images that are acquired from different viewpoints. The disadvantage of classical stereo vision systems is their need for a pair of precisely calibrated cameras, making it complex and costly for many applications. Therefore spatial scene reconstruction using monocular camera systems is often preferable. Structure-from-Motion (SfM) is such an alternative: From corresponding points in at least two images acquired sequentially at different camera positions the spatial positions of the points are recovered. The problem is that the scene can be reconstructed only up to a scaling factor as long as the camera positions are unknown.

Methods to establish point correspondences from different images require the detection and assignment of salient object features. Harris and Stephens (1988) propose image features that serve well for tracking algorithms. Widely used methods are SIFT features (Lowe, 2004), involving the extraction of scale invariant features using a staged filtering approach, or the Kanade-Lucas-Tomasi (KLT) feature detector (Shi and Tomasi, 1994) which is based on the Harris corner detector and takes into account affine deformation.

A different approach to scene reconstruction utilises position variant appearance, e.g. Depth-from-Defocus (Chaudhuri and Rajagopalan, 1999) and Depth-from-Focus (Subbarao, 1989; Ens and Lawrence, 1993; Subbarao and Choi, 1995). Depth-from-Focus uses images taken by a single camera at different focus settings to compute depth. The focus settings for the image depicting a point with minimal blurring are used to compute the absolute depth (Grossmann, 1987), relying on an appropriate calibration procedure. Depth-from-Defocus (DfD) methods rely on the fact that a real lens blurs the observed scene before the imaging device records it. The amount of blurring depends on the actual lens, but also on the distance of the observed object to the lens. Pentland (1982) uses this property to estimate depth simultaneously for all scene points from only one or two images. Depth information is extracted from a single image showing sharp discontinuities of intensity by Pentland (1987). A survey of existing methods is given by Chaudhuri and Rajagopalan (1999). Watanabe et al. (1995) propose a method that computes DfD in real-time using structured lighting.

So far, no attempt has been made to combine the precise relative scene reconstruction of SfM with the absolute depth data of DfD. A work related to this paper was published by Myles and da Vitoria Lobo (1998), where a method to recover affine motion and defocus simultaneously is proposed. However, the spatial extent of the scene is not reconstructed by their method, since it requires planar objects.

The main contribution of this article consists of a novel combination of SfM (a geometric method) with DfD (a real-aperture method). We will show that the combination of these methods yields a 3D scene reconstruction of high absolute metric accuracy based on an image sequence acquired with a monocular camera.

2 SfM and DfD

SfM recovers the spatial scene structure using a monocular camera. An initial step of SfM is the geometric calibration of the camera in terms of estimating the internal parameters, i.e. principal point, principal distance, pixel size, pixel skew, and lens distortion parameters (McGlone et al., 2004). An accurate value of the principal distance is required in Section 3 for calibration of the DfD method. Specifically, we use the semi-automatic approach for calibration rig detection proposed by Krüger et al. (2004). Subsequently, salient feature points are extracted and tracked across the sequence. The motion of these features relative to the camera is then used in a bundle adjustment (Triggs et al., 2000; McGlone et al., 2004) minimising the error term

$$E_{\mathrm{SfM}}\left(\{T_j\},\{X_i\}\right)=\sum_{i=1}^{N}\sum_{j=1}^{M}\left[\mathcal{P}\left(T_j X_i\right)-x_{ij}\right]^2 \tag{1}$$

with respect to the M camera transforms T_j and the N scene points X_i. Here, x_{ij} denotes the 2D pixel coordinates of feature i in image j. The function $\mathcal{P}$ denotes the projection of 3D scene points to image coordinates and T_j the transform of the camera coordinate system of image j with respect to an arbitrary world coordinate system.

DfD directly recovers the spatial scene structure using a monocular camera. The depth D of the tracked feature points is calculated by measuring the amount of defocus, expressed e.g. by the standard deviation σ of the Gaussian-shaped point spread function (PSF) that blurs the image. An exact description of the PSF due to diffraction of light at a circular aperture is given by the radially symmetric Airy pattern $A(r) \propto \left[J_1(r)/r\right]^2$, where $J_1(r)$ is a Bessel function of the first kind (Pedrotti, 1993). For practical purposes, however, particularly

when a variety of additional lens-specific influencing quantities (e.g. chromatic aberration) is involved, the Gaussian function is a reasonable approximation to the PSF (Chaudhuri and Rajagopalan, 1999). In the following, σ will be referred to as the "radius" of the PSF.

Measuring σ is the most important part of the depth estimation by DfD. The classical DfD approach uses two images of the same object taken at two different focal settings (Chaudhuri and Rajagopalan, 1999). Pentland (1987) shows that a-priori information about the image intensity distribution, e.g. the presence of sharp discontinuities (edges), allows the computation of the PSF radius σ based on a single image. This is achieved by estimating the value of σ that generates the observed intensity distribution from the known ideal intensity distribution.

Since in our scenario such a-priori information is not available, we suggest the empirical determination of a so-called Depth–Defocus Function. We assume that local features in the scene are tracked across a sequence of images and that for each feature the image is determined in which the feature appears best focused. Based on a calibration procedure, the radius σ of the Gaussian PSF which transforms the best focused version of the feature into the currently observed pattern is determined as a function of depth D.

3 Spatial scene reconstruction by combining SfM and DfD

3.1 The Depth–Defocus Function

The Depth–Defocus Function $\mathcal{S}(D) = \sigma$ expresses the radius σ of the Gaussian PSF as a function of depth D, i.e. the distance between the object and the lens plane. It is based upon the lens law (Pedrotti, 1993):

$$\frac{1}{v} + \frac{1}{D} = \frac{1}{f}. \tag{2}$$

An object at distance D is focused if the principal distance is v, with f denoting the focal length of the lens. Varying the principal distance by a small amount Δv causes the object to be defocused as the light rays intersect before or behind the image plane. In the geometric optics approximation, a point in the scene is transformed into a so-called circle of confusion of radius $c = |\Delta v|/(2\kappa)$ in the image plane, where κ is the f-stop number expressing the focal length in terms of the aperture diameter. Empirically, we found that for small $|\Delta v|$ the resulting amount F of defocus can be modelled by a zero-mean Gaussian,

which is symmetric in Δv:

$$F(\Delta v) = \frac{1}{\phi_1} e^{-\frac{1}{\phi_2}\Delta v^2} + \phi_3 \ . \tag{3}$$

The parameters ϕ_1 and ϕ_2 are normalising constants and ϕ_3 denotes the defocus level for very strongly blurred images. Setting $F = \sigma$ leads to the so-called Depth–Defocus Function as described below. To determine the best focused version of a tracked feature, F is represented by other measures such as the grey value variance or the high-frequency integral of the amplitude spectrum of the image or part of it (cf. Sections 3.2 and 3.3). The radius c of the circle of confusion and the PSF radius σ are related to each other in that σ is a monotonously increasing nonlinear function of c. Hence, the symmetric behaviour of $c(\Delta v)$ apparent from Fig. 1a implies a symmetric behaviour of $\sigma(\Delta v)$. Depending on the constructional properties of lenses different from those we used in our experiments, analytic forms different from Eq. (3) but also symmetric in Δv may better match the observed behaviour of the PSF.

However, the Depth–Defocus Function expresses the relation between the depth of an object and its defocus. I. e., the image plane is assumed to be fixed while the distance D of the object varies by the amount ΔD, such that $\Delta D = 0$ refers to an object that is well focused. The dependence of c on ΔD is asymmetric, as shown in Fig. 1b. Since neither D nor ΔD are known, the functional relation needs to be modelled with respect to Δv:

$$\frac{1}{v + \Delta v} + \frac{1}{D} = \frac{1}{f} \ . \tag{4}$$

A value of $\Delta v \neq 0$ refers to a defocused object point. Solving Eq. (4) for Δv and inserting Δv in Eq. (3) yields the Depth–Defocus Function

$$\mathcal{S}(D) = \frac{1}{\phi_1} e^{-\frac{1}{\phi_2}\left(\frac{fD}{D-f} - v\right)^2} + \phi_3 \ . \tag{5}$$

Calibrating the Depth–Defocus Function $\mathcal{S}(D)$ for a given lens corresponds to determining the parameters ϕ_1, ϕ_2, ϕ_3, and f in Eq. (5). This is achieved by taking a large set of measured (σ, D) data points and performing a least squares fit to Eq. (5), where D is the distance from the camera and σ the radius of the Gaussian PSF G used to blur the well focused image with a PSF of position-dependent radius σ. Let I_{if_i} represent a small region of interest (ROI) around feature i in image f_i in which this feature is best focused, and I_{ij} a ROI of equal size around feature i in image j. The ROIs I_{if_i} and I_{ij} are related by

$$I_{ij} = G(\sigma_{ij}) * I_{if_i}. \tag{6}$$

Ideally, I_{if_i}, I_{ij}, and $G(\sigma_{ij})$ are defined on an infinite domain, but in practice they are represented by image windows of finite size, e.g. 16×16 pixels. For the actual calibration of the Depth–Defocus Function refer to Section 3.2.

At this point it is useful to examine which focal length and lens aperture are required to obtain depth values of a given accuracy with the DfD method. Assume that for a lens of focal length f_1, an object is well focused at depth D_0, and a certain amount of defocus is observed at depth hD_0, where the factor h is assumed to be close to 1 with $|h-1| \ll 1$. The depth offset $\Delta D = (h-1)D_0$ implies a circle of confusion of radius c_1 with

$$c_1 = \frac{1}{2\kappa}\left(\frac{1}{f_1^{-1} - D_0^{-1}} - \frac{1}{f_1^{-1} - (hD_0)^{-1}}\right). \tag{7}$$

Now let the focal length be changed to a larger value f_2, and the object depth is set to a larger distance kD_0 with $k > 1$. The radius c_2 of the corresponding circle of confusion is readily obtained by

$$c_2 = \frac{1}{2\kappa}\left(\frac{1}{f_2^{-1} - (kD_0)^{-1}} - \frac{1}{f_2^{-1} - (hkD_0)^{-1}}\right). \tag{8}$$

The f-stop number κ and the pixel size remain unchanged. Since the radius of the circle of confusion is a monotonously increasing function of the PSF radius σ, we assume that observing the same amount of defocus in both scenarios implies an identical radius of the corresponding circle of confusion. With the abbreviations

$$\begin{aligned} K &= \frac{1}{f_1^{-1} - D_0^{-1}} - \frac{1}{f_1^{-1} - (hD_0)^{-1}} \\ L_1 &= \frac{1 - h^{-1}}{kD_0} \\ L_2 &= \frac{1 + h^{-1}}{kD_0} \\ M &= \frac{1}{hk^2 D_0^2}, \end{aligned} \tag{9}$$

setting $c_1 = c_2$ yields the focal length f_2 according to

$$f_2 = \left(\frac{L_2}{2} \pm \sqrt{\frac{L_1}{K} - M + \frac{L_2^2}{4}}\right)^{-1}. \tag{10}$$

Only the solution with the plus sign before the square root yields positive values for f_2. According to Eq. (10), the value of f_2 is approximately proportional

to $\sqrt{k}$ independent of the chosen value of h as long as $|h-1| \ll 1$. Hence, for constant f-stop number κ, constant relative variation $|h-1|$ of the object depth D, and constant pixel size, the required focal length and thus also the aperture of the lens are largely proportional to $\sqrt{D_0}$. Our experimental evaluation outlined in Section 4 will show that the DfD approach is favourably used in the close-range domain ($D \sim$ 1–2 m) as long as standard video cameras and lenses are used.

To facilitate the integration of defocus information into the SfM framework, the image sequences are acquired such that the object is blurred to a variable extent from image to image. The focal settings of the camera are adjusted according to the maximal and minimal distance of the object. It may be necessary to fully open the aperture in order to obtain a small depth of field.

3.2 Calibration of the DfD method

For calibration, an image sequence is acquired while the camera approaches at uniform speed a calibration rig displaying a chequerboard. The sharp black-and-white corners of the chequerboard are robustly and precisely detectable (Krüger et al., 2004) even in defocused images. Small ROIs of size 16×16 pixels around each corner allow the estimation of defocus using their grey-value variance χ. The better focused the corner, the higher is the variance χ. We found experimentally that the parameterised defocus model according to Eq. (5) is also a reasonable description of the dependence of χ on the depth D. For our calibration sequence the camera motion is uniform and the image index j is strongly correlated with the object distance D. Hence, Eq. (5) is fitted to the measured (χ, j) data points for each corner i. The location of the maximum of $\mathcal{S}$ yields the index f_i of the image in which the ROI around corner i is best focused. This ROI corresponds to I_{if_i}. The fitting procedure is applied to introduce robustness with respect to pixel noise. For non-uniform camera motion the index f_i can be obtained by a parabolic fit to the values of χ around the maximum or by directly selecting the ROI with maximal χ. The depth D of each corner is reconstructed by SfM from the pose of the complete rig according to Bouguet (1997).

For each tracked corner i, we compute for each ROI I_{ij} the amount of defocus, i.e. the σ value relative to the previously determined best focused ROI I_{if_i} according to Eq. (6). By employing the bisection method, we determine from a number of different values the value of σ for which the root mean square deviation between $G(\sigma) * I_{if_i}$, denoting the best-focused ROI convolved with a Gaussian PSF of radius σ, and I_{ij}, the currently observed ROI, becomes minimal. The Depth–Defocus Function is then obtained by a least mean squares fit to all determined (σ, D) data points. Two examples are shown in Fig. 2 for

lenses with focal lengths of 12 mm and 20 mm and f-stop numbers of 1.4 and 2.4, respectively. Objects at a distance of about 0.8 m and 0.6 m, respectively, are in focus, corresponding to the minimum of the curve.

3.3 Combining motion, structure, and defocus

The SfM analysis involves the extraction of salient features from the image sequence which are tracked using the KLT technique (Shi and Tomasi, 1994). For the integration of defocus information, a ROI of constant size is extracted around each feature point at each time step. For each tracked feature, the best focused image has to be identified in order to obtain the increase of defocus for the other images. We found that greyvalue variance as a measure for defocus does not perform well on features other than black-and-white corners. Instead we make use of the amplitude spectrum $|\mathscr{F}_I(\omega)|$ of the ROI extracted around the feature position. High-frequency components of the amplitude spectrum denote sharp details, whereas low-frequency components refer to large-scale features. Hence, the integral over the high-frequency components serves as a measure for the sharpness of a certain tracked feature. However, since the highest-frequency components are considerably affected by pixel noise and defocus has no perceivable effect on the low-frequency components, a frequency band between two empirically determined frequencies ω_0 and ω_1 is taken into account according to

$$H = \int_{\omega_0}^{\omega_1} |\mathscr{F}_I(\omega)| d\omega \tag{11}$$

with $\omega_0 = \frac{1}{4}\omega_{\max}$ and $\omega_1 = \frac{3}{4}\omega_{\max}$, where $\omega_{\max}$ is the maximum frequency for the ROI. The amount of defocus increases with decreasing value of H. The defocus measure H is used to determine the index of the best focused ROI for each tracked feature in the same manner as the greyvalue variance χ in Section 3.2. Fig. 3 illustrates that the value of H cannot be used for comparing the amount of defocus among different feature points since the maximum value of H depends on the image content. The same is true for the greyvalue variance. Hence, both the integral H of the amplitude spectrum as well as the greyvalue variance are merely used for determining the index of the image in which a certain feature is best focused.

The defocus, i.e. the radius σ of the Gaussian PSF, is then computed relative to the best focused ROI according to Section 3.2. The depth D is obtained by inverting the Depth–Defocus Function $\mathcal{S}(D)$ according to Eq. (5). The encountered two-fold ambiguity is resolved by using information about the direction of camera motion, which is obtained either based on a-priori knowledge or by

performing a SfM analysis according to Eq. (1), yielding information about the path of the camera. If the estimated value of σ is smaller than the minimum of $\mathcal{S}(D)$, the depth is set to the value at which $\mathcal{S}(D)$ is minimal. For an example feature, the calculated defocus and the inferred depth values are shown in Fig. 3.

We found experimentally that the random scatter of the feature positions extracted by the KLT tracker is largely independent of the image blur for PSF radii smaller than 5 pixels and is always of the order 0.1 pixels. However, more features are detected and less features are lost by the tracker when the tracking procedure is started on a well-focused image. Hence, the tracking procedure is repeated, starting from the "sharpest" image located near the middle of the sequence which displays the largest value of H averaged over all previously detected features, proceeding towards either end of the sequence and using the ROIs extracted from this image as reference patterns. The 3D coordinates X_i of the scene points are then computed by determination of the minimum of the combined error term

$$E_{\mathrm{comb}}\left(\{T_j\},\{X_i\}\right) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left[\left(\mathcal{P}\left(T_j X_i\right) - x_{ij}\right)^2 + \alpha\left(\mathcal{S}\left(\left[T_j X_i\right]_z\right) - \sigma_{ij}\right)^2\right] \tag{12}$$

with respect to the M camera transforms T_j and the N scene points X_i. The value of σ_{ij} corresponds to the estimated PSF radius for feature i in image j, α is a weighting factor, $\mathcal{S}$ the Depth–Defocus Function that calculates the expected defocus of feature i in image j, and $[\cdot]_z$ the z coordinate, i.e. the depth D, of a scene point. The correspondingly estimated radii σ_{ij} of the Gaussian PSFs define a regularisation term in Eq. (12), such that absolutely scaled 3D coordinates X_i of the scene points are obtained. The value of α denotes the relative weight of the two error terms in Eq. (12) and depends on the variances of the measurements x_{ij} and σ_{ij}. In the examples regarded in Section 4, a favourable choice is $\alpha = 0.5$, indicating that the variances of x_{ij} and σ_{ij} are similar. The values of X_i are initialised according to the depth values estimated based on the DfD approach. To minimise the error term E_{comb} we employ the Levenberg-Marquardt algorithm (Madsen et al., 1999). To reduce the effect of outliers, we use the M-estimator technique with the "fair" weighting function $w(x) = 1/(1 + |x|/c)$ with x as the error value, where $c = 1.3998$ is a favourable choice (Rey, 1983). A possible extension of our optimisation technique not regarded in the experiments described in Section 4 is to first weight down errors with a robust estimator, and if after some iteration steps some points are regarded as outliers, to repeat the weighting on the reduced set of observations. Furthermore, it might be favourable to compare the residuals with their individual covariance information, which provides information about how much larger a residual is than it is thought to be determined from

the given data. Such techniques are known to improve the convergence behaviour in many applications, but in the experiments regarded in Section 4 our simple robust estimator was always sufficient to obtain a solution after a few tens of iterations of the Levenberg-Marquardt scheme.

3.4 Online version of the algorithm

The 3D reconstruction method outlined so far is an offline (or "batch") algorithm since the error term Eq. (12) is minimised once for the complete image sequence. In this section we present an online version of the proposed combination of SfM and DfD which processes the acquired images instantaneously, thus generating a new 3D reconstruction result after acquisition of each image of the sequence. This is a desired property e.g. in the context of mobile robot navigation, SLAM, or in-situ exploration.

The online version starts by acquiring the current image. The feature tracker attempts to track the features present in the previous image and may add new features. Again, the KLT feature tracker (Shi and Tomasi, 1994) is used. The sharpness of each feature within the current frame is obtained based on the integral H over the amplitude spectrum of the ROI around the feature position (cf. Eq. (11)). The next step is the determination of the best focused frame for each feature. Since fitting the Depth–Defocus Function Eq. (5) imposes a considerable computational burden, a second-degree polynomial is fitted to the values of H instead. A threshold rating this fit selects possible candidates that may already have passed their point of maximum sharpness. The Depth–Defocus Function according to Eq. (5) is then fitted to the H values of the preselected feature candidates only. After determination of the sharpest frame, the initial depth values for the respective feature are computed by estimating the PSF radius σ as outlined in Section 3.2.

The depth values obtained by the DfD method are used to initialise the Levenberg-Marquardt scheme which determines the camera transforms and 3D feature points that minimise the error function given by Eq. (12) using an M-estimator as in the offline version. The current optimisation result is used as an initialisation to the subsequent iteration step involving the next acquired image.

4 Experimental evaluation

4.1 Offline algorithm

In all described experiments we used a Baumer 1032×776 pixels industrial CCD camera with Cosmicar-Pentax video lenses of focal length 12 mm (table-top scenes, Sections 4.1.1–4.1.3) and 20 mm (industrial quality inspection scenario, Section 4.1.4). In order to validate our approach we first reconstructed a planar object with reference points of precisely known mutual distance. A chequerboard as shown in Fig. 4 with 10×8 squares of size 15×15 mm^2, respectively, was used. The 99 corners serve as features and are extracted in every image using the method described by Krüger et al. (2004) to assure sub-pixel accuracy. The reference pose of the chequerboard is obtained according to Bouguet (1997) based on the given size of the squares. Note that Bouguet (1997) determines the reference pose of the chequerboard by applying a least mean squares fit on a single image, whereas the proposed algorithm estimates the 3D structure of a scene by means of a least mean squares fit applied to the whole image sequence. Comparing the obtained results with the determined reference pose of the object is therefore a comparison between two methods conducting different least mean squares fits.

The deviation E_{reconstr} of the reconstructed 3D scene point coordinates X_i from the reference values X_i^{ref} is given by

$$E_{\text{reconstr}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left\| X_i - X_i^{\text{ref}} \right\|^2}. \tag{13}$$

To determine an appropriate weight parameter α in Eq. (12) we computed E_{reconstr} for different α values in the range between 0 and 1. For $\alpha = 0$ the global minimisation is equivalent to SfM initialised with the calculated DfD values. One must keep in mind, however, that the absolute scaling factor is then part of the gauge freedom of the bundle adjustment, resulting in a corresponding "flatness" of the error function. Small α values therefore lead to an instable convergence. The value of E_{reconstr} levels off to 16 mm for $\alpha \approx 0.3$ and obtains its minimum value of 7 mm for $\alpha = 0.42$. The root mean square deviation of the reconstructed size of the squares from the true value of 15 mm then amounts to 0.2 mm or 1.3%. The most accurate scene reconstruction results are obtained with α between 0.3 and 0.5. The reconstructed 3D scene points X_i for $\alpha = 0.42$ are illustrated in Fig. 4, the dependence of E_{reconstr} on α in Fig. 5 (top).

In addition to the reconstruction error E_{reconstr}, a further important error

measure is the reprojection error

$$E_{\text{reprojection}} = \sqrt{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\mathcal{P}\left(T_j X_i\right) - x_{ij}\right)^2} \tag{14}$$

denoting the root-mean-square deviation between the measured 2D feature positions x_{ij} and the reconstructed 3D scene points X_i reprojected into the images using the reconstructed camera transforms T_j.

The defocus error denotes the root-mean-square deviation between measured and expected radii σ_{ij} of the Gaussian PSFs according to

$$E_{\text{defocus}} = \sqrt{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\mathcal{S}([T_j X_i]_z) - \sigma_{ij}\right)^2}. \tag{15}$$

Fig. 5 (bottom) shows the relation between the weight parameter α, the reprojection error $E_{\text{reprojection}}$, and the defocus error E_{defocus}. For $\alpha > 0.3$ the defocus error stabilises to 0.58 pixels per feature. Larger α values lead to a stronger influence of the DfD values on the optimisation result, resulting in an increasing reprojection error $E_{\text{reprojection}}$ due to the inaccuracy of the estimated σ_{ij} values.

Although the depth values derived by DfD are noisy, they are sufficient to establish a reasonably accurate absolute scale. Hence, this first evaluation shows that the combined approach is able to reconstruct scenes at absolute scale without prior knowledge. Our approach is favourably applied in the close-range domain ($D \sim 1$ m) using standard video cameras and lenses (f below ~ 20 mm, pixel size $\sim 10\mu$m, image size $\sim 10^6$ pixels). For larger distances around 10 m, the focal length required to obtain a comparable relative accuracy of absolute depth is proportional to $\sqrt{D}$, implying a narrow field of view of less than 7 degrees and thus rendering the application of our approach unfeasible from a practical point of view as SfM becomes unstable for small intersection angles.

Further experiments performed on real-world objects are described in the following paragraphs. Images from the beginning, the middle, and the end of the corresponding sequences, respectively, are shown in Fig. 6. In order to separate random fluctuations from systematic deviations, we computed the error measures for 100 runs for each example, respectively. For the utilised camera, the noise of the pixel grey values is proportional to the square root of the grey values themselves. Empirically, we determined for the standard deviation of a pixel with 8-bit grey value $I \in [0 \ldots 255]$ the value $\sigma_I = 0.22\sqrt{I}$. For each of the 100 runs, we added a corresponding amount of Gaussian noise to the

images of the sequence. The noise leads to a standard deviation of the feature positions x_{ij} obtained by the KLT tracker of 0.1 pixels.

4.1.1 Cuboid sequence

To demonstrate the performance of our approach on a non-planar test object of known dimensions we applied it to the cuboid-shaped object shown in Fig. 6a. This object displays a sufficient amount of texture to generate "good features to track" (Shi and Tomasi, 1994). In addition, black markers on white background with known mutual distances are placed near the edges of the cuboid. The 3D coordinates of the scene points are obtained by minimising the error term E_{comb} according to Eq. (12) with $\alpha = 0.5$ as the weight parameter. This value of α is used in all subsequent experiments. Tracking outliers are removed by determining the features with reprojection errors of more than $3E_{\text{reprojection}}$ and neglecting them in a subsequent second bundle adjustment.

The 3D reconstruction result for the cuboid sequence is shown in Fig. 7. We obtain for the average reprojection error $E_{\text{reprojection}} = 0.642$ pixels and for the defocus error $E_{\text{defocus}} = 0.64$ pixels. In order to verify the absolute scale, we compared the reconstructed pairwise distances between the black markers on the object (as seen e.g. in the top right corner of the front side) to the corresponding true distances. For this comparison we utilised a set of six pairs of markers with an average true distance of 32.0 mm. The corresponding reconstructed average distance amounts to 34.1 mm (cf. Table 1).

4.1.2 Bottle sequence

In order to demonstrate the performance of our approach on a real-world object, we applied it to a bottle as shown in Fig. 6b. No background features are selected since none of these feature obtains its maximum sharpness in the acquired sequence. The 3D reconstruction result is shown in Fig. 8. We obtained for the reprojection error $E_{\text{reprojection}} = 0.75$ pixels and $E_{\text{defocus}} = 0.39$ pixels. To quantify the accuracy of the determined absolute scale, we compared the diameter of the reconstructed object with that of the real bottle. We projected the reconstructed points into the xz plane of the camera coordinate system, in which the x axis is parallel to the image rows, the y axis is parallel to the image columns (and thus to the central axis of the cylinder), and the z axis is parallel to the optical axis. The circle fit to the projected 3D points as shown in Fig. 8 yields an average diameter of 82.8 mm, which is in good correspondence with the known parameter of 80.0 mm (cf. Table 1).

4.1.3 Lava stone sequence

As a further real-world object, we examined the lava stone shown in Fig. 6c. The 3D reconstruction result is shown in Fig. 9. The shaded view of the object is based on the triangulation of the reconstructed set of 3D points. The cusp visible in the left part of the reconstructed surface is due to three outlier 3D points generated by inaccurately determined feature positions. For this scene we have $E_{\text{reprojection}} = 0.357$ pixels and $E_{\text{defocus}} = 0.174$ pixels. Two points on the object with a true mutual distance of 60.0 mm were chosen as reference locations for estimation of the accuracy of the determined absolute scale. The reconstructed distance of the reference points amounts to 58.3 mm, which is consistent with the known distance of 60.0 mm (cf. Table 1).

4.1.4 Flange sequence: Raw cast iron surface

Another experimental evaluation of the offline algorithm addresses an industrial quality inspection scenario. We regard the 3D reconstruction of the raw cast iron surface of a flange. In our setting, the metal part is attached to a goniometer and can thus be rotated around two axes, while the camera is fixed. In an industrial inspection system, it would probably be more favourable to mount the camera on an industrial robot such that it can be moved with respect to a fixed part to be inspected. Three images of the acquired sequence are shown in Fig. 6d. Although the surface is rough, the extracted set of 3D points is rather sparse, which is due to specular reflections changing across time, leading to premature termination of tracks by the KLT tracker. The extracted set of points shows that the reconstructed surface region is essentially flat and inclined with respect to the image plane (Fig. 10a). We fitted a plane to the set of 3D points and determined a RMS distance of the 3D points from this reference plane of 1.46 mm. To examine the absolute scale of the reconstructed scene, we used two feature points situated at well-defined locations on the edge of small deformations of the surface (marked as 1 and 2 in Fig. 10b). According to our 3D reconstruction result, the mutual distance of the corresponding two 3D points amounts to 15.45 mm. The true distance, determined by tactile measurement, is 15.2 mm. Being too large by the small amount of 1.6%, the estimated absolute scale is in good agreement with the true value, again in very good correspondence with known distance.

In contrast to the previous examples, the ROIs around the extracted feature positions show strong small-scale intensity variations, such that the noise added to the ROIs has a negligible effect on the DfD result. On the other hand, the orientation of the surface with respect to the light source changes across the sequence, leading to an appearance of the ROIs that changes systematically over time. These variations are not of geometrical nature but are due to the strongly non-Lambertian reflectance behaviour of the surface. This is

not taken into account by the KLT tracker, which therefore tends to produce a slight drift of the feature positions relative to the object across the image sequence. At the same time the variations have a systematic influence on the amount H of defocus determined according to Eq. (11) and are presumably the main reason for the observed small discrepancy of 1.6% between the estimated and the true absolute scale of the scene. The very small standard deviation of the measured distance between the two reference points (cf. Table 1) is presumably due to the strong small-scale contrasts in the images, with many bright pixels being over-saturated. Hence, the results of the KLT tracker are barely affected by the Gaussian noise added to the pixel intensities.

The extracted set of 3D points is too sparse to reveal the small-scale deformations visible in the images. Hence, we used our sparse set of depth points shown in Fig. 10a as an input to an image-based framework for dense 3D surface reconstruction recently proposed by d'Angelo and Wöhler (2008). This technique is based on the combined analysis of reflectance, polarisation, and sparse depth data. An error functional consisting of several error terms related to the measured reflectance and polarisation properties and the depth data is minimised in order to compute a dense surface gradient field and in a subsequent step a dense depth map. The 3D profile obtained with this approach for the raw cast iron surface is shown in Fig. 10c. This 3D surface profile shows both the large-scale structure of the surface and the small surface deformations also visible in Fig. 6d.

4.2 Online algorithm

A systematic evaluation of the online algorithm was performed for the cuboid, bottle, and lava stone sequence. In the surface inspection scenario (flange sequence) we only employed the offline algorithm since here no knowledge about the object structure is necessary before the final dense reconstruction step performed after termination of image acquisition. The online algorithm generally starts with a very noisy set of 3D points, due to the small number of features already having reached their maximum sharpness at the beginning of the image sequence. After processing more and more images, the 3D reconstruction result starts to resemble the result of the offline algorithm. The results are not identical because generally a similar but not identical index f_i (cf. Eq. (6)) is determined for the sharpest ROI of each track by the offline and the online algorithm, respectively. The DfD results then differ correspondingly.

In the cuboid example shown in Fig. 11, the first 3D reconstruction result can be obtained after processing 21 images. After 5 iterations, still only 40 features are available, having passed their point of maximum sharpness. In this example, the same six pairs of reference points as in Section 4.1 were used

for evaluating the accuracy of the determined absolute scale. Fig. 12 shows the behaviour of the reconstruction accuracy in relation to the increasing number of iterations, averaged over 100 online runs carried out after adding Gaussian noise to the images as described in Section 4.1. In Figs. 12–14, the standard deviations across the 100 online runs are indicated by error bars. After 43 processed images, the measured average mutual distance of the six pairs of reference points differs by less than 1.8% from the true value, and this difference is smaller than 1.2 standard deviations. For less than 38 processed images, not all reference points have passed their point of maximum sharpness, and their 3D positions have therefore not yet been computed. The average relative scale error shown in Fig. 12 is derived from those pairs of reference points that already have passed their point of maximum sharpness.

Fig. 11 suggests that with increasing number of available features the reconstructed size and shape of the cuboid become more accurate. However, Fig. 12 shows that the maximum accuracy of the inferred absolute scale is obtained after 44 or 45 processed images. The reason is that the very last images of the sequence are strongly blurred. This leads to large inaccuracies of the depth values derived from the estimated PSF radii σ_{ij}, since far away from its minimum, the Depth–Defocus Function $\mathcal{S}(D)$ according to Eq. (5) is nearly horizontal. Furthermore, systematic deviations tend to arise since the observed behaviour $\sigma(D)$ of the PSF radius is best represented for small and intermediate values of σ by the analytic form chosen in Eq. (5) for the Depth–Defocus Function $\mathcal{S}(D)$ (cf. Section 4.3).

Analogous experimental evaluations were conducted for the bottle and the lava stone sequence. The results are shown in Fig. 13. For the bottle sequence, the average accuracy of the determined absolute scale (represented by the inferred diameter of the bottle as outlined in Section 4.1) is better than 3.0% already after 12 processed images. However, at the beginning of the sequence the random scatter across the 100 runs is about two times larger than near the end (after 23 processed images). The final difference between measured and true absolute scale corresponds to 1.9 standard deviations. For the lava stone sequence, the determined absolute scale is about 1.2% (corresponding to one standard deviation) too small after 12 processed images. The deviation becomes larger and appears to be of systematic nature when 13 and more images are processed. The last three images of the lava stone sequence are strongly blurred, which we assume to be the main reason for this behaviour (cf. Section 4.3). For this sequence it is favourable to adopt the 3D reconstruction result obtained after processing 12 images and to avoid utilising the last three, strongly blurred, images.

4.3 Analysis of random errors and systematic deviations

The main source of random errors is the pixel noise of the CCD sensor, which influences the estimation of the PSF radius according to the Depth–Defocus Function given by Eq. (5) and furthermore leads to a random scatter of the extracted feature positions of about 0.1 pixels. According to Table 1, the reprojection error is always significantly larger than the random scatter of the KLT tracker and amounts to several tenths of a pixel. Presumably, systematic deviations are introduced by the changing appearance of tracked features across the sequence which cannot be fully described by affine deformation and thus cannot be fully compensated by the KLT tracker. The 3D reconstruction results obtained with the offline algorithm for 100 runs over the cuboid, bottle, and lava stone sequences, respectively, show that the relative differences between the ground truth and the reconstructed absolute scale of the scene amount to a few percent and always correspond to between 1 and 2 standard deviations (cf. Table 1). Hence, the observed deviations are presumably due to a combination of random fluctuations and systematic errors.

Systematic errors may be introduced at the end of the sequence, where the images tend to be strongly blurred. For PSF radii smaller than 5 pixels we found that the random scatter of the feature positions extracted by the KLT tracker are of the order 0.1 pixels and independent of the PSF radius, such that the extracted feature positions do not introduce systematic errors. However, the observed relation $\sigma(D)$ between PSF radius and depth is accurately represented only for small and intermediate values of σ by the Depth–Defocus Function $\mathcal{S}(D)$ according to Eq. (5). Fig. 2a illustrates that the utilised 12 mm lens shows this effect for values of σ between 2 and 3 pixels on both sides of the minimum of $\mathcal{S}(D)$. Systematic errors might also be introduced by the nonlinearity of the Depth–Defocus Function $\mathcal{S}(D)$. Effectively, the estimation of the depth D is based on an inversion of Eq. (5). Even if we assume that the measurement errors of $\sigma(D)$ for a certain depth D can be described by a Gaussian distribution of zero mean (which is a good approximation to the observed behaviour), the statistical properties of the inverse relation $D(\sigma)$ generally cannot be described in terms of a zero-mean Gaussian distribution. Due to the nonlinear nature of $\mathcal{S}(D)$, the average deviation between the measured depth value D and its value predicted by the Depth–Defocus Function deviates from zero. For small PSF radii, i.e. close to the inflexion points of the Depth–Defocus Function $\mathcal{S}(D)$, where its curvature is close to zero and its shape is largely linear, this effect is only minor, but its importance increases for large PSF radii, where $\mathcal{S}(D)$ displays a strong curvature as apparent in Fig. 2b. For the lava stone sequence processed with the online algorithm, Fig. 14 illustrates the correlation between scale error and average PSF radius of the last processed image. The systematic effect is especially pronounced for this sequence since it comprises only 15 images (cf. Table 1). Measurement errors

obtained while processing the last three images, which are strongly blurred with $\sigma > 2$ pixels, thus have a substantial effect on the 3D reconstruction result. These findings suggest that features with large associated PSF radii should be excluded from the three-dimensional reconstruction process, where the range of favourable PSF radii depends on the Depth–Defocus Function $\mathcal{S}(D)$.

A further important source of systematic errors is the thermal expansion of the optical system. The body of the lens used for our experiments consists of Aluminium, having a relative thermal expansion coefficient of $\nu = 2.3 \times 10^{-5}\ \mathrm{K}^{-1}$. We assume that with a lens of focal length f at calibration temperature T_0 an image of maximum sharpness is observed at depth D_0 and that the focal length f is constant. The corresponding image distance v_0 is obtained according to the lens law given by Eq. (2). Assuming that the measurement is performed at temperature T, the thermal expansion of the lens body yields an image distance $v(T) = [1 + \nu(T - T_0)]v_0$, and the corresponding depth $D(T)$ for which an image of maximum sharpness is observed at temperature T is computed according to the lens law Eq. (2). As a result, the Depth–Defocus Function Eq. (5) is shifted by the amount $D(T) - D_0$ along the D axis (cf. Fig. 2), which introduces a corresponding systematic error. We find that for a given temperature difference $|T - T_0| \ll T_0$, the relative systematic deviation $[D(T) - D_0]/D_0$ of the DfD measurement is largely proportional to D_0, and for a given value of $D_0 \gg f$, it is largely proportional to $|T - T_0|$. As an example, for a focal length $f = 20$ mm, a depth $D_0 = 1000$ mm, and a temperature difference $|T - T_0| = 10$ K, we obtain a relative systematic deviation of the DfD measurements of 1.1%, which is of the same order of magnitude as the relative reconstruction errors observed in our experiments.

Further possible sources of systematic deviations are vibrations and shocks occurring after calibration (which we avoided during our experiments) and systematic variations of the appearance of the extracted ROIs across the image sequence especially for specular surfaces (cf. Section 4.1.4) or when the assumption of affine deformation does not hold. In general, such influences are difficult to quantify, but they may lead to systematic errors of at least the same order of magnitude as those inferred for thermal expansion.

5 Summary and conclusion

We have described a method for combining geometric and real-aperture methods for monocular 3D reconstruction of static scenes at absolute scale. The proposed algorithm is based on a sequence of images of the object acquired by a monocular camera of fixed focal setting from different viewpoints. Feature points are tracked over a range of distances from the camera, resulting in a

varying degree of defocus for each tracked feature point. After determining the best focused image of the sequence, we obtain information about absolute depth by a DfD approach. The inferred PSF radii for the corresponding scene points are utilised to compute a regularisation term for an extended bundle adjustment algorithm that simultaneously optimises the reprojection error and the absolute depth error for all feature points tracked across the image sequence. The proposed method yields absolutely scaled 3D coordinates of the object feature points without any prior knowledge about scene structure and camera motion. We have described the implementation of the proposed method as an offline and as an online algorithm.

Based on experiments with real-world objects, we have demonstrated that the offline version of the proposed algorithm yields absolutely scaled 3D coordinates of the feature points with typical relative errors of a few percent. For the online algorithm, the accuracy of 3D reconstruction increases with increasing number of processed images as long as the images do not become strongly blurred. At the end of the sequence, the reconstruction results of the online and the offline versions of the proposed algorithm are of comparable accuracy.

We have shown that the 3D reconstruction inaccuracies observed in our experiments can be explained by a combination of the random scatter of the extracted feature positions and the estimated PSF radii, which are both due to the noise of the pixel greyvalues, and systematic deviations of the order 1% due to thermal expansion of the optical system. Further systematic errors may be introduced if the image sequence contains strongly blurred images with an average PSF radius larger than about 2 pixels, due to deviations of the observed depth dependence of the PSF radius from the analytic model used for the Depth–Defocus Function. Since the PSF radius is continuously computed in the course of the 3D reconstruction process, it is possible and favourable to reject such strongly blurred images accordingly. Especially for specular surfaces, the changing appearance of the ROIs tracked across the sequence may introduce further systematic effects.

Possible application scenarios of our 3D reconstruction approach are in the domain of 3D reconstruction of objects and surfaces for industrial quality inspection. It might also be useful for simultaneous localisation and mapping (SLAM) in mobile robotic systems. Applications in which our approach is preferable to stereo camera systems are industrial machine vision systems with space limitations or where strong vibrations occur (the latter leading to the rapid deadjustment of the relative orientation of a pair of stereo cameras), e.g. a monocular camera on a moving machine.

Extensions of the presented approach might involve embedding the stable block of SfM equations into a frame given by DfD, similar to using low-quality control points (obtained by DfD) for a highly precise network (obtained by

SfM) – the internal relative accuracy of SfM is higher than that of DfD, since otherwise the reprojection error would not strongly increase for high values of the weight parameter α. In such a framework, the accuracy of the determination of each individual observation for SfM as well as for DfD (ideally the inverse covariance matrices of the observations) should be used for robust re-weighting.

References

Bouguet, J., 1997. Camera calibration toolbox for MATLAB. http://www.vision.caltech.edu/bouguetj/calib_doc.

Chaudhuri, S., Rajagopalan, A., 1999. Depth from Defocus: A Real Aperture Imaging Approach. Springer Verlag, Berlin.

d'Angelo, P., Wöhler, C., 2008. Image-based 3d surface reconstruction by combination of photometric, geometric, and real-aperture methods. ISPRS Journal of Photogrammetry and Remote Sensing 63 (3), 297–321.

Ens, J., Lawrence, P., 1993. An investigation of methods for determining depth from focus. IEEE Trans. on Pattern Analysis and Machine Intelligence 15 (2), 97–108.

Grossmann, P., 1987. Depth from focus. Pattern Recognition Letters 5 (1), 63–69.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proc. 4th Alvey Vision Conference, Manchester. pp. 147–151.

Krüger, L., Wöhler, C., Würz-Wessel, A., Stein, F., 2004. In-factory calibration of multiocular camera systems. In: Proc. of the SPIE, Photonics Europe, Automatic Target Recognition XIV. Vol. 5457. pp. 126–137.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. of Computer Vision 60 (2), 91–110.

Madsen, K., Nielsen, H. B., Tingleff, O., 1999. Methods for non-linear least squares problems. http://www2.imm.dtu.dk/pubdb/p.php?660.

McGlone, C., Mikhail, E., Bethel, J., 2004. Manual of Photogrammetry – Fifth Edition. American Society for Photogrammetry and Remote Sensing, Bethesda, MA.

Myles, Z., da Vitoria Lobo, N., 1998. Recovering affine motion and defocus blur simultaneously. IEEE Trans. on Pattern Analysis and Machine Intelligence 20 (6), 652–658.

Pedrotti, F. L., 1993. Introduction to Optics, 2nd Edition. Prentice Hall.

Pentland, A. P., 1982. Depth of scene from depth of field. In: Proc. Image Understanding Workshop. pp. 253–259.

Pentland, A. P., 1987. A new sense for depth of field. IEEE Trans. on Pattern Analysis and Machine Intelligence 9 (4), 523–531.

Rey, W. J. J., 1983. Introduction to Robust and Quasi-Robust Statistical Methods. Springer Verlag, Berlin, Heidelberg.

Scharstein, D., Szeliski, R., Zabih, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. of Computer Vision 47 (1), 7–42.

Shi, J., Tomasi, C., 1994. Good features to track. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition. pp. 593–600.

Subbarao, M., 1989. Efficient depth recovery through inverse optics. In: Freeman, H. (Ed.), Machine Vision for Inspection and Measurement. Academic Press, New York, pp. 101–126.

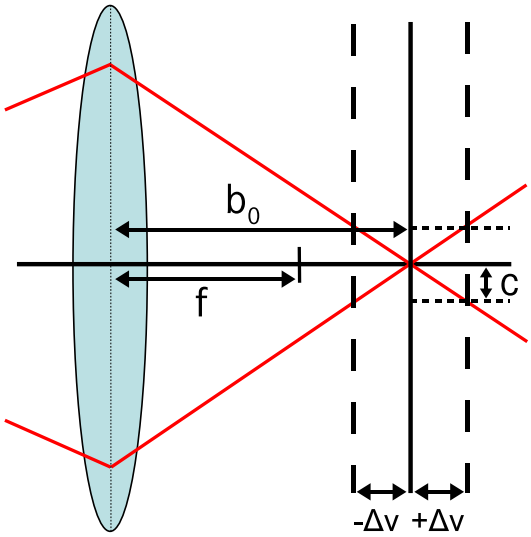
Subbarao, M., Choi, T., 1995. Accurate recovery of three-dimensional shape from image focus. IEEE Trans. on Pattern Analysis and Machine Intelligence 17 (3), 266–274.

Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A., 2000. Bundle adjustment – A modern synthesis. In: Triggs, W., Zisserman, A., Szeliski, R. (Eds.), Vision Algorithms: Theory and Practice. LNCS. Springer Verlag, Berlin, pp. 298–375.

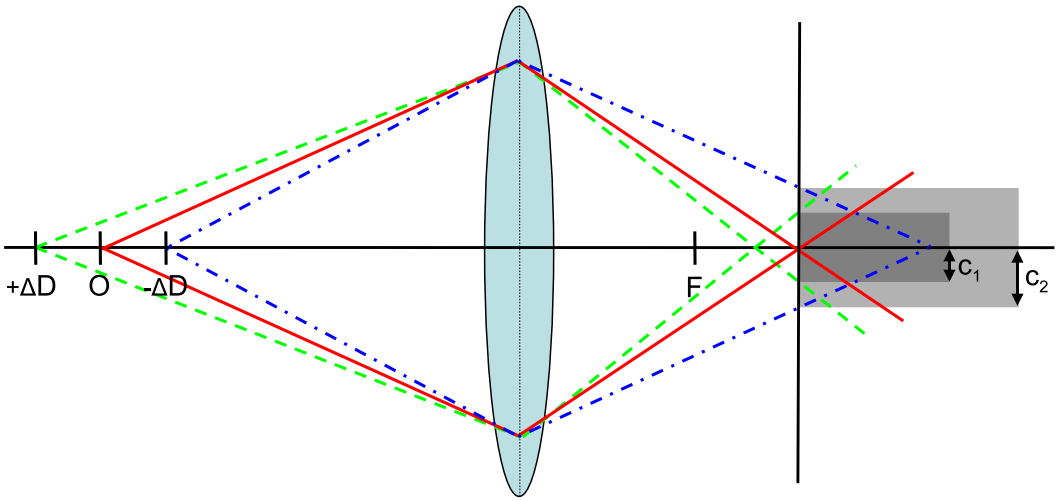
Watanabe, M., Nayar, S., Noguchi, M., 1995. Real-time computation of depth from defocus. In: Proc. of the SPIE, Three-Dimensional and Unconventional Imaging for Industrial Inspection and Metrology. Vol. 2599:A-03. pp. 14–25.

Table 1
Summary of the evaluation results for the offline algorithm.

Sequence	Length	$E_{\text{reprojection}}$ [pixels/point]	E_{defocus} [pixels/point]	Reference length [mm]	
				Ground truth	3D reconstruction
Cuboid	46	0.642	0.636	32.0	34.1 ± 1.6
Bottle	26	0.747	0.387	80.0	82.8 ± 1.4
Lava stone	15	0.357	0.174	60.0	58.3 ± 0.8
Flange	36	1.06	1.96	15.2	15.45 ± 0.01



(a) Symmetric dependence of c on Δv.



(b) Asymmetric dependence of c on ΔD.

Fig. 1. Dependence of the diameter c of the circle of confusion (a) on the offset Δv in image space and (b) on the offset ΔD in object space. The offsets Δv and ΔD are measured with respect to the principal distance v and the distance D between lens and object for the perfectly focused scenario described by the lens law Eq. (2). The value of c increases more strongly for motion towards the camera than for motion away from the camera.

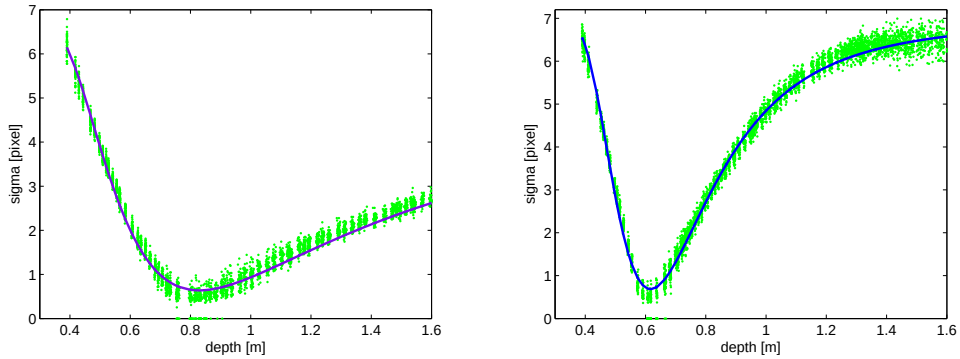


Fig. 2. Depth–Defocus Functions of two lenses with $f = 12$ mm (left) and $f = 20$ mm (right), fitted to the measured data points according to Eq. (5), respectively.

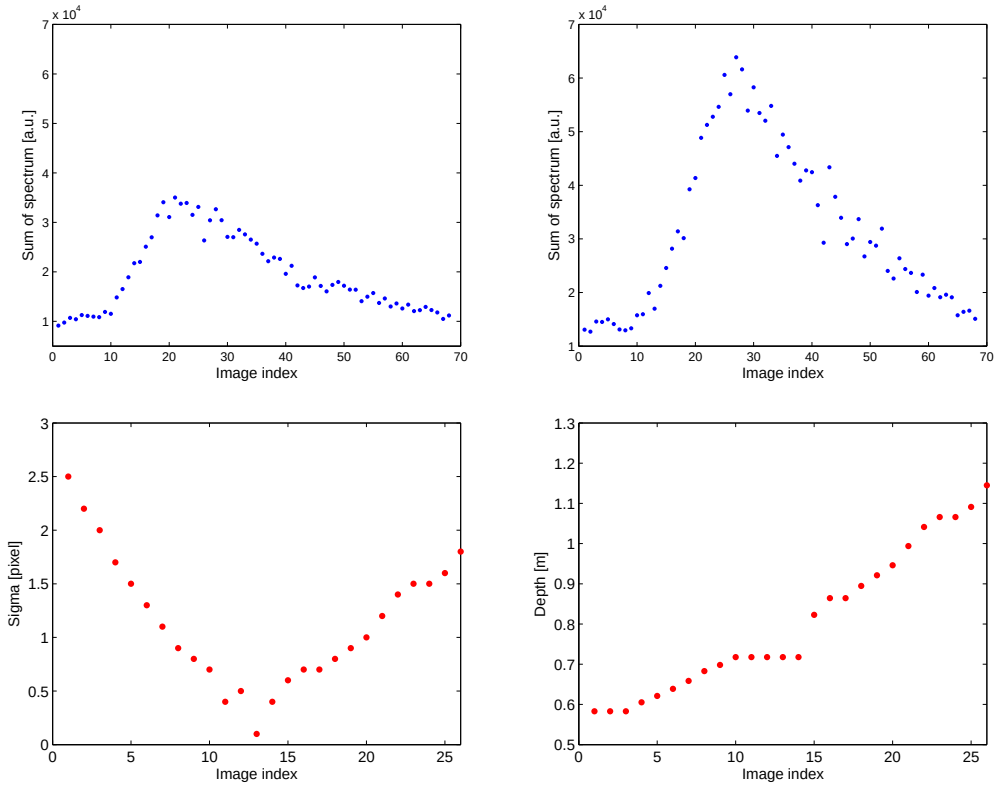


Fig. 3. Upper left and upper right: Image index vs. defocus measure H for two different tracked image features. Lower left: Image index vs. PSF radius σ. Lower right: Image index vs. inferred depth D.

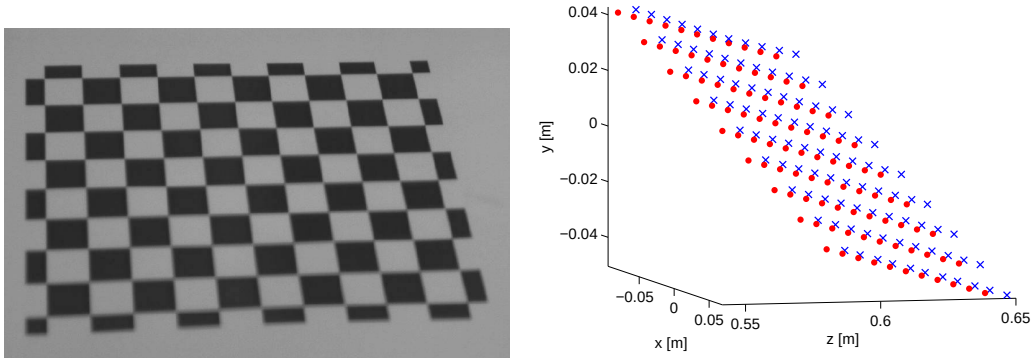


Fig. 4. True (dots) and reconstructed (crosses) 3D pose of the chequerboard ($\alpha = 0.42$).

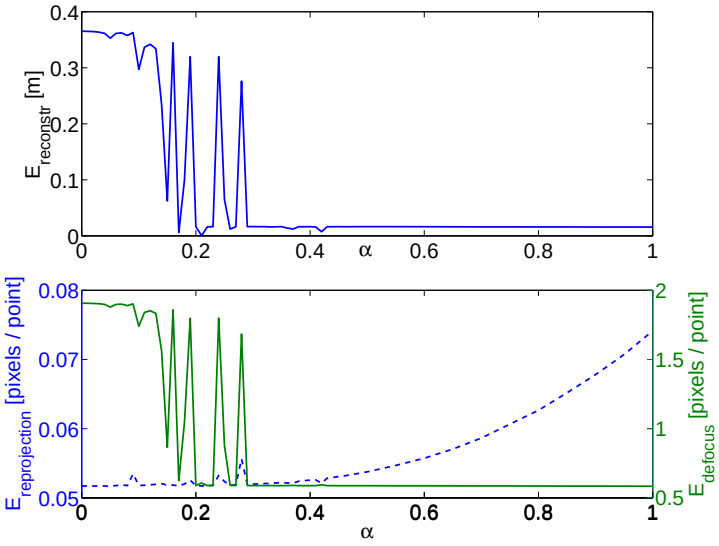


Fig. 5. Dependence of E_{reconstr} (upper diagram), $E_{\mathrm{reprojection}}$ (lower diagram, dashed curve, left axis), and E_{defocus} (lower diagram, solid curve, right axis) on the weight parameter α.

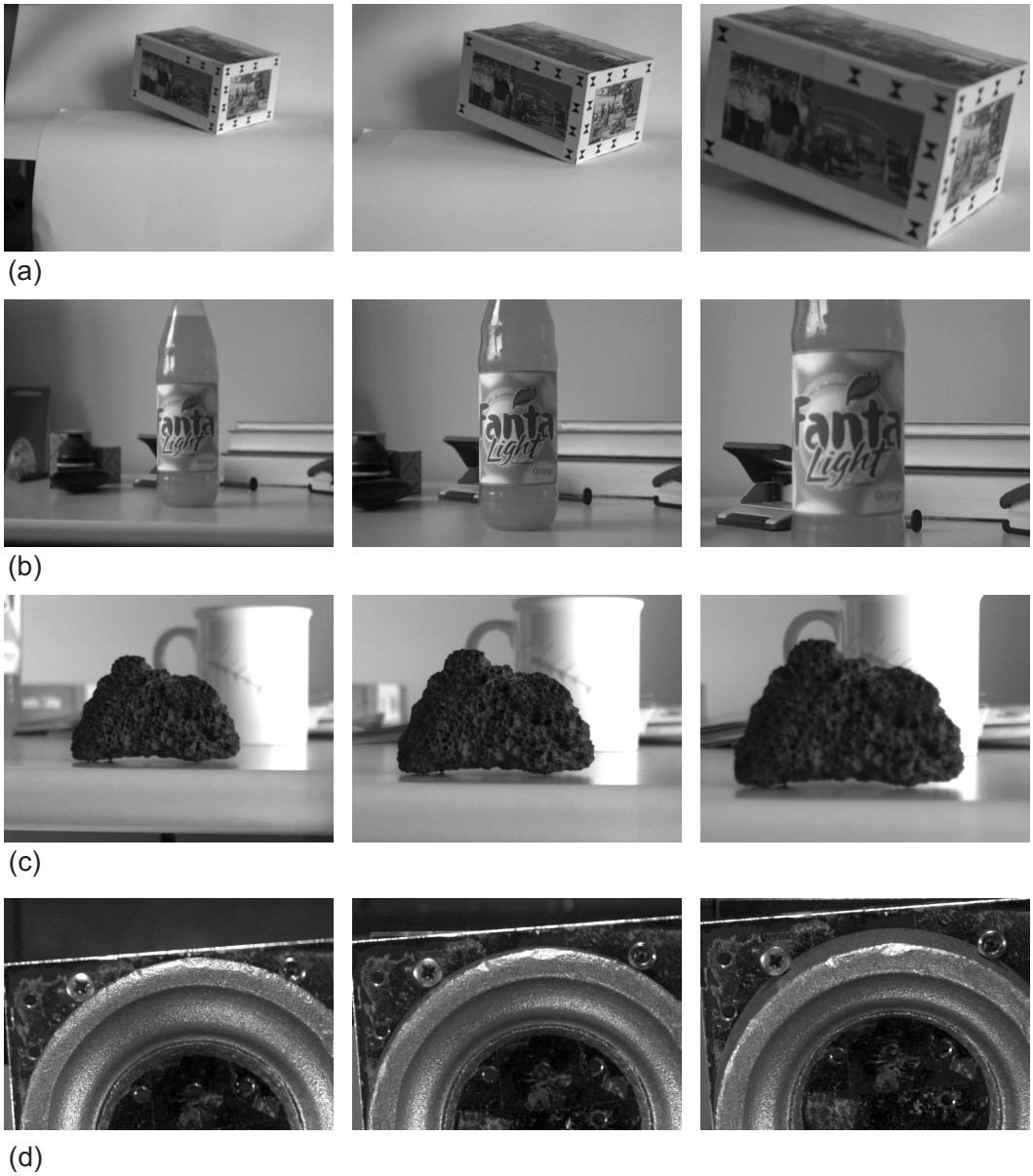


Fig. 6. Images from the beginning, the middle, and the end of (a) the cuboid sequence, (b) the bottle sequence, (c) the lava stone sequence, and (d) the flange sequence.

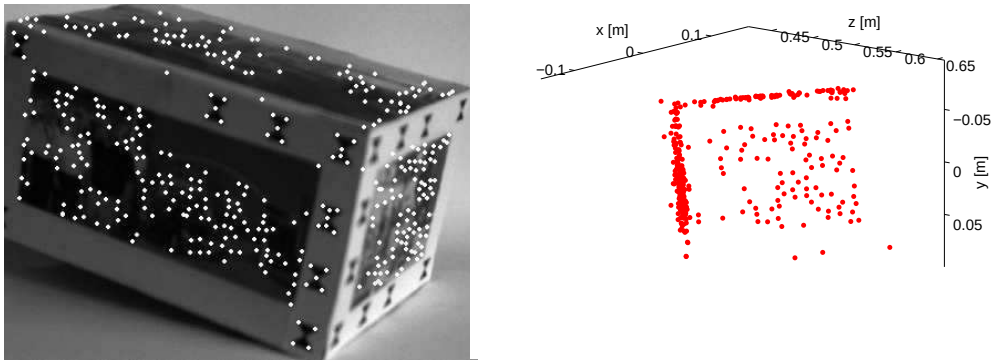


Fig. 7. 3D reconstruction of the cuboid.

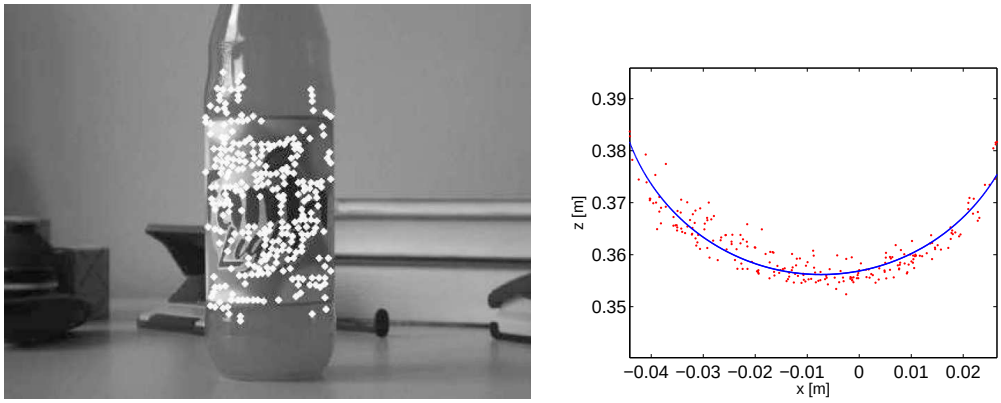


Fig. 8. 3D reconstruction of the cylindrical surface of the bottle.

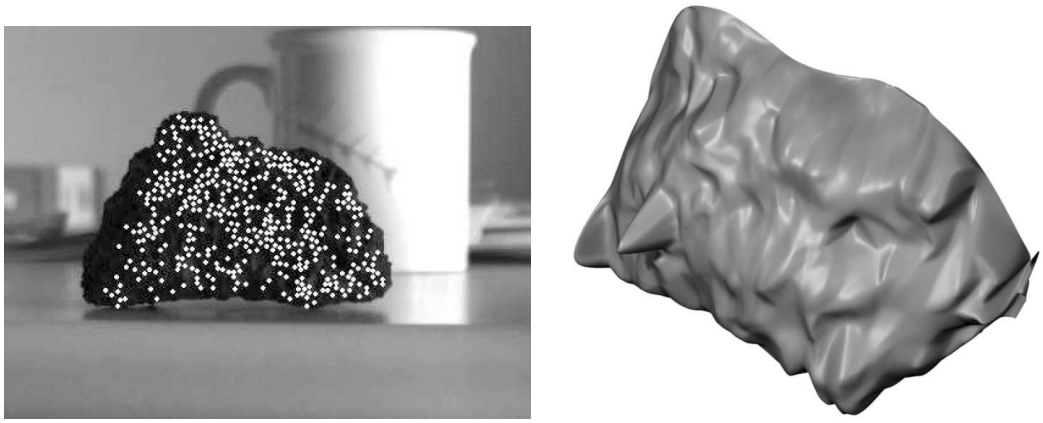


Fig. 9. 3D reconstruction of the lava stone. The cusp visible in the left part of the reconstructed surface is produced by three outlier points.

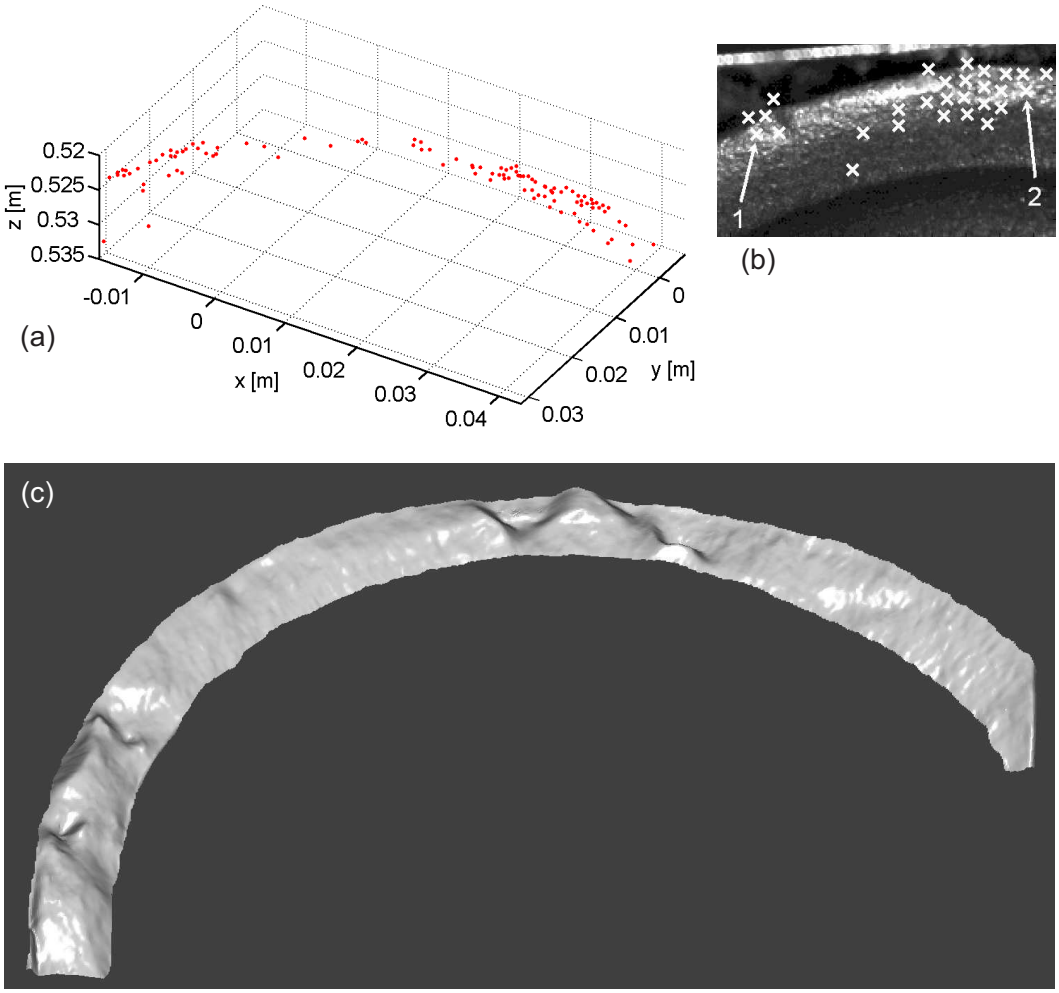


Fig. 10. 3D reconstruction of the raw cast iron surface of a flange. (a) Reconstructed 3D points. (b) Location of the reference points used to determine the accuracy of the estimated absolute scale. (c) 3D surface profile obtained by using the reconstructed 3D points as an input to the combined geometric and photometric approach by d'Angelo and Wöhler (2008). For further details, cf. Section 4.1.4.

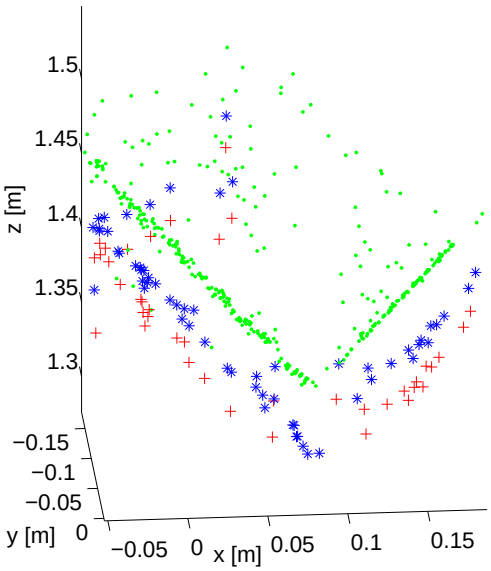


Fig. 11. 3D reconstruction of the cuboid, obtained with the online algorithm. The 3D reconstruction result is displayed after 5 (crosses), 9 (stars), and 25 (dots) iterations. The first iteration is performed after the 21st image.

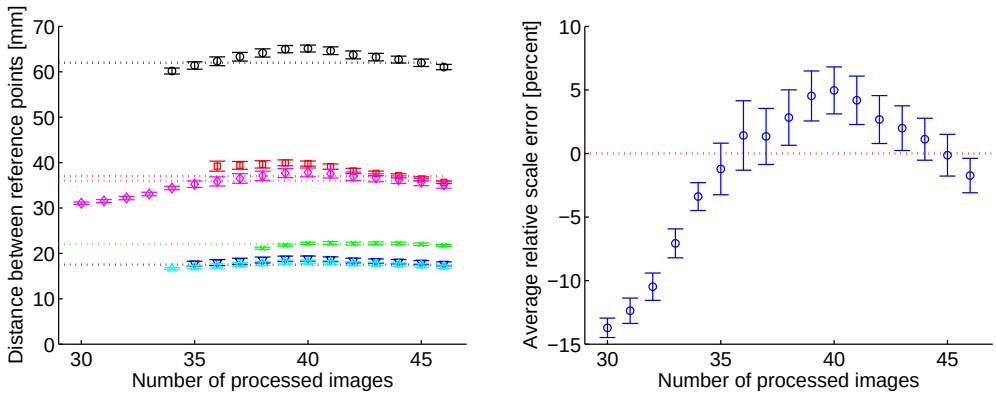


Fig. 12. Left: Behaviour of the distances between six pairs of reference points on the cuboid surface, obtained with the online algorithm, for increasing number of iterations, compared to the corresponding true values. The standard deviations across the 100 online runs are indicated by error bars. Right: Relative accuracy of the inferred absolute scale, given by the average relative deviation between the measured and true absolute distances between those pairs of reference points already having passed their point of maximum sharpness, for increasing number of iterations.

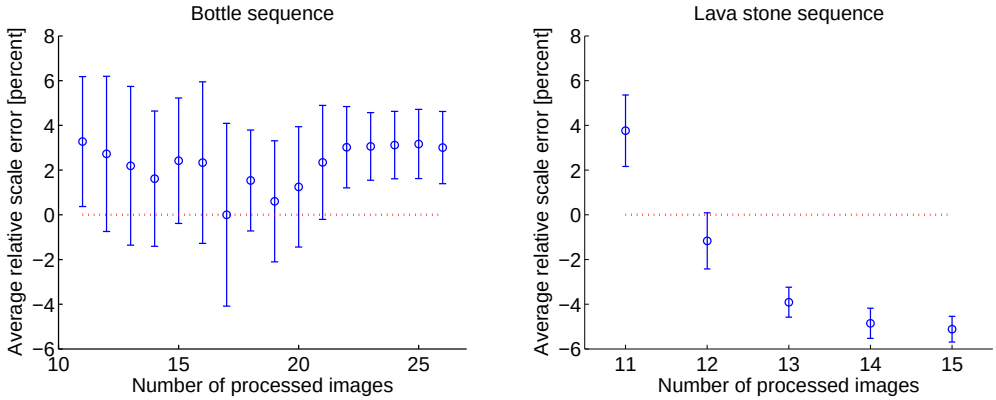


Fig. 13. Relative deviation between measured and true absolute scale for increasing number of processed images for the bottle sequence (left) and the lava stone sequence (right).

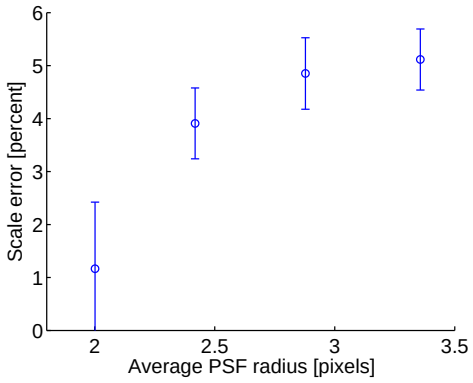


Fig. 14. Correlation between scale error and average PSF radius for the last 4 images of the lava stone sequence.