Pattern Recognition - Proc. 28th DAGM Symposium, Berlin, Germany. Lecture Notes in Computer Science 4174, pp. 607-616, Springer-Verlag Berlin Heidelberg, 2006

Monocular 3D Scene Reconstruction at Absolute Scales by Combination of Geometric and Real-Aperture Methods

Annika Kuhl[1,2], Christian Wöhler[1], Lars Krüger[1], Pablo d'Angelo[1], and Horst-Michael Groß[2]

[1] DaimlerChrysler AG, Group Research, Machine Perception
P. O. Box 2360, D-89013 Ulm, Germany
[2] Technical University of Ilmenau, Faculty of Computer Science and Automation
P. O. Box 100565, D-98684 Ilmenau, Germany

Abstract. We propose a method for combining geometric and real-aperture methods for monocular 3D reconstruction of static scenes at absolute scales. 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 the structure of the scene. Evaluating the algorithm on real-world data we demonstrate that it yields typical relative errors between 2 and 3 percent. Possible applications of our approach are self-localisation and mapping for mobile robotic systems and pose estimation in industrial machine vision.

1 Introduction

The knowledge of three-dimensional structure plays an important role in many fields of research such as navigation, obstacle avoidance, and object detection. Depth-from-Stereo [1] was one of the first methods for recovering depth information as it is inspired by human vision. Hereby 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 stereo vision is its 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 a preferrable solution. Structure-from-Motion is such an alternative: From corresponding points in at least two images acquired sequentially at different camera positions the spatial positions of the

K. Franke et al. (Eds.): DAGM 2006, LNCS 4174, pp. 607–616, 2006.

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. In [2] image features are proposed that serve well for tracking algorithms. Widely used methods are SIFT features [3], involving the extraction of scale invariant features using a staged filtering approach, or the Kanade-Lucas-Tomasi (KLT) feature detector described in [4] which is based on the Harris corner detector and takes into account affine motion.

A different approach to scene reconstruction utilises position variant appearance, e.g. Shape-from-Shading [5], Depth-from-Defocus [6], and Depth-from-Focus [7]. Depth-from-Defocus 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. In [8] this property is used to estimate depth simultaneously for all scene points from only one or two images. Depth information is extracted out of a single image showing sharp discontinuities (edges) [9]. A survey of existing methods is given in [6]. In [10] a method is proposed that computes Depth-from-Defocus in real-time using structured lighting. 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 [11]. Further work in this field includes Shape-from-Focus [12] and Inverse Optics [13].

So far, no attempt has been made to combine the precise relative scene reconstruction of Structure-from-Motion with the absolute depth data of Depth-from-Defocus. A work related to this paper was published in [14], where a method to recover affine motion and defocus simultaneously is proposed. However, the spatial extent of the scene is not reconstructed in [14], since planar objects are a requirement for the described method.

The main contribution of this paper consists of a novel combination of Structure-from-Motion (a geometric method) with Depth-from-Defocus (a real-aperture method). We will show that the combination of these methods yields a 3D scene reconstruction at absolute scales based on an image sequence acquired with a monocular camera.

2 Structure-from-Motion and Depth-from-Defocus

Structure-from-Motion recovers the spatial scene structure using a monocular camera. A pre-requisite for Structure-from-Motion is the geometric calibration of the camera in terms of estimating the internal parameters such as focal length, distortion parameters, etc. [15]. Subsequently, salient feature points are extracted and tracked across the sequence. The motion of these features relative to the camera is then used to minimise the Bundle Adjustment [16] error term

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

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. To facilitate the integration of defocus information into the Structure-from-Motion framework, the image sequences are acquired such that the object is blurred in the first image of the sequence, becoming increasingly focused in the middle and blurred again in the last images. 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.

Depth-from-Defocus 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 $A(r) \propto [J_1(r)/r]^2$, where $J_1(r)$ is a Bessel function of the first kind [17]. For practical purposes, however, 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 [6]. 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. The classical Depth-from-Defocus approach uses two images of the same object taken at two different focal settings [6]. In [9] it is shown 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 no such a-priori information is available, we suggest the emprirical determination of the so-called Depth-Defocus-Function, expressing the standard deviation σ of the Gaussian PSF as a function of depth D, based on a calibration procedure.

3 Spatial Scene Reconstruction by Combining Structure-from-Motion and Depth-from-Defocus

3.1 The Depth-Defocus-Function and Its Calibration

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 $v^{-1} + D^{-1} = f^{-1}$ [17]. An object at distance D is focused if the distance between lens and image plane is v, with f denoting the focal length of the lens. Varying the image plane distance v 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 diameter $|\Delta v|/\kappa$

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:

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

Here, the amount of defocus is described in terms of the radius σ of the Gaussian PSF. But since the Depth-Defocus-Function expresses the relation between the depth of an object and its defocus, 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. But since neither D nor ΔD are known, the functional relation needs to be modelled with respect to Δv:

$$\frac{1}{v + \Delta v} + \frac{1}{D} = \frac{1}{f} \ . \tag{3}$$

A value of $\Delta v \neq 0$ refers to a defocused object point. Solving Eq. (3) for Δv and inserting Δv in Eq. (2) yields the Depth-Defocus-Function

$$\mathcal{S}(D) = \frac{1}{\phi_1} e^{-\frac{1}{\phi_2}\left(\frac{fD}{D-f} - v\right)^2} + \phi_3 \ . \tag{4}$$

Calibrating the Depth-Defocus-Function $\mathcal{S}(D)$ for a given lens corresponds to determining the parameters ϕ_1, ϕ_2, and ϕ_3 in Eq. (4). This is achieved by obtaining a large set of measured (σ, D) data points and perform a least mean squares fit to Eq. (4), where D is the distance from the camera and σ the radius of the Gaussian PSF G used to blur the well focused image according to $I_{ij} = G(\sigma) * I_{if_i}$. Here, I_{if_i} represents a small region of interest (ROI) around feature i in the image f_i in which this feature is best focused, and I_{ij} a ROI of equal size around feature i in image j.

For calibration, an image sequence is acquired while the camera approaches at uniform speed a calibration rig displaying a checker board. The sharp black-and-white corners of the checker board are robustly and precisely detectable [15] even in defocused images. Small ROIs around each corner allow the estimation of defocus using their greyvalue variance χ. The better focused the corner, the higher is the variance χ. We found experimentally that the parameterised defocus model according to Eq. (4) 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. (4) is fitted to the measured (χ, j) data points for each corner i, such that 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 from the pose of the complete rig according to [18].

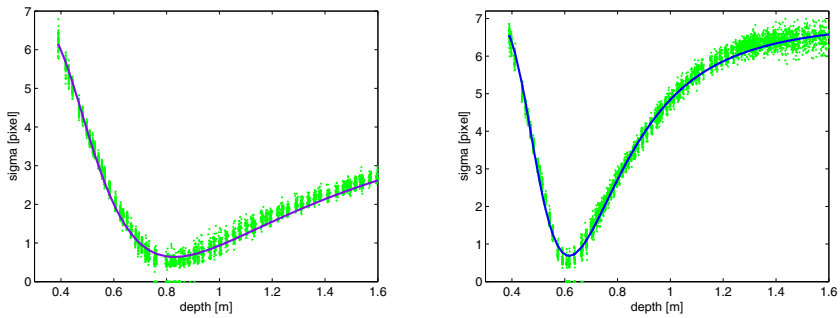

Fig. 1. Depth-Defocus-Functions of two lenses with $f = 12$ mm (left) and $f = 20$ mm (right), fitted to the measured data points according to Eq. (4), respectively

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}. By employing the bisection method, we determine the value of σ for which the root mean square deviation between $G(\sigma) * I_{if_i}$ and I_{ij} 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. 1 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.

3.2 Combining Motion, Structure, and Defocus

The Structure-from-Motion analysis involves the extraction of salient features from the image sequence which are tracked using the KLT technique [4]. To facilitate 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 the 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 ω_0 and ω_1 is taken into account according to $H = \int_{\omega_0}^{\omega_1} |\mathscr{F}_I(\omega)| d\omega$ with $\omega_0 = \frac{1}{4}\omega_{\max}$ and $\omega_1 = \frac{3}{4}\omega_{\max}$, where $\omega_{\max}$ is the maximum frequency. The amount of defocus increases with decreasing value of H. The defocus measure H

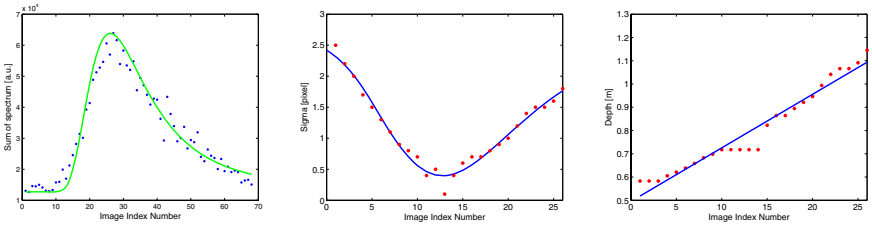

Fig. 2. From the left: Image index vs. defocus measure H for a tracked image feature; image index vs. PSF radius σ; image index vs. inferred depth D

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 3.1. 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 3.1. The depth D is obtained by inverting the Depth-Defocus-Function $\mathcal{S}(D)$ according to Eq. (4). 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 Structure-from-Motion 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. 2.

A general property of the KLT algorithm is that the accuracy of the feature tracker decreases with increasing defocus of the reference pattern. Hence, the feature positions are refined by repeating the tracking procedure for all features, starting from the "sharpest" image located near the middle of the sequence which displays the largest value of H averaged over all 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 searching for the minimum of the combined error term

$$E_{\text{comb}}\left(\{T_j\},\{X_i\}\right) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left[\left(\mathcal{P}\left(T_j X_i\right) - x_{ij}\right)^2 + \alpha\left(\mathcal{S}([T_j X_i]_z) - \sigma_{ij}\right)^2\right] \quad (5)$$

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

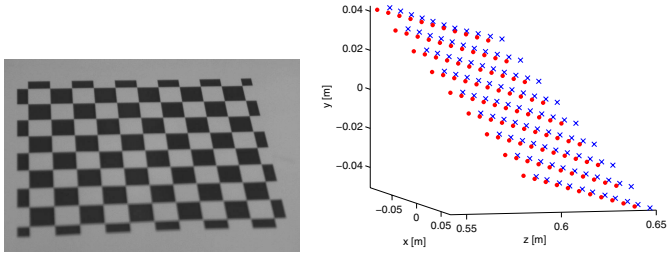

Fig. 3. True (dots) and reconstructed (crosses) 3D pose of the checker board ($\alpha = 0.42$)

a regularisation term in Eq. (5), such that absolutely scaled 3D coordinates X_i of the scene points are obtained. The values of X_i are initialised according to the depth values estimated based on the Depth-from-Defocus approach. To increase the accuracy of the reconstructed 3D scene points, we only make use of feature positions extracted from images in which the feature is not strongly blurred. To minimise the error term E_{comb} the Levenberg-Marquardt algorithm [19] is employed.

4 Experimental Evaluation

In order to validate our approach we first reconstructed a planar object with known ground truth, using a Baumer 1032×776 pixels CCD camera. A checker board as shown in Fig. 3 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 in [15] to assure sub-pixel accuracy. The true pose of the checker board is obtained according to [18] based on the given size of the squares. Note that in [18] the true pose of the checker board is determined 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 true pose of the object is actually a comparison between two methods conducting different least mean squares fits.

The deviation E_{rec} of the reconstructed 3D scene point coordinates X_i from the ground truth values X_i^{true} is given by $E_{\text{rec}} = \left[\frac{1}{N}\sum_{i=1}^{N}\left(X_i - X_i^{\text{true}}\right)^2\right]^{1/2}$. To determine an appropriate weight parameter α we computed E_{rec} for different α values in the range between 0 and 1. For $\alpha = 0$ the global minimisation is equivalent to Structure-from-Motion initialised with the calculated Depth-from-Defocus values. One must keep in mind, however, that the absolute scaling factor is then part of the gauge freedom of the Bundle Adjustment method, resulting in a corresponding "flatness" of the error function. Small α values lead to an instable convergence. The value of E_{rec} 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

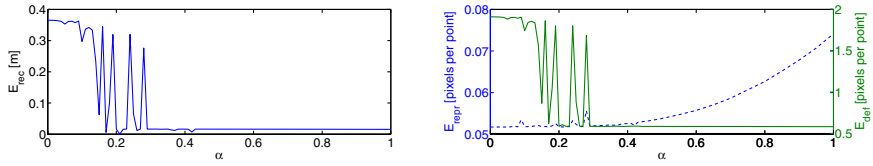

Fig. 4. Dependence of E_{rec} (left diagram), E_{repr} (right diagram, dashed curve, left axis), and E_{def} (right diagram, solid curve, right axis) on the weight parameter α

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. 3, the dependence of E_{rec} on α in Fig. 4 (left).

In addition to the reconstruction error E_{rec}, a further important error measure is the reprojection error $E_{\text{repr}} = \left[\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\mathcal{P}\left(T_j X_i\right) - x_{ij}\right)^2\right]^{1/2}$ 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 $E_{\text{def}} = \left[\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\mathcal{S}([T_j X_i]_z) - \sigma_{ij}\right)^2\right]^{1/2}$. Fig. 4 (right) shows the relation between the weight parameter α, the reprojection error E_{repr}, and the defocus error E_{def}. For $\alpha > 0.3$ the defocus error stabilises to 0.58 pixels per feature. Larger α values lead to a stronger influence of the Depth-from-Defocus values on the optimisation result, leading to an increasing reprojection error E_{repr} due to the inaccuracy of the estimated σ_{ij} values. Although the depth values derived by Depth-from-Defocus 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 scales without prior knowledge. For constant f-stop number, pixel size, and relative accuracy of the inferred depth D, it can be shown that the required focal length and aperture of the lens are largely proportional to $\sqrt{D}$ (proof omitted here). Hence, our approach is restricted to the close-range domain ($D \sim 1$ m) as long as standard video cameras and lenses are used.

In order to demonstrate the performance of our approach on a non-planar test object of known dimensions we applied our method to the cuboid-shaped object shown in Fig. 5. This object displays a sufficient amount of texture to generate "good features to track" [4]. In addition, black markers on white background with known mutual distances are placed near the edges of the cuboid. As described in Section 3.2, feature points are extracted and tracked using the KLT algorithm, and the 3D coordinates of the scene points are obtained by minimising the error term E_{comb} according to Eq. (5).

The reprojection error E_{repr} for $\alpha = 0.5$ amounts to 4.99 pixels. After removing tracking outliers (detected by their associated very large reprojection errors of more than $3E_{\text{repr}}$) the value of E_{repr} drops to 1.08 pixels while E_{def} amounts

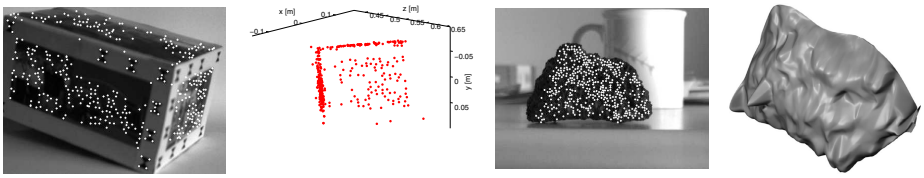

Fig. 5. 3D reconstruction of a cuboid and a lava stone ($\alpha = 0.5$)

to 0.24 pixels. In order to verify the absolute scale, we compared for $\alpha = 0.5$ 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 three pairs of markers with an average true distance of 23.3 mm. The corresponding reconstructed average distance amounts to 23.9 mm, which is 2.6% larger than the ground truth value.

As a real-world object, we examined the lava stone shown in Fig. 5. The resulting reprojection error E_{repr} amounts to 2.77 pixels. After outlier rejection, E_{repr} decreases to 0.96 pixels while E_{def} amounts to 0.19 pixels. The reconstructed shape of the lava stone was again obtained with $\alpha = 0.5$. The reconstruction is approximately 2.3% larger than the real object.

In all examples, the fact that the reconstructed absolute scale of the scene appears to be systematically somewhat too large is likely due to a slight deadjustment of the camera lens after calibration, which may readily occur for standard video lenses as a consequence e.g. of vibrations or variable ambient temperature.

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 scales. 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 Depth-from-Defocus 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 the scene structure. We have demonstrated experimentally that the proposed algorithm yields absolutely scaled 3D coordinates of the feature points with typical relative errors between 2 and 3 percent. Possible application scenarios of our approach are in the domains of self-localisation and mapping for mobile robotic systems as well as pose estimation in the context of industrial machine vision tasks.

References

1. Scharstein, D., Szeliski, R., Zabih, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: IEEE Workshop on Stereo and Multi-Baseline Vision. (2001)
2. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of The Fourth Alvey Vision Conference, Manchester. (1988) 147–151
3. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proc. of the International Conference on Computer Vision ICCV, Corfu. (1999) 1150–1157
4. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle (1994)
5. Horn, B.K.: Robot Vision. McGraw-Hill Higher Education (1986)
6. Chaudhuri, S., A.R.: Depth from Defocus: A Real Aperture Imaging Approach. Springer Verlag, Berlin (1999)
7. Ens, J., Lawrence, P.: An investigation of methods for determining depth from focus. IEEE Trans. Pattern Anal. Mach. Intell. **15** (1993) 97–108
8. Pentland, A.P.: Depth of scene from depth of field. In: Proc. Image Understanding Workshop. (1982) 253–259
9. Pentland, A.P.: A new sense for depth of field. IEEE Trans. Pattern Anal. Mach. Intell. **9** (1987) 523–531
10. Watanabe, M., Nayar, S., Noguchi, M.: Real-time computation of depth from defocus. In: Proc. of SPIE: Three-Dimensional and Unconventional Imaging for Industrial Inspection and Metrology. (1995)
11. Grossmann, P.: Depth from focus. Pattern Recogn. Lett. **5** (1987) 63–69
12. Subbarao, M., Choi, T.: Accurate recovery of three-dimensional shape from image focus. IEEE Trans. Pattern Anal. Mach. Intell. **17** (1995) 266–274
13. Subbarao, M.: Efficient depth recovery through inverse optics. In Freeman, H., ed.: Machine Vision for Inspection and Measurement. Academic Press, New York (1989) 101–126
14. Myles, Z., da Vitoria Lobo, N.: Recovering affine motion and defocus blur simultaneously. IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998) 652–658
15. Krüger, L., Wöhler, C., Würz-Wessel, A., Stein, F.: In-factory calibration of multiocular camera systems. In: Photonics Europe, Automatic Target Recognition XIV., Proceedings of the SPIE. Volume 5457. (2004) 126–137
16. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment – A modern synthesis. In Triggs, W., Zisserman, A., Szeliski, R., eds.: Vision Algorithms: Theory and Practice. LNCS. Springer Verlag, Berlin (2000) 298–375
17. Pedrotti, F.L.: Introduction to Optics, 2nd Edition. Prentice Hall (1993)
18. Bouguet, J.: Camera calibration toolbox for MATLAB. www.vision.caltech.edu/bouguetj/calib_doc (1997)
19. Madsen, K., Nielsen, H.B., Tingleff, O.: Methods for non-linear least squares problems. http://www2.imm.dtu.dk/pubdb/p.php?660 (1999)