Fusion of Combined Stereo and Focus Series for Depth Estimation

Ioana Gheţa[1], Christian Frese[1], Michael Heizmann[2]

[1]Lehrstuhl für Interaktive Echtzeitsysteme, Institut für Technische Informatik
Universität Karlsruhe (TH), Adenauerring 4, 76131 Karlsruhe, {gheta;frese}@ies.uni-karlsruhe.de

[2]Fraunhofer IITB Institut für Informations- und Datenverarbeitung
Fraunhoferstraße 1, 76131 Karlsruhe, heizmann@iitb.fraunhofer.de

Abstract: Estimating depth maps is an important step towards 3D reconstruction. In order to compensate the insufficiencies that different image based methods impose, a fusion of multiple information sources for depth estimation is proposed. This article concentrates on obtaining depth information by fusing combined stereo and focus series, which are efficiently gained using a multivariate camera array.

1 Introduction

A continuous challenge in visual inspection is the 3D reconstruction of scenes and, implicitly, the estimation of depth information. In literature, numerous approaches that use image series acquired by varying one imaging parameter can be found; such as depth from stereo (different camera positions) [FL01] or depth from focus (different focus settings) [PB97]. Each of them imposes constraints that can be overcome by the fusion of multivariate image series, i. e. images with more than one modified parameter. This contribution concentrates on the fusion of combined stereo and focus series.

For the image acquisition, we propose a camera array consisting of several sensors that can be configured differently. In our application, the varied parameters are the camera positions and the focus settings. Due to the improved information content and to the merging of algorithms for different parameter variations, more robust results are achievable. For extending the potential of the camera array, the system can be variably positioned by a robot, e. g. for gaining panoramic depth images.

2 Depth estimation from varying one acquisition parameter

Depth from stereo: This approach provides a dense depth map, i. e. reliable depth information is obtained for each image pixel, except for occlusions, unstructured regions or regions with periodical structures. A common way to formulate the stereo fusion employs energy functionals on the depth map f [KZ02]. The solution is achieved by minimizing

$$E_{\text{stereo}}(f) = E_{\text{data}}(f) + E_{\text{smoothness}}(f) + E_{\text{visibility}}(f)\,. \tag{1}$$

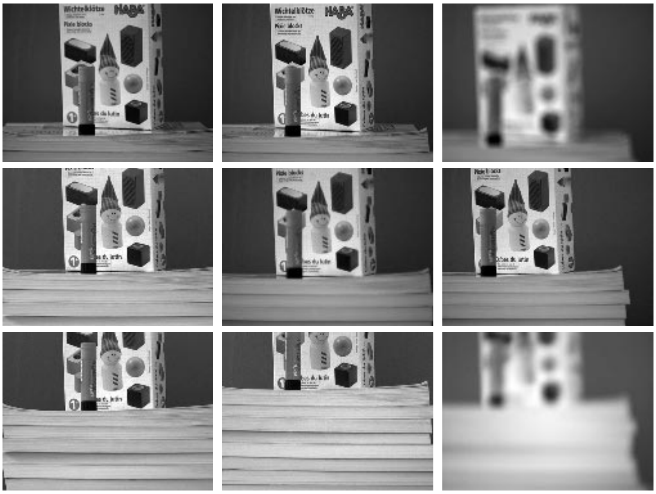

Figure 1: Combined stereo and focus series.

The data term $E_{\text{data}}(f)$ considers only corresponding pixels and ensures the compatibility with the data set by enforcing photo-consistency, i. e. corresponding pixels have similar gray values. The smoothness term $E_{\text{smoothness}}(f)$ is designed such that the balance is kept between the constraints that disparity is piecewise constant and that discontinuities often coincide with intensity edges by considering gray values and disparities of neighboring pixels. The visibility term $E_{\text{visibility}}(f)$ penalizes impossible geometrical configurations. The cost functions in the energy terms are modeled according to the epipolar constraints in the geometry of multiple views [FL01].

Depth from focus: Depth information from focus series is obtained by evaluating local blur in the images. Two approaches are common: depth from focus searches the sharpest image in the series, whereas depth from defocus models and evaluates the blur as a function of depth in at least one image.

In the depth from focus approach, the image with the optimal focus is determined. Since defocused image formation is a low-pass operator, the best focused image of the scene contains the maximum of high frequency components. For each position, the image of the series with a maximum focus measure represents the best focused view of the corresponding scene point. Depth is then reconstructed from the known camera parameters.

Depth from focus requires “dense” image series, i. e. the changes in the focus settings must be minimal from one image to the next. The literature provides a large number of algorithms, however for image series acquired with only one camera, i. e. collocated sensors [NN94, SC95]. In the case of the camera array, there are two problems to overcome: the number of focus positions is limited by the number of cameras and the images in the series are taken from different point of views. The first problem can be solved by using more cameras. Solutions for the second one are given in Sect. 3.

Depth from defocus: To use the depth from defocus approach, this contribution models

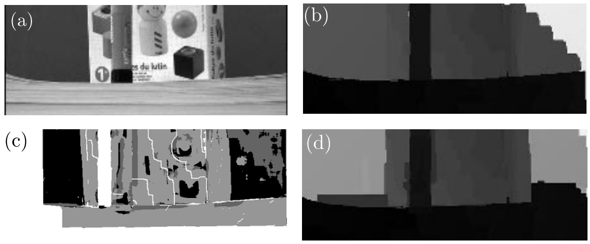

Figure 2: Fusion of the image series in Fig. 1: (a) Source image; (b) Depth map from stereo fusion (dark: near, light: remote); (c) Confidence of the depth map (b): regions without structures cause inconsistencies (labeled black) or low confidence (labeled dark gray); (d) Improved depth map obtained from fusion of stereo and focus series.

the blur on image edges as a space variant convolution with a pulse response. In practice, geometrical and chromatic lens aberrations cause the point spread function to differ from the ideal sharply bounded disk. Considering the central limit theorem of statistics, incorporation of the different sources of errors to the geometric model justifies a Gaussian point spread function, which has been proposed in literature [Pen87, SN88]:

$$h_\sigma(p_\mathrm{x}, p_\mathrm{y}) := \frac{1}{2\pi\sigma^2}\, \mathrm{e}^{-\frac{p_\mathrm{x}^2 + p_\mathrm{y}^2}{2\sigma^2}} \,, \tag{2}$$

where $\boldsymbol{p} = (p_\mathrm{x}, p_\mathrm{y})^\mathrm{T}$ are the pixel coordinates. An appropriate way to estimate the blur parameter σ in Eq. (2) is to use the blur information σ_x and σ_y along the coordinate axes [LFC92]: $\sigma = \frac{\sigma_\mathrm{x}\sigma_\mathrm{y}}{\sqrt{\sigma_\mathrm{x}^2 + \sigma_\mathrm{y}^2}}$.

Once the blur parameter σ is known, depth may be reconstructed by $z(\sigma) = \frac{k_1}{k_2 - \sigma}$. The constants k_1 and k_2 are calibrated using standard least-squares techniques [LFC92].

3 Fusion of combined image series

Fusion of depth from stereo and depth from focus: The absolute value of the focus measure strongly depends on the presence of object structures such as edges and texture. The approach requires that all considered focus measures correspond to the same scene point. Therefore, image warping techniques based on estimated disparities must be used to transform the images into other views [BG04].

The view with maximum focus measure can be predicted from the stereo disparity of a scene point. If it coincides with the actual maximum of the focus measure, the depth receives the highest confidence possible. However, there is some uncertainty in the method because of calibration errors and the depth of field. Thus, the focus maximum may instead be detected in a view with neighboring plane of focus, even if the estimated disparity is correct. A lower confidence level is assigned to these depth values obtained from the estimated disparity. If the focus maximum is neither in the predicted view nor in the neighboring one, the depth value is discarded.

The focus measure is integrated in the energy formulation of Eq. (1) by adding another

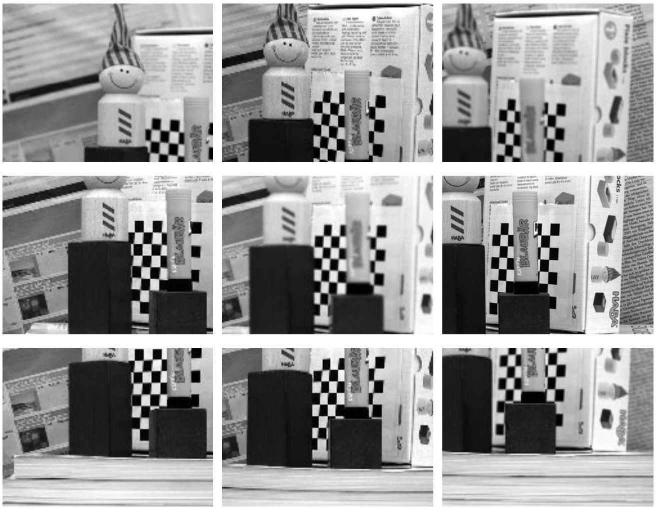

Figure 3: Combined stereo and focus series.

term:

$$E_{\text{fusion},1}(f) := E_{\text{stereo}}(f) + E_{\text{focus}}(f) \quad \text{with} \tag{3}$$

$$E_{\text{focus}}(f) := \sum_{\boldsymbol{p} \in \mathcal{P}} \begin{cases} 0 & \text{if } f(\boldsymbol{p}) \text{ has high confidence,} \\ \mu_1 & \text{if } f(\boldsymbol{p}) \text{ has medium confidence,} \\ \mu_2 & \text{if } f(\boldsymbol{p}) \text{ is inconsistent,} \end{cases}$$

where $\mu_2 > \mu_1 > 0$.

Fig. 2 shows a fusion example of the combined series in Fig. 1. The fusion result shows an improved depth estimation in areas with less structure.

Fusion of depth from stereo and depth from defocus: Given the depth estimate of an edge in one view from the depth from defocus approach, its position in a second view can be predicted. That way, each edge is transferred to any of the stereo images. The matching edge is searched in a list of candidates that are selected with the help of a distance function by using a windowing scheme based on the predicted positions. If no candidates are found, the edge remains unmatched, and is supposed to be occluded in the second view.

The actual fusion of stereo and depth from defocus is accomplished by adding another term to the energy functional of Eq. (1):

$$E_{\text{fusion},2}(f) := E_{\text{stereo}}(f) + E_{\text{edge}}(f)\,. \tag{4}$$

The term $E_{\text{edge}}(f)$ is non-zero only in the vicinity of edges. Near image edges, f is forced to a solution close to the one resulting from depth from defocus.

As an example, the combined series in Fig. 3 is fused to the depth maps in Fig. 4. The fusion result shows improvements in areas with periodic structures due to the additional information offered by depth from defocus.

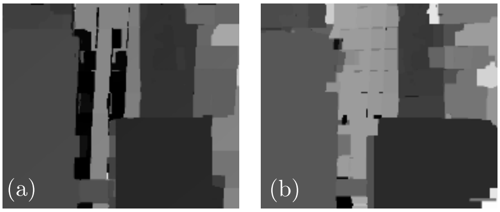

Figure 4: Fusion of the image series in Fig. 3: (a) Depth from stereo: periodic structures cause estimation errors; (b) Improved depth map obtained from fusion of stereo and defocus.

4 Conclusions

The article presents methods for fusing combined stereo and focus series in order to obtain improved depth information. Such multivariate image series can be efficiently acquired with a camera array. First, the depth estimation from stereo series is modeled using energy functionals. Depth from focus and depth from defocus, respectively, are then integrated into the model by formulating additional energy terms. Compared to the pure approaches, the fusion of the combined image series leads to better and more robust results in areas causing errors, such as structureless regions or regions with preriodic structures.

References

[BG04] Michael Bleyer and Margrit Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. In *Proc. of the 2004 International Conference on Image Processing*, pages 2997–3000, 2004.

[FL01] Olivier Faugeras and Quang-Tuan Luong. *The Geometry of Multiple Images*. MIT Press, 2001.

[KZ02] Vladimir Kolmogorov and Ramin Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *European Conference on Computer Vision*, pages 82–96, 2002.

[LFC92] Shang-Hong Lai, Chang-Wu Fu, and Shyang Chang. A Generalized Depth Estimation Algorithm with a Single Image. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(4):405–411, 1992.

[NN94] Shree K. Nayar and Yasuo Nakagawa. Shape from Focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(8):824–831, 1994.

[PB97] Fernando Puente León and Jürgen Beyerer. Datenfusion zur Gewinnung hochwertiger Bilder in der automatischen Sichtprüfung. *Automatisierungstechnik*, 45:480–489, 1997.

[Pen87] Alex Paul Pentland. A New Sense for Depth of Field. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(4):523–531, 1987.

[SC95] Muralidhara Subbarao and Tae Choi. Accurate Recovery of Three-Dimensional Shape from Image Focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(3):266–274, 1995.

[SN88] Muralidhara Subbarao and Gurumoorthy Natarajan. Depth Recovery from Blurred Edges. In *Conference on Computer Vision and Pattern Recognition*, pages 498–503, 1988.