A New Approach for Estimating Depth by Fusing Stereo and Defocus Information

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

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

[2]Fraunhofer Institut für Informations- und Datenverarbeitung (IITB)
Fraunhoferstraße 1, 76131 Karlsruhe, Germany, {heizmann,beyerer}@iitb.fraunhofer.de

Abstract: Several algorithms are common for estimating depth from stereo series, but many of them have difficulties when determining depth of objects having periodical structure. This contribution proposes a method to overcome the impediments by using defocus as additional information. The algorithm fuses depth from stereo and depth from defocused edges by analyzing and evaluating image series with simultaneously varied camera and focus positions. The problem is formulated by a comprehensive notation using energy functionals, which can be solved e. g. by applying graph cuts minimization.

1 Introduction

Even though there are several algorithms for depth estimation by means of image fusion, the problem still remains a challenge in the visual inspection domain. Most methods can only be applied under certain constraints. For example, depth from stereo is restricted to surfaces with visible, but non-periodic structure. Structureless regions and periodical structures can lead to false correspondences and further to false depth estimations [FG06]. In order to compensate these effects, additional information can be used, provided e. g. by homogeneous information sources such as (de)focus evaluation or by inhomogeneous sources such as triangulation or radar sensors.

Methods for fusing combined stereo and focus series to estimate depth were presented by Gheţa et. al. [GFH06] and Frese and Gheţa [FG06]. In [GFH06], a short general overview for fusing combined image series is presented, while [FG06] proposes a method for improving depth estimations by fusing depth from stereo and depth from focus.

In this contribution, another approach of image fusion is proposed: depth from stereo and depth from defocus. As input, combined stereo and focus series like the one shown in Fig. 1 are used. The resulting depth maps are dense (due to the depth from stereo approach) and more reliable than the results of any of the single approaches. To overcome the correspondence problem in stereo vision by means of defocus, a method for estimating depth from defocused edges requiring only one image is employed. In order to ensure real-time operability, a camera array whose cameras have different focus positions is deployed.

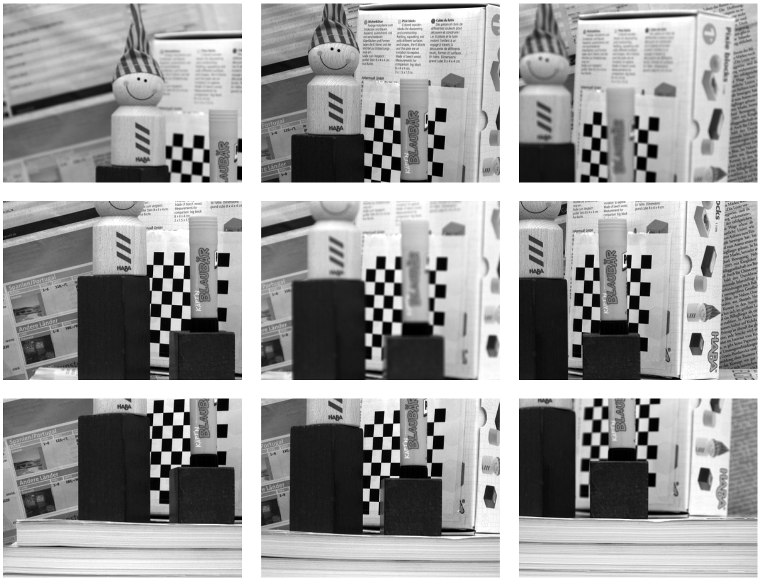

Figure 1: Combined stereo and focus series.

2 Fusion approach

The fusion problem is formulated using energy functionals. Their advantages are a clear specification of the problem, along with the possibility to explicitly introduce constraints. They allow to formally describe different kinds of information, e. g. stereo, focus, or defocus information. The fusion problem can then be solved by standard approaches, e. g. using graph cuts algorithms, which are a comprehensive tool to minimize energy functionals [SS02].

Depth from stereo: Estimating depth from stereo images can be described using the epipolar geometry (for details on basic principles of 3D reconstruction and depth estimation from multiple images, see e. g. [FL01] or [HZ03]).

The problem statement uses the function $f(\boldsymbol{p}) = \alpha$, which assigns each pixel $\boldsymbol{p} = (p_\mathrm{x}, p_\mathrm{y})^\mathrm{T} \in \mathcal{P}$ (with $\mathcal{P}$: the set of pixels in the entire image series) a *label* α in $\mathcal{L}$, where α is inversely proportional to the depth of a suitable projective plane. It can be seen as a generalization of the disparity notion in stereo vision and defines correspondences between rectified images, using the first two images in the series as reference:

$$f : \mathcal{P} \to \mathcal{L} : \quad f(\boldsymbol{p}_i) = \alpha \Leftrightarrow \begin{cases} \boldsymbol{p}_1 & \leftrightarrow \boldsymbol{p}_2 = (p_{1,\mathrm{x}} + \alpha, p_{1,\mathrm{y}})^\mathrm{T} \,, \\ \boldsymbol{p}_i & \leftrightarrow \boldsymbol{p}_1 \,, \end{cases} \tag{1}$$

where $\boldsymbol{p}_i$ denotes a pixel in the image i, and $\alpha = p_{2,\mathrm{x}} - p_{1,\mathrm{x}}$ is the disparity between the pixels of the rectified reference image pair.

The fusion task for pure stereo series can then be described by using the energy model:

$$E_\text{stereo}(f) = E_\text{data}(f) + E_\text{smoothness}(f) + E_\text{visibility}(f) \,. \tag{2}$$

The objective is to find a function f such that the total energy functional $E(f)$ is minimized, which is a standard task in computer vision and can be solved e. g. by means of graph cuts [KZ02].

Although this approach yields acceptable estimations in many situations, it often fails when periodical structures are present in the scene. In such cases, the association of corresponding features is not unique, leading to ambiguities. This drawback is avoided by using additional information from a depth from defocus approach.

Depth from defocus: In geometric optics, defocused image formation is described as a space-variant convolution of the sharp image with a circular disk, whose radius depends on the object distance and on the camera aperture and focal length [Mes77]. In the present approach, the defocus information is only used on edges as additional constraint in the process of estimating depth. As a consequence, the blur can be approximated without significant loss by a convolution of the sharp image with a Gaussian point spread function:

$$h_\sigma(\boldsymbol{p}) := \frac{1}{2\pi\sigma^2} \cdot \exp\left\{ -\frac{p_\mathrm{x}^2 + p_\mathrm{y}^2}{2\sigma^2} \right\}.$$

The algorithm for depth estimation shown here adapts and improves previously existing ones [Pen87, SN88, LFC92]. The first step is to detect intensity edges. In order to cover different degrees of blur, the Canny operator is applied several times, varying the spread parameter [Par97]. The next step is the modeling of the edge blur. The models are similar for horizontal (g_Mx) and vertical (g_My) edges. They are approximated analogously and independently, e. g.

$$g_\mathrm{Mx}(\boldsymbol{p}) = g_1 \Phi\left(\frac{a_\mathrm{x} p_\mathrm{y} + c_\mathrm{x} - p_\mathrm{x}}{\sigma_\mathrm{x}} \right) + g_2 \Phi\left(-\frac{a_\mathrm{x} p_\mathrm{y} + c_\mathrm{x} - p_\mathrm{x}}{\sigma_\mathrm{x}} \right), \tag{3}$$

where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution, σ_x is the standard deviation in x-direction, g_1, g_2 are the dominant gray values in the neighbouring regions $\mathcal{E}_1$, $\mathcal{E}_2$ of the edge (determined experimentally), and $a_\mathrm{x} p_\mathrm{y} + c_\mathrm{x} - p_\mathrm{x} = 0$ is the line equation of the edge. The line equation is determined by applying the Hueckel operator [Hue73], due to its performance in estimating edge orientation [LM88].

The model g_Mx is then fitted to the image data $g(\boldsymbol{p})$ by applying a nonlinear minimization method, with respect to σ_x, g_1, g_2, a_x, and c_x simultaneously, to the following cost function:

$$C := \sum_{(\boldsymbol{p}) \in \mathcal{E}_1 \cup \mathcal{E}_2} \left(g_\mathrm{Mx}(\boldsymbol{p}) - g(\boldsymbol{p}) \right)^2 . \tag{4}$$

Before the nonlinear minimization procedure can be started, suitable initial values for the parameters g_1, g_2, and σ_x are required. To this end, the average gray values within the regions $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively, are computed. Then, the lower and the upper quartil of the gray values are used as initial values in the regions having the smaller and the larger average, respectively. The blur parameter is initiated with the value from Subbarao's method [SN88].

The minimization procedure, employing the Levenberg-Marquardt algorithm [HZ03], is carried out with respect to a parameter vector consisting of the blur value σ_x, the gray levels g_1 and g_2 and the line parameters a_{x} and c_{x}. This is an extension of the algorithm of [LFC92] which did not consider the parameters of the line equation.

The Gaussian blur parameter σ is obtained from the blur information along the axes, i. e. σ_{x} and σ_{y} [LFC92]: $\sigma = \frac{\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}}{\sqrt{\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{y}}^2}}$. If the edge is almost parallel to one of the axes, only one blur parameter can be estimated; σ is then computed through: $\sigma = \sigma_{\mathrm{x}} \sin\xi$ or $\sigma = \sigma_{\mathrm{y}} \cos\xi$ respectively, where ξ denotes the angle between the edge line an the x-axis.

Since camera aperture, focal length and image distance might not be known with the required accuracy, edge depth is recovered through [LFC92]: $z = \frac{k_1}{k_2-\sigma}$, where the constants k_1 and k_2 are calibrated using standard least-squares techniques. For this purpose, an image series of a planar calibration pattern located at known distances to the camera is acquired.

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 using image warping. That way, candidate edge positions are identified in all stereo images. A matching edge is searched among the candidates that are selected by means of a distance function that assesses the displacement of the candidate from the predicted position. If no matching edge is found, the edge remains unmatched, and it 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:

$$E_{\text{fusion}}(f) := E_{\text{stereo}}(f) + E_{\text{edge}}(f)\,. \tag{5}$$

The term $E_{\text{edge}}(f)$ is non-zero only in neighbouring regions $\mathcal{E}_i(\boldsymbol{l}_i)$ of edges $\boldsymbol{l}_i$ and forces $f(\boldsymbol{p})$ to a disparity label close to the one resulting from depth from defocus:

$$E_{\text{edge}}(f) = \sum_{\boldsymbol{l}_i} \sum_{\boldsymbol{p}\in\mathcal{E}_i(\boldsymbol{l}_i)} \begin{cases} \frac{|f(\boldsymbol{p})-f(\boldsymbol{l}_i)|}{d(\boldsymbol{p},\boldsymbol{l}_i)+1}, & \text{if } \mid f(\boldsymbol{p}) - f(\boldsymbol{l}_i) \mid \leq S\,, \\ \frac{|f(\boldsymbol{p})-f(\boldsymbol{l}_i)|}{d(\boldsymbol{p},\boldsymbol{l}_i)+1} + K, & \text{if } \mid f(\boldsymbol{p}) - f(\boldsymbol{l}_i) \mid > S\,. \end{cases} \tag{6}$$

K is a penalty term for pixels having disparity values that differ greatly (given by the threshold S) from the disparity values estimated on edges; $d(\boldsymbol{p}, \boldsymbol{l}_i)$ is the distance of the pixel $\boldsymbol{p}$ to the edge $\boldsymbol{l}_i$. The denominators in Eq. (6) ensure that the influence of the edge decreases according to the distance $d(\boldsymbol{p}, \boldsymbol{l}_i)$ of the pixel to the edge.

3 Experimental results

As an example, Fig. 2 shows the $2\frac{1}{2}$D reconstructions obtained by applying a pure stereo approach and by fusing the combined series of Fig. 1. Especially at the periodic chessboard pattern behind the two objects, in the middle of the scene, the pure stereo approach finds false correspondences and hence yields erroneous depth estimations. In this case, it can be seen as two “black stripes” in the middle of the image, between the two front objects. In comparison, the depth estimation of the fusion approach shows considerable improvements in these areas. The scene reconstruction at the bottom of Fig. 2 reveals that

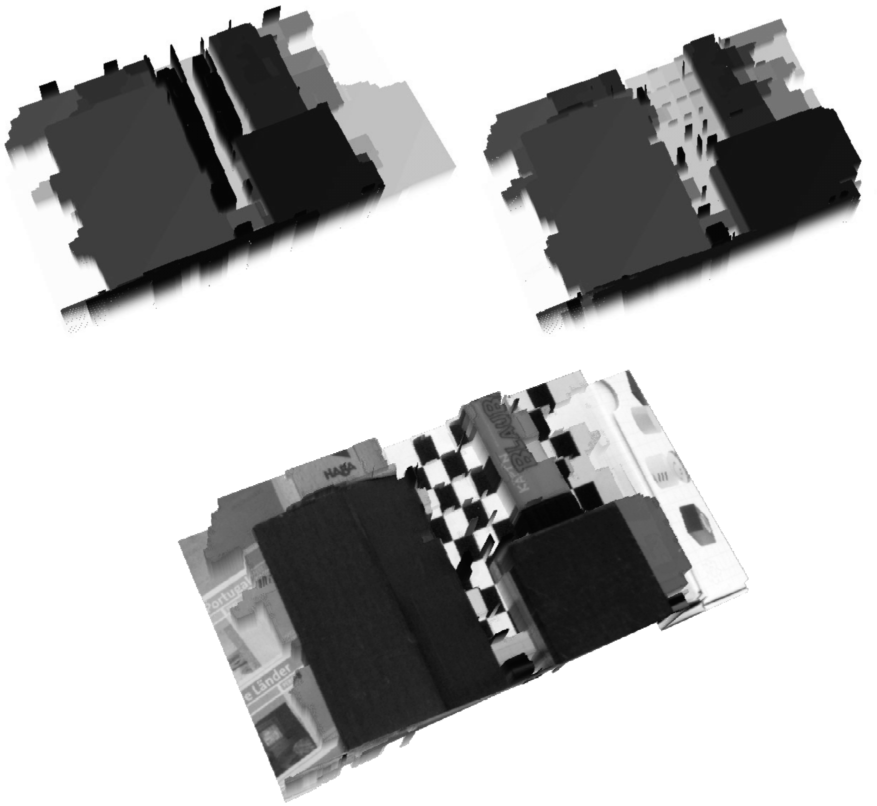

Figure 2: Fusion of the image series in Fig. 1: Top left: depth map using pure stereo series: periodic structures cause estimation errors (black strips in the middle); Top right: depth map obtained by fusing stereo and defocus information; the estimation is improved; Bottom: reconstruction using the fusion of combined series.

the depths of the borders of the two front objects and of the background pattern are almost correctly estimated.

Although this example shows a quite striking surface pattern where the depth from defocus approach shows its full potential whereas the pure stereo approach is handicapped by the periodic pattern, it is obvious that the complementary information provided by depth from stereo and depth from defocus improves the reliability of the depth estimation approach.

4 Conclusion

This contribution presents a new method for fusing combined stereo and focus series in order to improve the reliability and the accuracy of depth estimation. It uses the complementary information that is contained in the disparity of a stereo series and in the blur evaluated by means of a defocus approach. The fusion statement is based on a well-known approach of depth estimation from pure stereo series using energy functionals. The depth from defocus information is then integrated into the model by formulating an additional energy term which considers the depth estimation using a defocus approach in the vicinity of salient edges. Experimental results show that the presented fusion of the combined image series leads to better and more robust results compared to the pure approaches, especially in the presence of periodical structures.

References

[FG06] Christian Frese and Ioana Gheţa. Robust Depth Estimation by Fusion of Stereo and Focus Series Gained with a Camera Array. In *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI06)*, pages 243–248, 2006.

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

[GFH06] Ioana Gheţa, Christian Frese, and Michael Heizmann. Fusion of Combined Stereo and Focus Series for Depth Estimation. *Informatik 2006 – Informatik für Menschen*, 1:359–363, 2006.

[Hue73] Manfred H. Hueckel. A Local Visual Operator Which Recognizes Edges and Lines. *J. ACM*, 20(4):634–647, 1973.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.

[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.

[LM88] Edward P. Lyvers and O. Robert Mitchell. Precision Edge Contrast and Orientation Estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(6):927–937, 1988.

[Mes77] Franz Mesch. Systemtheoretische Beschreibung optisch-elektrischer Messsysteme. *Technisches Messen ATM*, 7(8):249–258, 1977.

[Par97] Jim R. Parker. *Algorithms for image processing and computer vision*. Wiley, 1997.

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

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

[SS02] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.