

Optimising Aperture Shapes for Depth Estimation

A. Sellent^{†1} and P. Favaro¹

¹Institut für Informatik und angewandte Mathematik, Universität Bern, Schweiz

Abstract

The finite depth of field of a real camera can be used to estimate the depth structure of a scene. While the distance of an object from the plane in focus determines the defocus blur size, the shape of the aperture determines the shape of the blur. This blur shape can be manipulated by introducing masks into the main lens aperture. We propose an intuitive criterion to design aperture patterns for depth estimation. Our design criterion imposes constraints directly in the data domain and optimises the amount of depth information carried by blurred images. As a quadratic function on the aperture transmission values, our criterion can be numerically evaluated to estimate optimised aperture patterns quickly. The proposed mask optimisation procedure is applicable for different depth estimation scenarios. We consider depth estimation from two images with different focus settings and depth estimation from two images with different aperture shapes.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis —Depth cues

1. Introduction

When a 3D object is imaged through a lens, objects at different distances to the camera are recorded with different sharpness of detail. The further away an object is placed from the focal plane, the blurrier it occurs in the image. Depth from defocus approaches make use of the finite depth of field to reconstruct the shape of the objects in the scene. For an arbitrary scene both depth and texture are unknown. It is therefore hard to distinguish between a low textured but focused object and a defocused, highly textured object. To avoid this ambiguity, depth from defocus approaches usually require several images. Traditionally, the images are acquired by an ordinary camera with different focus settings (DfD). More recently, also images acquired with different aperture masks but constant focus were used to solve the ambiguity between blur and texture [ZLN09], (CAP). Thereby, careful choice of a pair of masks allows to improve the depth estimation capacity of DfD algorithms.

Different properties of the masks that lead to desirable results have been analysed in the literature [ZLN09, Lev10]. However, non of these properties leads to a criterion that can be optimised within reasonable time. The idea of our approach is that a good mask for depth estimation should

produce blurs that are easily distinguishable. We evaluate that aim by generating blurred images at different depth and comparing their appearance as a function of the masks. The comparison is entirely defined in the spatial domain of images. By using a database of natural images, our approach naturally incorporates the statistic behaviour of images without explicit modelling. Focusing on mask dependent properties at different depths, our mask optimisation is also independent of the depth estimation method to be applied.

2. Determining Optimised Masks

To formalise our idea of good aperture masks, we consider a grid with $n \times n$ elementary apertures in the lens plane. According to the depth d of the scene, each elementary aperture forms a depth dependently blurred image on the image plane. The final image on the sensor plane is the sum of all elementary contributions. To compare the structure of the images independent of their brightness, we consider transmission values for each elementary aperture $\alpha_i \geq 0$ with $\sum_{i=1}^{n^2} \alpha_i = 1$.

In our applications, from a single sharp texture f we measure two images f_α^s and f_β^t . In DfD, the aperture mask is constant in both images, i.e. $\alpha = \beta$, but the blur scale is different, i.e. $s \neq t$. In CAP, the blur scales are constant, i.e. $s = t$, but the masks change, i.e. $\alpha \neq \beta$. In both cases we can stack the measurements into a vector $f_{\alpha,\beta}^d$. Our notion of

[†] Supported by DFG-Fellowship SE 2134/1

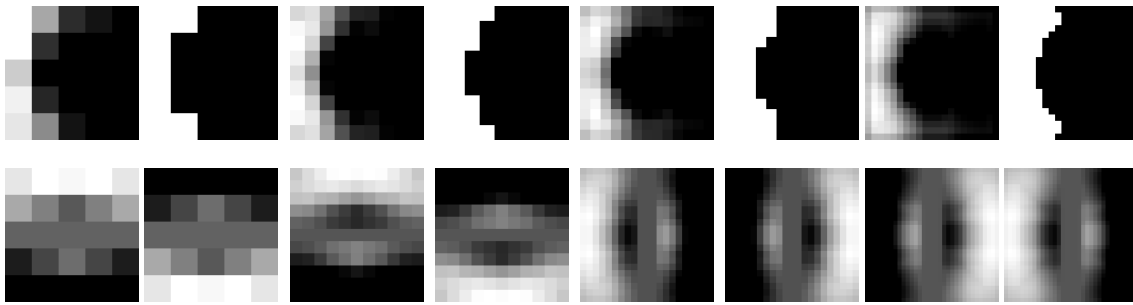


Figure 1: Optimised masks derived with our criterion: the first row shows depth from defocus masks with resolution $n = 5, 9, 13, 21$ together with their thresholded versions. The second row shows coded-aperture pairs of resolution $n = 5, 11, 21, 33$.

good apertures is that for different depth level $d_1 \neq d_2$, the difference $\|f_{\alpha,\beta}^{d_1} - f_{\alpha,\beta}^{d_2}\|^2$ should be as large as possible. I.e., the appearance of a fixed texture should depend as strongly as possible on the depth.

Closed elementary apertures decrease the signal to noise ratio. Therefore, we introduce an additional penaliser into our metric that encourages open apertures in the mask. In this setup, an optimal mask can be obtained by solving

$$\alpha^*, \beta^* = \arg \max_{\alpha, \beta} \sum_f \sum_{d_i \neq d_j} \|f_{\alpha,\beta}^{d_i} - f_{\alpha,\beta}^{d_j}\|^2 - \lambda(\|\alpha\|^2 + \|\beta\|^2) \quad (1)$$

with a constant $\lambda > 0$. Algebraic manipulations allow to simplify this problem into

$$\alpha^*, \beta^* = \arg \max_{\alpha, \beta} [\alpha; \beta]^\top M [\alpha; \beta] \quad (2)$$

with a symmetric matrix M . We solve this quadratic problem with the optimisation approach of Gill et al. [GMSW84]. For a 11×11 DfD mask with $\alpha = \beta$, solution of problem 2 with the MATLAB optimization toolbox requires 0.09s on a 2.90GHz quad-core computer. If $\alpha \neq \beta$ a set of optimised 11×11 CAP masks can be obtained in 2.40s.

As currently most actual implementations of aperture masks have $\alpha_i \in \{0, 1\}$, we also propose a thresholded version of the optimised masks where we set all aperture values smaller than $\frac{1}{n^2}$ to 0 and all others to 1.

3. Mask Analysis

We analyse the masks obtained in each setup via evaluation with a state-of-the art depth estimation algorithm [FS05]. This algorithm does not incorporate a model of the aperture shape and can thus be applied with the same parameters in all setups. As input structure we use a stair of 20 equally spaced steps with known depth and different textures. The maximal blur size in all experiments is fixed to 7 pixels. Depth estimation with optimised apertures on clean images is more accurate than with the reference masks in all cases. Adding noise according to the amount of incoming light, continuously valued masks turn out to be more susceptible to noise

than thresholded masks. For DfD the thresholded masks perform better than the traditional full aperture on all noise levels. For CAP our optimised, thresholded masks performs better than the same resolution mask from literature on all noise levels.

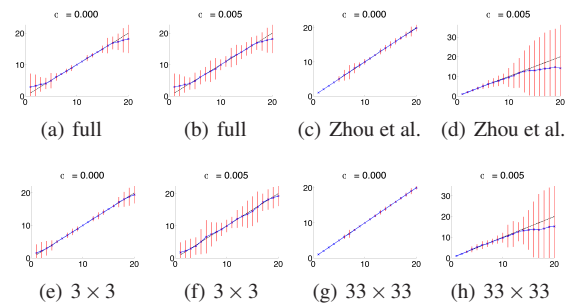


Figure 2: For different masks and noise level, we compare ground truth depth (black diagonal) to the estimated depth (blue), indicating three times the standard deviation with red bars. Optimised depth from defocus masks (e), (f), reach better depth estimation than fully open apertures (a) even on images with noise added proportionally to the incoming light. For coded aperture pairs, our optimised masks (g),(h) outperform the reference mask from Zhou et al.(c), (d)

References

- [FS05] FAVARO P., SOATTO S.: A geometric approach to shape from defocus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, 3 (2005), 406–417. 2
- [GMSW84] GILL P., MURRAY W., SAUNDERS M., WRIGHT M.: Procedures for optimization problems with a mixture of bounds and general linear constraints. *ACM Transactions on Mathematical Software (TOMS)* 10, 3 (1984), 282–298. 2
- [Lev10] LEVIN A.: Analyzing depth from coded aperture sets. In *Computer Vision—ECCV 2010*. Springer, 2010, pp. 214–227. 1
- [ZLN09] ZHOU C., LIN S., NAYAR S.: Coded aperture pairs for depth from defocus. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 325–332. 1