Single-Image Depth Map Estimation Using Blur Information

Dmitry Akimov
Department of Computational
Mathematics and Cybernetics
Moscow State University, Moscow, Russia
dakimov@graphics.cs.msu.ru

Dmitry Vatolin
Department of Computational
Mathematics and Cybernetics
Moscow State University, Moscow, Russia
dmitry@graphics.cs.msu.ru

Maxim Smirnov
YUVsoft Corp.

ms@yuvsoft.com

Abstract

This paper presents a novel approach for depth map estimation from a single image using information about edge blur. The blur amount at the edge is calculated from the gradient magnitude ratio of the input and re-blurred images. The final depth map can be obtained by propagating estimated information from the edges to the entire image using cross-bilateral filtering. Experimental results for real images and video sequences demonstrate the effectiveness of this method in providing a plausible depth map for 2D-to-3D conversion that generates comfortable stereo viewing.

Keywords: *Image Processing, Depth Map, Defocus Blur, Cross-bilateral Filtering.*

1. INTRODUCTION

3D images and videos have become more and more popular in everyday life, but the process of making 3D content is still very complicated and expensive. Many techniques for automatic 2D-to-3D conversion, and depth map estimation from 2D image sources plays an important role in this process.

Depth can be recovered either from binocular cues, such as stereo correspondence, or monocular cues, such as shading, perspective distortion, motion and texture. Conventional methods of depth map estimation have relied on multiple images, video sequences, calibrated cameras or specific scenes, or a huge training database. Stereo vision [3] measures disparities between a pair of images of the same scene taken from two different viewpoints, and it uses these disparities to recover the depth map. Shape from shading [4] offers another method for monocular depth reconstruction, but it is difficult to apply to scenes that do not have fairly uniform color and texture. Saxena et al. [2] presented an algorithm for estimating the depth map from numerous monocular image features using MRF modeling and a training database. The proposed algorithm in this paper focuses on a more challenging problem: recovering depth from a single defocused image captured by an uncalibrated camera without any prior information.

2. DEPTH ESTIMATION

The main part of the proposed algorithm is blur estimation at the edges and proper depth propagation from the edges to the entire image.

2.1 Defocus blur model

To estimate a blur value, the defocus blur model and special blur estimation method were introduced in [5].

Assume that the ideal edge is the step function $f(x)$ (Fig. 1(a)) with the jump at $x = 0$:
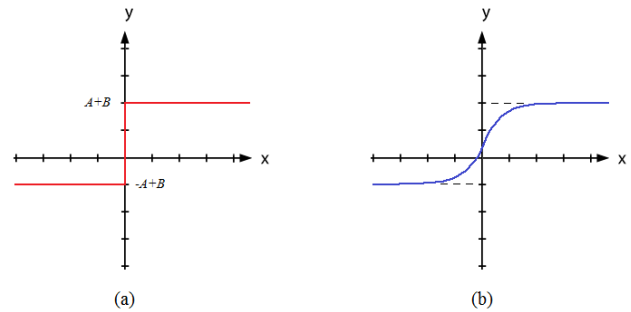
$$u(x) = \begin{cases} -1, & x \leq 0 \\ 1, & x \geq 0 \end{cases}$$

$$f(x) = Au(x) + B$$

Here, A and B is the amplitude and offset of the edge, respectively.

The blurred edge is the function $i(x)$ (Fig. 1(b)), which is the convolution of $f(x)$ and $g(x,\sigma)$ where the latter is a Gaussian function with an unknown parameter σ.

$$i(x) = f(x) \otimes g(x,\sigma) = \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau$$

Figure 1: (a) Ideal edge function $f(x)$.
(b) Blurred edge function $i(x)$.

2.2 Blur value estimation

A blurred edge with unknown parameter σ is re-blurred using a Gaussian function with a known parameter σ_1.

$$i_1(x) = i(x) \otimes g(x,\sigma_1)$$

Consider the gradient magnitude of $i_1(x)$:

$$\nabla i_1(x) = \nabla\big(i(x) \otimes g(x,\sigma_1)\big) =$$
$$= \nabla\Big(\big((Au(x) + B) \otimes g(x,\sigma)\big) \otimes g(x,\sigma_1)\Big) =$$
$$= \frac{A}{\sqrt{2\pi(\sigma^2 + \sigma_1^2)}} e^{-\frac{x^2}{2(\sigma^2+\sigma_1^2)}}$$

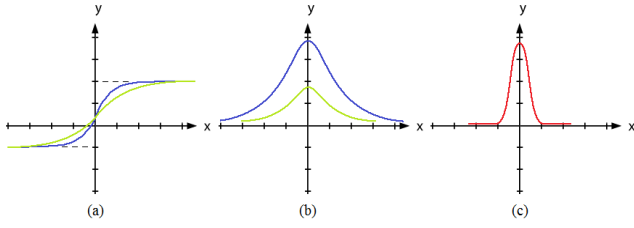The gradient magnitude ratio of the re-blurred edge and the original edge is the function $B_i(x)$.

$$B_i(x) = \frac{\nabla i_1(x)}{\nabla i(x)} = \sqrt{\frac{\sigma^2}{\sigma^2 + \sigma_1^2}} e^{\frac{x^2}{2\sigma^2} - \frac{x^2}{2(\sigma^2+\sigma_1^2)}}$$

The maximum of the function is at the edge location ($x = 0$):

$$B_i(0) = \frac{\nabla i_1(0)}{\nabla i(0)} = \sqrt{\frac{\sigma^2}{\sigma^2 + \sigma_1^2}}$$

This function depends only on blur parameters, and the unknown σ can be estimated from the gradient magnitude ratio:

$$\sigma = \frac{\sigma_1}{\sqrt{B_i^2(0)-1}}.$$

Figure 2: (a) Blurred edge $i(x)$ and re-blurred edge $i_1(x)$.
(b) Gradient magnitudes of $i(x)$ and $i_1(x)$.
(c) Gradient magnitude ratio function $B_i(x)$.

2.3 Blur value filtering

Blur values are estimated for each point of the image's edge mask. Assuming the entire scene in the image is beyond the camera's focal plane, the depth at the edges can be measured by the estimated parameter σ, forming a sparse map $D_0(x,y)$ and the mask of points with known depth values $Mask_0$. Because of noise in the image and inaccuracies in edge detection, however, the estimate of the sparse depth map $D_0(x,y)$ may contain some errors. To avoid such estimation errors, several prefiltering techniques are proposed.

The first step of filtration is outlier removal. Points with rare depth values relative to the depth value histogram are excluded from D_0 and $Mask_0$ (Fig. 3, row 1 and 2).

The second step is median filtering of D_0 according to mask and color difference. Figure 3 illustrates the result of median filtering; following noise reduction, two main depth profiles appeared.
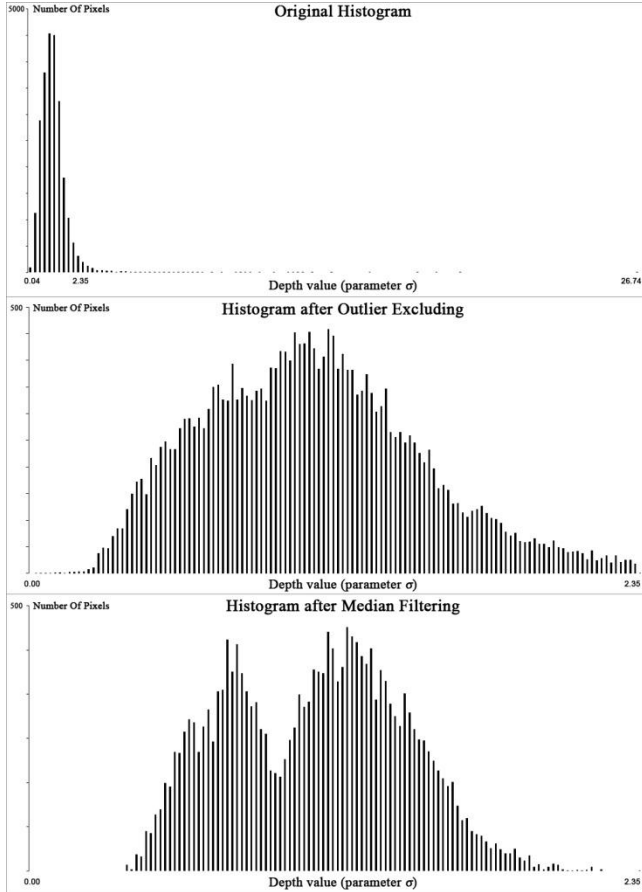
Figure 3: Histograms of depth values in sparse depth map.

2.4 Depth Propagation

The next important step in depth estimation is correct propagation of the blur information from the edges to the entire image. To achieve this, iterative cross-bilateral mask-oriented filtering was used:

$$D_i(x,y) =$$
$$= \frac{1}{2\pi\sigma_s\sigma_c\mathcal{N}} \iint\limits_{\Omega\cap Mask_{i-1}} D_{i-1}(\tau,\delta)e^{-\frac{(x-\tau)^2+(y-\delta)^2}{2\sigma_s^2} - \frac{(I(x,y)-I(\tau,\delta))^2}{2\sigma_c^2}} d\tau d\delta$$

where $D_i(x,y)$ is the depth map estimated at the iteration i of the propagation process, $Mask_i$ is the propagated region mask, σ_s and σ_c are spatial and color weights respectively and $\mathcal{N}$ is the normalization factor.

3. FINAL DEPTH MAP

The final depth map of the image is obtained from the propagated depth map. Following the sky detection process, the sky region is defined as the most distant area of the scene. The refined depth map is then filtered using cross-bilateral filtering.

3.1 Sky detection

To achieve more-accurate depth maps for outdoor scenes, a sky detection algorithm is introduced (Fig. 4). This algorithm is based on color segmentation of potential sky areas in HSV color space.

Figure 4: Original image and detected sky region.

The sky region is assumed to be in the upper area of the image, so the method recursively propagates the sky region from the image's upper edge. If a pixel color corresponds to a comparison criterion, the method marks it as a sky pixel and then checks the neighboring pixels. The algorithm stops if the current pixel doesn't correspond to the color criterion or if a color difference between the current and previous pixels exceeds a constant threshold.

The color criterion used here was chosen empirically and consists of three parts:

Figure 6: Comparison of proposed method (column 3) and [5] (column 2).

1. $(V(x, y) > 0.75) \cap (S(x, y) < 0.2)$
   to detect light tones in the sky

2. $(V(x, y) > 0.75) \cap (H(x, y) < 0.01)$
   to detect grey tones in the sky

3. $(V(x, y) > 0.3) \cap (H(x, y) > 0.4) \cap (H(x, y) < 0.8)$
   to detect blue, orange and yellow tones in the sky

4. RESULTS

In this work the Canny edge detection algorithm [1] was used and was tuned for uniform estimation of edges with different magnitudes.

The parameter σ_1 for re-blurring was set to $\sigma_1 = 0.5$. The estimated blur value at the edge location is then filtered using median filtering with a radius of 4 and using outlier exclusion according to histogram values.

For bilateral propagation, the parameters were set to $\sigma_s = 10$ and $\sigma_c = 7$, and Ω was defined as a circle of radius 30.
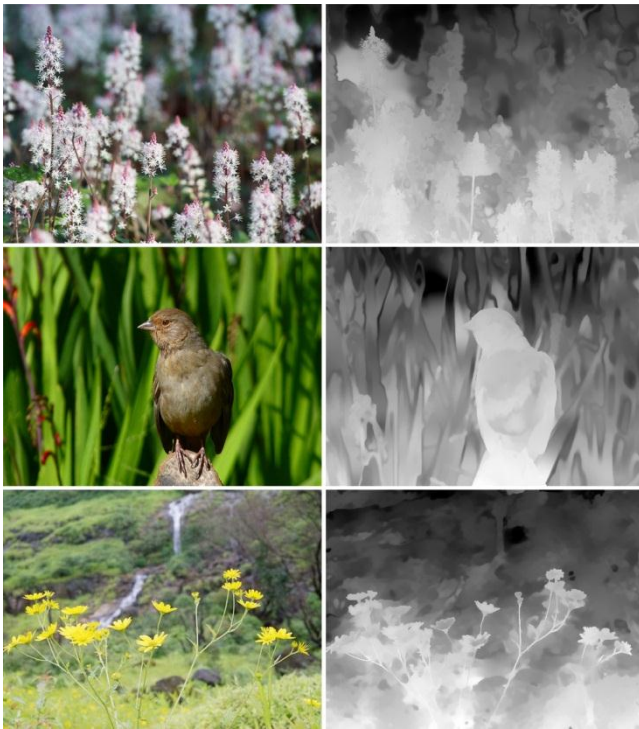
Figure 5: Source images (column 1) and estimated depth maps (column 2) using proposed method.

Figure 5 shows the results of the algorithm for real images. Rows 1 and 3 illustrates that the proposed method preserves all thin details of the source image, providing a correct depth map.

In figure 6 the proposed algorithm was compared with that of [5], which produced a coarse layered depth map with some layers not corresponding well to edges in the source image. By contrast, the results of the proposed algorithm are more accurate, continuous and detailed, all while preserving object boundaries.

5. CONCLUSION

This paper presents a novel approach for depth map estimation from a single image. The blur amount is calculated from the gradient magnitude ratio of the re-blurred and original images. A final depth map is then obtained using iterative cross-bilateral mask-oriented filtering. The proposed technique generates plausible depth maps that can be more accurate than those produced by existing methods.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Canny J., "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell. 8(6) (1986) 679-698*.

[2] Saxena A., Chung S. H., and Ng A. Y., "Learning depth from single monocular images," *in Neural Information Processing System (NIPS) 18, 2005.*

[3] Scharstein D. and Szeliski R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision (IJCV), vol. 47, 2002.*

[4] Shao M., Simchony T., and Chellappa R., "New algorithms for reconstruction of a 3-d depth map from one or more images", *In Proc IEEE CVPR, 1988.*

[5] Zhuo S. and Sim T., "On the Recovery of Depth from a Single Defocused Image," *In Computer Analysis of Images and Patterns. Springer, 2009.*

About the author

Dmitriy Vatolin (M'06) received his M.S. degree in 1996 and his Ph.D. in 2000, both from Moscow State University. Currently he is head of the Video Group at the CS MSU Graphics & Media Lab. He is author of the book Methods of Image Compression (Moscow, VMK MGU, 1999), co-author of the book Methods of Data Compression (Moscow, Dialog-MIFI, 2002) and co-founder of www.compression.ru and www.compression-links.info. His research interests include compression techniques, video processing and 3D video technics: optical flow, depth estimation - from motion, focus, cues, video matting, background restoration, high quality stereo generation. His contact email is dmitry@graphics.cs.msu.ru.

Maxim Smirnov is the chief technology officer at YUVsoft IT company. Maxim has graduated from Saint Petersburg State University of Aerospace Instrumention in 2001 and received a Ph.D. in data compression from the same university in 2005. His research interests include forecasting of processes in technical and economic systems, video and universal data compression, stereo video. His contact email is ms@yuvsoft.com.

Dmitry Akimov is a student at Moscow State University's Department of Computational Mathematics and Cybernetics. His contact email is dakimov@graphics.cs.msu.ru.