Novel trilateral approach for depth map spatial filtering

Alexander Voronov
Moscow State University
Graphics & Media Lab
avoronov@graphics.cs.msu.ru

Dmitriy Vatolin
Moscow State University
Graphics & Media Lab
dmitriy@graphics.cs.msu.ru

Maxim Smirnov
YUVsoft Corp.
ms@yuvsoft.com

ABSTRACT

In this paper, we present our approach for spatial filtering of depth map extracted from camera motion. Original depth map may have some artifacts because of non-perfect motion estimation. The challenge is to make depth map uniform on smooth areas and refine boundaries of objects, but not to blur edges. To solve this problem we propose the trilateral filter, which convolution kernel is composed from distance kernel, color-based kernel and depth-based kernel. Experiments demonstrate rather good results of this approach. Also we compare our results with results of a typical bilateral filter.

Keywords

depth map, disparity map, trilateral filtering, spatial filtering, post-processing, 3D video

1 INTRODUCTION

One of the widely used methods of creating 3D-video is changing of image parallax using depth map. For this purpose it is necessary to have information about the distance from the camera to objects in the scene. Depth map is a visualization of that distance to objects in every pixel of image: more distant spots are represented in a darker color. In common case creating of depth map from a single image is unsolvable problem. So until recently in most cases depth maps were painted manually by stereo artists and composers, what required much time and was rather expensive. But in some cases depth map can be created based on information from a scene. There are some works with application of machine learning algorithms where such information is extracted from rather large set of images in different scales [Sax06]. Also in [Zhu09] was proposed an approach using a property that usually the camera is focused on foreground objects, so that objects have sharp edges, and with increase of the distance from the camera boundaries of objects become more blurred. Another approach is restoration of depth from geometric properties of a scene [Bat04], for example taking into account vanishing point, horizon line, vertical lines and so on. And for scenes with camera motion we can create depth map applying an optical flow algorithm and analyzing how objects are moving in a scene. For example if the camera is panning displacement of an object in a given frame relative to the previous one depends on the distance of the object from camera. This approach was described in [Pou10] and [Kim07].

Application of an optical flow algorithm supposedly allows to obtain the best quality depth map estimation from camera motion. But results of this algorithm at that stage may be not good enough because of some problems. Firstly it is impossible to make accurate optical flow for two frames in regions of opening and occlusion that appears when objects are moving. In such regions depth can't be estimated correctly. Another problem is that it is impossible to detect true motion in smooth areas, particularly in case of noisy video. Also, considering high computational complexity of this stage, we often have to sacrifice quality of optical flow to increase processing speed, and that affects final results. So some postfiltering is required to reduce errors in a depth map or to make them less visible in the final result.

Such postfiltering may be performed as simple Gaussian smooth with some variation [Zha05] or as a more complex filtering, for example bilateral [Cha09]. There are being actively developed approaches where problem of depth estimation is reduced to energy optimization for the whole frame, what is rather expensive in terms of processing time, but makes possible to achieve better quality of final results [Zha08].

2 PROPOSED METHOD

To suppress the artifacts we propose trilateral filtering. Convolution kernel is being built for every pixel and is composed from the following components: Gaussian kernel G with specified radius, matrix $I(x,y)$ based on photometric difference from the source image for the current pixel with coordinates (x,y) with neighbour pixels and analogue matrix $D(x,y)$ calculated for this pixel from a depth map.

G responds to the distance from the currently processed pixel: the farther the pixel is from the center the lesser influence it has on the result. Weights i in image based component are linearly depending on difference between central pixel and other pixels:

$$i(m,n) =$$

$$= \begin{cases} \frac{thresh_{color} - IDiff_{xy}(m,n)}{thresh_{color}}, & IDiff_{xy}(m,n) \leq thresh_{color} \\ 0, & IDiff_{xy}(m,n) > thresh_{color} \end{cases}$$

where $IDiff_{xy}(m,n) = (|red(m,n) - red(x,y)| + |green(m,n) - green(x,y)| + |blue(m,n) - blue(x,y)|)/3$, $m \in [x-r, x+r], n \in [y-r, y+r]$

For depth component $D(x,y)$ linear dependence is not applicable. In some models depth has only few grades, so an error in one grade of depth may be too big in color range. Also we should penalize big differences in depth. So we decided to use logistic function, and weights of depth components are calculated in the following way:

$$d(m,n) = 1 - \frac{1}{1 - e^{-t \times DDiff_{xy}(m,n) - 6}}$$

where $DDiff_{xy}(m,n) = |D(m,n) - D(x,y)|, m \in [x-r, x+r], n \in [y-r, y+r]$

But when we take into account information from a rough depth map we may get a problem. All the artifacts on the depth map will have influence on the depth component in convolution kernel and consequently on the final result. So for calculating weights in the depth component we need a depth map mostly without artifacts but not blurred, with strong edges. To solve this problem we used bilateral filtering with adaptive threshold. If there are enough pixels of the same color near the current pixel, we set threshold low to preclude taking pixel from another depth level and blurring of an edge. But when there are few similar pixels, we set strength of filtering high enough to suppress the artifacts. The radius of filtering should be chosen depending on the size of artifacts that we want to suppress.

Then final convolution kernel $K(x,y)$ is calculated as element-wise multiplication of matrices G, $I(x,y)$, $D(x,y)$.

$$k(m,n) = g(m,n) \times i(m,n) \times d(m,n)$$

Resulting pixel value in the filtered depth map:

$$r(x,y) = \frac{\sum\limits_{m=x-r}^{x+r} \sum\limits_{n=y-r}^{y+r} k(m,n) \times z(m,n)}{\sum\limits_{m=x-r}^{x+r} \sum\limits_{n=y-r}^{y+r} k(m,n)}$$

In comparison to bilateral filtering the trilateral approach has some advantages. If we don't take into account information from the source image we are only blurring a depth map and not really enhancing it. But when we ignoring depth in building a convolution kernel we can blur a boundary between two objects of the same color. Also when we use only color component we can get wrong thin depth flowings on boundaries since boundaries usually have an average color of objects they divide.

3 RESULTS

Results of processing with the proposed method and its comparison to the bilateral approach are presented in Figure 1. In Figure 2 there is a visualization of depth maps before after filtering using proposed method. The proposed method is better than the image-based bilateral algorithm in preserving boundaries of objects detected by optical flow. Also it makes depth smoother in uniform areas than the depth-based bilateral approach.

4 FURTHER WORK

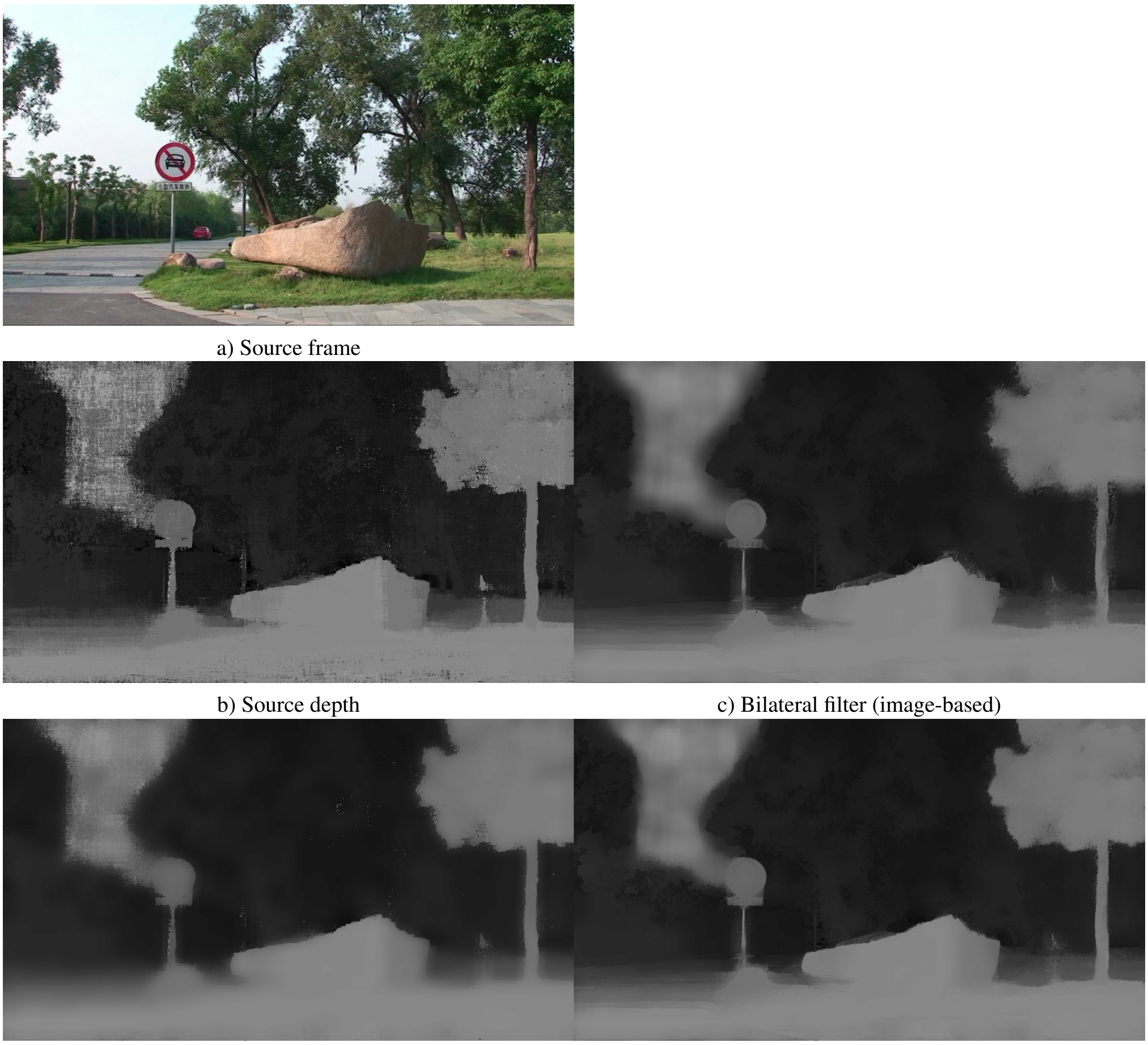
In short-term authors' plans there is a better integration of data from optical flow into postfiltering algorithm which will allow to improve quality of restoration of small details and also to build a confidence measure for each pixel. Using this measure it will be possible to estimate probability of artifacts appearance in certain regions and achieve better results.

Another direction in the development of the algorithm is utilizing temporal data from previous and next frames compensated according to optical flow. This approach will significantly increase computational complexity but should improve temporal stability of a depth map and increase number of correct details in a frame.

Also we plan to extract from the source image and optical flow data separate objects as structural units for more precise processing of object boundaries.

5 CONCLUSIONS

In this paper, was proposed an algorithm of trilateral postfiltering of depth maps created from camera motion. We made a comparison with other approaches, described and demonstrated advantages of



a) Source frame

b) Source depth

c) Bilateral filter (image-based)

d) Bilateral filter (depth-based)

e) Trilateral filter

Figure 1: Comparison of different methods of filtering on sequence "Road", frame 29.

our approach relative to other ones. Mentioning problems that appear when using an inaccurate depth map for calculating of a convolution kernel, we described our method how to solve these problems. Lastly, directions of further work were described.

ACKNOWLEDGMENTS

This research was partially supported by the grant number 10-01-00697-a by Russian Foundation for Basic Research.

REFERENCES

[Sax06] Ashutosh Saxena, Sung H. Chung, *Learning Depth from Single Monocular Images*. Advances in Neural Information Processing Systems, 2006.

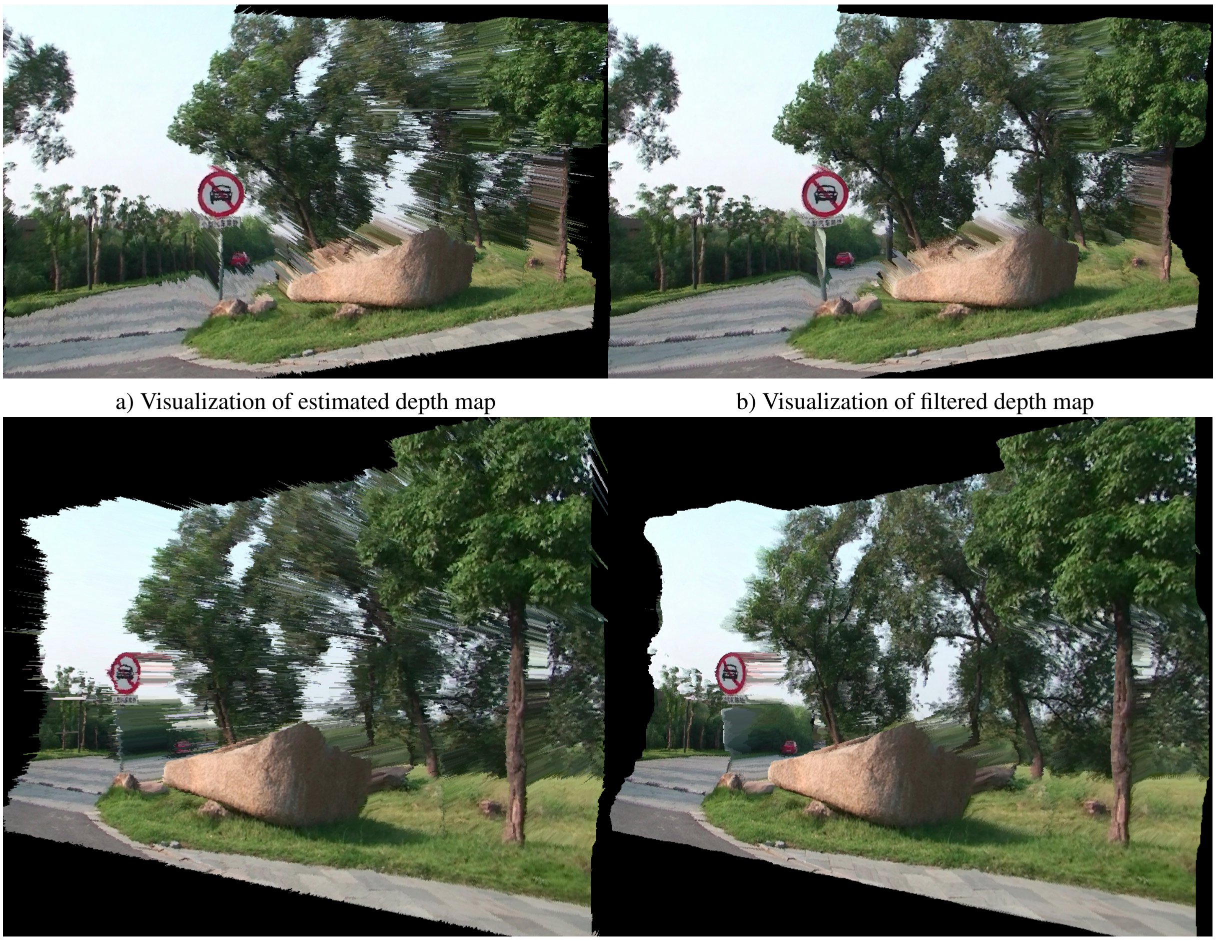
[Zhu09] Shaojie Zhuo, Terence Sim, *On the Recovery of Depth from a Single Defocused Image*. Proceedings of International Conference on Computer Analysis of Images and Patterns (CAIP), vol. 5702/2009, p. 889-897, 2009.

[Bat04] S. Battiatoa, S. Curtib, M. La Casciac, M. Tortorac, *Depth-Map Generation by Image Classification*. Proceedings of SPIE, vol. 5302, 95, 2004.

[Pou10] Mahsa T. Pourazad, Panos Nasiopoulos, Rabab K.Ward, *Generating the DepthMap from the Motion Information of H.264-Encoded 2D Video Sequence*. EURASIP Journal on Image and Video Processing, Volume 2010.

[Kim07] Donghyun Kim, Dongbo Min, Kwanghoon Sohn, *Stereoscopic Video Generation Method Using Motion Analysis*. Proc. of the 3DTV Conference, p. 1-4, 2007.

[Zha05] Zhang, L., Tam, W. J., *Stereoscopic Image*



a) Visualization of estimated depth map b) Visualization of filtered depth map

c) Visualization of estimated depth map d) Visualization of filtered depth map

Figure 2: 3D-visualization of depth map, sequence “Road”, frame 29

Generation Based on Depth Images for 3D TV. IEEE Trans. on Broadcasting, vol. 51, pp. 191-199, Jun. 2005.

[Cha09] Chao-Chung Cheng, Chung-Te Li, Po-Sen Huang, Tsung-Kai Lin, Yi-Min Tsai, and Liang-Gee Chen, *A Block-based 2D-to-3D Conversion System with Bilateral Filter*. International Conference on Consumer Electronics, p. 1-2, 2009.

[Zha08] Zhang, G., Jia, J., Wong, T., Bao, H., *Recovering Consistent Video Depth Maps via Bundle Optimization*. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, p. 1-8, 2008.