Post-Processing of Stereo Results Using Iterations of a Bilateral Filter

Phinehas Bynum and Peter Tsai
St. Olaf College
Northfield, MN 55057
bynum@stolaf.edu, tsai@stolaf.edu

Abstract

We implemented Yang, Yang, Davis and Nistér's [2006] approach to stereo image post-processing. Using successive iterations of a bilateral filter, this approach approximates the optimal disparity for a given pixel by aggregating and utilizing relevant data from the left and right color images.

The initial step creates a three dimensional representation of the data values from the input disparity map. This disparity space approximates the probability that a given pixel at *(x,y)* has disparity *z*.

Each layer of the disparity space is smoothed with an edge-preserving smoothing filter utilizing both color and distance data from the left and right color images. This bilateral filter allows each iteration to maintain the initial shapes while reducing pop-outs and sharpening edges. Finally, quadratic sub-pixel estimation is used to ensure the best possible disparity value, even if it occurs between pixel values. Yang, Yang, Davis and Nistér accomplished across the board improvements on the Middlebury stereo data set using this approach.

Introduction

Discarding irrelevant visual data is a deceivingly complex task for a computer. Giving a computer the tools to process stereo input is a challenge foundational to countless other technological applications. As there already exist stereo methods that consistently produce less than 5 percent overall difference from the ground truth, it seems reasonable to investigate post-processing steps designed to improve already excellent output as opposed to breaking new ground in a field with an already developed body of pre-existing research. "Spatial-depth super resolution for range images, " by Yang et. al caught our eye as an effective, simple and powerful method to achieve significantly better results.

Process

This algorithm successively approximates a final disparity map based on the current best. The initial step creates a three dimensional representation of the data values from the input disparity map. This disparity space approximates the probability that a given pixel at *(x,y)* has disparity *z*.

The initial depth map of *x* by *y* pixels is extended in the *z (disparity)* direction by exactly the value represented by pixel *(x,y)*, or *D(x,y)*. This three-dimensional location is assigned "cost" zero to denote the location of the current best guess for pixel *(x,y)*, and the other costs increase quadratically from this initial point, capped based on the search range, *L*.

$$C(x,y,d) = \min(\tfrac{1}{2}L,(d - D(x,y))^2$$

Next, each layer of the disparity space is smoothed with an edge-preserving smoothing filter utilizing both color and distance data from the left and right color images. Bilateral filtering allows each iteration to preserve the initial objects present in the disparity map while reducing pop-outs and sharpening edges. The formula for the bilateral filter used in this paper is as follows,

$$C(x,y,d) = \frac{\sum_{i=1}^{(2L+1)^2} f_c(x,y,u_i,v_i) f_d(x,y,u_i,v_i) C(x,y,d)}{\sum_{i=1}^{(2L+1)^2} f_c(x,y,u_i,v_i) f_d(x,y,u_i,v_i)},$$

$$f_c(x,y,u,v) = e^{-\left(\frac{W_c(x,y,u,v)}{\lambda_c}\right)},$$

$$f_d(x,y,u,v) = e^{-\left(\frac{W_d(x,y,u,v)}{\lambda_d}\right)},$$

$$W_c(x,y,u,v) = \frac{|R(u,v)-R(x,y)| + |G(u,v)-G(x,y)| + |B(u,v)-B(x,y)|}{3},$$

$$W_d(x,y,u,v) = \sqrt{(u-x)^2 + (v-y)^2},$$

where C is the cost volume. λ_c and λ_d define regions of interest for the color and distance weights, and are both set to 10 experimentally by the Spatial-depth researchers. In the case of one reference image, the new minimum cost per disparity is extracted for each pixel from the filtered cost volume, and sub-pixel estimation through a quadratic interpolation is used to ensure the best possible disparity value, even if it occurs between pixel values.

Two view modifications

When a second color reference image is present, further computation is added, most importantly the creation and filtering of a second cost volume. The primary difference between the first and second cost volumes is their method of creation: the first is computed from the disparity map, the second purely from the color images. Birchfield and Tomasi's pixel dissimilarity is used to reduce sampling noise[1] and fill the cost volume. The second bilateral filter uses data from both the left and the right image:

$$C(x,y,d) = \frac{\sum_{i=1}^{(2L+1)^2} f_c(x,y,u_i,v_i) f_d(x,y,u_i,v_i) f_c(x',y',u_i',v_i') f_d(x',y',u_i',v_i') C(x,y,d)}{\sum_{i=1}^{(2L+1)^2} f_c(x,y,u_i,v_i) f_d(x,y,u_i,v_i) f_c(x',y',u_i',v_i') f_d(x',y',u_i',v_i')},$$

where x,y,u,v are coordinates in the left view and x',y',u',v' coordinates in the right view.

In the case of two images, sub-pixel estimation is not performed after the first cost volume, but instead after the second. The final disparity guess and its adjacent values are fit to a parabola, and xmin is returned, allowing for a discrete data set to approximate a continuous function and improve accuracy. Though allowing potentially more average pixel accuracy, this estimation pixelates the output image, causing the next cycle to put points on the wrong disparity level. Thus, an adaptive box-car filter with search window 9x9 is passed across the disparity map, removing the pixelation effect but preserving disparity level data and edges. Yang, Yang, Davis and Nistér accomplished across-the-board improvements on the Middlebury stereo data set using this process, however we were unable to replicate their results.

Optimization/Performance

We ran through many thousands of iterations using permutations of λ_c and λ_d with over 30 cycles per combination, checking our results with our own implementation of the Middlebury stereo error computation code which was extensively checked against the original Middlebury algorithm to ensure the same output. Using our own version allowed us to automate multiple file checks at once and facilitate data aggregation for thousands of generated disparity maps.

Results

Experimentally, we determined that for our implementation the values of λ_c and λ_d should be set to 6 and 5 respectively, with L as 7. These data gave us a final result for the Eriol output as follows. Eriol is the current best stereo algorithm at St. Olaf College, developed by Professor Olaf Hall-Holt.

Before (Error threshold=1)
Non-occluded: 1.18331%
All: 1.48581%
Discontinuities: 6.24446%

After
Non-occluded: 1.08149%
All: 1.38775%
Discontinuities: 5.83914%

Conclusion

There is still much work to be done, primarily in replicating the results of Yang et. al. Moreover, as each parameter is so sensitive, we would like to automate parameter selection as well. Our experiment is a success in that it has shown the potential to improve stereo matching data, but less so in that it is not as consistent or robust as the original algorithm developed by Yang et. al.

References

[1] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In ICCV, pages 839–846, 1998.

[2] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. CVPR 2007.

[3] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. In CVPR (2), pages 2347–2354, 2006.