

Variable Window Methods for Stereo Disparity Determination

C. Chris Erway
cce3@cornell.edu

Ben Ransford
bar13@cornell.edu

Ithaca, New York, May 2000

Abstract:

Local stereo-matching methods such as *sum of squared differences* (SSD), while effective in approximating disparities between images, employ fixed-size windows for pixel correlation and are thus too often influenced by irrelevant pixels within the area of consideration. Choosing a variable-window approach provides many benefits, but introduces a “chicken-and-egg” problem: to choose “perfect” window sizes, we must already know the disparities of the scene; to find correct disparities, we must be using perfect window sizes. We introduce a flexible variable window size technique, which uses an Expectation Maximization-like approach to approximate both quantities.

I. Introduction

When using local methods like SSD for stereo correlation which use fixed-size windows for comparing the neighborhood of a pixel, one of our main problems is that of irrelevant information – that is, not pertaining to the current pixel – contaminating our data. Our problem is this: what is the best window size and shape that will provide the best stereo correlation?

Ideally, if we already had knowledge as to the ground truth of the disparities in our scene, we would be able to answer this readily: the best window size for a pixel in a scene would be those pixels connected to it at roughly the same disparity in the scene. Using this neighborhood of similar-disparity pixels would provide “perfect” correlations.

Herein lies our “chicken-and-egg” problem. Given the true disparities of a scene, we would easily be able to pick variable window sizes allowing good correlation matches. Similarly, given a “perfect” set of variable window sizes for each pixel, we would then easily be able to perform stereo matching and extract the correct disparities.

However, in vision, none of this information is given. We have no ground truth, nor do we have pre-sliced local neighborhoods per pixel. For that reason the simplest approach has been to naively assume that a square window around each pixel would provide reasonable results: this is the basis for SSD, SAD, and various other methods.

Kanade and Okutomi approach this problem through using a statistical model of the “uncertainty” of a given neighborhood of pixels around a specified point [1]. This statistical model, based on both the intensities of the image and iterative disparity estimates, works quite well in rating different possible sizes of windows for a pixel, but falls short in the selection of each pixel’s window – they must still, for each pixel, search for an “optimum” window size. In their implementation, Kanade and Okutomi maintain a strict rectangular-shaped window, weakening the success of their method.

We propose an approach similar to the expectation-maximization (EM) principles that have been applied to other areas of vision – a good example being Birchfield and Tomasi’s modified multiway cut algorithm for slanted surfaces in stereo matching [2]. Their application of EM to its own “chicken-and-egg” problem (if they had good segmentations, they could find slanted surfaces, and vice versa) is simple, but yields satisfactory results.

II. Application

Our algorithm, then, involves the application of EM principles to variable window sizing using an initial estimation of the stereo disparities. Restating our problem, as we see it:

- If we knew the correct window sizes, then we could correctly find the disparities by performing stereo correlation.
- If we had the disparities, then we could similarly find correct window sizes for each pixel.

So we apply a mixture model. Our algorithm:

1. Start off with an initial estimate of the disparities (say, using SSD or SAD)
2. Find new disparities using this estimate:
 - a. Assuming these disparities correct, estimate optimal window sizes for each pixel
 - b. Assuming these window sizes, find a new disparity estimate (perform stereo matching)
3. Repeat step 2 as desired, continually feeding the disparities produced from 2b into 2a.

Step 2a bears some explanation. When we use fixed-size windows to determine information about the neighborhood of a pixel, we must “take the bad with the good” and allow pixels dissimilar to the current pixel to contaminate our measurements. Thus our definition of a good variable-sized window for a pixel p , given an estimate of the disparities in a scene, is: find all pixels, within some radius from p , that have similar disparities (within some threshold) as p . These pixels must also be “connected” to p through some neighbors also satisfying this test. These pixels and p make up the mask describing our variable window in 2a.

The intuition behind this choice for sizing the variable windows is that it will allow us to perform correlation on only other elements in the local window that are also at the same disparity as our pixel. The effect of this is that these windows will not be able cross disparity edges, sharpening the disparity edges in our result. Fixed-size windows, however, blur disparity edges, as they use windows that contain pixels that average across these edges.

III. Experimental Results

Our method consistently improved the disparity estimates from SSD. While the SSD method alone tended to produce “blurry” disparity images, with non-uniform disparities at object borders, our method was able to produce somewhat sharper disparity edges that make better intuitive sense than those found by SSD only. The following capture of a local window sizing

shows how the blurry original disparity calculation (shown on the right) is sharpened.

```

running iteration 1 (taking initial estimate from SAD calculation)
for pixel 287, 94:
  local window mask          current disparity estimate
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
1 1 1 1 1 1 1 0 0          7 7 7 7 7 7 7 5 5
1 1 1 1 1 1 1 1 1          7 7 7 7 7 7 7 7 7
0 0 0 0 0 1 1 1 1          8 8 8 8 8 7 7 7 7
0 0 0 0 0 1 1 1 1          8 8 8 8 8 7 7 7 7
0 0 0 0 0 0 1 1 1          8 8 8 8 8 8 7 7 7
0 0 0 0 0 0 0 1 1          8 8 8 8 8 8 8 7 7

running iteration 2 (taking in new disparities from last correlation)
for pixel 287, 94:
  local window mask          current disparity estimate
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
0 0 0 0 0 0 0 0 0          5 5 5 5 5 5 5 5 5
1 1 1 1 1 1 1 0 0          8 8 8 8 8 8 8 5 5
1 1 1 1 1 1 1 1 1          8 8 8 8 8 8 8 8 8
1 1 1 1 1 1 1 1 1          8 8 8 8 8 8 8 8 8
1 1 1 1 1 1 1 1 1          8 8 8 8 8 8 8 8 8
1 1 1 1 1 1 1 1 1          8 8 8 8 8 8 8 8 8
1 1 1 1 1 1 1 1 1          8 8 8 8 8 8 8 8 8

```

We see a blurry disparity edge in the original SAD calculation (upper right), shown by the 7-valued disparities produced by the SAD's fixed window set on the area of the disparity edge. Our local pixel is in this 7-valued area; its local window size mask (upper left) is variably chosen as its neighboring 7-valued disparities. Using this window for correlation causes the 7-valued area to "decide" their label ought to be 8; this produces a sharper disparity edge.



Figure 1 - SAD correlation with window radius 3

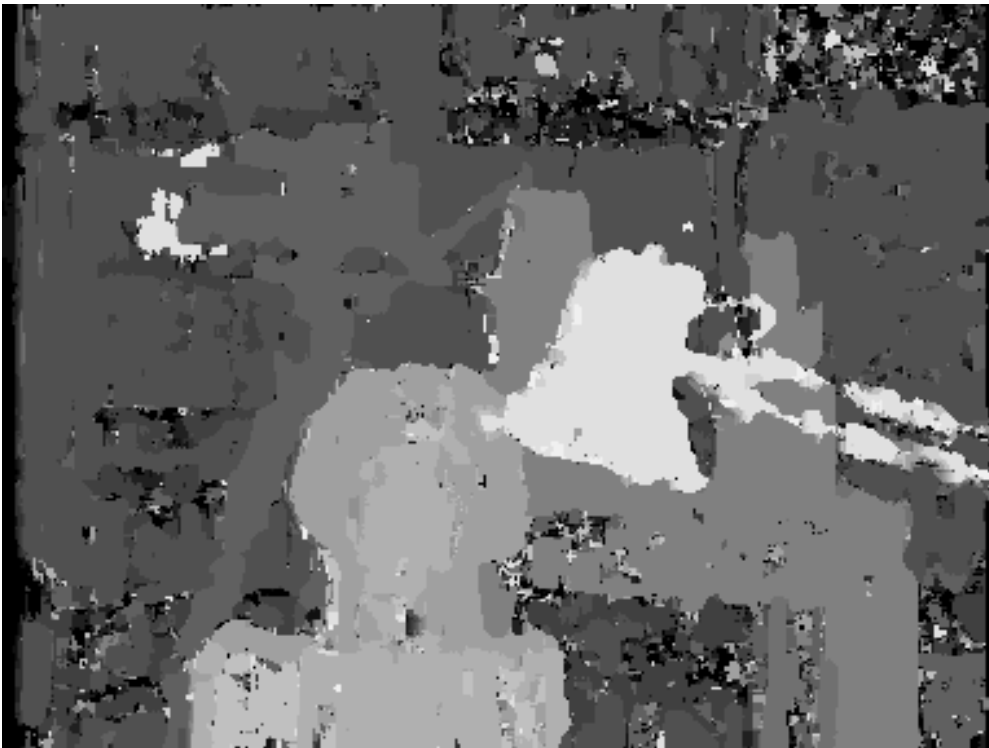
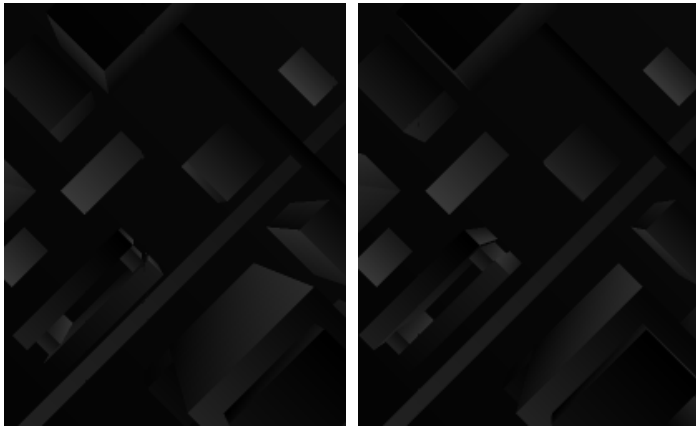


Figure 2 – variable window sizing applied 10 times with search window radius 3

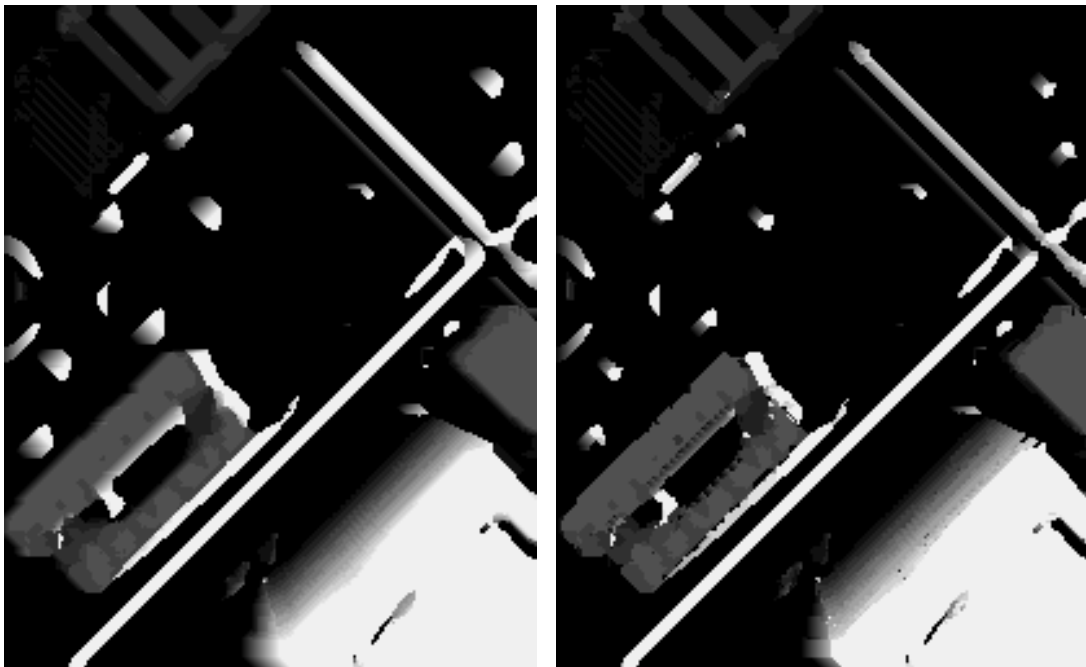
Figures 1 and 2 show the results of our algorithm on the popular Japanese scene with provided ground truth, with disparities represented as brighter intensities for better visibility. Comparing the ground truth with the SAD result, and summing differences between the results gave us a sum difference of 3738544 for figure 1. Comparing the ground truth with the result from our algorithm gave us a sum difference of 800651, considerably lower. Just upon inspection of the image, this it is also evident – edges appear much sharper in our result.

We received similar results from the following pair of synthetic images of diagonally oriented, rectangular-shaped objects:



left scene

right scene



SAD result, window radius 3

variable-window size result

Again, our operation resulted in crisp disparity edges and more precise detail.

Opportunities for Future Work

Though our application of EM to the problem is clean and simple, our method of choosing local windows at each pixel is rather naive. Our method of finding similar-disparity pixels in the neighborhood (step 2a) does not take into account, for example, intensity data when making this decision, and enforces strict smoothness constraints on the neighborhood it uses for local support. One can imagine a statistical method that considers both intensity and disparity information when making this decision.

Also, the correlation itself could benefit from improvement: it only concerns itself with intensity differences, and does not consider the disparity estimates we have available. We could perhaps use SSD on the disparity weights as well; or, we could devise a scheme of weighted masking, different from the unweighted (either a local pixel is in the window, or it is not) scheme we currently use. Associating a statistically-based weighted value with each pixel (2a), and then passing that weighted local mask to the correlation operation (2b) to create a sort of “weighted SSD” matching might produce more accurate results.

References

- [1] T. Kanade and M. Okutomi, “A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 9, pp 920-931, Sep. 1994.
- [2] S. Birchfield and C. Tomasi, “Multiway Cut for Stereo and Motion with Slanted Surfaces,” *Proceedings of the Seventh IEEE International Conference on Computer Vision*,” Sep. 1999.