Stereo matching using Semi-global matching

Robert Rentea
dept. Computer Science
Technical University of Cluj-Napoca
Cluj, Romania
robert.rentea@student.utcluj.ro

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Acurate stereo matching is important for many applications that work with 3D reconstruction from 2D images. Fast calculations are often required, either because of real-time aplications or because of large datasets of images that need to be processed efficiently.

A. Objectives

Given a set of left and right rectified stereo images, I have to determine the depth of objects from each image. I will do this by finding the disparity of each pixel from the left image to the right image using Semi-Global matching.

II. Related Work

There is a wide range of stereo matching algorithms with different properties, but they all fall into two categories: local and global aproaches [1]. Local (or area based) approaches perform matching solely on a local search window and are ideal for hardware implementation due to the implicit parallelism and little data dependencies. However those approaches struggle with processing areas that have low quality or repetitive textures [2]. Among the global aproches are *Graph Cuts* [3] which computes a graph structure where the nodes of the graph are pairs of pixels, then it determines what is known in the literature as a minimum cut, i.e. a cut in the graph that minimizes the cost [4]. Another method is *Belief Propagation* [5] which uses Markov networks and Bayesian approach to infer scene structure by using multiple sensors, e.g, stereoscopic structure, shape from shading, shape from shadows, shape from focus, shape from silhouette, and shape from texture. Both *Graph Cuts* and *Belief Propagation* offer solution to the difficult problems of stereo vision: noise, textureless regions, depth discontinuities and occlusions. But they are both memory intensive and *Graph Cuts* is rather slow.

Traditional stereo algorithms often struggle over changes in lightning between their two views. *Mutual Information* [6] has been introduced as a solution to this problem. This mathod relies on the entropy of the images to deduce propability densities. It has been shown [7] that it is robust against many complex intensity transformations and even reflections.

III. Proposed Solution

A. Semi-Global Matching

The idea of Semi-Global Matching(SGM) method is based on the idea of pixelwise matching and apoximating a global by performing line optimisations along multiple directions. The number of directions affects the run time of the algorithm, and while 16 directions usually ensure good quality, a lower number can be used to achieve faster execution. A typical 8-direction implementation of the algorithm can compute the cost in two passes, a forward pass accumulating the cost from the left, top-left, top, and top-right, and a backward pass accumulating the cost from right, bottom-right, bottom, and bottom-left.

B. Pixelwise Cost Calculation

The first step of SGM is to compute a disparity map by considering individual pixels from the left image and finding the apropriate match in the right image. This is done by traversing the epipolar line on which the matching pixel can be found, in the case of rectified images the epipolar line is the row of the base pixel.

When performing pixelwise image matching, the measure of dissimilarity between pairs of pixels from different images is affected by differences in image acquisition such as illumination bias and noise.

The Birchfield–Tomasi disimilarity [8] measure compensates for the sampling effect by considering the linear interpolation of the samples. Pixel similarity is then determined by finding the best match between the intensity of a pixel sample in one image and the interpolated function in an interval around a location in the other image.

Considering the stereo matching problem for a rectified stereo pair, where the search for correspondences is performed in one dimension, given two columns x_l and x_r along the same scanline for the left and right image respectively, it is possible to define two symmetric functions:

$$d_l(x_l, x_r) = \min_{x_r - \frac{1}{2} \leq x \leq x_r + \frac{1}{2}} \mid I_l(x_l) - \widehat{I}_r(x) \mid \tag{1}$$

$$d_r(x_r, x_l) = \min_{x_l - \frac{1}{2} \leq x \leq x_l + \frac{1}{2}} \mid \widehat{I}_l(x) - I_r(x_r) \mid \tag{2}$$

Where $\hat{I}_l$ and $\hat{I}_r$ are the linear interpolation functions of the left and right image intensity I_l and I_l along the scanline. The Birchfield–Tomasi dissimilarity can then be defined as:

$$d(x_l, x_r) = \min\{d_l(x_l, x_r), d_r(x_l, x_r)\} \tag{3}$$

In practice the measure can be computed with only a small and constant overhead with respect to the calculation of the simple intensity difference, because it is not necessary to reconstruct the interpolant function. Given that the interpolant is linear within each unit interval centred around a pixel, its minimum is located in one of its extremities. Therefore, $d_l(x_l, x_r)$ can be written as

$$d(x_l, x_r) = \max\{0, I_l(x_l) - I_{max}, I_{min} - I_l(x_l)\} \tag{4}$$

Where,

$$I_{max} = \max\{I_r(x_r), I_r^+(x_r), I_r^-(x_r)\} \tag{5}$$

$$I_{min} = \min\{I_r(x_r), I_r^+(x_r), I_r^-(x_r)\} \tag{6}$$

Fig. 1. gives a visualisation of how these values are computed.

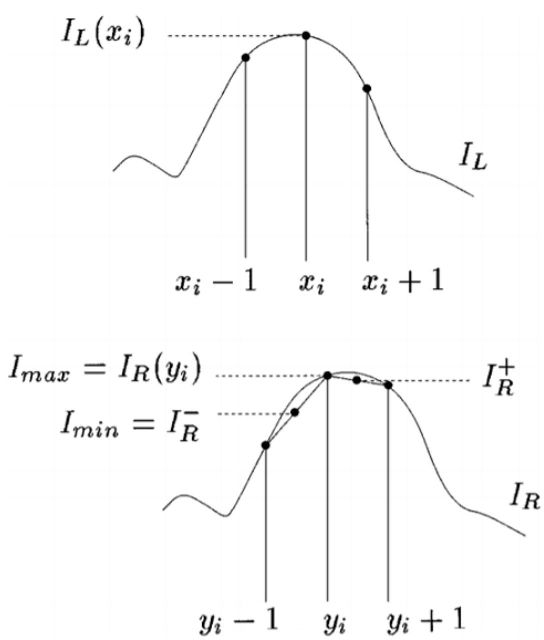


Fig. 1: Definition and computation of $d(x_l, x_r)$ [8]

Denoting with $I_r^+(x_r)$ and $I_r^-(x_r)$ the values of the interpolated intensities at the rightmost and leftmost extremities of a one-pixel interval centred around x_r

$$I_r^+(x_r) = \frac{1}{2}(I_r(x_r) + I_r(x_r + 1)) \tag{7}$$

$$I_r^-(x_r) = \frac{1}{2}(I_r(x_r - 1) + I_r(x_r)) \tag{8}$$

C. Aggregation Of Cost

Pixelwise cost calculation is generally ambiguous and wrong matches can easily have a lower cost than correct ones, due to noise, etc. Therefore, an additional constraint is added that supports smoothness by penalizing changes of neighboring disparities. To the cost of a pixel p it is added a constant P1 for every pixel q in the neighborhood N_p of p, for which the disparity changes by 1, and a larger constant P2 for every pixel q of N_p that have a larger disparity change.

The idea behind cost aggregation is to add the cost from all directions equally. The aggregated (smoothed) cost S(p,d) for a pixel p and disparity d is calculated by summing the costs of all 1D minimum cost paths that end in pixel p at disparity d (Figure 1). It is noteworthy that only the cost of the path is required and not the path itself.

The accumulated cost $S(p,d) = \sum_r L_r(p,d)$ is the sum of all costs $L_r(p,d)$ to reach pixel p with disparity d along direction r. Each term can be expressed recursively as

$$\begin{aligned} L_r(p,d) = D(p,d) + \min\{&L_r(p-r,d), L_r(p-r,d-1) + P_1, \\ &L_r(p-r,d+1) + P_1, \\ &\min_i L_r(p-r,i) + P_2\} - \min_k L_r(p-r,k) \end{aligned}$$

Lastly, the disparity of pixel p is given by finding the disparity d that corresponds to the minimum cost, i.e. $\min_d S(p,d)$.

D. Complexity and Implementation

The computation of pixelwise cost starts by iterating through every pixel and calculating the cost for every disparity and selecting the minimum. This process has a time complexity of O(WHD), where W, H is the width and height of the image and D is the maximum disparity that a pixel can have.

The next step is computing for every pixel the agregated cost from R paths, for each disparity. In this step each pixel is visited R times, which is a constant, resulting in the time complexity of O(WHD).

IV. Experimental Results

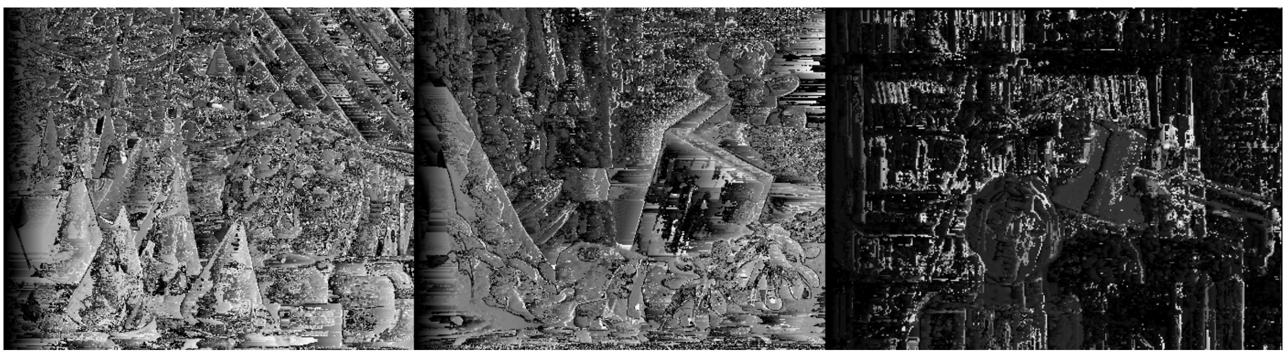
The implementation of the Semi Global Matching algorithm has been tested on Cones, Teddy and Tsukuba images that are used in the literature to evaluate the performance of stereo matching algorithms.

In Fig. 3 is the result of computing the cost volume using Birchfield-Tomasi dissimilarity. We can see that it alone struggles to estimate good disparity at the boudaries of objects. Also it is noticeble that in the left of the image for some objects that are closer to the camera we can't estimate a disparity because they dissapear from the right image, but objects that are further behind remain in the image so we can get a result.

In Fig. 4 is presented the resulting disparity map after computing aggregation of cost from 8 paths. We can see that the disparity map is much more clear and we can easily distinguish objects further from the camera than those closer to the camera.

(a) Cones (b) Teddy (c) Tsukuba

Fig. 2: Sample images



(a) Cones cost volume (b) Teddy cost volume (c) Tsukuba cost volume

Fig. 3: Birchfield-Tomasi cost volume

(a) Cones disparity map (b) Teddy disparity map (c) Tsukuba disparity map

Fig. 4: Disparity maps

V. Conclusion

The resulting Semi-Global Matching (SGM) method performs much better matching than local methods and is almost as accurate as global methods. However, SGM is much faster than global methods. A near real-time performance on small images has been demonstrated as well as an efficient calculation of huge images.

References

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.

[2] H. Hirschm¨uller and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," Computer Vision and Pattern Recognition. IEEE Conference on, pp. 1–8, 2007.

[3] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In International Conference for Computer Vision, pages 508 515, 2001

[4] L. Ford and D. Fulkerson. Flows in Networks. Princeton University Press, 1962.

[5] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7):787 800, July 2003

[6] G. Egnal. Mutual information as a stereo correspondence measure. Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, Philadelphia, USA, 2000.

[7] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In International Conference on Computer Vision, 2003.

[8] S. Birchfield and C. Tomasi. Depth discontinuities by pixelto-pixel stereo. In Proceedings of the Sixth IEEE International Conference on Computer Vision, pages 1073–1080, Mumbai, India, January 1998.