Confidence-based cost modulation for stereo matching

Riccardo Gherardi
Dipartimento di Informatica, University of Verona
Strada Le Grazie, 15 - 37134 Verona, IT
riccardo.gherardi@univr.it

Abstract

We present a novel operator to be applied at raw matching costs in the context of low level vision tasks such as stereo matching or optical flow. It aims at improving matching reliability by efficiently modulating pixel-wise pairing costs, injecting a confidence backed bias before the aggregation step. It works analyzing a noisy estimate of the correspondances in order to favor or prune potential matches. We test the operator by developing a local, realtime stereo matching algorithm and showing that our solution can drastically clean the resulting depth map while also reducing border bleeding. Its good performance is also evaluated quantitavely by testing the algorithm against the popular Middlebury benchmark where our local greedy implementation is able to obtain results comparable to those of näive global approaches.

1. Introduction

We describe a novel pixel-wise operator aimed at refining and improving the reliability of a underlying matching cost in the context of low level vision. The current trend for tasks like stereo matching and optical flow computation has been an ever-increasing sofistication, exacerbated and fueled by the publication of common dataset and benchmarks [9, 1]. The top performers in each category are often composed of several complex modules like plane fitting, edge-preserving smoothing, image segmentation and many others.

It's easy to see that any improvement in the earliest step of the matching computation, namely in the calculation of the first matching cost, can have profound and beneficial effects on the remainder of a algorithm pipeline.

In this paper we propose a simple and efficient operator capable of drastically pruning potential correspondances for a pixel. It works analyzing (or in a sense, refining) a noisy initial approximation of the depth or flow map, smoothly inhibiting matching pairs without sufficient support in a local, unstructured neighbourhood.

To support our claim that incorporating such operator into existing algorithms could provide additional reliability while allowing a simplification in the regularization tecniques, we implement a local, greedy stereo matching implementation whose results are comparable to näive global approaches at a fraction of the sofistication and time complexity.

The rest of the paper is organized as follows: in section 2 we briefly cover the related literature; section 3 will describe our neighbourhood confidence operator, and the following section the algorithm developed to test it. Section 5 will detail the completed experiments. Last section will present our conclusions.

2. Related work

Since the main topic of this article deals with matching measures and aggregation strategies we refer the interested reader to the following papers [6, 10] for some recent and fair comparaison.

Regarding the representation of confidence, literature reports several successfull approaches in stereo matching research. Historically, autocorrelation or the left-right consistency constraint have been used to characterize the ambiguity of a pixel, but several other methods exist like for example image entropy or curvature metric [3, 4].

The notion of distinctiveness maps [7], recently reinterpreted by [11], or that of stability [8] are also reconducible to confidence measures.

Confidence is usually employed to guide the matching process or the constraint enforcement in a high confidence first fashion, or as a weighting function in depth map fusion [5].

Our approach, instead of computing confidence as a by-product of the matching process, extrapolate it *a posteriori* from a initial, given, possibly noisy disparity

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

estimate and use it to directly modulate the underlying matching costs.

3. Confidence-based cost modulation

In the past, confidence measures have usually been calculated as a function of the entire x, y, d space. We propose instead to infer a confidence measure from a initial, possibly noisy estimate of the sought flow or disparity map and to use it to modulate the underlying matching cost function, as follows:

$$C'_{x,d} = \frac{\sum_{y \in \mathcal{N}} e^{-\frac{|d_x - d_y|}{k}}}{||\mathcal{N}||} \cdot [C_{x,d} - \mathbf{P}] + \mathbf{P}$$

$C_{x,d}$ and $C'_{x,d}$ are respectively the old and new native matching costs for pixel x at disparity d, $\mathcal{N}$ is the neighbourhood of x and d_w represents the disparity value of location w in the given initial estimate of the disparity map.

The value assumed by the first fraction is proportional to the ratio of pixels with a similar disparity value found in the chosen neighbourhood (the notion of "similarity" is controlled by the parameter k). This ratio is then used to modulate a linear interpolation between the actual cost $C_{x,d}$ and the penalization constant $\mathbf{P}$.

We purposely not inserted any locality principle or distance based penalty because we wanted our operator to be able to non-uniformly incentivate similar regions even if distant or unconnected. The global effect of the operator, when properly configured, is to enable the self-organization of the support regions, favoring compactness and inhibiting small or isolated areas. Thin structures, once established, usually provide themselves enough support to thrive.

4. Stereo algorithm

In order to evaluate our confidence modulation we developed as a testbed a simple, local stereo algorithm based on a greedy, fixed window correlation algorithm. Such methods are simple to implement and well-understood, letting us concentrate on assessing and factoring out the properties and the effects of our proposal.

We stress that, even if the resulting algorithm is capable of realtime performance and overall produces decent results it was never meant to be compared with the current state of the art but just as a evaluation platform.

4.1. Initial disparity estimation

We start by calculating a initial approximate disparity needed for our confidence operator. We choose as cost matching a truncated version of the popular Birchfield and Tomasi sampling insensitive measure [2]. To aggregate matching cost, we use a 5x5 gaussian filter with σ=2. The resulting disparity map, shown in fig.1, displays all the typical shortcomings of fixed-window correlation algorithms.

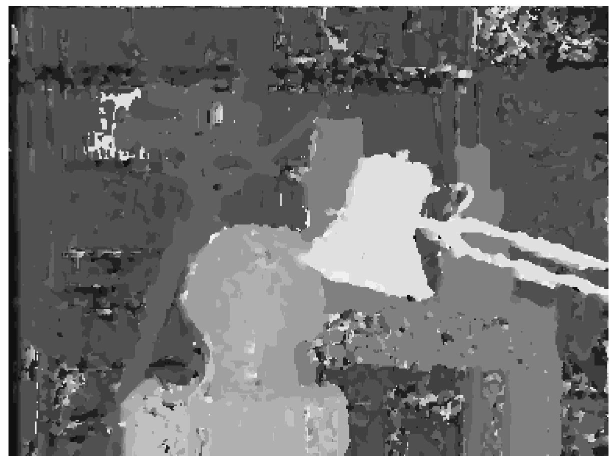

Figure 1. Initial disparity estimation.

4.2. Aggregation with modulated costs

Subsequently, we compute a novel disparity map using the modulated values computed from the estimate built in the previous step. Our neighbourhood choice is a uniform disk of radius 7.

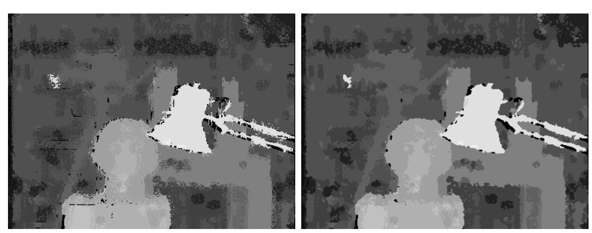

Figure 2. Raw and aggregated output of the confidence estimator.

The left side of figure 2 shows the map obtained when not using any form of cost aggregation: each pixel then assumes the disparity value that minimizes its cost. The picture presents some curious visual artifacts near discontinuities, caused by the influence of pixels across the depth gap. On the right the same cost volume is shown but aggregated with a small 3x3, σ=1 gaussian filter.

What both pictures have in common is a drastic decrease of the noise levels with respect to the initial disparity estimation. Other effects include the reduction of border bleeding and the minor entropy of untextured region which are now filled with still wrong yet more uniform disparities.

4.3. Disparity cleaning

In this step we apply some common and simple heuristics to remove small and untextured regions. To remove this second category we compute an estimate of the noise magnitude and variance and use them to threshold the sum of pixel-wise matching costs (fig.3). The resulting regions are then assigned to the best overall disparity for the entire group. Small holes caused by removing small regions are filled with the minima between the neighbouring left and right disparity. On the right of figure 3 is shown the resulting depth map.

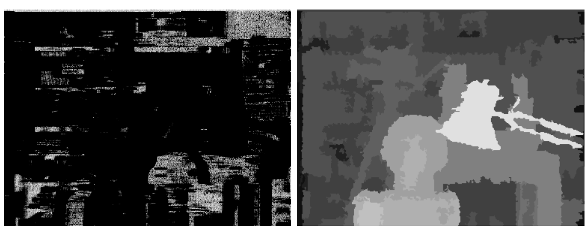

Figure 3. Untextured regions and the cleaned depth map.

4.4. Final regularization step

Since in the previous step we have obtained a new disparity estimate, we can now use it again to produce a confidence modulated depth map. The resulting depth map is further checked for consistency using the unicity constraint. The final result is shown in fig.4. It is surprisingly good considering it was produced from a standard winner-take-all window correlation algorithm.

5. Experiments

In order to obtain quantitative results we have run the algorithm described in the previous section on all the four couples in the Middlebury stereo benchmark with and without the proposed confidence based cost modulation. The obtained results are reported in table 1.

Figure 4. Final disparity map for the Tsukuba dataset.

Our complete results on the Middlebury dataset are shown in figure 5: skipping the analysis of the Tsukuba set that was already covered in section 4 we proceed to notice in the Venus couple some of the shortcomings of our näive approach: most of the bad pixels (marked in black) come from disparity holes erroneously filled across disparity boundaries and over untextured region.

The Teddy and Cones couples share the same problems, but moreover they show difficulties in filling image borders. Regarding Teddy, the whole ground plane is missing, because its steep angle and fine texture are completely incompatible with fixed-window correlation algorithms.

Overall, we can state that the obtained results are surprisingly good considering that they were computed with a local algorithm just by modulating its cost function. The method, while stabilizing itself in the lower half of the Middlebury table is at par with low-end global implementations.

Table 1. Comparaison of stereo results.

Image pair	with Cost modulation	without
Tsukuba	1.77_{23}	8.7_{41}
Venus	2.33_{32}	22.1_{41}
Teddy	15.0_{34}	24.3_{41}
Cones	10.0_{33}	21.4_{41}
Overall	30.9	76.5

6. Conclusions

We presented a novel confidence-based operator aimed at improving the reliability of an underlying

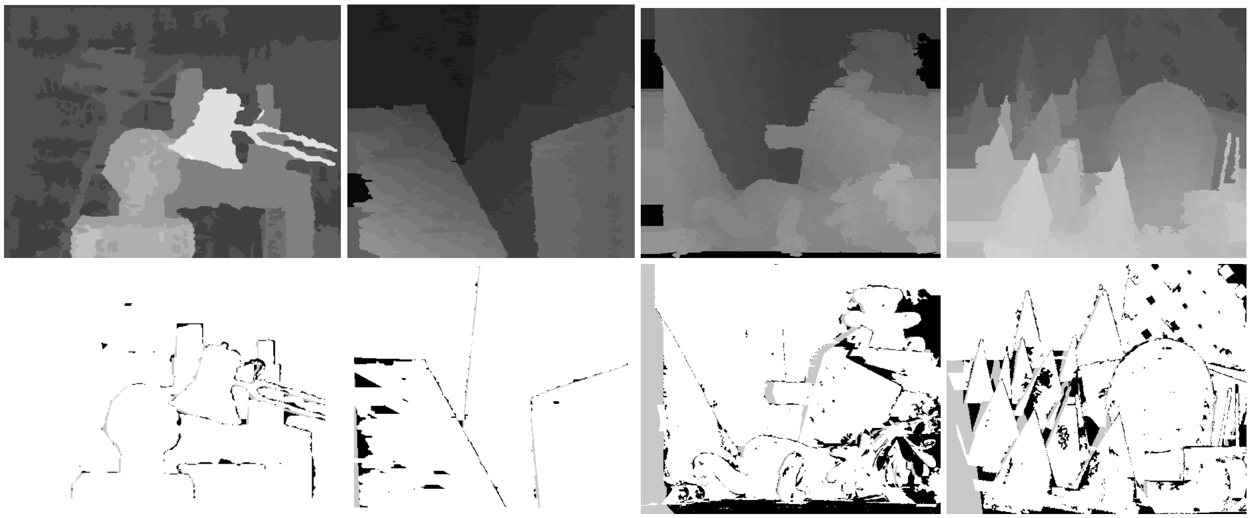

Figure 5. Results on the Middlebury dataset: Tsukuba, Venus, Teddy and Cones.

matching cost. The performance improvement has been demonstrated on a local, greedy stereo algorithm based on window correlation: our proposal was shown to dramatically improve the signal to noise ratio as well as to help better localize the borders. The performed quantitave evaluation demostrated that using our improved, confidence backed matching costs a local approach can obtain results surprising for a correlation-based algorithm and comparable to those of low-end global approaches, at a fraction of the sofistication and time complexity.

References

[1] S. Baker, S. Roth, D. Scharstein, M. Black, J. Lewis, and R. Szeliski. A database and evaluation methodology for optical flow. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007.

[2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, Apr. 1998.

[3] G. Egnal, M. Mintz, and R. P. Wildes. A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image Vision Comput.*, 22(12):943–957, 2004.

[4] A. Fusiello, V. Roberto, and E. Trucco. Symmetric stereo with multiple windowing. *IJPRAI*, 14(8):1053–1066, 2000.

[5] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *CVPR (2)*, pages 2402–2409, 2006.

[6] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *CVPR*, 2007.

[7] R. Manduchi and C. Tomasi. Distinctiveness maps for image matching. *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pages 26–31, 1999.

[8] R. Sara. Finding the largest unambiguous component of stereo matching. *Proceedings 7th European Conference on Computer Vision (ECCV2002)*, 2:900–914, may 2002.

[9] D. Scharstein and R. Szeliski. A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, IJCV*, 47(1):7–42, april 2002.

[10] L. Wang, M. Gong, M. Gong, and R. Yang. How far can we go with local optimization in real-time stereo matching. *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 129–136, June 2006.

[11] K.-J. Yoon and I. S. Kweon. Stereo matching with the distinctive similarity measure. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–7, 14-21 Oct. 2007.