ORDER STATISTIC FILTERS FOR IMAGE MATCHING

Jasmine E. Banks

Fraunhofer Institute for Computer Graphics,
Rundeturmstr. 6, Darmstadt 64283, Germany.
Email: jbanks@ieee.org

ABSTRACT

The rank and census are two filters based on order statistics which have been applied to the image matching problem for stereo pairs. Advantages of these filters include their robustness to radiometric distortion and small amounts of random noise, and their amenability to hardware implementation. In this paper, a new matching algorithm is presented, which provides an overall framework for matching, and is used to compare the rank and census techniques with standard matching metrics. The algorithm was tested using both real stereo pairs and a synthetic pair with ground truth.

The rank and census filters were shown to significantly improve performance in the case of radiometric distortion. In all cases, the results obtained were comparable to, if not better than, those obtained using standard matching metrics. Furthermore, the rank and census have the additional advantage that their computational overhead is less than these metrics. For all techniques tested, the difference between the results obtained for the synthetic stereo pair, and the ground truth results was small.

1. INTRODUCTION

Stereo vision is one technique for perception of the 3D environment, in which two or more images of a scene are taken from different perspectives, and depth information is obtained by triangulating corresponding points in the images. A fundamental problem faced by stereo vision algorithms is that of locating corresponding points in the images. This is known as the image matching or correspondence problem.

The rank and census are two filters based on order statistics which have been applied to the matching problem. Advantages of these filters include their robustness to radiometric distortion and small amounts of random noise, and their amenability to hardware implementation. This paper presents a new matching algorithm, which provides an overall framework for matching, and is used to compare standard matching metrics with rank and census based techniques.

This paper is organised as follows. Section 2 describes the rank and census filters. The new matching algorithm is then outlined in Section 3. Results obtained for a number of test images, using both standard matching metrics, and the rank and census, are presented in Section 4. In Section 5, results obtained for a synthetic stereo pair are compared with a known ground truth. Section 6 provides some discussion of these results and concludes.

2. RANK AND CENSUS FILTERS

The rank and census filters were first proposed for the stereo matching problem in [8]. Their reliance on the ordering of pixel values, rather than pixel values themselves, means that they are robust to radiometric distortion. This is significant because radiometric distortion is a problem which can often arise in the field, particularly where low cost cameras are used[2]. Furthermore their low computational complexity lends itself to hardware implementation, therefore they have potential for real-time applications.

2.1. Rank Filter

A window of size $M \times N$ is passed over the image. At each location in the image, the number of pixels less than the centre pixel are counted. This becomes the value of the rank image at that location. Two rank filtered image regions may be compared using the SAD (Sum of Absolute Differences) metric[1].

2.2. Census Filter

Again, a window of size $M \times N$ is passed over the image. At each location in the image, the pixel neighbourhood is mapped to a bit string. If a pixel is less than the centre pixel, the corresponding position in the bit string is set to 1, otherwise it is set to 0. Each location in the census transformed image therefore consists of a bit

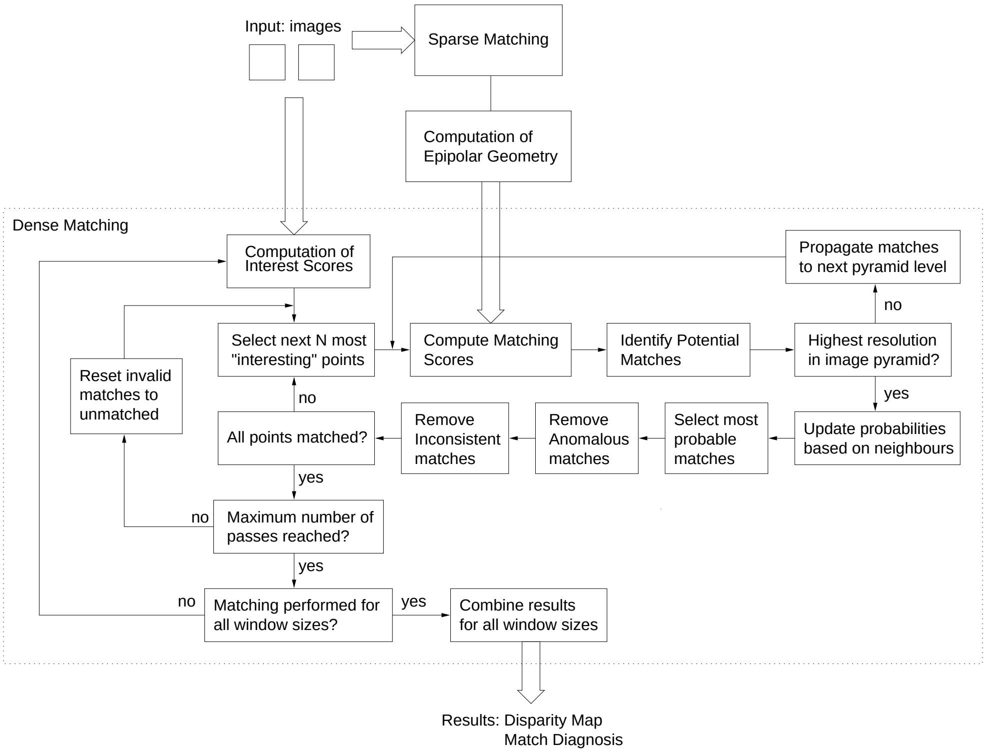

Figure 1: Overall matching algorithm.

string, which characterises the pixel neighbourhood at that image location. Regions in census filtered images may be compared by counting the number of bits which differ in the bit strings — effectively an XOR operation — and summing over the region.

Two hardware implementations of matching using the census filter are described in [3, 7].

3. MATCHING ALGORITHM

Figure 1 depicts the overall matching algorithm. The various components of this algorithm are described as follows:

3.1. Sparse Matching and Epipolar Geometry

The input to the algorithm consists of an image pair, and the fundamental matrix, which encapsulates the geometric relationship between the two images which comprise the image pair. For completeness, Figure 1 shows that the main algorithm is preceded by a sparse matching stage, which automatically detects a set of robust matches in the images, and a stage which computes the epipolar geometry (in this case performed using the INRIA FMatrix software[4]). The sparse matching algorithm could itself be decomposed into a block diagram and described in detail, however this is beyond the scope of this paper.

3.2. Computation of Interest Scores

An interest score is computed for every location in the images. Points whose interest score is below a threshold are flagged as "low interest".

3.3. Selection of Points

The N most interesting points are selected for matching, where N is typically set to one tenth the size of the image. The algorithm progressively matches the next N lesser important points, until all points are matched.

3.4. Computation of Matching Scores

A template window centered on each point in the first image is shifted in integer increments along the epipolar line in the second image. The value of the match score is computed at each candidate position, using match metrics[1], or the rank and census techniques. This results in an array of match scores, in which potential matches are identified as local maxima or minima, depending on the metric used. The match scores are used to estimate an initial probability for each match.

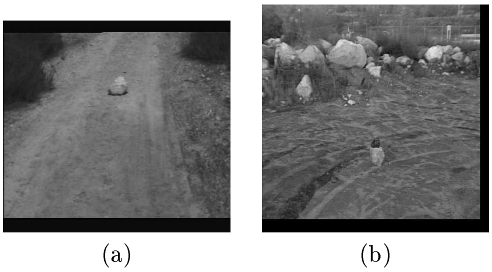

Figure 2: Test image pairs (a) road (b) hmmwv2.

3.5. Image Pyramid

The algorithm allows for the use of an image pyramid, where match scores are computed from multiple resolution images. If the highest resolution level has not been reached, the potential matches are propagated to the next level.

3.6. Update of Match Probabilities

The match probabilities are increased proportional to the number of neighbouring points already matched, which have a similar disparity.

3.7. Removal of Invalid Matches

For each point, the match having the highest probability is selected. The match score array is interpolated to determine the disparity to sub-pixel accuracy. A number of techniques are then used to cull matches likely to be invalid. Locally anomalous matches, which differ from their neighbours more than a given threshold, are flagged as "anomalous". Matching is then performed in reverse, that is, the matched location in the second image is itself matched back to the first image. If the matched location matches back to a location other than the original point, the match is flagged as "inconsistent".

3.8. Multiple algorithm passes

It is possible to run the algorithm more than once — on each "pass", all points flagged as having invalid matches are set to unmatched, and it is attempted to match these points again.

3.9. Multiple Match Window Sizes

The matching process may be repeated for a number of different window sizes. The disparity results obtained for each window size are then combined into a single disparity result. This is done by initially setting the global disparity to that obtained using the largest window size. The global disparity is updated in turn by the disparity results of successively smaller window sizes. It is updated depending on the existence of enough neighbours in the global disparity which are within a specified threshold of the smaller window disparity value. This strategy assumes that the disparity results from larger window sizes are the most reliable, and ensures that obviously incorrect disparity values from smaller window sizes do not influence the global disparity result. The aim of this technique is to improve the accuracy of the disparity result and remove the "smoothing" effect introduced by using a window of pixel values for matching.

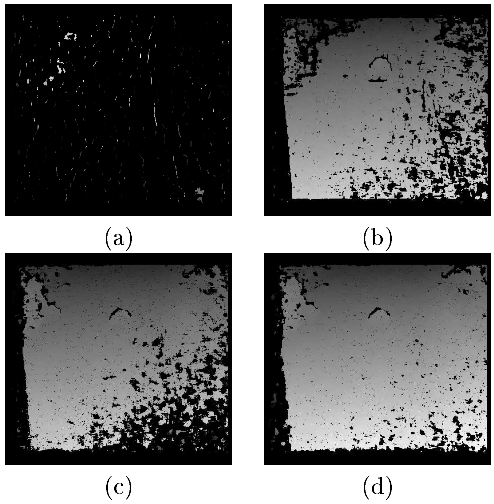

Figure 3: Disparity results for "road" stereo pair (a) SAD (b) NCC (c) RANK (d) CENSUS.

3.10. Output of the Algorithm

The outputs of the matching algorithm consist of x, y and absolute value of disparity images, as well as an image of "match flags". Each point in the "match flags" image is one of the following values:

matched - point successfully matched.

low interest - interest score below threshold.

not found - there exists no local maxima or minima of the match scores array.

inconsistent - match failed the consistency test.

anomalous - match was locally anomalous.

border - location was a border region.

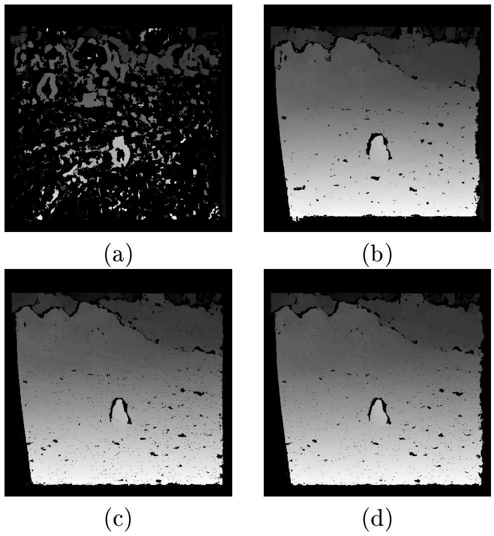

Figure 4: Disparity results for "hmmwv2" stereo pair (a) SAD (b) NCC (c) RANK (d) CENSUS.

4. RESULTS FOR TEST IMAGES

Figure 2 shows the left image of two different test pairs, obtained from [5]. Both image pairs suffer from radiometric distortion, for the "road" pair, the right image is approximately 30% brighter than the left, while for the "hmmwv2" pair, the left image is 15% brighter than the right.

Disparity results for the test pairs, using the SAD (Sum of Absolute Differences) and NCC (Normalised Cross Correlation) metrics [1], and the rank and census techniques, are shown in Figures 3 and 4. In each case, the results were produced using the simplest case — one pyramid level, one pass of the algorithm, and only one match window size (in this case 11). This was in order to focus on the comparison between standard matching metrics and rank and census based techniques, and to avoid the combinatorial explosion of alternative combinations of match algorithm input parameters.

Table 1 shows, for each test image pair, the proportion of points matched, and the number of matches returned for each value of match flag. The computation of the proportion matched does not include points flagged as "border" or "low interest", for which matching was not attempted.

5. COMPARISON WITH GROUND TRUTH RESULTS

Comparison with ground truth, ie, "correct" results, provides a method of testing the accuracy of the results returned by the matching algorithm. However, dense ground truth data is generally not available for real stereo pairs. The "corridor" stereo pair[6] is one synthetic stereo pair for which ground truth disparity data is known. Figure 5 shows the left image of this stereo pair, and the ground truth disparity image.

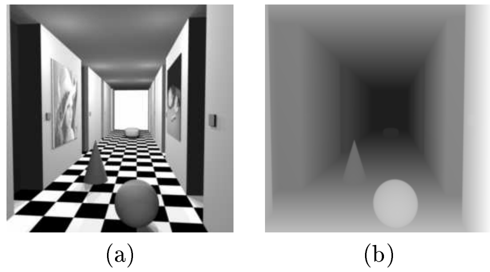

Figure 5: Corridor images (a) left image (b) ground truth disparity.

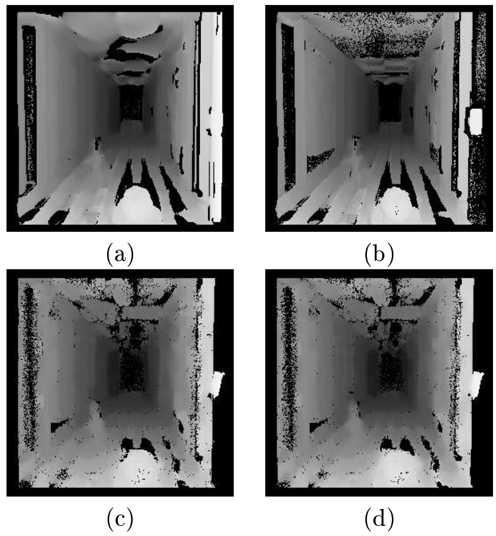

Figure 6: Disparity results for corridor images (a) SAD (b) NCC (c) RANK (d) CENSUS.

The matching algorithm was run with the parameters as for the test pairs in Section 4, except that the threshold "interest score" was set to zero. This meant that matching would be attempted for all image points (apart from border areas). Figure 6 shows the disparity results obtained using the SAD and NCC metrics, and the rank and census techniques. The proportion of points matched, and the number of points returned with each match flag, are given in Table 1. For the corridor images, Table 1 also shows the computed rms difference been the disparity of the matched pixels, and the ground truth disparity.

image	match method	prop. matched	matched	low interest	not found	inconsistent	anomalous	border	rms diff.
road	SAD	0.0208	4372	20050	7497	191374	6851	15616	-
	NCC	0.6942	145852	20050	3674	54473	6095	15616	-
	RANK	0.7221	150451	17971	3297	43769	10832	19440	-
	CENSUS	0.8662	180477	17971	2880	20782	4210	19440	-
hmmwv2	SAD	0.1849	43585	10335	22559	138294	31243	16128	-
	NCC	0.7845	184903	10335	21518	20945	8315	16128	-
	RANK	0.7915	185005	8325	21933	21104	5697	20080	-
	CENSUS	0.8010	187217	8325	22537	18310	5675	20080	-
corridor	SAD	0.8192	47182	0	1773	5795	2848	7936	0.3990
	NCC	0.7559	43539	0	3101	8044	2916	7936	0.5449
	RANK	0.8157	45431	0	4019	4796	1450	9840	0.4148
	CENSUS	0.8262	46016	0	4139	4189	1352	9840	0.4100

Table 1: Results of matching for test images.

6. DISCUSSION

From Figures 3 and 4 it can be clearly seen that the SAD is not robust to radiometric distortion. Other metrics, such as the NCC, are robust to this type of distortion, however introduce more computational overhead. It can be seen from Table 1, and from results from of other test pairs not shown in this paper, that the performance of the rank and census is generally comparable to, if not better than, metrics such as the NCC. Furthermore, they have the advantage that their computational overhead is much less than these metrics.

The matching metrics and the rank and census have also been compared using a known ground truth pair. In Table 1 shows this is the only pair for which the SAD metric has out-performed all other techniques. It is supposed that this is due to the "perfect" nature of a synthetic stereo pair — in that it does not suffer from noise or radiometric distortion. However, it can be seen from Table 1 that the rank and census have performed comparable to the SAD, and that for all matching techniques, the difference between the computed disparity and ground truth disparity is small.

Further work would involve more extensive testing of the presented matching framework — for example, using multiple "passes" of the algorithm to rematch points flagged as "not found", "inconsistent" or "anomalous", and combining results obtained from multiple window sizes.

7. ACKNOWLEDGEMENTS

This work was carried out with the support of the European Union through the CAMERA (ERB FMRX-CT97-0127) research network.

8. REFERENCES

[1] P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In Förstner and Ruwiedel, editors, Robust Computer Vision, pages 268–289. Wickmann, 1993.

[2] J. Banks, M. Bennamoun, P. Corke, and K. Kubik. Suitability of non-parametric stereo matching techniques for mining automation. Australian Journal of Intelligent Information Processing Systems, 5(2):111–119, 1998.

[3] P. Dunn and P. Corke. Real-time stereopsis using FPGAs. In Proceedings 7th International Workshop on Field-Programmable Logic and Applications, pages 400–409, London, September 1997.

[4] Softwares FMatrix & Fdiff. http://www-sop.inria.fr/robotvis/personnel/zzhang/software-fmatrix.html.

[5] Vision and Autonomous Systems Center's Image Database. http://www.vasc.ri.cmu.edu/idb/.

[6] Stereo Images with Ground Truth Disparity and Occlusion. http://www-dbv.cs.uni-bonn.de/ ft/stereo.html.

[7] J. Woodfill and B. Herzen. Real-time stereo vision on the PARTS reconfigurable computer. In Proceedings IEEE Workshop in FPGAs for Custom Computing Machines, pages 242–250, 1993.

[8] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In Proceedings ECCV'94, pages 151–158, Stockholm, May 1994.