EUSIPCO 2013 1569743475

EFFICIENT DISPARITY CALCULATION BASED ON STEREO VISION WITH GROUND OBSTACLE ASSUMPTION

Zhen Zhang, Xiao Ai, and Naim Dahnoun

Department of Electrical and Electronic Engineering
University of Bristol
Merchant Venturers Building
Woodland Road, Bristol
BS8 1UB, United Kingdom
Email: zhen.zhang@bristol.ac.uk

ABSTRACT

This paper presents a fast local disparity calculation algorithm on calibrated stereo images for automotive applications. By utilizing the ground obstacle assumption for a typical road scene, only a small fraction of disparity space is required to be visited in order to find a disparity map. It works by using the neighbourhood disparities of the pixels in the lower image line as supporting points to determine the search range of its upper vicinity line. Unlike the conventional seed growing based algorithms that are only capable of producing a semi-dense disparity map, the proposed algorithm utilises information provided by each pixel rather than trusting only the featured seeds. Hence, it is capable of providing a denser disparity output with low errors in homogeneous areas. The experimental results are also compared to a normal exhaustive search (block matching) algorithm, showing a factor of ten improvement in speed, whilst the accuracy is enhanced by 20% without constraint to the maximum possible disparity.

Index Terms— Stereo Vision, Ground plane, Obstacle, Block-based, Real-time

1. INTRODUCTION

Disparity calculation based on stereo cameras is a popular and growing research area [1]. It provides important depth information for object detection as the distance is inversely proportional to the disparity at each pixel in the image. In order to find the disparity, the cost or likelihood function must be decided in order to measure the similarities between the image pixels or patches of image pixels taken from the left and right images. Then, global or local optimisation is performed, based on the calculated costs, to determine the final disparities.

The global optimization algorithm [2] processes the matching as a problem of minimisation of energy functions, for example, the graph cut (GC) method [3], the belief propagation (BP) algorithm [4] and the dynamic programming method (DP) [5]. GC and BP have a higher matching accuracy but at the same time a higher algorithm complexity, while DP has a higher computational efficiency and a better matching in practical applications, but it is still computationally intensive. Hence, they are not suitable for real-time applications.

The most important step for the disparity calculation is to find the correspondence between the pixels or patches on the left and right images. For the local based methods [2, 6–10], the correspondence is calculated based on the information provided by a block of pixels surrounding the pixel of interest (cost functions). However, their disadvantage is that a large amount of errors can be introduced in the occluded and homogeneous regions. To improve the performance in homogeneous regions, cost aggregation methods [2] are utilised, in which the correspondence is found by averaging the neighbourhood cost functions. Such algorithms can outperform many global optimisation algorithms, evidenced by our previous work [11]. However, they cannot achieve real-time performance as they require exhaustive searches (a full search range) and exhaustive dense matching is usually not satisfactory because the same search range (typically 50 to 100) is used for the entire image, potentially yielding many false matches. Cech et al. [12] proposed disparity detection based on a small set of corresponding seeds (GCS). The correspondence is then found in a small neighbourhood around an initial set of seed correspondences and grown iteratively. However, the major problem in this approach is that, when a seed is incorrectly selected, the error will cumulate in the matching procedure and cannot be recovered. So there are many mismatches in the final result which could not satisfy our requirements of obtaining a dense disparity map.

For a typical road scene, the disparity of the road decreases from the bottom to the top and the disparities of the obstacle remain the same. Fig. 1 shows a typical road scene with its disparity map. In the disparity map, we can observe

(a)

(b)

Fig. 1: Stereo vision disparity map example. (a): original reference image of a typical road scene. (b): disparity map of the scene.

that the base part of most obstacles intercepts with the road, showing similar disparities. The other part of the obstacle which protrudes from the road, showing higher disparities than the road disparity of the same image line, whilst still share similar disparities with the base part of the obstacle. By using these observations, we assume that all obstacles are on the road. That means, for an obstacle, the bottom part would have the same disparity as the road.

Exploring the road obstacle assumption and inspired by the GCS algorithm, we proposed our algorithm. In the proposed algorithm, the disparities of the pixels in the image line v are calculated with a shorter search range. This search range is generated by using the neighbourhood disparities of the pixels in the image line $v-1$. Hence, a much faster calculation compared to the traditional block-based algorithm can be achieved. Unlike the seed based algorithm [12], the proposed algorithm used all the pixels in the lower lines as support points to achieve a better performance.

In the reminder of this paper, Section 3 introduces details of the proposed algorithm, Section 4 compares the experimental results of the proposed algorithm with two similar algorithms and Section 5 concludes the paper.

2. CONVENTIONAL SEED GROWING ALGORITHM

Because disparity errors mostly occur in homogeneous areas of the images, to determine robust seed disparity points, conventional seed growing algorithms such as the GCS and Lhillier et al. [12, 13] use the Harris corner detector (acquisition of feature points) on both the right and left images and

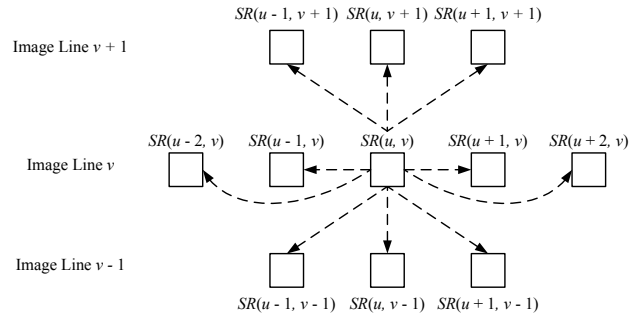

Fig. 2: Conventional seed growing algorithm example. $SR(u, v)$ denotes the search range for pixel (u,v) and it propagates the search range to neighbouring pixels.

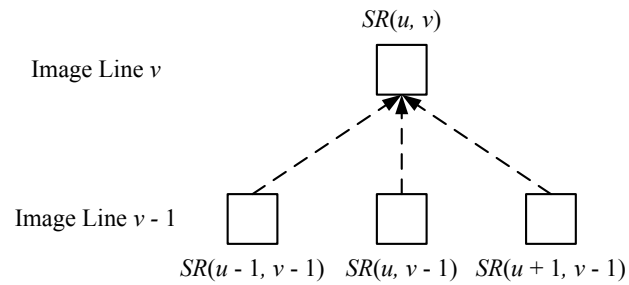

Fig. 3: Search range generation example.

correspond such feature points to obtain their disparities.

These algorithms then grow an initial seed to its neighbouring region and propagate the search range to 11 neighbours of the seed, as shown in Fig.2. Each neighbour shall then obtain a disparity value that minimises the cost (similarity) function within the propagated search range. Then, it is added to the seed list, unless the minimum cost is below a threshold (an indication of unreliable matches) or the disparity value violates the uniqueness constraint, that is it fails the left and right consistency check (indication of occluded area). Another seed is drawn in the order of the image similarity and the process is repeated until the set of seeds is empty. In this way, matching relationships spread from seeds to neighbouring regions of the entire image. Fig. 2 illustrates this process.

3. PROPOSED ALGORITHM

For this paper, we assume the input images are rectified and co-planar, so that epipolar lines are aligned with corresponding scanlines. In this case, the correspondence can only exist on the same scanline. If $p(u, v)$ and $q(u', v)$ are corresponding pixels in the left and right images respectively, then the disparity d between $p(u, v)$ and $q(u', v)$ is defined as $d = u - u'$.

The proposed algorithm consists of three steps: matching cost computation, search range recalculation and, finally, dis-

parity enhancement. Firstly, it is assumed that every obstacle is on the road surface and, then, that the bottom part of the obstacle has the same disparity as the road surface of the same image line.

3.1. Matching cost computation

The Normalized Cross-Correlation (NCC) has been chosen as the cost function for many traditional algorithms [1]. Unlike the Sum of Absolute Difference (SAD), it is less sensitive to intensity changes between the left and right images, which can occur when two physical cameras are used. The NCC cost volume (C) is calculated as

$$C(u,v,d) = \frac{\sum_{(u,v)\in W} \left[I_r(u,v) - \bar{I}_r\right] \cdot \left[I_l(u+d,v) - \bar{I}_l\right]}{\sqrt{\sum_{(u,v)\in W} \left[I_r(u,v) - \bar{I}_r\right]^2 \cdot \sum_{(u,v)\in W} \left[I_l(u+d,v) - \bar{I}_l\right]^2}} \tag{1}$$

where I_r and I_l represent the intensity of pixels on the right and the left images, and u and v represent the image row and column number, respectively. The shift distances (d) of the maximum values in the cost volume are selected as the disparities for the pixels. By calculating the NCC cost, the cost volume $C(u,v,d)$ for each pixel is obtained. For each pixel's matching cost function, a distinct peak may exist which indicates the correct correspondence. However, this process consumes a large amount of computational power during the exhaustive correspondence searching and, when complex situations are encountered, errors are likely to be introduced in homogeneous regions. To solve this problem, we proposed that our algorithm should generate a controlled search range.

3.2. Controlled search range

Giving the road obstacle assumption, the disparity of the current pixel $d(u,v)$ would have connections with the disparities of the pixels of the lower line $d(u-1,v-1)$, $d(u,v-1)$ and $d(u+1,v-1)$. By exploring these connections, the neighbourhood disparities of the pixels in the image line $v-1$ were used to generate a much smaller search range of the pixels in the image line v. By controlling the search range $SR(u,v)$ according to the neighbourhood support points, the potential matching ambiguities were reduced. Hence, a much faster and more accurate calculation compared to the exhaustive search algorithms can be achieved. The proposed algorithm generates a controlled search range according to Eq. (2).

$$SR(u,v) = SR(u-1,v-1) \cup SR(u,v-1) \cup SR(u+1,v-1) \tag{2}$$

Algorithm 1 Proposed disparity calculation algorithm

```
disp(u, v) - disparity for pixel (u, v)
for (u, v) ∈ image size do
    SR(u − 1, v − 1) =
    (disp(u − 1, v − 1) − τ) : (disp(u − 1, v − 1) + τ)
    SR(u + 1, v − 1) =
    (disp(u + 1, v − 1) − τ) : (disp(u + 1, v − 1) + τ)
    SR(u, v − 1) =
    (disp(u, v − 1) − τ) : (disp(u, v − 1) + τ)
    SR(u, v) =
    unique(SR(u−1, v−1), SR(u, v−1), SR(u+1, v−1))
    for d_i ∈ SR(u, v) do
        Calculate cost(u, v, d_i) over
        central square window of size w
    end for
    disp(u, v) = arg min(cost(u, v, d_i))
end for
```

where

$$SR(u-1,v-1) \in \{(d(u-1,v-1)-\tau)\ldots(d(u-1,v-1)+\tau)\} \tag{3}$$

$$SR(u,v-1) \in \{(d(u,v-1)-\tau)\ldots(d(u,v-1)+\tau)\} \tag{4}$$

$$SR(u+1,v-1) \in \{(d(u+1,v-1)-\tau)\ldots(d(u+1,v-1)+\tau)\} \tag{5}$$

$d(u,v)$ represents the disparity value of the pixel in position (u,v) and τ denotes the bound of the search range. By minimising τ, the $SR(u,v)$ can be minimised and is shown in Section 3.3. The proposed algorithm is described in the pseudo code above.

3.3. Computational complexity

For the v_{max} line which is at the bottom line of the image, a full search range (SR) from 0 to d_{max} is used and the Winner-Take-All (WTA) technique is used to select the disparities. For the v_n line, SR is controlled by supporting points in v_{n-1}, as described in Eq. (2). By using the controlled SR, the number of operations required for correspondence matching is significantly reduced, while the sharpness of the edges in the disparity map is still preserved.

$$O_{full} = N \times M \times d_{max} \tag{6}$$

$$O_{proposed} = \left(\sum_{u=0}^{N}\sum_{v=0}^{M-1} SR(u,v)\right) + (N \times d_{max}) \tag{7}$$

O_{full} denotes the operation required for the exhaustive search algorithm to calculate the disparity, while $O_{proposed}$

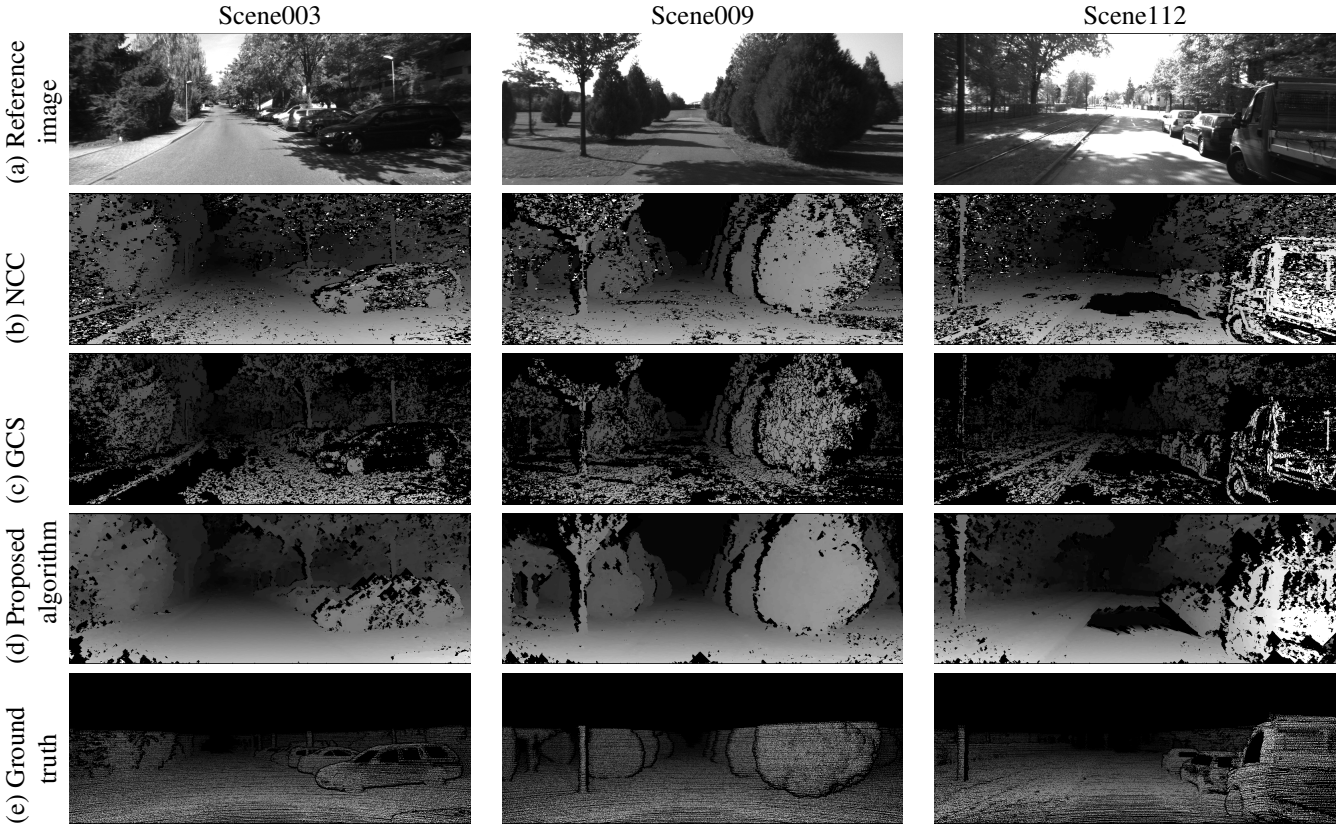

Fig. 4: Stereo vision evaluation results using different road scenes [14]. The first row is the reference image. The second row is the resultant disparity map for NCC. The third row is the resultant disparity map for the GCS algorithm [12]. The fourth row is the resultant disparity map for the proposed algorithm. The last column is the ground truth disparity map.

denotes the number of operations required for the proposed algorithm to calculate the disparity. N and M are the width and height of the image I. In the experiment, $\tau = 2$ is used to minimise the controlled search ranges and $d_{max} = 100$. While in most cases, $SR(u, v) = 5$, regardless of implementation used, since if the search range changed, this would impact on the results. Compared with the exhaustive search methods, more than 90% of the operation can be saved by the proposed algorithm.

4. EXPERIMENTAL RESULTS

In our experiments, the KITTI image datasets [14] are used. The resolution of the test image is 1242×376 pixels. A total number of 194 different road image pairs with their ground true disparity maps are used to evaluate the performance of the algorithm in real world situations. Two algorithms, NCC [2] and GCS [12], have been used as a comparison. Table 1 summarises the comparison between the proposed algorithm with two disparity estimation algorithms. In this case, $W = 5$, $d_{max} = 100$ and $\tau = 2$ have been chosen to achieve optimum results for all test data sets. The error rates are generated by comparing the results with the ground truths in non-occluded areas, while the ground truths are generated using the laser scanner [14]. An absolute error threshold of 1 was chosen to evaluate the performance of our algorithm, whilst only allowing less than 1 in the disparity step variation. As the table illustrates, the proposed algorithm achieved a better error rate.

	NCC %	GCS %	Proposed algorithm %
Scene099	6.96	18.10	2.33
Scene128	5.72	14.08	2.76
Scene148	4.62	16.79	2.50
Average	10.94	20.03	7.56

Table 1: Percentage of error pixels (absolute disparity error > 1) for three image pairs in non-occluded areas.

In Table 1, some of the frames have been selected and the last row is the average error rate of the total 194 scenes. The proposed algorithm achieved significant improvement compared to two similar algorithms, while maintaining a low computational cost.

Fig. 4 shows the results of a road scene using the proposed methods. The first row is the reference image. The second row is the resultant disparity map using the WTA technique directly from the NCC costs, with a Left-Right consistency check. The third row is the resultant disparity map for the GCS algorithm. The fourth row is the resultant disparity map for the proposed algorithm with a Left-Right consistency check applied. The last column is the ground truth disparity map for these image pairs.

By comparing with the NCC, we can see that the proposed algorithm has better accuracy on planar surfaces and preserves the sharpness of the edges, whilst there are only errors occuring in the occluded areas. Because the proposed algorithm is a local block-based algorithm, it also has the same disadvantages as most of the other block-based algorithms, such as NCC in this case. There are some errors occurring in the homogeneous areas because a small search window was applied. All three methods also generated large errors from the highly saturated areas, indicated in the last column of Fig. 4. These errors could potentially be extrapolated by global optimization techniques but these are not the focus of this work.

Comparing with GCS methods, the proposed algorithm provides a much denser disparity map, especially for ground surface detection and localization of small obstacles that would be missed out by a semi-dense disparity map. This is very important for the automotive applications. It also outperforms the GCS algorithm in homogeneous areas, as shown in Fig. 4(c). Because of its accuracy and low computational complexity, the proposed algorithm is suitable for automotive applications, which require real-time performance on resource limited embedded systems. Because of the unique property of the results from the proposed algorithm, which favours planar surfaces, the disparity result is suitable for on-road clear space detection, obstacle detection, 3-D reconstruction applications and other applications.

5. CONCLUSION

In this work, an efficient disparity calculation algorithm is presented and the road obstacle scenario has been exploited to reduce the computation and improve the accuracy of the disparity calculation. The computation reduction is due to the controlled correspondence search range that is propagated from the lower image line to its vicinity upper line. With the reduction in the search range, the cost function peak ambiguity can be minimized, which yields the attained accuracy improvements. By comparing the results with the conventional exhaustive search block matching algorithm and a state-of-the-art region growth algorithm using the KITTI stereo datasets, the experimental results indicate that the proposed algorithm has the advantage of providing much denser and more accurate disparity maps. The benefit of low errors in homogeneous areas and the high preservation of edge details can be especially useful in automotive applications, such as road detection and modelling, and obstacle detection and classification. It is also capable of real-time implementation on to the embedded systems due to the relaxed memory requirement and highly parallelable (by image lines) code architecture. This is our immediate future work.

6. REFERENCES

[1] N. Lazaros, G.C. Sirakoulis, and A. Gasteratos, "Review of Stereo Vision Algorithms: From Software to Hardware," *International Journal of Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[2] 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.

[3] Li Hong and G. Chen, "Segment-based stereo matching using graph cuts," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June-2 July, vol. 1, pp. I–74–I–81 Vol.1.

[4] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient belief propagation for early vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June-2 July, vol. 1, pp. I–261–I–268 Vol.1.

[5] Jae Chul Kim, Kyoung-Mu Lee, Byoung-Tae Choi, and Sang-Uk Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, June, vol. 2, pp. 1075–1082 vol. 2.

[6] S. Yoon, S.K. Park, S. Kang, and Y.K. Kwak, "Fast correlation-based stereo matching with the reduction of systematic errors," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2221–2231, 2005.

[7] L.D. Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," *Image and vision computing*, vol. 22, no. 12, pp. 983–1005, 2004.

[8] Zhen Zhang, Yifei Wang, Jason Brand, and Naim Dahnoun, "Real-time obstacle detection based on stereo vision for automotive applications," in *5th European DSP Education and Research Conference (EDERC 2012)*, Amsterdam, The Netherlands, September 2012, pp. 281–285.

[9] Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, and Margrit Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3017–3024.

[10] Leonardo De-Maeztu, Stefano Mattoccia, Arantxa Villanueva, and Rafael Cabeza, "Linear stereo matching," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1708–1715.

[11] Zhen Zhang, Xiao Ai, N. Canagarajah, and N. Dahnoun, "Local stereo disparity estimation with novel cost aggregation for sub-pixel accuracy improvement in automotive applications," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, June 2012, pp. 99 –104.

[12] Jan Cech and Radim Sara, "Efficient Sampling of Disparity Space for Fast And Accurate Matching," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.

[13] Maxime Lhuillier and Long Quan, "Match propagation for image-based modeling and rendering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 8, pp. 1140–1146, 2002.

[14] Bernd Kitt, Andreas Geiger, and Henning Lategahn, "Visual Odometry based on Stereo Image Sequences with RANSAC-based Outlier Rejection Scheme," in *IEEE Intelligent Vehicles Symposium*, San Diego, USA, June 2010, pp. 486–492.