# An alternative confidence measure for local matching stereo algorithms

*Thulani Ndhlovu*

Mobile Intelligent Autonomous Systems
CSIR
South Africa
Email: tndhlovu@csir.co.za

*Fred Nicolls*

Department of Electrical Engineering
University of Cape Town
South Africa
Email: fred.nicolls@uct.ac.za

***Abstract*—We present a confidence measure applied to individual disparity estimates in local matching stereo correspondence algorithms. It aims at identifying textureless areas, where most local matching algorithms fail. The confidence measure works by analyzing the correlation curve produced during the matching process. We test the confidence measure by developing an easily parallelized local matching algorithm, and use our confidence measure to filter out unreliable disparity estimates. Using the Middlebury dataset and our own evaluation scheme, the results show that the confidence measure significantly decreases the disparity estimate errors at a low computational overhead.**

## I. Introduction

Stereo vision is an actively researched topic in computer vision. In robotic systems, stereo vision provides a low-cost alternative for range imaging compared to expensive laser range-finders for applications such as 3D reconstruction and obstacle avoidance. The major issue in such a system is the correspondence problem: given two or more images of the same scene from different viewpoints, find corresponding pixels and the distance by which the pixel in one view is translated relative to its corresponding pixel in the other view. A number of solutions have been produced to the extent that an online evaluation has been developed [1].

Solutions for the stereo correspondence problem consist of complex modules such as plane-fitting, edge-preserving smoothing and image segmentation. Among these solutions are local matching algorithms, which can be easily parallelized for real-time applications. Although these algorithms are applicable in real-time systems, they generally produce more errors compared to other more complex non-real-time approaches.

We present a method of assigning a confidence to a disparity estimate for local matching algorithms. Our approach is expected to give low confidences to disparity estimates in textureless regions, where many local matching algorithms fail. While our approach is similar to a number of previously developed confidence measures, in that the confidence of a disparity estimate is a by-product of the matching process, our analysis focuses on the basin of convergence (refer to Fig. 3) of a disparity estimate.

To evaluate our confidence measure, we implement a local matching algorithm. This confidence measure is expected to be applicable across the different variations of these algorithms because of the uniform structure of local matching algorithms. We run our algorithm on the widely used Middlebury dataset [1] in order to evaluate the performance of our confidence measure using our evaluation scheme.

The remainder of the paper is structured as follows. Section II briefly covers the related literature. Section III discusses the local matching algorithm implemented. Section IV discusses the confidence measure and how we use it for disparity refinement. Section V discusses our evaluation methodology and the results of experiments. The paper is concluded in Section VI.

## II. Related work

In stereo vision research, there have been several successful approaches in representing the confidence of a disparity estimate. The left-right consistency constraint [2]–[6] has been traditionally used to characterize pixel ambiguity. The constraint checks the left image reference disparity estimate and compares it to the inverse mapping of the right image reference disparity estimate. This approach is successful in detecting occluded regions. There have been approaches that analyze the matching score of the disparity estimate [7], [8]. The confidence of a pixel is based on the magnitude of the similarity value between the pixel in the reference image and the matching pixel in the target image. Other approaches analyze the curvature of the correlation curve [9], [10] and assign low confidences to disparity estimates resulting from a flat correlation curve. Approaches such as [11], [12] estimate the confidences of pixels with two similar match candidates. Research has also been conducted in determining pixel confidence based on image entropy [13], [14]. Low confidence scores are assigned to low entropy points in the reference image. Recently, a new approach has been developed which extrapolates confidence *a posteriori* from an initial, given and possibly noisy disparity estimate [15].

## III. Stereo algorithm

According to [16], stereo vision algorithms generally perform the following steps:

1) matching cost computation, where a matching cost used to quantify pixel similarity is formulated,
2) cost aggregation, where a support region is defined to spatially aggregate the matching cost,
3) disparity computation, where the best disparity hypothesis for each pixel is computed to minimize a cost function, and
4) disparity refinement, where the computed disparity maps are post-processed to remove mismatches or to produce sub-pixel disparity.

We are interested in local matching algorithms, which generally perform the steps 1, 2 and 3. We include step 4 as our confidence measure. Our interest in local-matching algorithms is motivated by the following:

- they can be easily parallelized which allows them to be implemented on graphics processing units or field programmable gate arrays for real-time computation,
- they are generally used as an initial estimate for a number of the state-of-the-art algorithms, and
- their uniform structure, shown in Algorithm 1, which includes steps 1-4 mentioned above, allows our confidence formulation to be used across the different variations of these algorithms.

However, local matching algorithms generally fail because of a lack of texture in an image and occluded regions.

**Algorithm 1** Stereo algorithm

*INPUT: Stereo images, window size, disparity range.*
*OUTPUT: Disparity map, confidence map.*

```
for each pixel in the left frame do
    set support region around the pixel (left frame)
    set search window in the right frame
    for each pixel in the search window (right frame) do
        set correlation window around the pixel
        correlate support region with correlation window
    end for
    find best match
    calculate disparity
    calculate disparity confidence
end for
```

Local matching algorithms generally differ in steps 1 and 2, matching cost computation and cost aggregation. For a comprehensive study of matching costs and cost aggregation, the reader is referred to [17] and [18].

For our purposes, we use Birchfield and Tomasi's sampling insensitive matching cost [19]. We also perform the left-right consistency check to detect occluded regions and filter them out. Fig. 1 shows the Tsukuba image pair and its ground truth. The resulting disparity map from our algorithm for the Tsukuba image pair with a $5 \times 5$ aggregation window and 15 disparities followed by a $5 \times 5$ median filter, is shown in Fig. 2. It should be noted that the algorithm implemented is to be used as a testbed for our confidence measure and is not meant to be compared with the state of the art.

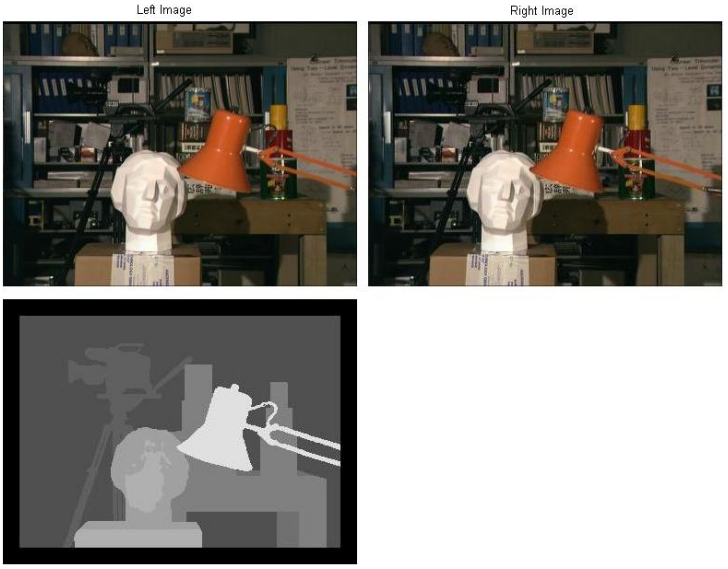


Fig. 1. Tsukuba image pair at the top and its ground truth at the bottom with the left image as reference.

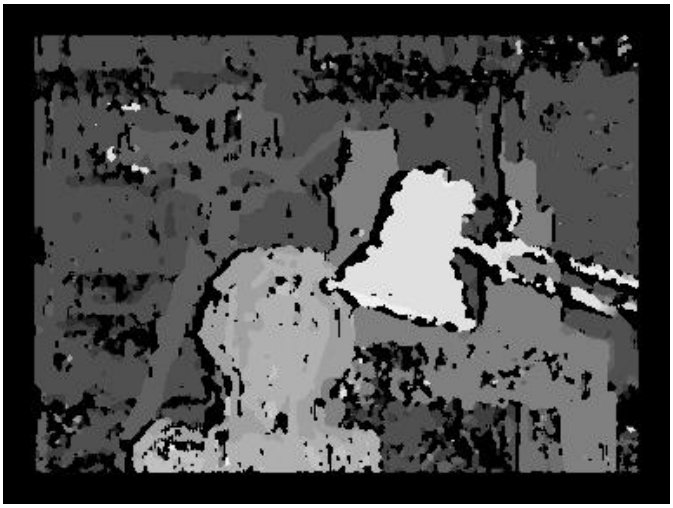

Fig. 2. Computed disparity map of the Tsukuba image pair. Close objects are bright while distant objects are darker.

## IV. Confidence measure

Our confidence measure is calculated as a function of $x, y, d$, where $(x, y)$ are the image coordinates and $d$ is the disparity. A typical correlation curve is shown in Fig. 3. Local matching algorithms aim to find the disparity which minimizes the error of this curve. In this instance the sum of absolute differences (SAD) of the intensity values is used as the error

measure. By analyzing the basin of convergence, $B$, of the disparity estimate, a confidence of the estimate can be inferred. Given a disparity $d$, the confidence of a disparity estimate can be computed as follows:

$$C(d) = \frac{B}{d_{max} - d_{min}}.$$

Here $C(d)$ is the confidence for a given disparity, $B$ is the basin of convergence of the disparity estimate $d$, and $d_{max} - d_{min}$ is the disparity range. It is expected that in textureless regions the correlation curve will have multiple local minima with small $B$ values, and since $C(d)$ is proportional to $B$ we expect low confidences. A high confidence value would have few local minima in the correlation curve and a fully confident disparity estimate would arise where the local minimum is the global minimum of the correlation curve.

Our algorithm uses gradient ascent to determine $B$. Given a disparity estimate $d$, we perform gradient ascent on both sides of the estimate and determine the two local maxima. The number of disparities covered by the two local maxima is defined as the basin of convergence $B$.

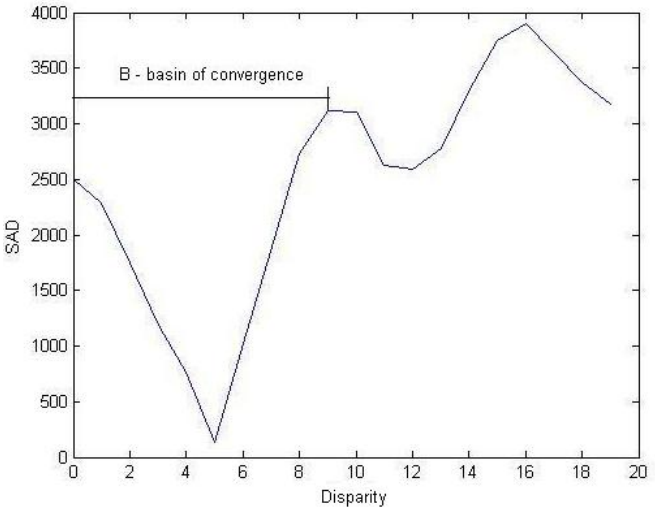


Fig. 3. Correlation curve and the basin of convergence.

### A. Disparity refinement

After computing confidences for all the disparity estimates, we empirically select a threshold, $T$, to create a mask of acceptable and unacceptable estimates. Acceptable disparity estimates are defined as those satisfying $C(d) > T$. The refined disparity map for $T = 2$ is shown in Fig. 4. By comparing Fig. 2 and Fig. 4 one can visually see that most of the noisy estimates arising from the local matching algorithm are filtered out successfully.

Fig. 4. Refined disparity map for the Tsukuba image pair with $T = 2$.

## V. Experiments

The Middlebury stereo benchmark provides a testbed to quantitatively evaluate stereo algorithms. Although the testbed is widely used in the computer vision community, it requires a dense disparity map. Generally, algorithms which perform disparity refinement would also include a hole filling step. Our algorithm does not perform hole filling because of the errors it might introduce, which leaves a sparse disparity map. Evaluating our sparse disparity map on the Middlebury stereo benchmark would not be appropriate because most errors would arise from the filtered out disparities. Thus we use our own evaluation scheme.

Pixels are classified as containing no information, unreliable information, or good information. We define occluded pixels as containing no information, pixels with $C(d) \leq T$ as containing unreliable information, and the rest of the pixels as containing good information. In our evaluation we only consider pixels containing good information.

We calculate the root mean square error ($RMSE$) as follows:

$$RMSE = 100 \times \sqrt{\frac{1}{N_p} \sum_{(x,y) \in p} (d(x,y) - d_g(x,y))^2},$$

where, $p$ is the set of all pixels containing good information, $N_p$ is the number of pixels containing good information, $d(x,y)$ is the estimated disparity at pixel $(x,y)$, and $d_g(x,y)$ is the ground truth disparity at pixel $(x,y)$.

Included in our evaluation is the percentage computational overhead for a chosen value of $T$. We include this metric instead of time in seconds because the actual time depends on the processor used to carry out the experiments. Percentage computational overhead on the other hand is independent on

TABLE I
TABLE SHOWING THE $RMSE$ WITH A CHOSEN WINDOW SIZE, NUMBER OF DISPARITIES, AND THRESHOLD $T = 0$ FOR THE MIDDLEBURY DATASET.

| Image pair | Window size | Number of disparities | $T$ | $RMSE$ |
|---|---|---|---|---|
| Tsukuba | $5 \times 5$ | 15 | 0 | 7.56 |
| Venus | $5 \times 5$ | 19 | 0 | 29.84 |
| Teddy | $9 \times 9$ | 59 | 0 | 37.43 |
| Cones | $9 \times 9$ | 59 | 0 | 30.13 |

TABLE II
TABLE SHOWING THE $RMSE$ FOR A CHOSEN WINDOW SIZE, AN OPTIMAL $T$ VALUE AND THE PERCENTAGE COMPUTATIONAL OVERHEAD FOR THE MIDDLEBURY DATASET.

| Image pair | Window size | T | RMSE | Computational overhead(%) |
|---|---|---|---|---|
| Tsukuba | $5 \times 5$ | 2 | 6.56 | 2.33 |
| Venus | $5 \times 5$ | 4 | 21.94 | 18.35 |
| Teddy | $9 \times 9$ | 2 | 32.79 | 19.23 |
| Cones | $9 \times 9$ | 6 | 23.38 | 17.60 |

the processing power.

The Middlebury dataset is used for evaluation. The results for the Middlebury dataset on the different image pairs with $T = 0$ are shown in Table I, while results with an optimal value of $T$ are shown in Table II.

In our experiments it was noted that the $RMSE$ starts increasing after a certain value of $T$ for a selected window size. This is due to the errors introduced by the window size. Local matching stereo algorithms assume constant disparity throughout the aggregation window, therefore errors known as the "foreground fattening" effect [16] arise. Also, since the images have pixel resolution, a window size greater than a pixel affects the resolution of our disparity estimates. Errors are introduced where the image details are smaller than the window size. Since our algorithm does not filter out these errors, they are fixed with a changing value of $T$. The larger the value of $T$, the smaller the value of $N_p$ while the errors remain fixed. A plot showing the relationship between $T$ and $N_p$ with a $5 \times 5$ window size for the Tsukuba image pair is shown in Fig. 5.

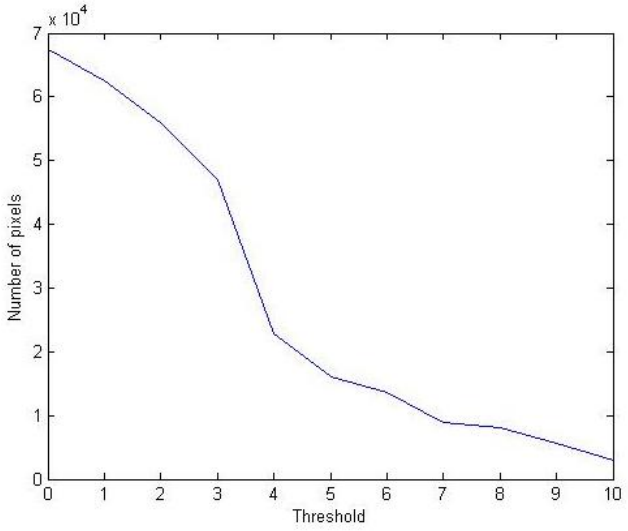


Fig. 5. Plot of number of pixels $N_p$ versus Threshold ($T$) with a $5 \times 5$ window size for the Tsukuba image pair.

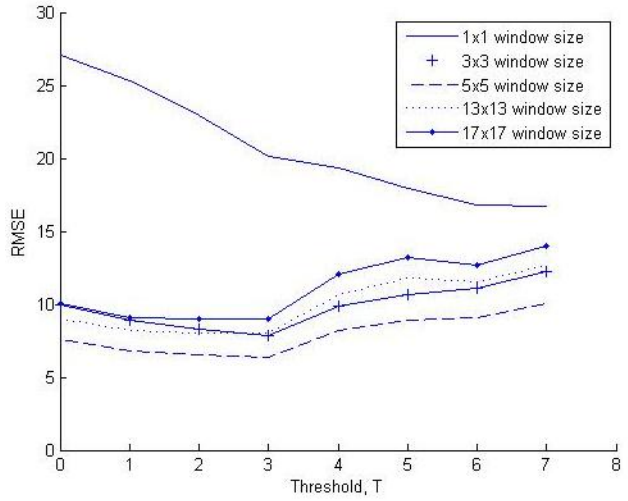


Fig. 6. Plot of Threshold ($T$) versus RMSE for varying window sizes.

To show the effect of the window size on our evaluation, Fig. 6 contains a plot of $T$ versus $RMSE$ for varying window sizes. Different window sizes tends to shift the curve up or down. As the window size increases, the curve shifts downwards until a point where a larger window introduces more errors, causing the curve to shift upwards.

## VI. CONCLUSIONS AND FUTURE WORK

We present a confidence measure to detect textureless regions for local matching algorithms. The effectiveness of our approach was demonstrated by implementing a local matching algorithm and filtering unreliable depth estimates. Our quantitative evaluation demonstrates that the confidence measure decreases the disparity estimate errors at a small computational cost.

We plan to use the developed confidence measure to address the problem of temporal stereo, which entails using previously computed disparity maps to seed new disparity maps. The confidence measure will aid in identifying good points to seed successive disparity estimates in the hope of decreasing computation time for stereo reconstruction.

## ACKNOWLEDGMENT

The author thanks the Council for Scientific and Industrial Research (CSIR) and the Mobile Intelligent Autonomous Systems (MIAS) group for their support on this work.

## REFERENCES


[1] D. Scharstein and R. Szeliski, "Middlebury stereo vision page," August 2009, `vision.middlebury.edu/stereo/`.

[2] G. Egnal and R. P. Wildes, "Detecting binocular half-occlusions: Empirical comparisons of five approaches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1127–1133, 2002.

[3] K. Konolige, "Small vision systems: hardware and implementation," in *Eighth International Symposium on Robotics Research*, 1997, p. 111–116.

[4] J. J. Little and W. E. Gillett, "Direct evidence for occlusion in stereo and motion," in *ECCV '90: Proceedings of the First European Conference on Computer Vision*. London, UK: Springer-Verlag, 1990, pp. 336–340.

[5] A. Luo and H. Burkhardt, "An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions," *Int. J. Comput. Vision*, vol. 15, no. 3, pp. 171–188, 1995.

[6] R. Trapp, S. Drüe, and G. Hartmann, "Stereo matching with implicit detection of occlusions," in *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*. London, UK: Springer-Verlag, 1998, pp. 17–33.

[7] M. J. Hannah, "Computer matching of areas in stereo images." Ph.D. dissertation, Stanford, CA, USA, 1974.

[8] D. Smitley and R. Bajcsy, "Stereo processing of aerial, urban images," *7ICPR*, vol. 84, pp. 433–435, 1984.

[9] P. Anandan, "Computing dense displacement fields with confidence measures in scenes containing occlusion," *IUW*, vol. 84, pp. 236–246, 1984.

[10] ——, "A computational framework and an algorithm for the measurement of visual," Amherst, MA, USA, Tech. Rep., 1987.

[11] J. J. Little and W. E. Gillett, "Direct evidence for occlusion in stereo and motion," in *ECCV '90: Proceedings of the First European Conference on Computer Vision*. London, UK: Springer-Verlag, 1990, pp. 336–340.

[12] D. Scharstein, "View synthesis using stereo vision," Ph.D. dissertation, Ithaca, NY, USA, 1997.

[13] G. Leclerc and Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," 1994.

[14] D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc, "Variable Albedo Surface Reconstruction from Stereo and Shape from Shading," in *Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina*, 2000.

[15] R. Gherardi, "Confidence-based cost modulation for stereo matching," in *ICPR*, 2008, pp. 1–4.

[16] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *SMBV '01: Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV'01)*. Washington, DC, USA: IEEE Computer Society, 2001, p. 131.

[17] H. Hirschmller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582–1599, 2009.

[18] L. Wang, M. Gong, M. Gong, and R. Yang, "How far can we go with local optimization in real-time stereo matching," in *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 129–136.

[19] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401–406, 1998.