

A Stereo Confidence Metric Using Single View Imagery

Geoffrey Egnal, Max Mintz
GRASP Laboratory
University of Pennsylvania
{gegnal,mintz}@grasp.cis.upenn.edu

Richard P. Wildes
Centre for Vision Research
York University
wildes@cs.yorku.ca

Abstract

Although stereo vision research has progressed remarkably, stereo systems still need a fast, accurate way to estimate confidence in their output. In the current paper, we explore using stereo performance on two different images from a single view as a confidence measure for a binocular stereo system incorporating that single view. Although it seems counterintuitive to search for correspondence in two different images from the same view, such a search gives us precise quantitative performance data. Correspondences significantly far from the same location are erroneous because there is little to no motion between the two images. Using hand-generated ground truth, we quantitatively compare this new confidence metric with five commonly used confidence metrics. We explore the performance characteristics of each metric under a variety of conditions.

1 Introduction

1.1 Overview

We present a new method to diagnose where a stereo algorithm has performed well and where it has performed badly. All stereo systems estimate correspondences, but not all of these correspondences are correct. Many systems do not give an accurate estimate of how trustworthy their results are. Our new confidence method addresses many causes of stereo error, but it focuses on predicting stereo error caused by low-texture regions. This error source is particularly bothersome when operating in urban terrain or when the imagery contains large regions of sky.

The new confidence metric is based upon correspondence in images from a single view. Although it seems counterintuitive to search for correspondence in two different images from the same view, such a search provides valuable information. It gives us precise quantitative performance data; disparities significantly far from zero are erroneous because there is no motion between the two images. We use the term Single View Stereo (SVS) to refer to the disparity map produced by a stereo system applied to two images from one view, separated in time.

Specifically, we use single view stereo performance data as a confidence metric that predicts how a binocular stereo

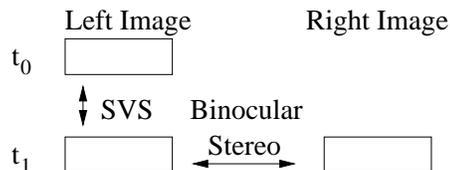


Figure 1: Single View Stereo. SVS searches for correspondence in two images from the same view. We use SVS failure as a confidence metric to predict binocular failure.

system which incorporates the single view would perform. The SVS output shows empirically where the stereo system has failed on the current scenery, and for this reason, SVS failure predicts failure pixelwise in the binocular case (see Figure 1). In practice, many stereo cameras are in motion, and the SVS disparity will not be zero at each pixel. We discuss how to modify the static SVS algorithm to accommodate images taken from a moving camera.

We compare the performance of SVS as a confidence metric with five commonly-used confidence metrics: (i) left/right consistency (LRC) predicts stereo error where the left-based disparity image values are not the inverse mapping of the right-based disparity image values. (ii) The matching score metric (MSM) directly bases confidence upon the magnitude of the similarity value at which the stereo matcher finds that left and right image elements match. (iii) The curvature metric (CUR) marks disparity points resulting from flat correlation surfaces with low confidence. (iv) The entropy-based confidence score (ENT) predicts stereo error at points where the left image has low image entropy. (v) The peak ratio (PKR) estimates error for those pixels with two similar match candidates. A brief comparison of the five metrics can be found in Table 1.

1.2 Previous Work

Unlike much research on stereo performance, we do not attempt to compare different stereo algorithms to see which is best [10, 5, 24, 29, 9]. Also, we do not attempt to find theoretical limits to stereo performance [4, 8, 19].

Instead, our current research deals with on-line confidence metrics which predict errors within seconds for a

	Error Source	SVS	LRC	MSM	CUR	ENT	PKR
Hardware	Lens Distortion		D	D			
	Sensor Noise	D	D	D			
	Quantization		D	D			
Algorithm (Software)	Resampling		D	D			
	Similarity Metric	D	D				
	Window Size Search Range		D D	D D			
External (Scene)	Half-Occlusion		D	D			
	Foreshortening		D				
	Periodic Structure	D	D				D
	Lighting Change Low Texture	D	D	D	D	D	D

Table 1: A Comparison of Five Confidence Metrics. The table lists various error sources and indicates with a 'D' which confidence metrics can detect errors due to the error source.

given system. Within the confidence metric field, there has been much research. Research that has considered left/right consistency includes [33, 18, 7, 20, 15, 32]. Previous research into the matching score metric includes [12, 30]. Researchers that have examined curvature metrics include [13, 1, 2], and research looking at entropy-like stereo confidence metrics includes [16, 26]. Previous work with the peak ratio includes [18, 28]. More general research into on-line error diagnosis includes [34, 17, 31]. In perhaps the most similar approach to ours, Leclerc, Luong and Fua [17] have shown the effectiveness of the self-consistency approach to estimate confidence. SVS differs from self-consistency in that SVS has a ground truth and does not rely on the agreement of multiple matching processes to predict performance. Other stereo performance research focuses on modeling the effect of sensor errors on stereo [14] and statistical techniques to estimate stereo disparity [22, 21].

In order to verify stereo performance, one needs ground truth. The oldest source of ground truth comes from artificial imagery that has additive noise from a known distribution [23, 11]. While easy to generate and accurate, this imagery does not accurately model the real world situations that most stereo systems face. A second approach manually estimates disparities in laboratory images [25, 10, 27]. This ground truth measures stereo performance better, but does so at a greater labor cost and a corresponding small loss in accuracy. A third approach measures a 3D scene from the actual location in which the stereo system will operate. Most avoid this option due to the extremely high labor cost, but laser scans have made this task easier [24]. Recently, Szeliski [31] has proposed a new approach using view-prediction as a measurement scheme with ground truth. This approach caters specifically to the view synthesis application and compares a synthesized image with the ac-

tual image in that location. Although some have used SVS as ground truth [3], SVS has not been tested as a confidence metric. In this paper, we use a manually-obtained ground truth to find the actual stereo errors so that we can verify the confidence metrics' estimates.

In light of previous work, our main contributions in this report are (i) to show how SVS can be used as an on-line confidence metric and (ii) to quantitatively evaluate and compare this new confidence metric with five more-traditional confidence metrics on a uniform basis. We use three datasets, each comprised of a stereo pair of images taken in a laboratory setting, to gain understanding of the different performance characteristics of each confidence metric. Using manually-obtained ground truths, we check where the stereo actually fails, and we verify how well each confidence metric predicts the failures and correct matches.

2 Implementation

2.1 Single View Stereo (SVS)

We propose using single view stereo results as a pixel-wise confidence metric for binocular stereo. The algorithm has three steps: (i) run the stereo algorithm on two different images taken from the same view in close temporal succession. Since the epipolar geometry is singular with only one view, we rectify both images as if they were left images in the binocular case and search along a horizontal scan line. (ii) Label the errors in the output from step one, where an error is a match outside a threshold interval around zero disparity. (iii) Each error in the SVS output predicts failure at the same pixel location for a binocular stereo system incorporating the view used in the first step. For example, if the left camera takes two pictures in quick temporal succession, we can run stereo on the two left images to generate confidence data for the same location in the left/right stereo. The reverse example uses the right-based SVS results to predict right/left stereo performance.

Sensor noises, such as CCD noise and framegrabber jitter, are the primary difference between the two SVS images. Since these noise sources also exist between the binocular cameras, failure on SVS imagery should predict failure on binocular imagery at the same location. If we define signal as image structure that arises from the scene and noise as image structure that arises from other sources, then one can view SVS as a measure of signal to noise. In areas with low signal where noise dominates the scene-based image structure, SVS should return a low confidence score for the stereo matcher. Any stereo system can use this confidence metric because SVS uses the same stereo algorithm that the binocular stereo system uses. Moreover, the computational expense is tolerable for many applications; SVS requires one extra matching process.

In practice, many stereo systems are in motion, and the disparity of the two successive monocular images may not be zero. Still, the SVS disparity values should be small due to the limited motion between the two images. To compen-

sate for this motion, one can relax the threshold in labeling SVS errors. For example, the disparities in the single view stereo results might be allowed to vary up to 3 pixels in either direction, beyond which the system labels the matches as erroneous. Of course, this disparity relaxation depends on the system parameters. A fast frame rate and distant scenery will reduce the disparity between successive cameras. In this paper, we test the SVS confidence metric on still cameras, and the disparity should be exactly zero at all SVS image pixels.

In our study, we employ a basic scanline-search stereo algorithm. Since we only aim to measure *internal confidence*, and not compare our stereo approach with others, the absolute performance of our algorithm is less relevant. Following rectification, the images are filtered with a Laplacian operator [6], accentuating high frequency edges. Next, the matcher calculates 7x7 windowed modified normalized cross-correlation values (MNCC) for integral shifts, where

$$\text{MNCC}(X, Y) \triangleq \frac{2 \text{Cov}(X, Y)}{\text{Var}(X) + \text{Var}(Y)}.$$

The disparity is the shift of each peak MNCC value. Subpixel peak localization is made via interpolation with a quadratic polynomial. For the tests, both the binocular stereo and SVS algorithm search along scanlines in a disparity range of ± 20 pixels.

2.2 Left/Right Consistency (LRC)

Our implementation of left/right checking relies on the consistency between left-based and right-based matching processes. The matching processes should produce a unique match, and left/right checking attempts to enforce this. Formally, if x_R matches $x'_L = x_R + d_{x_R}^R$, where x_R is a right coordinate and x'_L is an estimated left match at right-based horizontal disparity $d_{x_R}^R$, then LRC is defined as

$$\text{Error} = x_R - (x'_L + d_{x'_L}^L),$$

with $d_{x'_L}^L$ the left-based horizontal disparity.

When the error is high, then the confidence is low. A low score corresponds to the situation when the left and right-based matches disagree significantly.

2.3 Matching Score Metric (MSM)

The matching score metric uses the similarity score, S_{Max} , at the disparity where the matcher has chosen a match. MSM estimates whether the match was successful or not based on this score. Intuitively, a low similarity score means that the matcher itself does not believe its best match was very good. The range of MSM is limited to the range of the similarity metric, which in the case of normalized cross-correlation is $[-1, 1]$.

2.4 Curvature Metric (CUR)

The curvature metric measures the curvature of the similarity function around the maximal similarity score S_{Max} and uses the formula:

$$\text{Curvature} = 2S_{Max} - S_{Left} - S_{Right}$$

where S_{Left} and S_{Right} are the similarity scores to the left and right of S_{Max} . When curvature is low, CUR predicts a false match. A low curvature score happens in low-texture regions, where nearby values appear to be similar.

2.5 Image Entropy (ENT)

The image entropy metric relies on the intuition that stereo works best in high texture areas. Entropy, much like MDL [16], measures image texture. If we assume the pixel values X are discrete random variables (RVs) with point mass function P , then we can define the entropy H :

$$H(X) \triangleq -E_X[\log(P(X))]$$

An intuitive description of entropy is that it measures the randomness of a RV. A low entropy means that the average probability over the support set for a given RV is low. For example, a constant region in an image has a one point distribution, and thus a lower entropy than a highly textured region. Since low texture causes problems for many stereo systems, a low entropy score can be used to predict stereo error. In our implementation, we use a 20 bin histogram to measure the probability function of a 7x7 support window around each pixel.

2.6 Peak Ratio Metric (PKR)

Ideally, a match should be unique, and PKR attempts to enforce this by rejecting any match with more than one similar candidate. The peak ratio metric computes the ratio of the two most similar matches, $S_{SecondMax}$ and S_{Max} . When the ratio is close to one, PKR predicts a false match. When the ratio is small, PKR indicates a unique match, which is ostensibly an accurate match. Typically, the peak ratio is high in areas of repetitive structure and low texture.

2.7 Comparison Methodology

For the comparisons, we first obtain and rectify a stereo pair, from which we manually construct a ground truth at integer resolution (see Figure 2). We use this ground truth to verify our confidence match predictions. First, we measure where our stereo algorithm has matched badly, counting any match within 1 pixel of the truth as valid. Unmatchable half-occlusion points, dead pixels and boundary points are excluded from the statistics. We elect to use a binary hit/miss scoring system because in the current context we feel that once a match is erroneous, it is matching a wrong partner, and the direction and magnitude of the error is irrelevant. After calculating where the matcher actually fails, we calculate where each confidence metric estimates failure.

We characterize the performance of a confidence metric using two numbers: the error and match prediction probabilities - the conditional probability that the binocular stereo actually produces an error (match) given the confidence metric predicts an error (match). We choose these probabilities rather than overall number of prediction hits because they measure false positives and false negatives and because they normalize the results for easy comparison. The probability is a realistic measure because it quantifies the impact if a system always trusts the error estimates of a particular confidence metric.

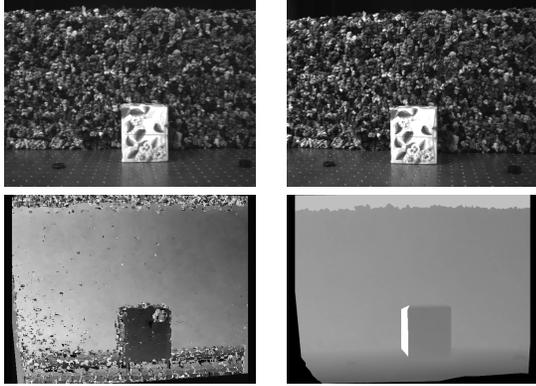


Figure 2: Stereo Error Measurement. Using ground truth, we measure where the stereo system fails. The top row contains the first test stereo pair. The stereo results are on bottom left, with a manually-obtained ground-truth to the right. The ground truth labels and excludes unmatchable points, including half-occlusion and boundary points.

3 Results

This section shows quantitative and qualitative comparisons of SVS performance relative to other confidence metrics. The dataset for these comparisons consists of three stereo pairs showing a box, a doll, and a pair of shoes. Each dataset demonstrates different aspects of the confidence metrics’ performance. The box demonstrates a planar scene with high texture. The doll test set demonstrates how the confidence metrics react to geometric difficulties, such as occlusion boundaries and a search range that is small with respect to the attendant disparity. The shoes dataset demonstrates how the metrics react to repetitive structure and large areas of low texture.

For each case, we run the stereo on two left images to get SVS data. Then, we run both the left-based and right-based matching processes on the binocular stereo pair. We calculate ENT on the rectified left image before the stereo matching, and the MSM, PKR and CUR metrics are generated during the left-based stereo run. LRC is executed after the two stereo processes have finished.

For a given test, we calculate false positives and false negatives. In this case, a false positive is where the confidence metric predicts a stereo failure, but does so incorrectly. A false negative is where the confidence metric predicts a stereo hit, but does so incorrectly. False positives decrease the error prediction accuracy, while false negatives decrease the match prediction accuracy. The purpose of SVS is to predict stereo failure, not success. However, we include the false negatives for completeness.

The quantitative comparison rank orders each confidence metrics’ estimates of binocular failure and plots how well the metrics predict stereo performance as the x axis progresses from best to worst estimates. We compare the metrics at a given *participation* rate, by which we mean the percent of the possible matches that a confidence metric es-

timates that the binocular stereo will fail. A higher participation rate includes lower confidence points. Since SVS and LRC do not predict as many errors as others, they will not produce results at higher participation rates. For example in case 1, SVS only predicts that up to 7% of the stereo matches will be failures. Although this termination seems premature, this is not an error; the metric can not predict any more errors without jumping to the useless case where every point is an error prediction. Similarly, ENT produces many estimates, even at the lowest threshold of zero.

3.1 Case 1: Box

The first test set shows a box on top of an optical bench in front of a textured background (see the Box set in Figure 3). Although this test set does exhibit features which cause stereo error, such as a low-texture patch at top and occluding boundaries near the box, the scene has high texture and planar scenery. Both factors aid the stereo performance and allow the confidence metrics to predict fewer errors.

Qualitatively, one can see the differences between the confidence metrics at the top of Figure 3. The figure displays in white the pixels at which each metric predicts the stereo system will err. For an even comparison, we threshold each metric so that 7% of the possible matches are error predictions. The SVS metric correctly predicts the stereo errors in the textureless patch at top, and also correctly labels points in the lower portion of the image as errors. SVS is not able to label repetitive structure in this test because the holes on the table are distinctive enough that SVS matches correctly at those regions. LRC labels many points in the top patch as true matches, finding that both the left and right-based matching process agree on the false match. LRC counts errors made in either of the two views; so, some right-based errors are introduced into the left-based matching error results. CUR and PKR mistakenly predict errors at correct matches that happen to have smooth correlation curves. MSM and ENT performs well at this participation rate, but mistakenly label a few more errors within the lower portion of the image than SVS.

The quantitative results for the Box test are shown at the left of Figure 4. We present two graphs: error prediction at top and correct match prediction at bottom. Ideally, one would like to see two traits in these graphs: a high-probability point at high participation and also a flat curve followed by a sharp descent. The high probability point indicates that the metric can accurately predict binocular stereo error. The flat curve indicates that the metric is relatively insensitive to thresholding.

For error prediction, SVS exhibits both positive traits. It has a higher error prediction probability than the other metrics. Although SVS has lower participation, it consistently estimates errors with above 99% accuracy, even up to a 7% participation. SVS only reaches 7% participation because the SVS threshold ranges from ± 2 to ± 20 pixels away from zero. The binocular stereo actually produces errors for 32% of possible matches, and so the confidence metrics would need at least a 32% participation rate to predict ev-

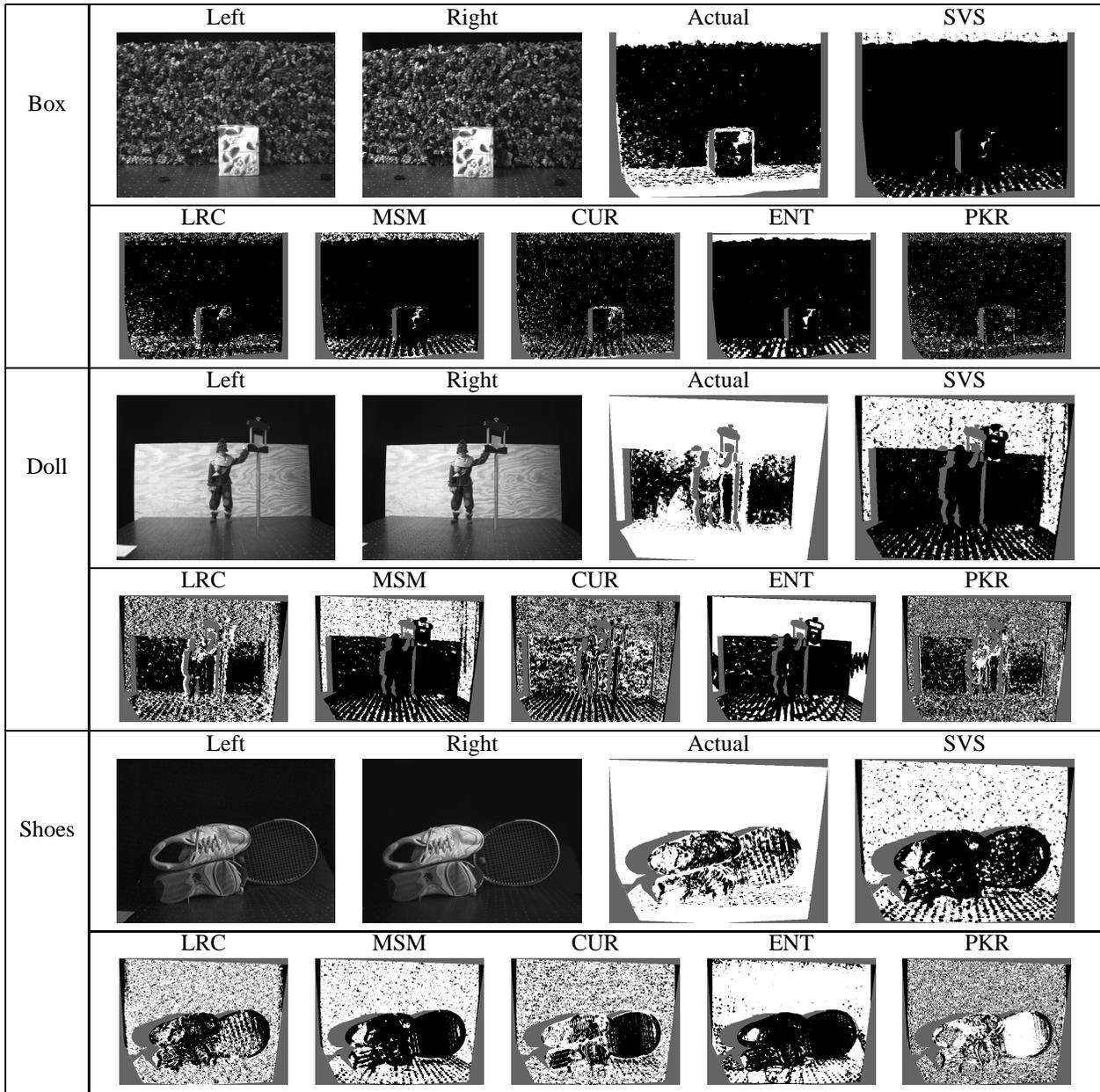


Figure 3: The Misses Predicted by the Confidence Metrics. The images display the pixels that each confidence metric would predict as bad matches in white, and the predictions of correct matches in black. The results for the Box test are at top, the Doll test in middle and the Shoes set at bottom. Each display shows the stereo pair, the actual stereo misses, and the misses predicted by each confidence metric.

ery error. SVS does not predict all errors, but the errors it predicts, it predicts well. The closest competitor to SVS, MSM, performs at a 5% lower probability. This difference may seem negligible, but with repeated application, such as that used in hierarchical stereo, even small prediction error compounds. ENT performs well, correctly labeling the low texture regions. LRC performs well, but spuriously predicts more errors than SVS in the middle of the image. CUR and

PKR are relatively inefficient at participation rates above 5%. Also, all confidence algorithms aside from SVS depend heavily upon their threshold. SVS' flat probability of error prediction makes its implementation much easier.

SVS does not fare as well when predicting correct binocular matches as it did when predicting errors. Since the error prediction rates of SVS are low, SVS inherently predicts many stereo matches. Interestingly, SVS performs well at

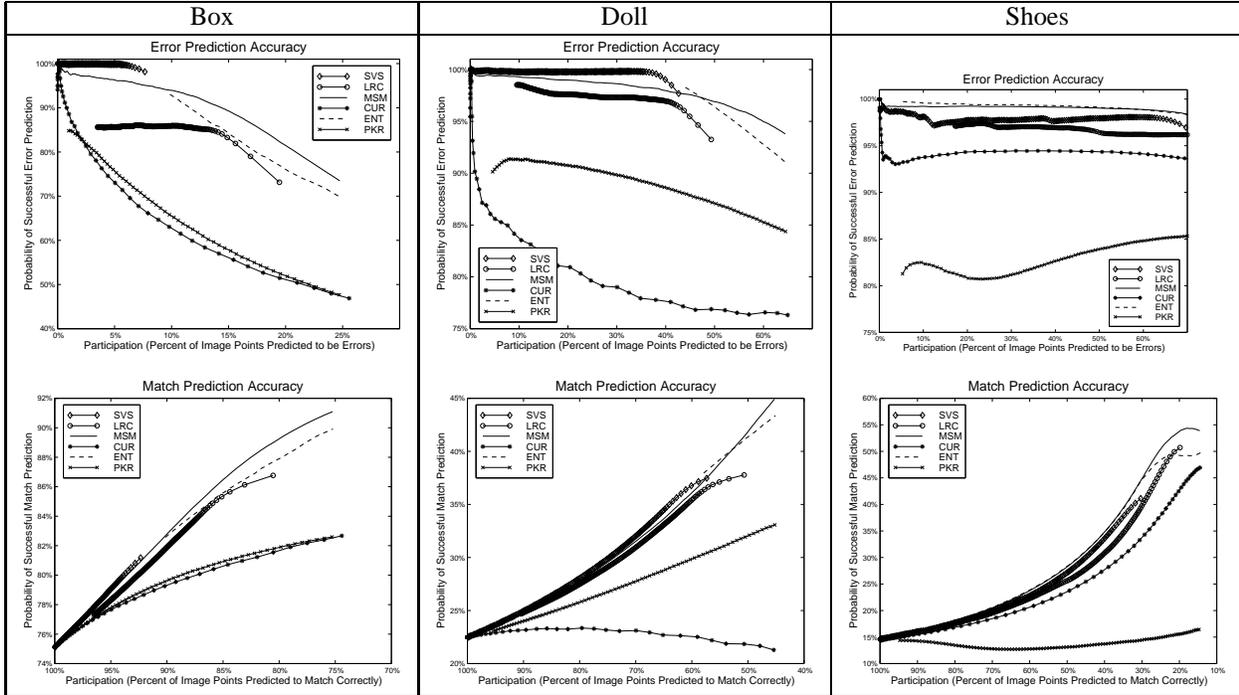


Figure 4: Quantitative Results. The top row contains charts that show how well each confidence metric predicts stereo error, while the bottom row contains charts showing how each metric predicts stereo success. As the metrics predict more binocular error, their error probability decreases. As they do this, they predict fewer stereo hits, and the hit probability increases. We compare the metrics at all possible LRC threshold levels. Note that SVS and LRC do not predict as many errors as the other metrics. The next available threshold would force the two metrics to indiscriminately predict all pixels as erroneous.

this high level of match prediction, higher than the other metrics. However, the performance gain is marginal as all of the metrics converge to the same point.

3.2 Case 2: Doll

The second test set shows a doll next to a pole (see the Doll set in Figure 3). This test differs from the previous one in that the subject is non-planar, the background has less texture, and the pole is thin with many occluding borders. More importantly, the head and chest of the doll figure lie outside the binocular stereo search range of ± 20 pixels. The case demonstrates how the metrics behave around geometric error sources. The error sources lead the binocular stereo to achieve only a 22% match rate.

The qualitative data lies in the middle of Figure 3. For an even comparison, we threshold each metric to show the case where 36% of the possible matches are predicted errors. As before, SVS and ENT best predict the stereo errors in the textureless black patch, and also correctly labels points in the lower portion of the image as errors. Unlike LRC and PKR, the other metrics are not able to label points on the doll's chest and head where the match was out of the stereo search range. The same is true for the stereo errors on the floor that are out of range and the stereo errors near the occluding boundary of the pole. MSM and ENT perform well at this participation rate, but mistakenly labels a few more errors within the backboard than SVS.

The quantitative results for the Doll test are shown in the middle of Figure 4. SVS has a higher error prediction probability than the other metrics, bolstered by its performance around the black background. The closest competitors to SVS, MSM and ENT, perform at a slightly lower probability. LRC performs well, but spuriously predicts more errors than SVS in the middle of the image. Again, CUR and PKR are inefficient. As before, SVS has the flattest curve, which indicates that SVS does not need its threshold fine-tuned. At the bottom of the figure, SVS is the top performer in predicting correct stereo matches, albeit at a low rate of error prediction. SVS outperforms the other metrics only in the area of the black background, and even though it predicts too many matches (too few errors) in the area of the doll, its overall rate is highest. MSM and ENT have a high match prediction rate for similar reasons. LRC is able to correctly predict matches in the area of the doll, but over-predicts matches in the black background, putting LRC third to MSM and SVS.

3.3 Case 3: Shoes

The last test set shows a pair of sneakers in front of a racquet (see the Shoes set in Figure 3). This test differs from the previous cases in that the objects exhibit significant repetitive structure and the image has a large area without texture. Such test imagery may realistically simulate the low texture patches found in outdoor urban terrain or scenes

with large patches of sky. These error sources cause the binocular stereo to achieve only a 15% match rate.

The qualitative data is at the bottom of Figure 3. For an even comparison, we threshold each metric so that the images display the case where 62% of the possible matches are predicted errors. The SVS and ENT metrics perform best in the textureless black patch. Unlike MSM and ENT, SVS predicts stereo error in the repetitive pattern of the racquet strings. SVS performs well in this area, but not as well as it normally does in the low-texture regions. Because the repetitive structure is rarely an exact repeat, while the same structure at the same point is, the zero disparity solution has a strong tendency to beat out the shift-by-one-cycle solution for SVS. PKR is able to label the repetitive structure of the tennis racquet. LRC detects errors both in the background and in the repetitive structure of the racquet. CUR has high confidence in the racquet face, but overpredicts error in the front of the top shoe.

The quantitative results for the Shoes test are shown at the right of Figure 4. SVS error predictions in the repetitive structure of the racquet are correct overall, but at a lower probability than error predictions in the textureless areas. For this reason, SVS drops to 98% probability and scores third to MSM and ENT in predicting stereo error. LRC predicts errors well in both the area of the racquet face and the background, but produces false error predictions which puts its error prediction probability near 97%, or fourth. All algorithms have a flat curve, which signals their insensitivity to thresholding. When predicting matches, the methods perform similarly. SVS does not achieve the same match prediction rate of MSM and ENT because the repetitive error structure of the racquet face causes SVS to err slightly.

3.4 Grouped Results

To quantify exactly which types of error sources the metrics are estimating, we manually segmented the actual stereo misses into four groups: out-of-range errors, low-texture errors, repetitive structure errors, and other errors (see Figure 5). We measure the successful error estimates for each metric in each of these categories. The plots compare the metrics when working at participation rates of 7, 36, and 62% for the Box, Doll, and Shoes cases, respectively.

The plots show that SVS predicts the most low-texture errors correctly in all but the last case, with MSM and ENT behind SVS in the first two cases. PKR is effective at detecting repetitive structure, especially in the shoes case where the racquet face presents difficulty for the matcher. LRC has the most even performance across error categories, picking up all four kinds of error effectively.

4 Discussion

4.1 Algorithm Results

Each algorithm has different performance characteristics. SVS performs best in regions of low texture. LRC detects most types of stereo error (see Table 1), but does not predict error when the two matching processes mistakenly

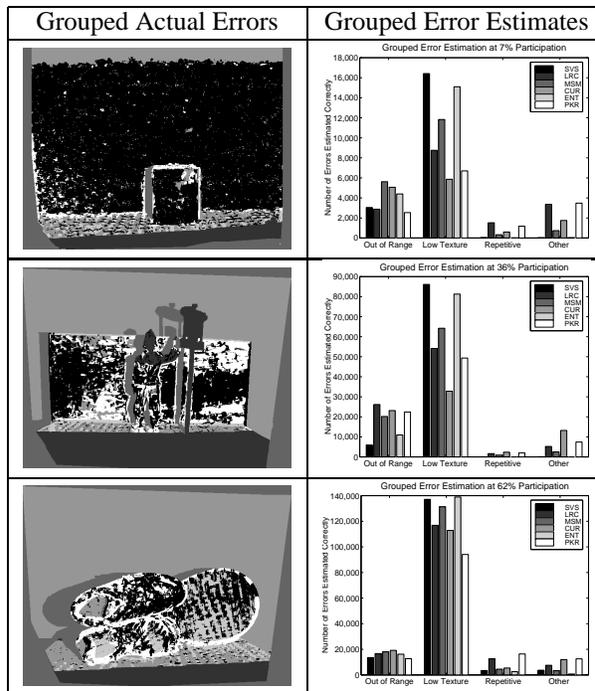


Figure 5: Grouped Error Estimates. For each test case, the left image depicts the manually grouped stereo errors, while the right plot shows the correctly labeled subset of errors when each metric estimates 7, 36, and 62% (respectively) of possible matches to be errors. In the manually grouped errors image, the grayscale corresponds to the type of error, where 50 means 'Out of Range', 150 means 'Low Texture', 200 means 'Repetitive Structure', and 255 means 'Other'.

agree. As evidenced by the black regions for LRC in Figure 3, this happens often. When both matching processes are accurate, LRC predicts fewer false errors. Like SVS, MSM and ENT react to a subset of the possible error sources, but do so with a high probability of success. All three are particularly sensitive to low-texture areas, but MSM is also able to detect errors from more sources than SVS and ENT. Future work includes testing non-MNCC MSM metrics, such as SSD, which may not measure low texture regions as well. PKR is particularly effective in repetitive structure areas, but produces many false positives for correct matches in areas of repetitive structure (e.g., shoes dataset) and fails in low texture regions where erroneous matches are unimodal. CUR has a significant false positive rate, labeling as errors the correct matches that have low curvature.

The confidence metrics have other advantages and disadvantages. Most notably, the non-SVS methods predict *more* error. Labeling more errors with less accuracy may sometimes be desirable. However, when every match is needed, the error prediction probability is more important than the number of error predictions because it would discount fewer matches and do so more accurately. Another difference lies in the metrics' implementation. SVS and LRC do not de-

pend critically on thresholding, which makes implementation easier. The other confidence metrics are sensitive to their threshold and need to be more finely tuned for optimal accuracy. The running time for the metrics' is also different. The fastest metric, MSM, requires no additional computation since it relies on the stereo matching process itself. CUR and PKR require an incremental calculation during the stereo maximization phase. SVS and LRC require an additional matching process, while ENT requires a costly histogramming operation.

4.2 Summary

In this paper, we have presented a new confidence metric, SVS, which delivers high probability of error and match prediction, especially around low-texture regions. We have compared the metric against five other metrics, LRC, MSM, CUR, ENT and PKR, under a variety of laboratory settings.

The choice of confidence metric ultimately depends upon the scenery at hand and the operational requirements of the system. When low-texture areas dominate the scenery, SVS has a high probability of success; it also can diagnose difficulties arising from sensor noise and periodic structure. When the matching is generally accurate, LRC will be a good confidence metric because it reacts to most error source types and both of its matching processes should succeed. MSM suits systems that require extreme speed because the system already has to calculate a match score. In the final analysis, it may be desirable to combine these metrics to take advantage of their different strengths.

References

- [1] P. Anandan. Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion. In *SPIE Intelligent Robots and Computer Vision*, volume 521, pages 184–194, 1984.
- [2] P. Anandan. A Computational Framework and an Algorithm for the Measurement of Visual Motion. *IJCV*, 2:283–310, 1989.
- [3] D.N. Bhat and S.K. Nayar. Ordinal Measures for Image Correspondence. *PAMI*, 20(4):415–423, 1998.
- [4] S.D. Blostein and T.S. Huang. Error Analysis in Stereo Determination of 3-D Point Positions. *PAMI*, 9(6):752–765, 1987.
- [5] R. Bolles, H. Baker, and M. Hannah. The JISCT Stereo Evaluation. In *DIUW*, pages 263–274, 1993.
- [6] P. Burt and E. Adelson. The Laplacian Pyramid as a Compact Image Code. *Trans. Computers*, 31:532–540, 1983.
- [7] C. Chang, S. Chatterjee, and P.R. Kube. On an Analysis of Static Occlusion in Stereo Vision. In *CVPR*, pages 722–723, 1991.
- [8] S. Das and N. Ahuja. Performance Analysis of Stereo, Vergence and Focus as Depth Cues for Active Vision. *PAMI*, 17(2):1213–1219, 1995.
- [9] G. Egnal and R.P. Wildes. Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches. *PAMI*, 2002. To Appear.
- [10] O. Faugeras, P. Fua, B. Holz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and Qualitative Comparison of Some Area and Feature-Based Stereo Algorithms. In Forstner and Ruwiedel, editors, *Robust Computer Vision: Quality of Computer Vision Algorithms*, pages 1–26. Wichmann, 1992.
- [11] W.E.L. Grimson. A Computer Implementation of a Theory of Human Stereo Vision. In *Phil. Trans. Royal Soc. London*, volume B292, pages 217–253, 1981.
- [12] M.J. Hannah. *Computer Matching of Areas in Stereo Images*. PhD thesis, Stanford University, May 1974.
- [13] M.J. Hannah. Detection of Errors in Match Disparities. In *DIUW*, pages 283–285, 1982.
- [14] G. Kamberova and R. Bajcsy. Sensor Errors and the Uncertainties in Stereo Reconstruction. In Kevin W. Bowyer and P. Jonathan Phillips, editors, *Empirical Evaluation Techniques in Computer Vision*, pages 96–116. IEEE Computer Society Press, 1998.
- [15] K. Konolige. Small Vision Systems: Hardware and Implementation. In *Intl. Symp. Robotics Research*, 1997.
- [16] Y.G. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *IJCV*, 3(1):73–102, 1989.
- [17] Y.G. Leclerc, Q.-Tuan Luong, and P. Fua. Self-Consistency: A Novel Approach to Characterizing the Accuracy and Reliability of Point Correspondence Algorithms. In *DIUW*, pages 793–807, 1998.
- [18] J. Little and W. Gillett. Direct Evidence for Occlusion in Stereo and Motion. In *ECCV*, pages 336–340, 1990.
- [19] X. Liu, T. Kanungo, and R.M. Haralick. Statistical Validation of Computer Vision Software. In *DIUW*, pages 1533–1540, 1996.
- [20] A. Luo and H. Burkhardt. An Intensity-Based Cooperative Bidirectional Stereo Matching with Simultaneous Detection of Discontinuities and Occlusions. *IJCV*, 15:171–188, 1995.
- [21] R. Mandelbaum, G. Kamberova, and M. Mintz. Stereo Depth Estimation: A Confidence Interval Approach. In *ICCV*, 1998.
- [22] L. Matthies. Stereo Vision for Planetary Rovers: Stochastic Modeling to Near Real-Time Implementation. *IJCV*, 8:71–91, 1992.
- [23] J.E.W. Mayhew and J.P. Frisby. Psychophysical and Computational Studies Towards a Theory of Human Stereopsis. *Artificial Intell.*, 17:349–385, 1981.
- [24] J. Mulligan, V. Isler, and K. Daniilidis. Performance Evaluation of Stereo for Tele-presence. In *ICCV*, 2001.
- [25] Y. Ohta and T. Kanade. Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. *PAMI*, 7(2):139–154, 1985.
- [26] D. Samaras, D. Metaxas, P. Fua, and Y.G. Leclerc. Variable Albedo Surface Reconstruction from Stereo and Shape from Shading. In *CVPR*, pages 480–487, 2000.
- [27] Kiyohide Satoh and Yuichi Ohta. Occlusion Detectable Stereo - Systematic Comparison of Detection Algorithms. In *ICPR*, 1996.
- [28] D. Scharstein. *Stereo Vision for View Synthesis*. PhD thesis, Cornell University, Jan. 1997.
- [29] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *IJCV*, 47(1):7–42, 2002.
- [30] T. Smitley and R. Bajcsy. Stereo Processing of Aerial Urban Images. In *ICPR*, pages 405–409, 1984.
- [31] R. Szeliski. Prediction Error as a Quality Metric for Motion and Stereo. In *ICCV*, pages 781–788, 1999.
- [32] R. Trapp, S. Drüe, and G. Hartmann. Stereo Matching with Implicit Detection of Occlusions. In *ECCV*, pages 17–33, 1998.
- [33] J. Weng, N. Ahuja, and T.S. Huang. Two-View Matching. In *Proc. Int'l Conf. Computer Vision*, pages 64–73, 1988.
- [34] Y. Xiong and L. Matthies. Error Analysis of a Real-Time Stereo System. In *CVPR*, 1997.