

# The Self-Aware Matching Measure for Stereo

Philippos Mordohai  
Stevens Institute of Technology  
Hoboken, New Jersey, USA  
mordohai@cs.stevens.edu

## Abstract

We revisit stereo matching functions, a topic that is considered well understood, from a different angle. Our goal is to discover a transformation that operates on the cost or similarity measures between pixels in binocular stereo. This transformation should produce a new matching curve that results in higher matching accuracy. The desired transformation must have no additional parameters over those of the original matching function and must result in a new matching function that can be used by existing local, global and semi-local stereo algorithms without having to modify the algorithms. We propose a transformation that meets these requirements, taking advantage of information derived from matching the input images against themselves. We analyze the behavior of this transformation, which we call Self-Aware Matching Measure (SAMM), on a diverse set of experiments on data with ground truth. Our results show that the SAMM improves the performance of dense and semi-dense stereo. Moreover, as opposed to the current state of the art, it does not require distinctiveness to match pixels reliably.

## 1. Introduction

All stereo algorithms begin by computing a simple similarity or dissimilarity metric between pixels in the reference image and pixels in the target image. Stereo processing proceeds by aggregating and optimizing these pixel-wise metrics to obtain the best disparity map according to some criteria, which typically include accuracy. A comparative evaluation of numerous binocular stereo algorithms can be found at the Middlebury Stereo Vision Page (MSVP) ([vision.middlebury.edu/stereo](http://vision.middlebury.edu/stereo)) and a categorization of the algorithms in [20]. The common theme among these methods is that they take the cost function for granted. In fact, the last breakthrough in this area was arguably the sampling-insensitive dissimilarity measure of Birchfield and Tomasi in 1998 [2]. (The adaptive support weight method of Yoon and Kweon [23] is, from our perspective, not a strictly local

cost function.) In this paper, we revisit the matching functions for stereo aiming at extracting more information from the image pair, before applying a stereo algorithm. Note that the terms cost and similarity are used interchangeably in the paper, since one can trivially be converted to the other via a monotonically decreasing function. We use the term matching measure to refer to both cost and similarity.

We focus our investigation to the core of the stereo correspondence problem: the estimation of dense or semi-dense disparity maps from rectified stereo pairs. (By dense or semi-dense we refer to an attempt to estimate correspondences for all pixels without a priori feature extraction. The result may be semi-dense, if correspondences are rejected according to some criterion.) We seek a transformation that operates on the matching measures between pixels and results in a new matching function, which has the same form as the original and is better than the original according to widely accepted criteria. The requirements for the transformation are that it must have *no additional parameters* over those of the original matching function and that the new matching function must behave exactly as the original in the sense that *existing stereo algorithms must be able to use it without any modifications*. We do not attempt to estimate sub-pixel matches, do not model occlusion or depth discontinuities and do not apply constraints, such as ordering or uniqueness, other than the epipolar constraint. We also do not attempt to rank the various matching functions.

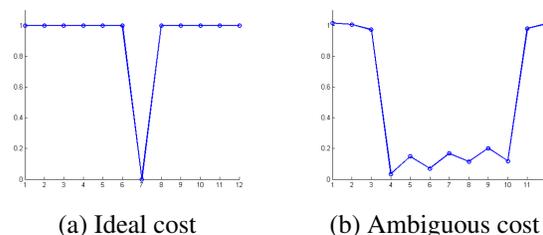


Figure 1. The ideal cost curve for stereo has one distinct minimum. A less ideal cost curve has several local minima and/or flat regions of low cost.

See recent surveys in [12] and [10] for evaluations of cost functions and aggregation schemes respectively.

The “ideal” cost curve as a function of disparity for a pixel is shown in Fig. 1(a). It has a single, distinct minimum. The curve in Fig. 1(b) is less attractive. However, if we knew, that, when this curve is observed, the true match is, for instance, two thirds of the distance between the two ends of the valley, then we could unambiguously find it.

Our approach was inspired by the observation that *the cross- and self-matching curves should behave similarly around the correct match and zero disparity, respectively*. By cross-matching curve we refer to the graph of matching cost between pixels of the reference and target images as a function of disparity. The self-matching curve is a similar graph, but now the matching takes place between pixels of the same image. A perfect match can always be found at zero disparity when a pixel is compared with itself.

We claim that a stereo algorithm can only benefit when the structure of the self-matching function is taken into account. We accomplish this by measuring the *correlation coefficient* between the two curves, which should be high when the curves are aligned at their true minima and low otherwise. Because of this reliance on the self-matching function, we use the term *Self-Aware Matching Measure (SAMM)* for our new measure. Previous research that uses information from self-matching to reason about the reliability of cross-matching was published by Manduchi and Tomasi [15] and Yoon and Kweon [25]. Both these approaches base their analysis on pixel *distinctiveness*, which is not a requirement for us. As shown in Fig. 3, unambiguous results can be obtained by the SAMM for pixels with low distinctiveness.

## 2. Motivation and Related Work

One of the motivations for our work is the observation that conventional matching functions (SAD, SSD, NCC) do not assign the lowest cost or highest similarity to the most unambiguous matches. For example, see Fig. 2(c) and (d) which show the minimum cost value, computed by the Sum of Absolute Differences (SAD) and the negative of the Normalized Cross Correlation (NCC)<sup>1</sup>, for each pixel of the Teddy stereo pair from MSVP. Careful observation of such cost maps reveals that the cost value itself is not very useful as a criterion for selecting reliable pixels. It is still of some value, as evidenced by the success of methods that detect ground control points [3] or seed matches, but we claim that improvements are possible. Methods that would benefit include progressive stereo [7, 15, 28, 14, 6], cooperative stereo [29, 27] and certain multiple-view methods [9, 17, 4, 5] that require confidence estimates for the matches. The consistency between the left-to-right and the

right-to-left match is often used for the detection of reliable matches [3, 14]. This does not meet our requirements as it results in a binary decision and not a ranking.

Our new function should assign a value to each potential pixel match that can be used to rank matches from most to least reliable. Matthies [16] proposed a posterior probability estimation technique, which can achieve this goal. The need, however, to select a value for  $\sigma$  according to the expected noise variance, violates our requirement for no additional parameters. More importantly, this choice is not trivial because the expected noise variance is not the same for all pixels, but depends on local texture as explained in Fig. 2. The SAMM effectively models the individual characteristics of each pixel and overcomes this difficulty.

Among the first methods for computing the reliability of pixels based on monocular information is the one of Salari and Strong [19]. The “matchability” of a pixel is computed as a function of intensity correlation and local image variance. Later, Manduchi and Tomasi [15] distinguish interest from distinctiveness, showing that one does not imply the other. They define distinctiveness as the perceptual distance between the pixel and the most similar other pixel in the search range. Egnal et al. [8] investigate several confidence measures, including single-view stereo, in which different images of an identical scene from the same viewpoint are

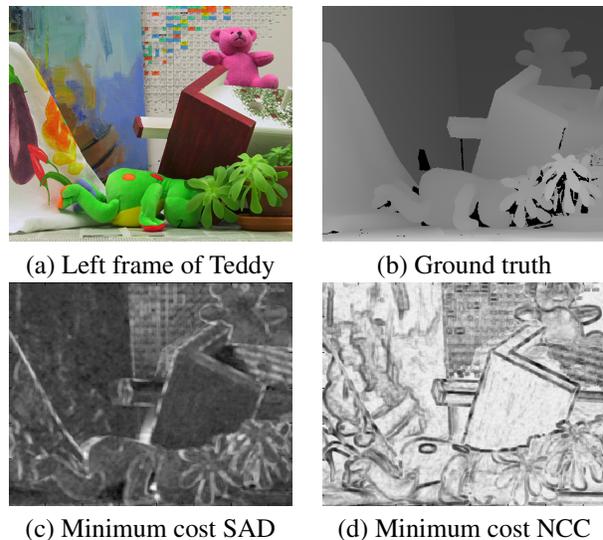


Figure 2. The left image of the Teddy stereo pair from MSVP, the ground truth disparity map and the minimum cost maps for SAD and negative NCC in  $5 \times 5$  windows. High intensity corresponds to large cost. Pixels in uniform areas have lower cost values under SAD since the cost can approach 0, while larger values are observed at textured pixels, such as those on the periodic table, due to quantization artifacts during image formation or calibration errors. Under NCC, pixels at uniform areas have large costs because, after subtracting the mean, correlation operates on noise.

<sup>1</sup>NCC always refers to zero-mean NCC in this paper

matched to estimate the expected noise level at each pixel, and the peak ratio metric that is equal to the ratio of the similarity values of the two best matching candidates for each pixel. These methods meet one of our requirements by enabling the ranking of pixels. The methods of [15] and [8] also require no additional parameters. They fall short of our overall requirements because they only characterize the reliability of the best match for each pixel according to the matching function. In other words, unlike [16], they cannot produce a new matching function for all disparities.

An approach for estimating whether a pixel belongs to a small disparity band (the foreground) was proposed by Agarwal and Blake [1]. They estimate the likelihood ratio of a pixel being in the foreground or the background. The total foreground likelihood is computed by cross-matching over a small foreground disparity band and then marginalizing out disparity. To reduce computational cost, the total stereo likelihood is approximated by using the left image as a proxy for the right image and computing the self-matching likelihood over a small disparity range that includes zero. The background likelihood is approximated by subtracting the foreground likelihood from the approximate total likelihood. This scheme approximates the foreground likelihood by the fraction of total likelihood it has explained and can assign high confidence values to pixels that do not necessarily have unique, sharp peaks in their similarity functions, such as the one in Fig. 1(b), but does not compute disparity.

The key observation from [1] that inspired our work is that the *shape* of the matching function around the correct match should resemble the shape of the self-matching around the ideal match at zero disparity. While the Sum of Squared or Absolute Differences (SSD or SAD) is zero when matching a pixel with itself, the shape of the self-matching function as it moves away from the true match should be similar to that of the cross-matching function as it moves away from the correct match between the left and right image. A sharp valley in self-matching cost should correspond to a sharp valley in the cross-matching function. Similarly, a flat region should correspond to a flat region.

Recently, Yoon and Kweon [24, 25] presented an approach that meets all our requirements. They consider both the probability of mismatches and the probability of good matches for a pixel to compute the *distinctive similarity measure (DSM)* for each potential pixel correspondence. Their definition of distinctiveness, as in [15], is the minimum dissimilarity between the pixel under consideration  $p$  and all other pixels of the same image that compete with  $p$  for matches in the other image. The DSM is large for a correspondence between two pixels that are similar to each other and distinct from all other pixels in their neighborhoods. The DSM can be used to transform the stereo cost volume according to our requirements and treats both images symmetrically. Results in [25] demonstrate improve-

ments in semi-dense and dense stereo due to the DSM. In Section 4, we show similar improvements using the SAMM. It is important to point out here that the main difference between the DSM and SAMM is that the former considers the appearance of a pixel and the latter considers its cost function.

### 3. The Self-Aware Matching Measure

In this section, we define our proposed transformation and demonstrate its benefits.

#### 3.1. Notation and Preliminaries

We refer to the two images of the rectified stereo pair as left and right ( $I_L$  and  $I_R$ ), without loss of generality. We distinguish between the *reference image* for the pixels of which the disparity map is computed, and the *target image*. Not all pixels of the target image are guaranteed to have matches after a dense disparity map has been computed for the reference image. The disparity  $d(x, y)$  of a potential match on an epipolar line  $y$  is always defined as the difference between the  $x$ -coordinates of the pixels in the reference and target images:  $x_{tgt} = x_{ref} - d(x_{ref}, y)$ . If the disparity range for the left image is  $\{d_{min}, d_{max}\}$ , then the disparity range for the right image is  $\{-d_{max}, -d_{min}\}$ .

The cost function for a pixel in the left image, when compared to pixels in the right image, is denoted by  $c_{LR}(x_L, y, d)$ . The cost volume is denoted by  $C_{LR}$  and contains cost values for all permissible matches within the disparity and image boundaries. Cost values for matches to pixels outside the image boundaries are assigned a very large constant and ignored. The cost function for matching a pixel in the left image with pixels also on the left image is denoted by  $c_{LL}(x_L, y, d)$ . When the right image is used as reference, the cross- and self-matching functions are denoted by  $c_{RL}(x_R, y, d)$  and  $c_{RR}(x_R, y, d)$ , respectively.

We have used the following matching functions in our experiments. In all cases, aggregation is performed within square windows  $W$  with sides of odd length in pixels. The width of the window,  $N$ , is the only adjustable parameter in most of our experiments.

- The Sum of Absolute Differences (SAD):  
 $SAD(x, y, d) = \sum_{i \in W} |I_L(x_i, y_i) - I_R(x_i - d, y_i)|$ .
- The Sum of Squared Differences (SSD):  
 $SSD(x, y, d) = \sum_{i \in W} (I_L(x_i, y_i) - I_R(x_i - d, y_i))^2$ .
- Normalized Cross Correlation (NCC), which is also referred to as Zero-Mean NCC by some authors:  
 $NCC(x, y, d) =$

$$\frac{\sum_{i \in W} (I_L(x_i, y_i) - \mu_L)(I_R(x_i - d, y_i) - \mu_R)}{\sigma_L \sigma_R},$$

where  $\mu_L$  and  $\sigma_L$  are the mean and standard deviation of all pixels in the square window in the left image. ( $\mu_R$  and  $\sigma_R$  are defined in the same way for the right image.)

- Modified Normalized Cross Correlation, which was proposed by Moravec [18]:  $MNCC(x, y, d) =$

$$\frac{\sum_{i \in W} 2(I_L(x_i, y_i) - \mu_L)(I_R(x_i - d, y_i) - \mu_R)}{\sigma_L + \sigma_R}.$$

Note that to unify computations in the remainder, we convert NCC and MNCC to cost functions by using  $1 - NCC$  and  $1 - MNCC$  respectively.

The above definitions are for monochrome images. We, however, carry out all experiments on color images. To process color images we sum over the color channels for SAD and SSD. For NCC and MNCC we treat each pixel as a 3D vector. We also compute a 3D mean vector over the matching window. We compute, however, a single variance for a total of  $3 \times N^2$  zero-mean measurements, with the appropriate mean used for each color channel. We do not compute separate variances for each color channel because it is susceptible to noise from uninteresting color channels.

Due to space constraints and for clarity, we only show results for SAD and NCC. Results under SSD and MNCC are very similar to SAD and NCC respectively. In the future, we would like to verify these results using the sampling insensitive measure of Birchfield and Tomasi [2], rank and census transforms [26] and adaptive support weight [23] as matching functions. Results in [25], suggest that no surprises should be expected for the latter function.

### 3.2. Self-Aware Matching Measure Computation

We begin by examining the computation of the SAMM for a potential match estimated by one of the above methods. The motivation behind our approach was the observation that the cross- and self-matching curves should behave similarly around the correct disparity and zero disparity respectively. Our initial thoughts were to match the curves, possibly as in [22] or [21]<sup>2</sup>, or by modeling the transformation from one curve to the other in a non-parametric way.

The purpose of computing a transformation between the two curves would be to account for distortions, which include an offset due to the fact that the self-matching curve has a minimum at zero and does not suffer from quantization effects at integer disparity steps. Other forms of distortion are present, but they are hard to model without making further assumptions. To keep our method as general as possible, we measure similarity between curves by computing their correlation coefficient. Specifically, we com-

pute the correlation coefficient between the two curves after aligning them so that the match suggested by the cross-matching function corresponds to the ground truth of the self-matching function. This results in *invariance to offset and uniform scaling*, which has the desirable property of canceling the expected offset. While the unmodeled distortions are not necessarily equivalent to a uniform scaling, this invariance turns out to be sufficient for our purposes.

Formally, the *Self-Aware Matching Measure (SAMM)* for assigning disparity  $d_o$  to a pixel  $(x_L, y)$  is defined as:

$$SAMM(x_L, y, d_o) = \frac{\sum_d (c_{LR}(x_L, y, d - d_o) - \mu_{LR})(c_{LL}(x_L, y, d) - \mu_{LL})}{\sigma_{LR}\sigma_{LL}}. \quad (1)$$

where the summation occurs only over valid values of the disparity, that is when both cost functions compare pixels inside the image boundaries.  $\mu_{LR}$  and  $\sigma_{LR}$  are the mean and standard deviation of the cross-matching function over the valid disparity range and  $\mu_{LL}$  and  $\sigma_{LL}$  are defined similarly for the self-matching function. Self-matching is computed over a disparity range twice as large of the range used for cross-matching centered at  $d = 0$ . We assign minimum SAMM score to matches for which equation 1 cannot be computed with at least 11 terms in the sum to avoid trivial high SAMM scores. Note that the SAMM is a similarity, not a cost, but can easily be converted by taking its negative. Also note that the SAMM can be computed only after an underlying or base cost function has been computed.

SAMM values are high when the variations of  $c_{LR}$  and  $c_{LL}$  correspond after alignment. As long as they move on the same surface and perspective effects are moderate, this remains true. When either function jumps to a different surface, the signals should be uncorrelated. This is undesirable, but, obviously, if the disparity discontinuities were known a priori, the problem of stereo would be much simpler. In practice, local minima of the cost function may have large SAMM values when aligned with the ground truth of the self-matching function. More often than not, however, the correct match results in the maximal SAMM.

The advantage of the SAMM over the DSM [25] is that it works even for relatively ambiguous  $c_{LR}$  and  $c_{LL}$ , which are likely to still be *highly correlated* when aligned at the correct match. Repeated patterns cause no difficulties as long as the same instances are visible in both images. This condition is required for all stereo algorithms in this situation. The SAMM will disambiguate between multiple local minima by aligning the first valley of the cross-matching curve to the first valley of the self-matching curve, the second with the second and so forth.

Computing the SAMM only at the disparity with minimum cost results in a confidence value for each pixel, which

<sup>2</sup>Here the curves would be matching cost over disparity, not intensity or intrinsic curves, respectively.

can be used for ordering matches according to reliability and does not require parameter tuning. See Sec. 4.2 for an evaluation of SAMM as a confidence measure.

Of course, one does not need to restrict the computation of SAMM to a fixed  $d_o$ . Equation 1 can be computed while varying  $d_o$  to produce a new cost volume with the same dimensions as the original. This volume,  $C_{SAM,L}(x, y, d)$ , fulfills all the requirements of Section 1: its computation requires no additional parameters over that of the base cost function and it can be used by any type of stereo algorithm without modifying the algorithm. Figure 3 contains exam-

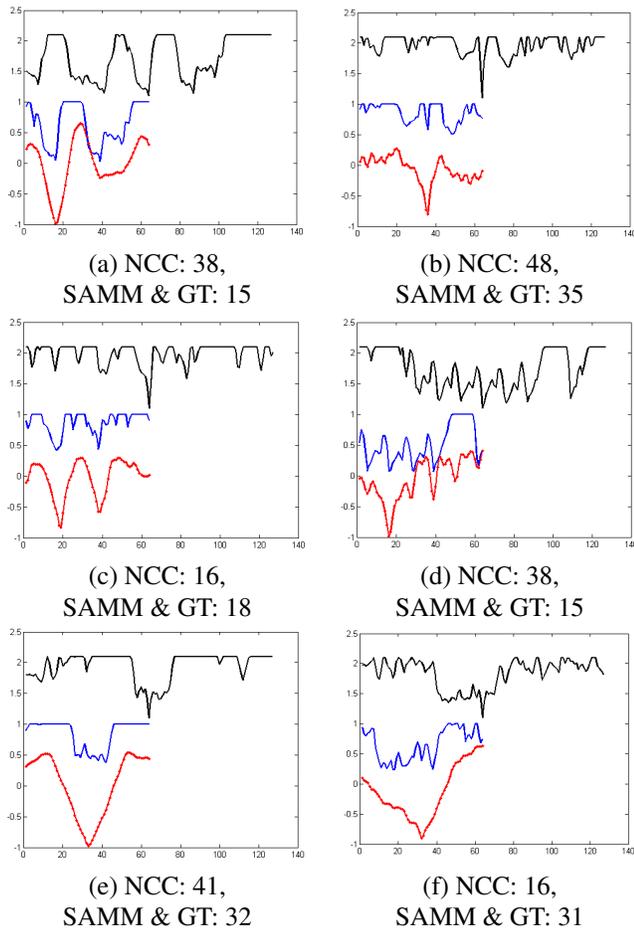


Figure 3. Examples of corrections made by the SAMM on results produced by NCC aggregated in  $5 \times 5$  windows for the Teddy stereo pair. The top (black) curve is the self-matching function, which is computed over twice the regular disparity range and has a minimum at 64 (the midpoint). The middle (blue) curve is NCC and the bottom (red) curve is  $-SAMM$ . The curves have been offset for visualization. GT denotes the ground truth disparity. Notice that: in (a) and (b) NCC selected the wrong local minimum; in (c) NCC selected the correct valley, but was not precise; the texture in (d) is repetitive; in (e) and (f) SAMM unambiguously selects the correct match from very flat NCC curves.

ples of SAMM computations for pixels of the Teddy stereo pair. Of particular interest is the last row that demonstrates the ability of the SAMM to select the correct matches with high confidence without requiring distinctiveness. Figure 4 shows a few examples where the SAMM fails and NCC works. It should be noted that these cases are much fewer than those of Fig. 3. See Section 4.3 for quantitative results.

### 3.3. Symmetric SAMM

The definition of the SAMM in Eq. 1 is effective, but does not exploit any information from the self-matching function of the right image. Note that the cost volume with respect to the right image contains the same values as the cost volume for the left image, re-arranged such that  $C_{LR}(x, y, d) = C_{RL}(x - d, y, -d)$ . This does not hold for the SAMM, since computing the SAMM for the right image brings in additional information from  $C_{RR}$ .

Having computed two SAMM volumes, one for the left and one for the right image, the next step is combining them. There are typically two approaches for this: to consider SAMM values as scores or costs and add them, or to consider them as probabilities (after normalization) and multiply the corresponding values to estimate the joint probability of a match for  $(x, y)$  at  $d$ . We are uncomfortable with the independence assumption required for the second option and chose to add corresponding SAMM values to produce a symmetric SAMM volume that combines information from  $C_{LR}$ ,  $C_{RL}$ ,  $C_{LL}$  and  $C_{RR}$ . (In practice, we have not found any significant difference between the two approaches in a number of experiments.)

The symmetric SAMM (SSAMM) with  $I_L$  as reference is defined as:

$$SSAMM(x, y, d) = C_{SAM,L}(x, y, d) + C_{SAM,R}(x - d, y, -d). \quad (2)$$

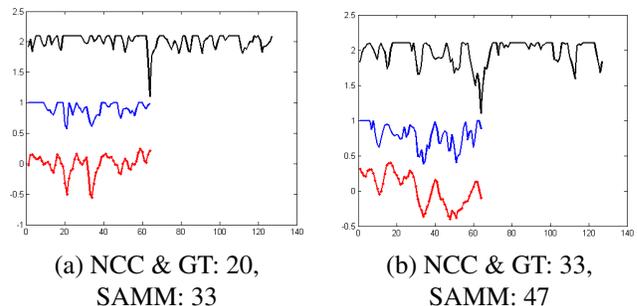


Figure 4. Examples in which the SAMM results in an error, while NCC is correct. See the caption of Fig. 3 for description of visualization. In (a) SAMM selects the wrong local minimum, while in (b) there is little similarity between the self- and cross-matching curves.

## 4. Experimental Results

In this section, we present a sequence of experimental results on stereo pairs with ground truth from the Middlebury Stereo Vision page. Our objective is not to compete with the top algorithms, but rather to show that the SAMM can serve as an effective data term for these algorithms. **In all experiments the SAMM is computed on the underlying matching function it is compared against.**

### 4.1. Evaluation Criteria: Error Rate and ROCs

We begin by introducing the criteria used for evaluating matching functions. The first criterion is the error rate (ER) as defined in [20]. That is, an error is counted for every *non-occluded* pixel assigned a disparity that is off by more than one level from the ground truth.

The second criterion tests the ability of a matching function to rank matches according to their reliability. For this purpose, we rank all pixels in ascending cost order and generate semi-dense disparity maps by selecting the top  $L^{th}$  percentile of matches. Then, we plot the error rate versus the density of the disparity map and measure the area under the curve (AUC). Better performance results in a smaller area under this curve, which indicates that more good matches are ranked high and more wrong matches are

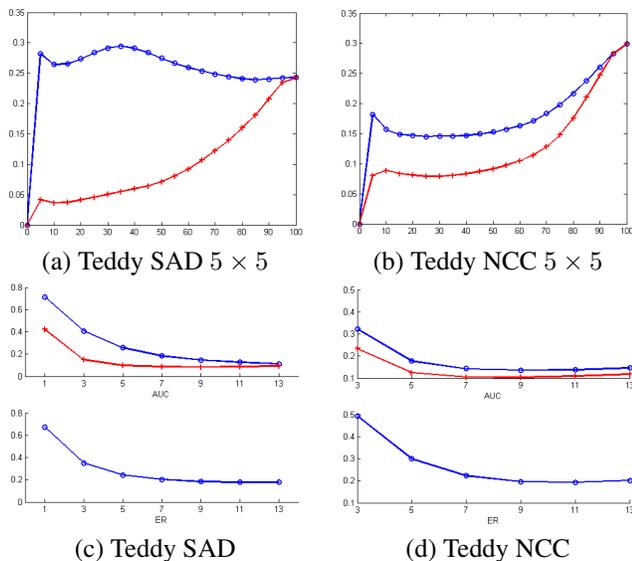


Figure 5. Evaluation of the SAMM as confidence measure for Teddy. The blue curve with circles corresponds to the original cost function and the red curve with crosses corresponds to *non-symmetric* SAMM. (a) and (b): error as a function of disparity map density for SAD and NCC in  $5 \times 5$  windows. The AUC in (a) is 0.257 for SAD and 0.097 for SAMM. (c) and (d) AUC and ER as functions of window size for SAD and NCC. There is a single error rate since SAMM is not used to select disparities here. The  $x$ -coordinate is the width  $N$  of the  $N \times N$  window.

ranked low.

We opted for the simple ROC criterion of Gong and Yang [11], instead of a similar criterion proposed by Kostliwa et al. [13]. Our concern about the latter is that errors can be forgiven if they are caused by other errors. Since we are mostly dealing with noisy disparity maps, we felt that certain types of errors may not be accounted for using [13].

### 4.2. Evaluation of SAMM as Confidence Measure

The first experiment evaluates the SAMM as a confidence measure. We compute disparity values for the MSVP data by the winner-take-all (WTA) strategy of assigning to each pixel the disparity with minimum cost. We, then, compute the SAMM only for the *disparities selected by the underlying cost function*. The evaluation is in terms of the AUC approximated by generating disparity maps at 5% increments in density, including the most reliable matches according to the original cost function or the SAMM.

Results on the Teddy example under SAD, NCC and non-symmetric SAMM for both cases can be seen in Fig. 5. Figure 5(a) and (b) show density-error curves for particular matching functions and the resulting SAMM, while (c) and (d) show how AUC varies as a function of window size. In all cases the SAMM outperforms the original cost function and results in an increasing error curve as density increases, which is the desirable performance. The total error rate at full density is equal in both cases, since we do not select new disparity values according to SAMM.

Results using the symmetric implementation of the SAMM (see Section 3.3) are qualitatively similar and quantitatively slightly better. In the interest of space we do not present any such results for this experiment. The relative improvement in AUC by using the symmetric form for Teddy using SAD over the single-image implementation shown in Fig. 5(c) is approximately 1%.

### 4.3. Evaluation of Winner-Take-All Stereo

The second set of experiments evaluates the SAMM as a matching function by computing WTA disparity maps using one of the matching functions of Section 3.1 and SAMM. We only show results using symmetric SAMM (Eq. 2). Single-image SAMM behaves in a similar manner, with slightly worse quantitative results. Figure 6 shows density-error curves for representative cost functions on the MVSP datasets, while Fig. 7 reports AUC and ER as the window size varies from  $1 \times 1$  to  $13 \times 13$ . Note that NCC is applicable in  $3 \times 3$  windows or larger and that values of  $N$  above 7 result in disparity maps of very low visual quality, which is not reflected in ER.

Yoon and Kweon [25] (Table 4) report an average relative improvement in ER of 0.8% on the MVSP datasets for  $5 \times 5$  SAD. We carried out a similar comparison for the symmetric-SAMM. The improvements over SAD are

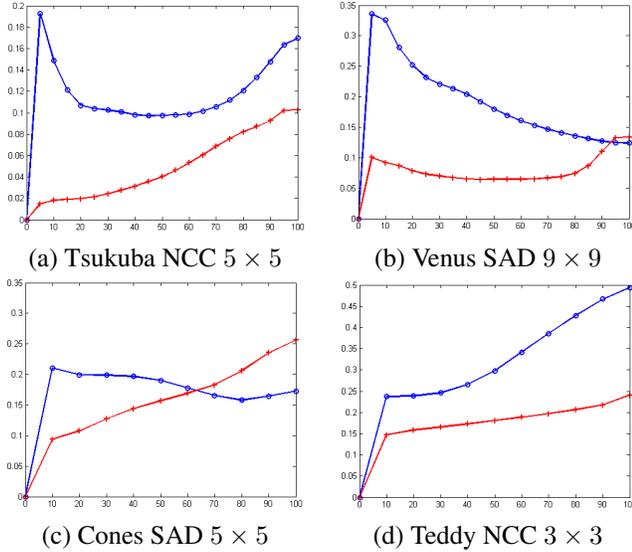


Figure 6. Representative density-error ROC curves for the MSVP data set using SAD, NCC and symmetric-SAMM. Color coding as in Fig. 5. (c) shows an aberration in which SAMM results in higher ER, but still maintains a lower AUC for SAD  $5 \times 5$ .

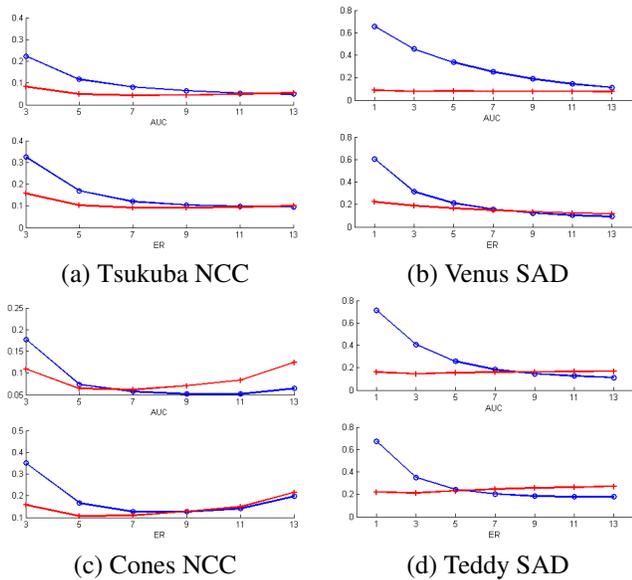


Figure 7. Representative AUC and ER curves for the MSVP data set using SAD, NCC and symmetric-SAMM as the window size varies. The  $x$ -coordinate is the length  $N$  of the  $N \times N$  window. Color coding as in Fig. 5.

135%, 33%, -2.2% and -16% for windows with width 1, 3, 5 and 7. The improvements over NCC are 61%, 25%, 11% and 3.5% for windows with width 3, 5, 7 and 9. In line with all our experiments, the SAMM is more beneficial for small window sizes.

Dataset	SSAMM vs. SAD		SSAM vs. NCC	
	ER(%)	D(%)	ER(%)	D(%)
Tsukuba	-0.60	0.57	2.58	-0.45
Venus	6.10	3.14	1.35	1.82
Teddy	2.15	5.87	0.15	4.61
Cones	7.80	1.68	1.10	1.11
Overall	3.86	2.82	1.30	1.77

Table 1. Comparison on the effects of traditional matching measures and SSAMM on error rate (ER) and density (D) of GCS stereo [6]. SSD was run on  $1 \times 1$  windows and NCC in  $3 \times 3$ . Each entry of the table is the relative *improvement* of SSAMM over the traditional method. We report  $(ER_{SAD} - ER_{SSAMM})/ER_{SAD}$  and  $(D_{SSAMM} - D_{SAD})/D_{SAD}$ . Positive numbers indicate that SSAMM outperforms SAD or NCC.

#### 4.4. Results on GCS Stereo

The Growing Correspondence Seeds (GCS) algorithm [6] is a progressive stereo algorithm with performance guarantees. Source code can be downloaded from: [cmp.felk.cvut.cz/~cechj/GCS/](http://cmp.felk.cvut.cz/~cechj/GCS/). GCS is a semi-dense algorithm that grows disparity components from seeds. It is very robust and can generate meaningful results starting from random seeds. There is, however, some sensitivity to seed selection. We performed the following experiment to compare the capability of traditional matching functions and SAMM to generate useful seeds for GCS. For each of the four MSVP datasets, we select the top 10% of the matches according to the original cost function and use them as seeds for GCS. Then, we apply SSAMM on the original cost volume and rank the matches according to SSAMM. The top 10% of the new matches are used as seeds for GCS. Note that each seed is only a point in disparity space and GCS computes its own matching table. We used default values for all parameters. In Table 1, we report the improvement by SAMM over the results of SAD and NCC in terms of error rate and density of the disparity map.

#### 5. Conclusion

We have presented a transformation that converts stereo matching curves into new curves. The benefits from this transformation include a confidence measure that generates a meaningful ordering of the matches, as well as disparity maps that are usually better than those computed by traditional matching functions according to our criteria. Experimental results show that, as expected, quantitative results using the SAMM depend on the underlying matching function, but also that the SAMM causes an improvement in general, especially when the size of the cost aggregation window is small. Small windows are more attractive for algorithms that do further processing because they do not oversmooth the disparity map and suffer less from foreground fattening.

Moreover, the SAMM offers more predictable perfor-

mance as the window size varies. This is a valuable property when the appropriate window size cannot be estimated a priori. Even when the total error rate in the WTA setting (Sec. 4.3) is larger after SAMM, the resulting cost volume is still very useful for semi-dense and progressive stereo. This is due to the AUC being much smaller for SAMM, especially at low to medium levels of density. (See ROC curves in Figs. 5(a) and (b) and 6.)

The advantage of the SAMM over the methods of [16] and [25] is that it works well even on ambiguous cross- and self-matching curves, which are still *highly correlated* when aligned at the correct match. This is, to the best of our knowledge, a unique property of our method. The SAMM would work for the hypothetical cost function of Fig. 1(b) as opposed to the methods of [16, 15, 25] which would reject this pixel as unreliable a priori. See Fig. 3 for representative real examples in which the SAMM resolves errors and ambiguities of the underlying matching function.

## References

- [1] A. Agarwal. and A. Blake. Dense stereo matching over the panum band. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009.
- [2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998.
- [3] A. Bobick and S. Intille. Large occlusion stereo. *Int. Journal of Computer Vision*, 33(3):1–20, 1999.
- [4] D. Bradley, T. Boubekeur, and W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2008.
- [5] N. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *European Conf. on Computer Vision*, pages 766–779, 2008.
- [6] J. Cech and R. Sara. Efficient sampling of disparity space for fast and accurate matching. In *BenCOS*, 2007.
- [7] Q. Chen and G. Medioni. A volumetric stereo matching method: Application to image-based modeling. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages I: 29–34, 1999.
- [8] G. Egnal, M. Mintz, and R. Wildes. A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and Vision Computing*, 22(12):943–957, 2004.
- [9] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 2402–2409, 2006.
- [10] M. Gong, R. Yang, L. Wang, and M. Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *Int. Journal of Computer Vision*, 75(2):283–296, 2007.
- [11] M. Gong and Y. Yang. Fast unambiguous stereo matching using reliability-based dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):998–1003, 2005.
- [12] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009.
- [13] J. Kostliva, J. Cech, and R. Sara. Feasibility boundary in dense and semi-dense stereo matching. In *BenCOS*, 2007.
- [14] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(8):1140–1146, 2002.
- [15] R. Manduchi and C. Tomasi. Distinctiveness maps for image matching. In *Int. Conf. on Image Analysis and Processing*, pages 26–31, 1999.
- [16] L. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *SPIE*, 1570:187–200, 1991.
- [17] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nistér, and M. Pollefeys. Real-time visibility-based fusion of depth maps. In *Int. Conf. on Computer Vision*, 2007.
- [18] H. Moravec. Rover visual obstacle avoidance. In *Int. Joint Conf. on Artificial Intelligence*, pages 785–790, 1981.
- [19] E. Salari and J. Strong. On the reliability of correlation based stereo matching. In *IEEE Int. Conf. on Systems Engineering*, pages 559–561, 1990.
- [20] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision*, 47(1-3):7–42, 2002.
- [21] C. Tomasi and R. Manduchi. Stereo matching as a nearest-neighbor problem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):333–340, 1998.
- [22] P. Viola and W. Wells, III. Alignment by maximization of mutual information. *Int. Journal of Computer Vision*, 24(2):137–154, 1997.
- [23] K. Yoon and I. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.
- [24] K. Yoon and I. Kweon. Stereo matching with the distinctive similarity measure. In *Int. Conf. on Computer Vision*, 2007.
- [25] K. Yoon and I. Kweon. Distinctive similarity measure for stereo matching under point ambiguity. *Computer Vision and Image Understanding*, 112(2):173–183, 2008.
- [26] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conf. on Computer Vision*, pages B:151–158, 1994.
- [27] Y. Zhang and C. Kambhamettu. Stereo matching with segmentation-based cooperation. In *European Conf. on Computer Vision*, pages II: 556–571, 2002.
- [28] Z. Zhang and Y. Shan. A progressive scheme for stereo matching. In *Lecture Notes in Computer Science*, 2018, pages 68–85, 2001.
- [29] C. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(7):675–684, 2000.