Accepted for publication, June 2003.
To appear in IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)

Sampling the Disparity Space Image

Richard Szeliski
Microsoft Research
Redmond, WA 98052
szeliski@microsoft.com

Daniel Scharstein
Middlebury College
Middlebury, VT 05753
schar@middlebury.edu

Abstract

A central issue in stereo algorithm design is the choice of *matching cost*. Many algorithms simply use squared or absolute intensity differences based on integer disparity steps. In this paper we address potential problems with such approaches. We begin with a careful analysis of the properties of the continuous *disparity space image* (DSI) and propose several new matching cost variants based on symmetrically matching interpolated image signals. Using stereo images with ground truth, we empirically evaluate the performance of the different cost variants and show that proper sampling can yield improved matching performance.

1 Introduction

The last few years have seen a dramatic improvement in the quality of dense stereo matching algorithms [13]. A lot of this improvement can be attributed to better optimization algorithms and better smoothness constraints [6, 4, 17]. However, a remarkable amount of the improvement has also come from better matching metrics at the input [3]. In fact, Birchfield and Tomasi's sampling-insensitive dissimilarity measure is used by a number of today's best performing algorithms [6, 4].

Using something better than just pixel-sampled intensity differences is not a new idea. For example, Matthies *et al.* interpolated scanlines by a factor of 4 using a cubic interpolant before computing the SSD score [10]. Tian and Huhns wrote an even earlier survey paper comparing various algorithms for sub-pixel registration [18]. In fact, some stereo and motion algorithms have always evaluated displacements on a half-pixel grid, but never mentioned this fact explicitly.

The set of initial matching costs that are fed into a stereo matcher's optimization stage is often called the *disparity space image* (DSI) [20, 5]. However, while the concept of stereo matching as finding an optimal surface through this space has been around for a while [20, 2, 5], relatively little attention has been paid to the proper sampling and treatment of the DSI itself.

In this paper, we take a more careful look at the structure of the DSI, including its frequency characteristics and the effects of using different interpolators in sub-pixel registration. Among the questions we ask are: What does the DSI look like? How finely do we need to sample it? Does it matter what interpolator we use? We also propose a number of novel modifications to the matching cost that produce a better set of initial high-quality matches, at least in textured, unoccluded areas. It is our contention that filling in textureless and occluded areas is best left to a later stage of processing [5, 6, 4, 17], which is why we do not consider global optimization techniques in this paper.

The remainder of the paper is structured as follow. Section 2 presents our analysis of the DSI and discusses minimal sampling requirements. Section 3 develops several novel matching costs based on our analysis. The utility of these novel costs is validated experimentally in Section 4. We conclude with some ideas for future research.

2 Matching costs

In this section, we look at how matching costs are formulated. In particular, we analyze the structure of the DSI and its sampling properties and propose some improvements to commonly used

matching costs.

2.1 The continuous disparity space image

Given two input images, $I_L(x, y)$ and $I_R(x, y)$, we wish to find a disparity map $d_L(x, y)$ such that the two images match as closely as possible

$$I_L(x, y) \approx I_R(x - d_L(x, y), y). \tag{1}$$

In this paper we assume that the images have been rectified to have a horizontal epipolar geometry [11, 8], i.e., that the images have been pre-warped so that corresponding pixels are on the same scanline. Stereo correspondence can under such circumstances be reduced to one-dimensional search. Note that our paper proposes interpolating images to a higher resolution after rectification. However, these two steps could be combined into a single resampling operation to reduce aliasing artifacts.

We define the 3D *signed difference image* (SDI) as the intensity (or color) difference between the shifted left and right images,

$$SDI_L(x, y, d) = I_L(x, y) - I_R(x - d, y). \tag{2}$$

We also define the raw *disparity space image* (DSI) as the squared difference (summed over all the color bands),

$$DSI_L(x, y, d) = \|SDI_L(x, y, d)\|^2. \tag{3}$$

Alternate metrics such as absolute differences or robust functions are also possible. However, the quadratic case is easiest to analyze and also corresponds to the case of Gaussian noise, as we will discuss shortly. In the ideal (continuous, noise-free) case with no occlusions, we expect $DSI_L(x, y, d_L(x, y))$ to be zero.

Unfortunately, we do not actually have access to continuous, noise-free versions of $I_L(x, y)$ and $I_R(x, y)$. Instead, we have sampled noisy versions, $\hat{I}_L(x_i, y_i)$ and $\hat{I}_R(x_i, y_i)$,

$$\hat{I}_L(x_i, y_i) = [I_L * h](x_i, y_i) + n_L(x_i, y_i) \tag{4}$$

$$\hat{I}_R(x_i, y_i) = [I_R * h](x_i, y_i) + n_R(x_i, y_i), \tag{5}$$

where $h(x, y)$ is the combined point-spread-function of the optics and sampling sensor (e.g., it incorporates the CCD fill factor [19]), and n_L is the (integrated) imaging noise.

Given that we usually only evaluate the DSI at the integral grid positions (x_i, y_i), we have to ask whether this sampling of the DSI is adequate, or whether there is severe aliasing in the

resulting signal. We cannot, of course, reconstruct the true DSI since we have already band-limited, corrupted, and sampled the original images. However, we can (in principle) reconstruct continuous signals from the noisy samples, and then compute their continuous DSI.

The reconstructed signal can be written as

$$\begin{aligned} \overline{I}_L(x,y) &= \sum_i \hat{I}_L(x_i,y_i) g(x-x_i, y-y_i) && (6) \\ &= \tilde{I}_L(x,y) + \tilde{n}_L(x,y), && (7) \end{aligned}$$

where $g(x,y)$ is a reconstruction filter, $\tilde{I}_L(x,y)$ is the sampled and reconstructed version of the *clean* (original) signal, and $\tilde{n}_L(x,y)$ is an interpolated version of the noise. This latter signal is a band-limited version of continuous Gaussian noise (assuming that the discrete noise is i.i.d. Gaussian).

We can then write the interpolated SDI and DSI as

$$\begin{aligned} \overline{SDI}_L(x,y,d) &= \overline{I}_L(x,y) - \overline{I}_R(x-d,y) \quad \text{and} && (8) \\ \overline{DSI}_L(x,y,d) &= \|\overline{SDI}(x,y,d)\|^2. && (9) \end{aligned}$$

What can we say about the structure of these signals?

2.2 Frequency analysis and adequate sampling

The answer can be found by taking a Fourier transform of the SDI. Let us fix y for now and just look at a single scanline,

$$\mathcal{F}\{\overline{SDI}\} = \mathcal{F}\left\{\overline{I}_L(x) - \overline{I}_R(x-d)\right\} = H_L(f_x) - H_R(f_x) e^{j2\pi(f_x - f_d)}, \qquad (10)$$

where H_L and H_R are the Fourier transforms of $\overline{I}_L$ and $\overline{I}_R$, and f_x and f_d are the x and d frequencies.

Figure 1 shows the SDIs and DSIs and their Fourier transforms for two scanlines taken from the 38th and 148th row of a test image pair containg newsprint on a slanted surface. The first term in (10) corresponds to the horizontal line in the SDI's Fourier transform (second column of Figure 1), while the second term, which involves the disparity, is the slanted line.

Squaring the SDI to obtain the DSI (third column in Figure 1) is equivalent to convolving the Fourier transform with itself (fourth column in Figure 1). The resulting signal has twice the bandwidth in x and d as the original SDI (which has the same bandwidth as the interpolated signal). It is also interesting to look at the structure of the DSI itself. The thin diagonal stripes are spurious

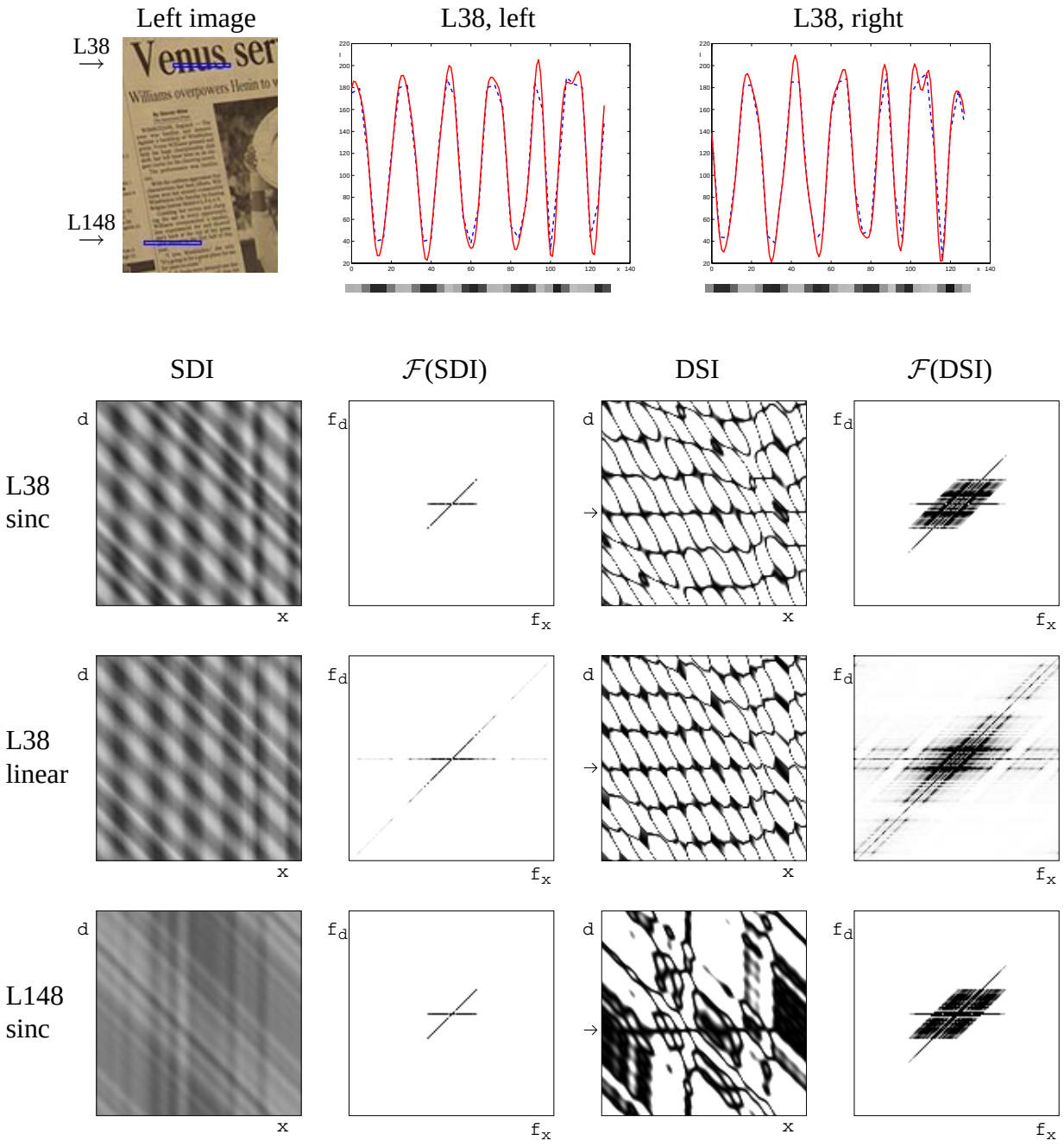


Figure 1: *Sample SDIs and DSIs and their Fourier transforms. Top row: Original image with two selected scanlines and intensity profiles of the first selected scanline (L38); notice how the sinc-interpolated signals (solid) are more similar than the linearly interpolated ones (dashed). Bottom rows: Signed Difference Image (SDI) and its transform, and Disparity Space Image (DSI) and its transform; first for L38 using perfect (sinc) interpolation, then for L38 using piecewise linear interpolation, then for L148 using perfect interpolation. The correct disparity in the DSI images is marked with an arrow.*

bad matches (dark-light transitions matching light-dark transitions), while the horizontal stripes are good matching regions (the straighter and darker the better).

What can we infer from this analysis? First, the continuous DSI has significant frequency content above the frequencies present in the original intensity signal. Second, the amount of additional content depends on the quality of the interpolator applied to the signal. Thus, when perfect band-limited reconstruction (a sinc filter) is used, the resulting DSI signal only has twice the frequency of the image. It is therefore adequate (in theory) to sample the DSI at 1/2 pixel intervals in x and d. When a poorer interpolant such as piecewise linear interpolation is used, the sampling may have to be much higher. The same is true when a different non-linearity is used to go from the SDI to the DSI, e.g., when absolute differences or robust measures are used. This is one of the reasons we prefer to use squared difference measures. Other reasons include the statistical optimality of the DSI as the log likelihood measure under Gaussian noise, and the ability to fit quadratics to the locally linearized expansion of the DSI (see Section 3.3).

We can summarize these observations in the following Lemma:

Lemma 1: *To properly reconstruct a Disparity Space Image (DSI), it must be sampled at* least *twice the horizontal and disparity frequency as the original image (i.e., we must use at least 1/2 pixel samples and disparity steps).*

It is interesting to note that if a piecewise linear interpolant is applied between image samples before differencing and squaring, the resulting DSI is piecewise quadratic. Therefore, it suffices in principle to simply compute one additional squared difference between pixels, and to then fit a piecewise quadratic model. While this does reconstruct a continuous DSI, there is no guarantee that this DSI will have the same behavior near true matches as a more properly reconstructed DSI. Also, the resulting minima will be sensitive to the original placement of samples, i.e., a significant bias towards integral disparities will exist [14].

For example, if the original signal is a fairly high-frequency chirp (Figure 2a), applying a piecewise linear interpolant will fail to correctly match the signal with a fractionally shifted version. Figure 2b and c show the results of aggregating the original raw DSIs with a 7-pixel filter (see Section 3). Clearly, using the linear interpolant will result in the wrong disparity minimum being selected in the central portion (where the central horizontal line is weak). One might ask whether such high-frequency signals really exist in practice, but it should be clear from Figure 1 that they do.

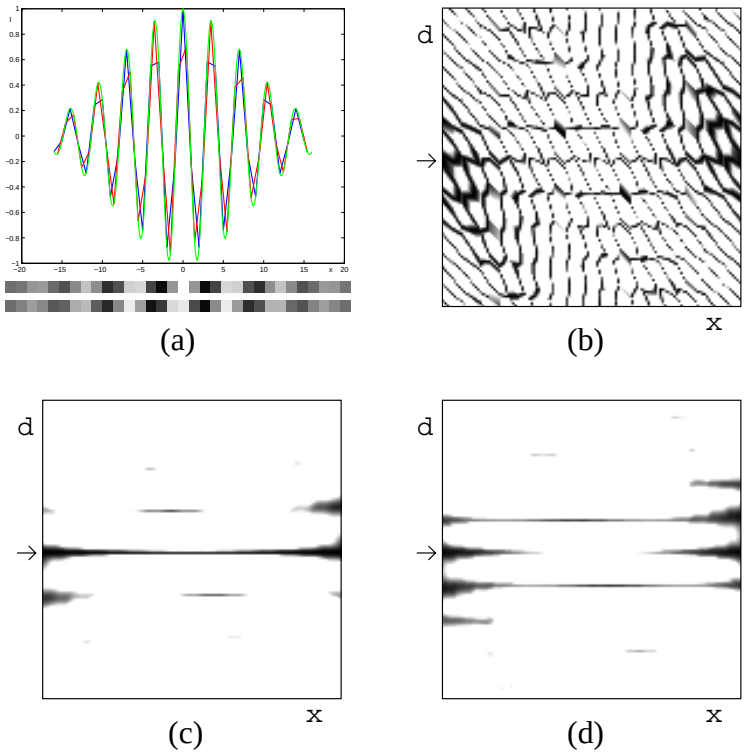


Figure 2: *Chirp signal matching: (a) a continuous signal and its shifted and discretely sampled versions; (b) Disparity Space Image (DSI) for linear interpolation; (c) horizontally aggregated DSI for sinc interpolation, showing the correct minimum; (d) horizontally aggregated DSI for linear interpolation, with incorrect minima near the center. The correct disparity in the DSI images is marked with an arrow.*

3 Improved matching costs

Given the above analysis, how can we design a better initial matching cost? Birchfield and Tomasi [3] and Shimizu and Okutomi [14] have both observed problems with integral DSI sampling and have proposed different methods to overcome this problem.

Birchfield and Tomasi's *sampling-insensitive dissimilarity measure* compares each pixel in the reference image against the linearly interpolated signal in the matching image, and takes the minimum squared error as the matching cost. It then reverses the role of the reference and matching images, and takes the minimum of the resulting two cost measures. In terms of our continuous DSI analysis, this is equivalent to sampling the DSI at integral x locations, and computing the minimum value vertically and diagonally around each integral d value, based on a piecewise linear reconstruction of the DSI from integral samples.

Shimizu and Okutomi [14] compute "cancellation costs" using half-pixel interpolated signals, and add these to the original cost measure to reduce the bias towards integral estimates (which they call "pixel locking").

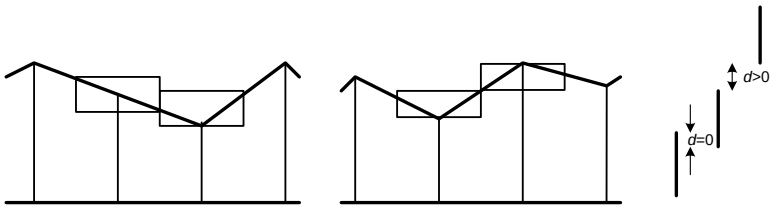


Figure 3: *Interval analysis: (a–b) two signals with their corresponding half-sample intervals; (c) three intervals being compared (differenced). The difference between the first two intervals is* $d = 0$ *because their ranges overlap. The difference between the second two intervals is* $d > 0$*, i.e., the difference between the nearest two points in the two intervals.*

In this paper, we define a family of improved matching costs including generalizations of Birchfield and Tomasi's matching measure.

3.1 Symmetric matching of interpolated signals

First, we interpolate both signals up by a factor s using an arbitrary interpolation filter. In this paper, we study linear ($o = 1$) and cubic ($o = 3$) interpolants. (The cubic interpolant is a compact approximation to a sinc filter [15] and is often used as the standard high-quality interpolant in many image-processing applications.) We then compute the squared differences between *all* of the interpolated and shifted samples, as opposed to just between the original left (reference) image pixels and the interpolated and shifted right (matching) image samples. This difference signal is then reduced back to the original horizontal image sampling rate (i.e., to a single value per original pixel) using a symmetric box (moving average) filter of width s and then downsampling. A higher-order filter could potentially be used, but we wish to keep discontinuities in depth sharp in the DSI, so we prefer a simple box filter.

3.2 Interval matching

If we wish to apply the idea of a sampling-insensitive dissimilarity measure [3], we can still do this on the interpolated signals before downsampling. However, rather than treating the reference and matching images asymmetrically and then reversing the roles of reference and matching (as in [3]), we have developed the following related variant that is based on interval analysis.

Figure 3 shows two signals that have been interpolated to yield the set of discrete intensity samples shown as vertical lines. (A piecewise linear interpolant is used here since we expect the original interpolation stage to take care of aliasing.) The original Birchfield-Tomasi measure com-

pares a pixel in the reference image with the interval in the matching image defined by the center pixel and its two 1/2-sample interpolated values (rectangular boxes in Figure 3a–b). (This difference is 0 if the pixel falls within the intervals, else it is the smaller of the differences from the two endpoints.) It then performs this same computation switching the reference and matching images, and takes the minimum of the resulting two costs. Our version of the algorithm simply compares the two *intervals*, one from the left scanline, the other from the right, rather than comparing values against intervals. The unsigned difference between two intervals is trivial to compute: it is 0 if the intervals overlap (Figure 3c), else it is the gap between the two intervals. A signed difference could also be obtained by keeping track of which interval is higher, but in our case this is unnecessary since we square the differences after computing them. When working with color images, we currently apply this interval analysis to each color band separately. In principle, the same sub-pixel offset should be used for all three channels, but the problem then becomes a more complicated quadratic minimization problem instead of simple interval analysis.

3.3 Local minimum finding (quadratic fit)

An alternative to doing such interval analysis is to directly compute the squared differences, and to then fit a parabola to the resulting sampled DSI. This is a classic approach for obtaining sub-pixel disparity estimates [18, 1, 10], although applying it directly to integer-valued displacements (disparities) can lead to severe biases [14].

When the DSI has been adequately sampled, however, this is a useful alternative for estimating the analytic minimum from the (fractionally) sampled DSI. Note that we use the parabola fit here *not* to obtain sub-pixel disparities, but rather to reconstruct the minimum DSI value, i.e., the actual smallest *matching cost* in the vicinity of the sampled value.

In order to reduce the noise in the DSI before fitting, we apply spatial aggregation (averaging with neighbors) first. In this paper, we use fixed uniformly weighted square windows (i.e., box filters) , which perform well in textured areas, as long as the window does not straddle a depth boundary. While the use of shiftable windows (windows offset from the center pixel) [13] can improve the performance of matching near depth discontinuities, it makes the analysis more difficult, and is not the main focus of our paper.

3.4 Collapsing the DSI

Finally, once the local minima in the DSI at all pixels have been adequately modeled, we can collapse the DSI back to an integral sampling of disparities. This step is often not necessary, as many stereo matchers do their optimization at sub-pixel disparities. It does, however, have several potential advantages:

- For optimization algorithms like graph cuts [6] where the computation complexity is proportional to the square of the number of disparity levels, this can lead to significant performance improvements.
- Certain symmetric matching algorithm (e.g., dynamic programming) require an integral sampling of disparity to establish two-way optima.

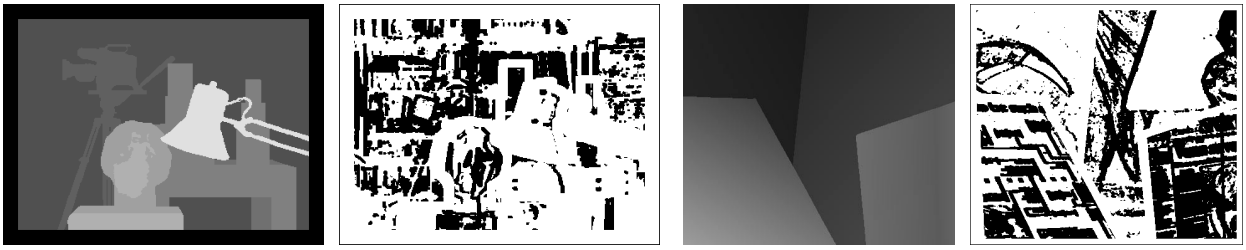
To collapse the DSI, we find the lowest matching score within a $\frac{1}{2}$ disparity from each integral disparity, using the results of the parabolic fitting, if it was used. We also store the relative offset of this minimum from the integral disparity for future processing and for outputting a final high-accuracy disparity map, as well as the local certainty in the match, which can be determined from the parabolic fit [1, 10]. Alternately, sub-pixel estimates could be recomputed at the end around each winning disparity using one of the techniques described in [18], e.g., using a Lucas-Kanade gradient-based fit [9] to nearby pixels at the same disparity.

4 Experimental evaluation of matching costs

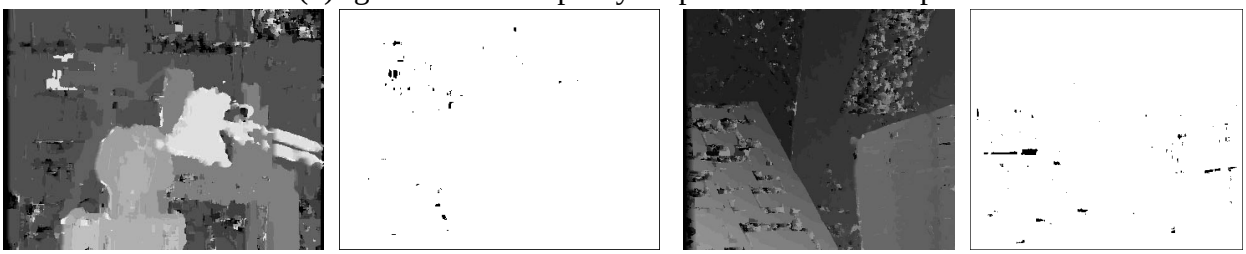
Since there are so many alternatives possible for computing the DSI, how do we choose among them? From theoretical arguments, we know that it is better to sample the DSI at fractional disparities and to interpolate the resulting surface when looking for local minima. However, real images have noise and other artifacts such as aliasing and depth discontinuities. We therefore evaluate our new techniques using the Sawtooth, Tsukuba, and Venus stereo test sequences with ground truth from [13], which are available at http://www.middlebury.edu/stereo. Two of these sequences are shown in Figure 4a. We should note that the Sawtooth and Venus data sets have high-quality sub-pixel accurate ground-truth estimates (computed using piecewise planar surface fitting), while the Tsukuba ground truth has only integer disparities. We are not using the Map data set, which can be solved almost perfectly, making it ill-suited for comparing different matching costs.

In this paper, we focus on the accuracy of these techniques in unoccluded textured areas. The effect of different matching costs in textureless areas is harder to evaluate, since the results depend

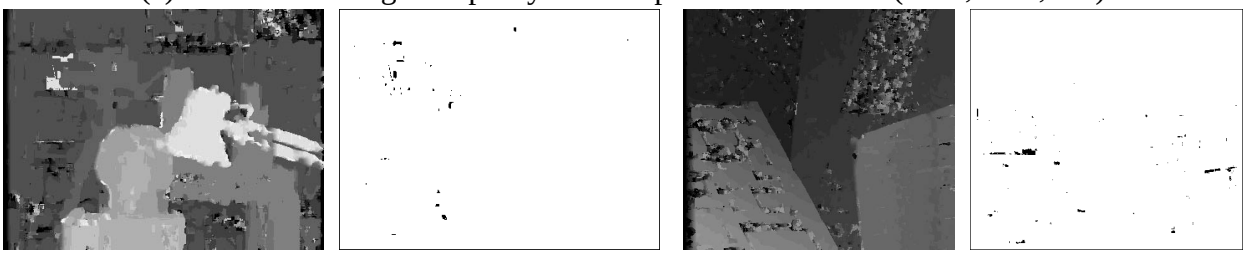
(a): Tsukuba Venus



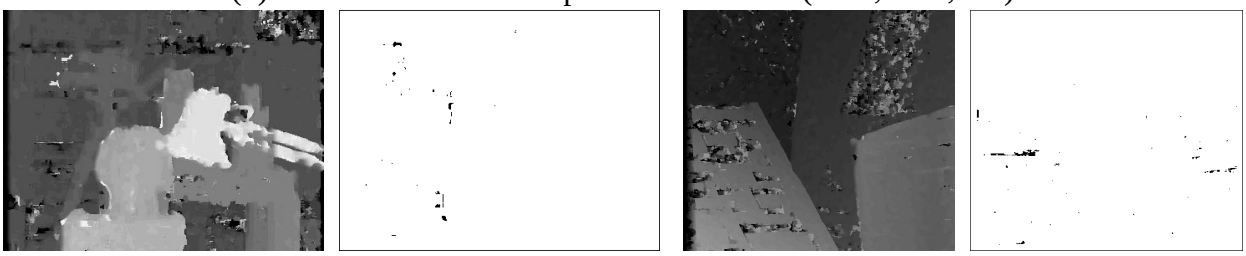
(b): ground truth disparity maps and textured maps



(c): traditional integer-disparity SSD disparities and errors ($s=1$, $i=0$, SD)



(d): Birchfield-Tomasi disparities and errors ($s=1$, $i=0$, BT)



(e): half-pixel symmetric interval-difference results ($s=2$, $i=1$, ID)

Figure 4: *Test images and selected results: (a) input images; (b) true disparity maps and textured nonoccluded regions (shown in black) in which errors are being evaluated; (c) traditional SSD results (disparity maps and error maps); (d) Birchfield-Tomasi results; (e) fractional disparities with symmetric matching and interval difference. The error maps in (c–e) show in black the "bad" matches in textured, nonoccluded regions, i.e., pixels whose floating point disparity differs from the ground truth by more than 1.*

strongly on the aggregation or global optimization algorithm. We therefore restrict our analysis to textured areas and use a simple window-based correspondence algorithm. Untextured areas can be handled in many ways; for example, after establishing "certain" matches in textured areas, untextured areas can be filled in using diffusion [12], aggregation with successively larger windows [16], or global optimization methods [6]. We also exclude pixels near depth discontinuities, which present problems for window-based methods.

For our analysis, we select textured pixels by computing the squared horizontal gradient at each pixel (averaging the left and right values to remain symmetrical). These values are then averaged in a 3×3 neighborhood and thresholded, using a threshold of 6 gray levels squared. Occluded pixels are found by forward-warping the true disparity maps, and depth discontinuity regions are selected by dilating the locations of strong disparity jumps and occlusion [13]. The resulting textured unoccluded pixels are shown as black pixels in Figure 4b.

The parameters that we vary in our experiments are as follows:

- $s = 1, 2, 4$: interpolation rate (inverse of fractional disparity);
- $o = 1, 3$: interpolation order – linear or cubic;
- $i = 1, 0$: symmetric matching of interpolated scanlines (Section 3.1) – on or off;
- $d =$ SD, ID, BT: dissimilarity metric (Section 3.2) – squared differences, interval differences, or Birchfield-Tomasi measure;
- $f = 1, 0$: parabola fit for minimum cost estimation (Section 3.3) – on or off.

We have also varied other parameters, including window size (which is 7×7 in all experiments reported here), and using absolute (rather than squared) differences. The effect of changing these parameters is discussed below. In order to be able to evaluate subpixel disparity performance, we do not collapse the DSI to an integer sampling in this study. The statistics we gather for each experiment are the RMS disparity errors and the percentage of "bad" matches, i.e., pixels whose floating point disparity differs from the ground truth by more than 1.

Table 1 shows the numerical results of some of our experiments. The top three rows list the different values of parameters s, i, and d; the other parameters are held constant at $o = 3$ (cubic interpolation), $f = 0$ (no cost fitting), and a window size of 7×7. Each of the remaining five rows compares the matching performance under the different parameter settings in the textured, unoccluded regions of a given data set. The lowest score for each data set is highlighted in boldface.

While there is no single setting that consistently outperforms the others, our new cost variants generally do better than the original costs. We now evaluate the effect of the different parameters,

Interpolation rate s :		$s=1$			$s=2$				$s=4$			
Symmetric matching i :		0			0		1		0		1	
Dissimilarity metric d :		SD	ID	BT	SD	ID	SD	ID	SD	ID	SD	ID
Bad pixel %	Sawtooth	2.55	3.19	2.96	1.81	1.94	1.78	2.15	1.66	1.75	**1.65**	1.74
	Tsukuba	1.07	0.82	0.87	1.25	1.01	1.07	**0.71**	1.55	1.44	1.39	1.09
	Venus	1.68	1.37	1.30	0.91	0.88	0.86	0.88	0.93	0.88	0.82	**0.79**
RMS error	Venus	0.85	0.73	0.68	0.62	**0.55**	0.62	0.59	0.62	0.58	0.59	**0.55**
	Venus / subpix	0.80	0.68	0.62	0.60	**0.53**	0.60	0.56	0.61	0.57	0.58	0.55

Table 1: *Matching performance as a function of parameters s, i, and d. (SD=squared differences, ID=interval differences, BT=Birchfield-Tomasi.) Parameters o and f are held constant at $o=3$ (cubic interpolation) and $f=0$ (no cost fitting). The middle three rows show the percentages of bad matching pixels for the three data sets tested; the last two rows show the RMS disparity errors for the Venus data set without and with a final subpixel fitting step. The lowest number in each row is highlighted in boldface. The underlined numbers correspond to the results shown in Figure 4.*

focusing first on the bad pixel percentages, which give a good indication of the overall performance of the different cost variants.

- SD vs. ID — Interval differences outperform squared differences on the Tsukuba and Venus data sets accross other parameter variations. On the Sawtooth images, however, they result in decreased performance. Careful analysis of the images and the error maps reveals that there is a small vertical misregistration present in the original images, and that the errors occur in areas with near-horizontal lines. This suggest that interval differences are a good idea, but may amplify aliasing problems caused by misaligned images. For comparison, we have included the results for the Birchfield-Tomasi measure (BT, third column). Given integer sampling ($s=1$), BT and ID yield similar results; using interpolation (discussed next), however, we can clearly improve upon BT's performance.

- Interpolation rate s — Interpolating the images ($s=2$ and $s=4$) yields an obvious improvement over integer-based costs ($s=1$), verifying our theoretical results from Sections 2 and 3. Quarter-pixel steps ($s=4$) perform similar to half-pixel steps ($s=2$); the numbers are slightly better on Sawtooth and Venus images, but worse on Tsukuba (reasons for this anomaly are discussed below). This suggests that for most practical applications half-pixel steps are sufficient, as long as a good interpolant (e.g., cubic) is used.

- Symmetry ($i=1$ vs. $i=0$) — Symmetric matching ($i=1$) yields slightly better results in most cases.

Figure 4c–e shows the results corresponding to the underlined numbers in the first ($s=1$, SD), third ($s=1$, BT), and seventh ($s=2$, $i=1$, ID) column of Table 1. Note that the seventh column consistently outperforms columns 1–3.

Given the small differences in the numerical scores, the question arises whether the results are statistically significant. Careful examination of the error maps for the different parameter settings (including those in Figure 4c–e) shows that our new costs do indeed result in a significant reduction of errors in high-frequency image regions, as predicted by our theoretical analysis. This is most apparent for the Venus images, which contain many such regions. Errors are also reduced in other areas affected by aliasing, such as strong intensity discontinuities or near-horizontal edges. Other errors, however, are not a direct result of the matching cost, and can obscure the numerical results. The Tsukuba images in particular contain fewer high-requency regions, but several areas with repetitive patterns and fine disparity variations that are challenging for a window-based method, and thus result in spurious errors that are not directly a function of the matching cost used.

Although not shown in Table 1, we have also analyzed the effect of changing other parameters:

- Using linear interpolation ($o=1$) gives clearly inferior results than cubic interpolation, again validating our observations from Section 3.

- Refining the cost values by local fitting ($f=1$) results in minor differences, and does not yield a clear improvement.

- Using absolute differences rather than squared differences yields comparable results, and in some cases even small improvements.

- Decreasing the window size increases the errors overall, but does not significantly change the relative performance of the different matching cost variants.

An interesting question is to what extent the new cost variants improve the quality of subpixel disparity estimation. The last two rows of Table 1 show that the RMS disparity errors on the Venus data set decrease slightly when subpixel (floating-point) disparity estimation is turned on (using a standard parabola fit around the winning cost values). Note that the RMS numbers are contaminated by gross errors; visual inspection of the disparity maps shows an obvious improvement over the typical "stair-casing effect" exhibited by our discrete matching algorithm (which is noticeable even at quarter-pixel steps).

In summary, it can be seen that symmetric interpolated matching ($i=1$ and $s=2$ or $s=4$) usually outperforms traditional, integer-based matching, in particular in high-frequency image

regions. Cubic interpolation should always be used. Interval differences help as well, but seem less tolerant to calibration errors. The benefit of interval matching also depends on the winner selection strategy (for example, it can cause problems for algorithms that analyze cost distributions, because good matches often yield a matching cost of 0). Cost refinement by parabola fitting does not seem to increase matching performance. However, using the same fitting technique to refine the winning (half or quarter-pixel) disparities into true floating-point disparities generally further reduces the remaining disparity errors and results in smoother disparity maps.

5 Conclusion

In this paper we have presented novel matching costs based on interpolated image signals. The need for such costs was motivated by a frequency analysis of the continuous disparity space image (DSI). We have explored several symmetric cost variants, including a generalized version of Birchfield and Tomasi's matching criterion [3]. While there is no clear winner among the different variants, we have demonstrated that our new matching costs result in improved performance, particularly in high-frequency image regions, and that they also yield improved subpixel disparity estimates.

An interesting generalization of our approach is to use a smaller interval from each image, e.g., to only interpolate $\pm\frac{1}{4}$ pixel away (or in general ϵ away). This could be used to compensate for small unmodeled shifts in the images, e.g., residual vertical parallax. We call this dilation of a pixel value to an interval determined by its neighbors' values a *partial shuffle*, since it is related to Kutulakos' *shuffle transform* [7].

Another major direction for future work is to determine which pixels can be matched with high certainty (negligible error), and to use these matches as a set of anchors points for resolving the remaining ambiguous matching regions [5, 16, 17]. It is our hope that this approach could be used to produce high-quality correspondence maps without the higher computational requirements of global optimization methods. In general, we believe that paying close attention to the quality of local evidence (matching costs) will play a significant role in computing high-quality stereo reconstructions.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, January 1989.

[2] P. N. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3):237–260, August 1996.

[3] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, April 1998.

[4] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *Seventh International Conference on Computer Vision (ICCV'99)*, pages 489–495, Kerkyra, Greece, September 1999.

[5] A. F. Bobick and S. S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, September 1999.

[6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001.

[7] K. N. Kutulakos. Approximate N-view stereo. In *Sixth European Conference on Computer Vision (ECCV 2000)*, volume I, pages 67–83, Dublin, Ireland, June/July 2000. Springer-Verlag.

[8] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, volume I, pages 125–131, Fort Collins, June 1999.

[9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision. In *Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, pages 674–679, Vancouver, 1981.

[10] L. H. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.

[11] M. Okutomi and T. Kanade. A multiple baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, April 1993.

[12] D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, 28(2):155–174, July 1998.

[13] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, May 2002.

[14] M. Shimizu and M. Okutomi. Precise sub-pixel estimation on area-based matching. In *Eighth International Conference on Computer Vision (ICCV 2001)*, volume I, pages 90–97, Vancouver, Canada, July 2001.

[15] R. Szeliski and M. R. Ito. New Hermite cubic interpolator for two-dimensional curve generation. *IEE Proceedings E*, 133(6):341–347, November 1986.

[16] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. In *Seventh European Conference on Computer Vision (ECCV 2002)*, volume II, pages 525–540, Copenhagen, May 2002. Springer-Verlag.

[17] H. Tao, H.S. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *Eighth International Conference on Computer Vision (ICCV 2001)*, volume I, pages 532–539, Vancouver, Canada, July 2001.

[18] Q. Tian and M. N. Huhns. Algorithms for subpixel registration. *Computer Vision, Graphics, and Image Processing*, 35:220–233, 1986.

[19] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of CCD imaging process. In *Eighth International Conference on Computer Vision (ICCV 2001)*, volume I, pages 480–487, Vancouver, Canada, July 2001.

[20] Y. Yang, A. Yuille, and J. Lu. Local, global, and multilevel stereo matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pages 274–279, New York, June 1993. IEEE Computer Society.