A Pixel Dissimilarity Measure That Is Insensitive to Image Sampling

Stan Birchfield and Carlo Tomasi

Abstract—Because of image sampling, traditional measures of pixel dissimilarity can assign a large value to two corresponding pixels in a stereo pair, even in the absence of noise and other degrading effects. We propose a measure of dissimilarity that is provably insensitive to sampling because it uses the linearly interpolated intensity functions surrounding the pixels. Experiments on real images show that our measure alleviates the problem of sampling with little additional computational overhead.

Index Terms—Dissimilarity, stereo matching, correspondence.

1 INTRODUCTION

WHEN a point in the world is imaged by a stereo pair of cameras, the intensity values of the corresponding pixels are in general different. Many factors contribute to this difference, such as the fact that the light reflected off the point is not the same in the two directions, the two cameras have different gains and biases, the intensities of the pixels are quantized, and noise exists in the camera and framegrabber electronics. Moreover, a pixel value is actually not the image of a point but of a surface patch, and two pixels that contain corresponding world points integrate light reflected off two different surface patches due to foreshortening, depth discontinuities, lens blur, and image sampling.

Although some researchers have proposed measures of pixel dissimilarity that are insensitive to gain, bias, noise, and depth discontinuities [6], [9], [10], [11], [13], there seems to be no work on explicitly achieving insensitivity to image sampling. Yet this latter phenomenon can significantly change the intensity value of a pixel where the intensity function is changing rapidly and where the disparity is not an integral number of pixels (see Fig. 1). Although this may not be a problem if one is only interested in finding the best match for a given pixel, it *is* a problem if a threshold is used to determine matching failure or if the dissimilarities between the pixels are added to other quantities.

For example, there has recently emerged a class of stereo algorithms [1], [2], [5], [7], [8] in which epipolar scanlines are matched by minimizing a cost function that sums the absolute or squared differences of pixel intensities with penalties for occlusions. With the exception of [1] and [2], all of these algorithms work at pixel resolution, and therefore a measure of pixel dissimilarity that is insensitive to sampling would eliminate the errors that they experience due to sampling effects [5]. Moreover, because these algorithms explicitly search over all possible disparities using dynamic programming, working at subpixel resolution is often infeasible because it results in an unacceptable increase in the computational burden.

In this paper, we propose a measure of pixel dissimilarity that compares two pixels using the linearly interpolated intensity functions surrounding them. However, because it does not explicitly reconstruct those functions, the computation required is only slightly more than that of taking the absolute difference in intensity. Our measure is provably insensitive to sampling and is shown to improve the results of a stereo algorithm on real images.

The paper is organized as follows. We define the dissimilarity measure and describe its computation in Section 2. In Section 3, we present two theorems that guarantee that our measure will exhibit the desired behavior under certain general conditions, and we show that it behaves reasonably even when those conditions are not met. The measure is incorporated into a stereo algorithm to demonstrate the improved results in Section 4, followed by a discussion in Section 5 comparing our dissimilarity measure with working at subpixel resolution.

2 DEFINITION AND COMPUTATION OF DISSIMILARITY

Assume that we have a rectified stereo pair of cameras, so that the scanlines are the epipolar lines. Along two corresponding scanlines, let i_L and i_R be the one-dimensional continuous intensity functions that result from convolving the amount of light incident upon the two image sensors with a box function whose support is equal to the width of one pixel. This convolution is due to the fact that a real image sensor can be modeled as an integration of intensity over each pixel followed by an ideal sampler—thus, to allow us to concentrate on ideal sampling, we remove the integration at the outset. The functions i_L and i_R are sampled at discrete points by the ideal sampler of the image sensor, resulting in two discrete one-dimensional arrays of intensity values, I_L and I_R, as shown in Fig. 2. Our goal is to compute the dissimilarity between a pixel at position x_L in the left scanline and a pixel at position x_R in the right scanline; the other pixels shown in the figure are adjacent to these two. First, we define $\hat{I}_R$ as the linearly interpolated function between the sample points of the right scanline, then we measure how well the intensity at x_L fits into the linearly interpolated region surrounding x_R. That is, we define the following quantity:

$$\bar{d}(x_L, x_R, I_L, I_R) = \min_{x_R - \frac{1}{2} \le x \le x_{R+\frac{1}{2}}} \left| I_L(x_L) - \hat{I}_R(x) \right|.$$

Defining $\hat{I}_L$ similarly, we obtain a symmetric quantity:

$$\bar{d}(x_R, x_L, I_R, I_L) = \min_{x_L - \frac{1}{2} \le x \le x_{L+\frac{1}{2}}} \left| \hat{I}_L(x) - I_R(x_R) \right|.$$

The dissimilarity d between the pixels is defined symmetrically as the minimum of the two quantities:

$$d(x_L, x_R) = \min\left\{ \bar{d}(x_L, x_R, I_L, I_R),\ \bar{d}(x_R, x_L, I_R, I_L) \right\}. \tag{1}$$

Since the extreme points of a piecewise linear function must be its breakpoints, the computation of d is straightforward. First, we compute

$$I_R^- \equiv \hat{I}_R\left(x_R - \frac{1}{2}\right) = \frac{1}{2}\left(I_R(x_R) + I_R(x_R - 1)\right),$$

the linearly interpolated intensity halfway between x_R and its neighboring pixel to the left, and the analogous quantity

$$I_R^+ \equiv \hat{I}_R\left(x_R + \frac{1}{2}\right) = \frac{1}{2}\left(I_R(x_R) + I_R(x_R + 1)\right).$$

Then, we let $I_{\min} = \min\left\{I_R^-, I_R^+, I_R(x_R)\right\}$ and $I_{\max} = \max\left\{I_R^-, I_R^+, I_R(x_R)\right\}$. With these quantities defined,

$$\bar{d}(x_L, x_R, I_L, I_R) = \max\left\{0, I_L(x_L) - I_{\max}, I_{\min} - I_L(x_L)\right\}.$$

This computation, along with its symmetric counterpart $\bar{d}(x_R, x_L, I_R, I_L)$, takes only a small, constant amount of time more than the absolute difference in intensity. In practice, we have found the total computing time of our stereo algorithm to increase by less than 10 percent.

• *The authors are with the Computer Science Department, Stanford University, Stanford, CA 94305. E-mail: {birchfield; tomasi}@cs.stanford.edu.*

Manuscript received 16 May 1997; revised 6 Mar. 1998. Recommended for acceptance by R. Szeliski.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106566.

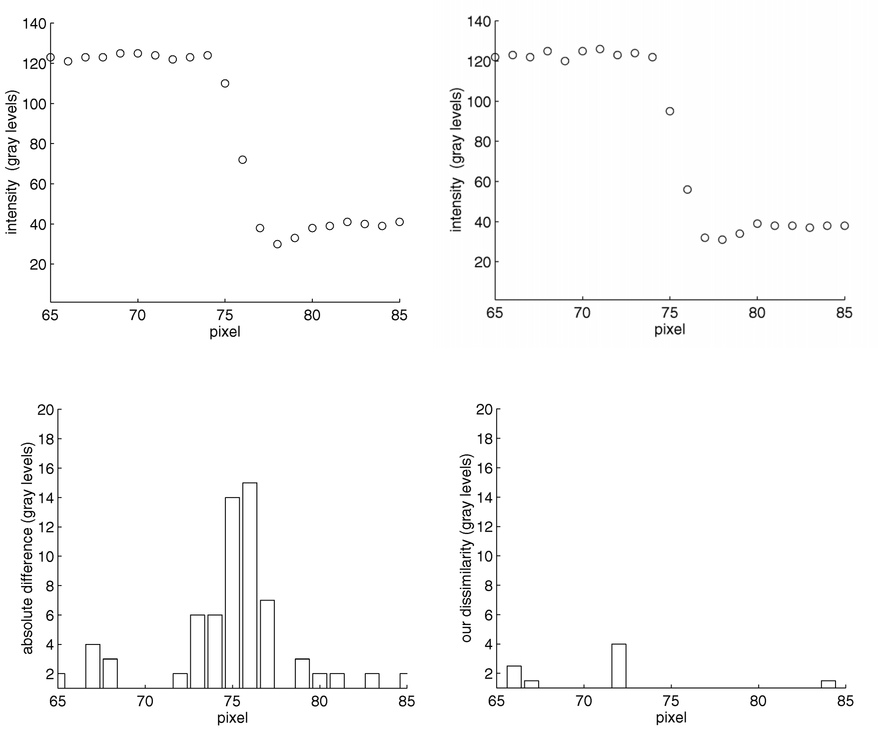


Fig. 1. Top: Two scanlines differing by a disparity of 0.4 pixel. Bottom: The absolute difference in intensity (left) compared with our dissimilarity measure (right). The scanlines are from images of a fronto-parallel, planar object viewed by a camera mounted on a precise translation stage.

3 Analysis

In the following two subsections, the dissimilarity measure of (1) is analyzed, first, by mathematics and simulation and, then, by experiments on real images.

3.1 Theoretical Analysis

For now, let us assume that our cameras are ideal samplers, there is neither photometric nor geometric distortion or shift between the two intensity functions, and there is no noise, so that $i_L = i_R = i$. Keep in mind that, in the following analysis, this restriction is only required in a very small neighborhood surrounding each pixel. We will show by theorems and simulations that the dissimilarity measure defined in (1) is relatively insensitive to sampling if the lenses are slightly defocused to remove aliasing from the sampled intensity functions.

We model the imaging process as a blur function followed by an ideal sampler. The blur results from both lens defocus (which is always present, even in the best of lenses) and integration over the pixel area, and it causes the intensity function to be bandlimited. Since any continuous signal, and hence a bandlimited one, can be broken up into a series of alternating convex and concave sections,[1] two situations are possible: Either both corresponding sampling points lie within a convex or concave region or they straddle one or more inflection points. We will now examine these two situations in turn.

First, wherever the continuous intensity function is either convex or concave in the vicinity of the pixels x_L and x_R, these pixels are correctly assigned a dissimilarity of zero if they should correspond (that is, they are closer to each other than they are to any other sampling points). This is stated by Theorem 1.

THEOREM 1. *Let i be either convex or concave on an interval A, and let x_L and x_R be sufficiently inside A so that $\left[x_L - \frac{1}{2}, x_L + \frac{1}{2}\right] \subseteq A$ and $\left[x_R - \frac{1}{2}, x_R + \frac{1}{2}\right] \subseteq A$. If $\left|x_L - x_R\right| \le \frac{1}{2}$, then $d\left(x_L, x_R\right) = 0$.*

In addition, wherever the intensity function is also linear with nonzero slope (recall that linear functions are by definition both convex and concave), a second theorem guarantees the dissimilarity of two noncorresponding pixels to be nonzero:

THEOREM 2. *Let i be linear and have nonzero slope on an interval A, and let x_L and x_R be sufficiently inside A so that $\left[x_L - \frac{1}{2}, x_L + \frac{1}{2}\right] \subseteq A$ and $\left[x_R - \frac{1}{2}, x_R + \frac{1}{2}\right] \subseteq A$. Then $d(x_L, x_R) = 0$ if and only if $\left|x_L - x_R\right| \le \frac{1}{2}$.*

For interested readers, the proofs to these two theorems can be found in [3].

In the second situation, when the pixels are near inflection points, the behavior of our dissimilarity measure is more difficult to analyze. As long as the inflection points are spaced far enough

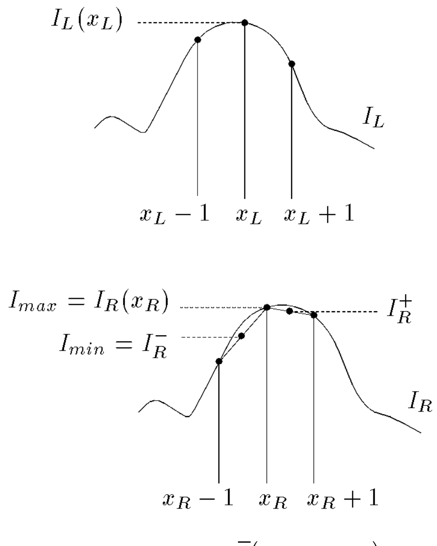


Fig. 2. Definition and computation of $\bar{d}\left(x_L, x_R, I_L, I_R\right)$.

1. Recall that a function is convex if no chord lies below the function and concave if no chord lies above it [12].

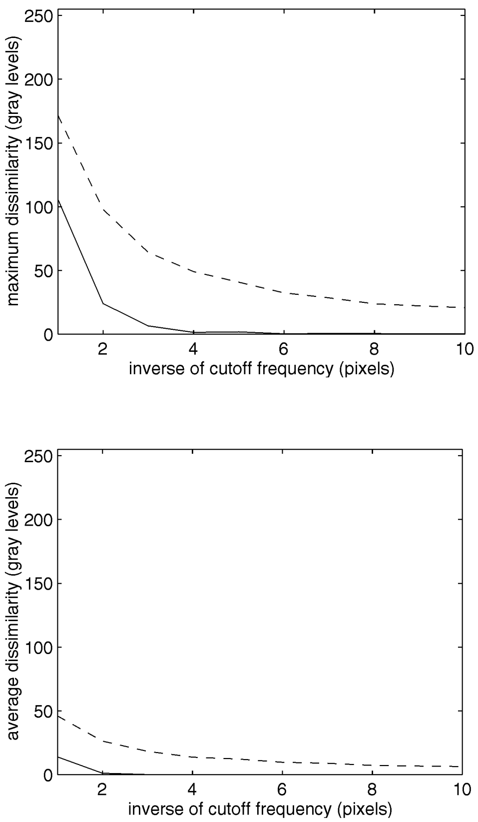


Fig. 3. Our measure (solid) compared with the absolute difference in intensity (dashed) for synthetic pairs of bandlimited images. As the cutoff frequency decreases, our measure quickly goes to zero while the absolute difference remains high. In these images, the dissimilarities were measured for a disparity of zero pixels, while the actual disparity was 0.4 pixel.

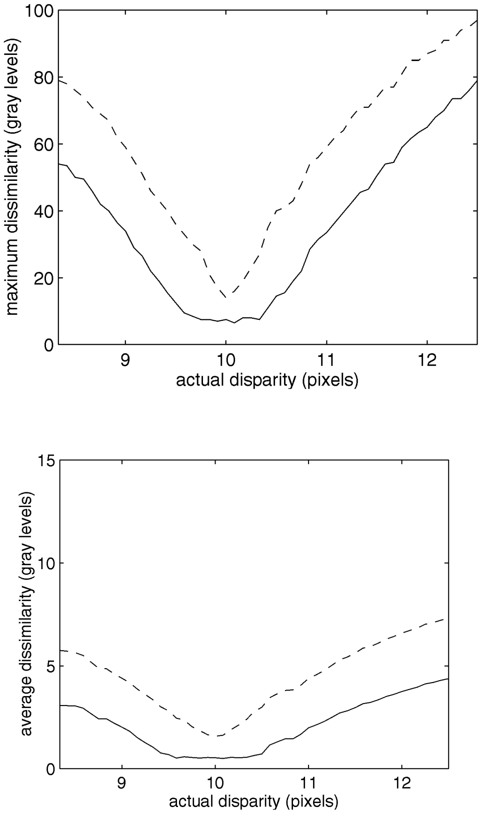


Fig. 4. Our measure (solid) compared with the absolute difference in intensity (dashed) at a hypothesized disparity of 10 pixels.

apart, then the regions surrounding them will be approximately linear, and there will likely be no significant error because of Theorem 2. Although the minimum distance between inflection points cannot be guaranteed, bandlimited signals have the property that as the inflection points get closer together, so do the values of adjacent maxima and minima. In other words, inflection points that are close together have little effect upon the shape of the signal. Therefore, it seems intuitive that the dissimilarity measure will work well as long as there are no high frequencies present in the intensity function.

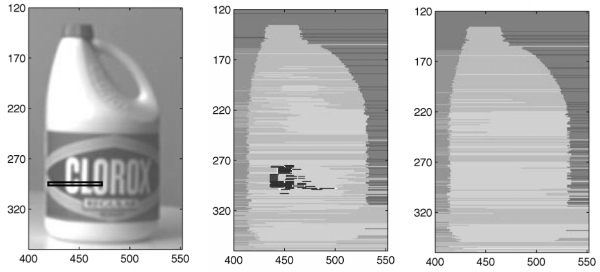
To test this hypothesis, we used Matlab to generate a large number of random bandlimited one-dimensional intensity functions. These functions were shifted by exactly 0.4 pixel to produce corresponding functions. Rounding to the nearest integer, then, these pairs of functions had a true disparity of zero. Fig. 3 shows the dissimilarities at a disparity of zero computed using the absolute difference in intensity and our measure, plotted versus $T_c = 1/f_c$, where f_c is the cutoff frequency (i.e., the maximum frequency) of the intensity function. We see that when the imaginary lenses are slightly defocused to remove aliasing ($T_c = 2$ pixels), the maximum dissimilarity using our measure is 24 gray levels, and the average dissimilarity is just barely one gray level. Compare this to the absolute difference, which yields 98 gray levels for the maximum and 26 gray levels for the average. If the lenses are defocused slightly more, so that $T_c = 4$ pixels, our dissimilarity reduces the effects of the sampling problem to the equivalent of quantization noise, whereas the absolute difference still yields errors on the order of 20 gray levels.

3.2 Experimental Analysis

To obtain quantitative analysis on real images, we used a single Pulnix camera mounted on a precise translation stage to take 52 images of a frontoparallel, planar object that was about 600 mm from the camera. The camera translated 4.00 mm between the first two images and 0.04 mm between subsequent images, the translation being roughly parallel to the scanlines. Empirically, it was determined that each pixel of disparity was equivalent to about 0.48 mm of translation. Therefore, with respect to the reference image (the first image), the disparities of the images ranged from about 8.5 to about 12.5 pixels. The lens was slightly defocused to remove aliasing.

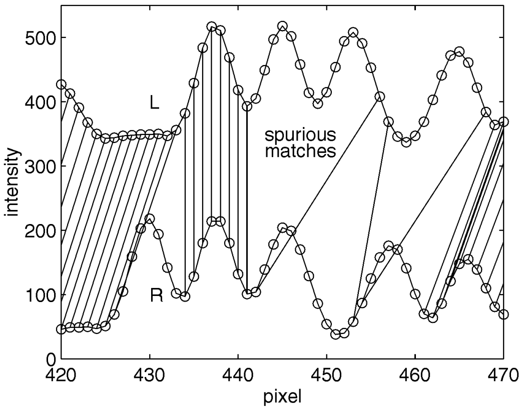
Each pixel in the reference image was compared with its corresponding pixels in all the other images, assuming a disparity of 10 pixels. Maximum and average values were then computed for each image, along with the values obtained by the absolute difference. The results, shown in Fig. 4, are significant for two reasons. First, they verify the validity of the theorems and simulations of the previous subsection. That is, our measure indeed yielded relatively flat behavior for disparities between 9.5 and 10.5, with the maximum dissimilarity varying by no more than two gray levels between disparities of 9.58 and 10.32. In contrast, in this same region the dissimilarities obtained by the absolute difference spanned a range of 19 gray levels. Therefore, meaning is readily attached to the dissimilarity obtained by our measure, while the dissimilarity obtained by the absolute difference is hard to interpret because a large dissimilarity between two pixels may or may not indicate that they correspond.

Second, these plots answer a question left open by the previous analysis, that is, how much discriminating power is lost. Because the slopes outside of the flat region are nearly the same for our measure and the absolute difference, our measure still yields rela-

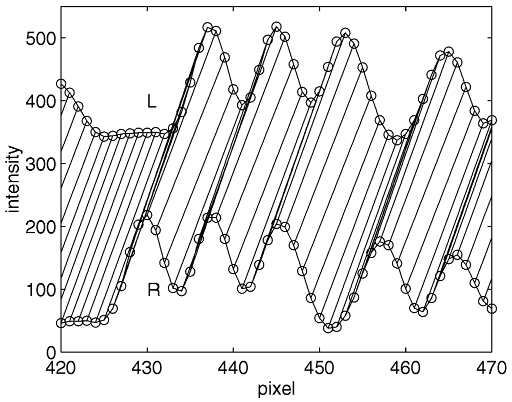


(a) (b) (c)

Fig. 5. (a) An object with a true disparity of about 7.5 pixels. (b) The absolute difference in intensity made a number of mistakes, while (c) our dissimilarity measure correctly assigned disparities of seven or eight throughout the object.



(a)



(b)

Fig. 6. Left and right scanlines from row 294 of the previous figure, along with the matches found by the stereo algorithm. (a) Rather than match the correct pixels, the absolute difference in intensity preferred to match a few random, unrelated pixels and to declare five occlusions. (b) Our dissimilarity measure correctly yielded a disparity of eight throughout.

tively high values for incorrect disparities and thereby allows a matching algorithm to find the best disparity.

4 Performance on Real Stereo Images

To ascertain the performance improvement in the context of stereo matching, we used the scanline-based dynamic programming algorithm described in [4], which computes the disparities along each scanline by minimizing a cost function that adds the dissimilarities of the matched pixels to penalties for the occlusions. We compared the results of the algorithm using our dissimilarity measure with the results of the algorithm using the absolute difference in intensity.

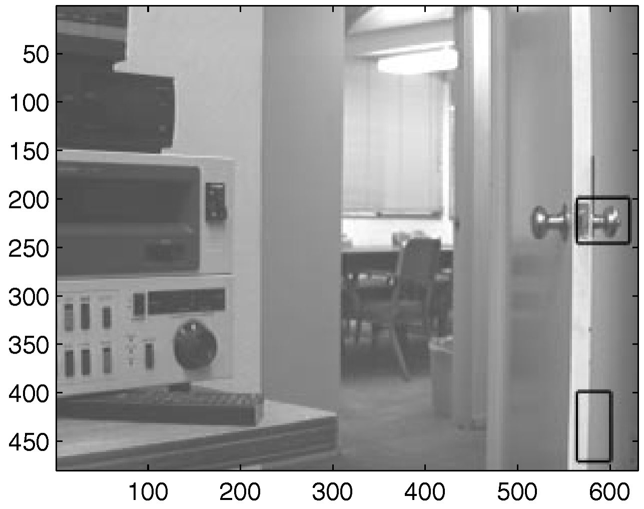


Fig. 7. Regions in which our dissimilarity measure outperformed absolute differences.

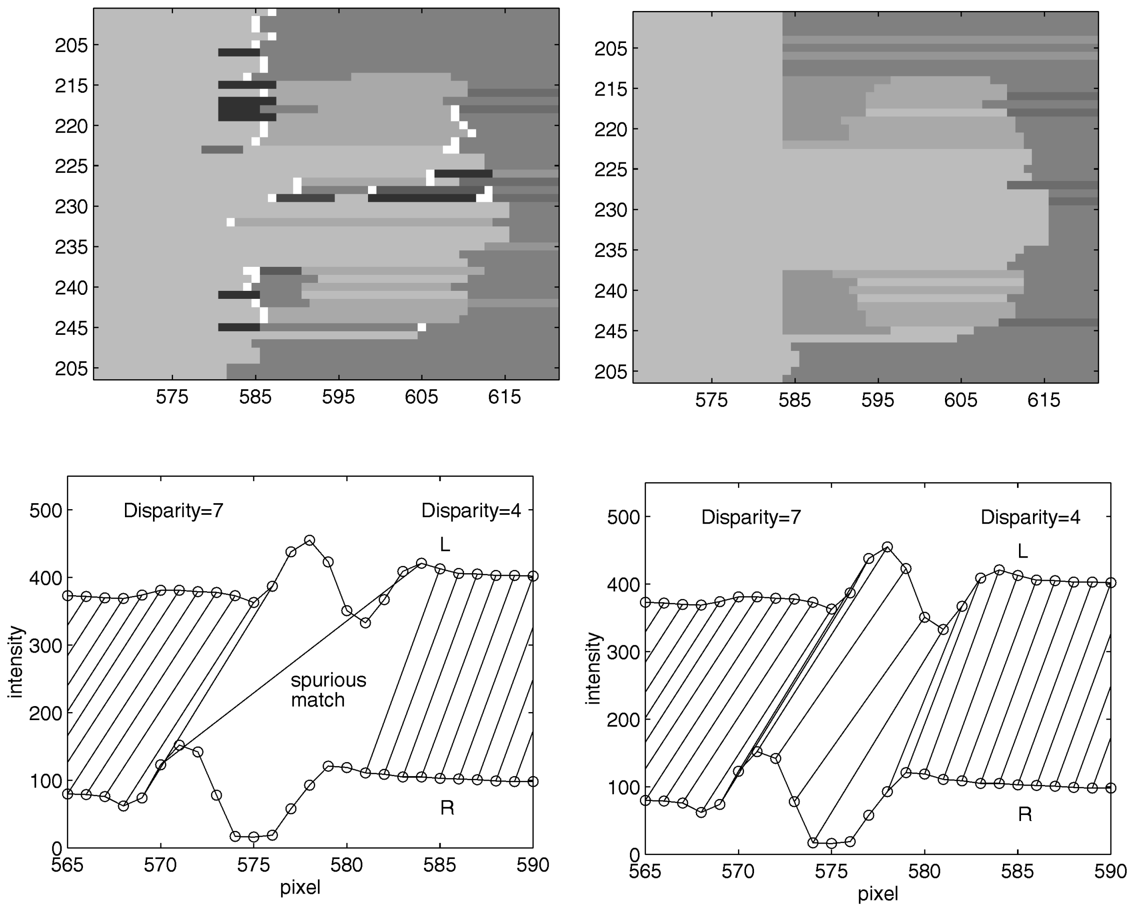
On the six pairs of images that we tested, the absolute difference measure was usually adequate. In fact, when comparing the two measures, we found that fewer than 10 percent of the pixels in the disparity maps changed. However, wherever the intensity function was changing rapidly and the disparity was not an integral number of pixels, our measure was crucial to recovering accurate disparities. In this section, we will highlight three situations in which this behavior was achieved.

One situation was that of the object shown in Fig. 5. Since the actual disparity was about 7.5 pixels, the absolute difference in intensity yielded erratic results in some places while our measure correctly assigned disparities of seven or eight throughout. In particular, along scanline 294 (shown in Fig. 6), a disparity of either seven or eight pixels created such a large absolute difference in intensity that the algorithm preferred to declare five occlusions and assign various disparities to parts of the object. Our measure, on the other hand, correctly assigned a disparity of eight pixels throughout.

The other two situations are shown in Fig. 7. In the region around the doorknob, the boundary between the door and the wall was not cleanly found when using absolute differences, instead large numbers of pixels were skipped (see Fig. 8). This behavior was again due to sampling effects, which caused the corresponding pixels to have very different intensities. In the final situation, along the door edge our dissimilarity measure produced a fairly straight edge, while the absolute difference measure caused a jagged edge whose location was subject to sampling noise, as shown in Fig. 9.

Although the first two errors mentioned above could have been eliminated by simply increasing the overall penalty for occlusions, the wiggly door edge remained no matter how the parameters[2] were chosen. (In contrast, our measure did not yield any of these errors, regardless of the parameter values.) Moreover, the parameter values that alleviated the sampling problem for absolute differences on one image were different from the parameter values that worked on other images. In short, for our

2. In the stereo algorithm, two parameters govern the occlusion penalty.



(a) (b)

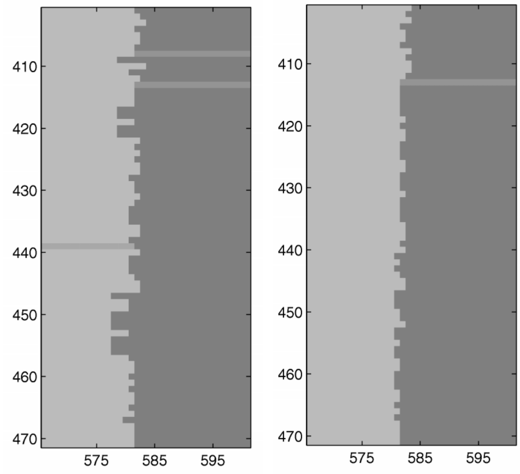
Fig. 8. Top: Disparity map of doorknob using (a) absolute differences, and (b) our dissimilarity measure. Bottom: Portions of the scanlines of row 205, with matches. (a) Absolute differences refused to match the correct pixels and instead declared a spurious match with two large sets of occluded pixels. (b) In contrast, our measure correctly found a disparity of seven on the door and four on the wall, with three occluded pixels in between.

six images, there was no single choice of parameter values that yielded results using absolute differences that were as good as those obtained using our dissimilarity measure with a particular choice of parameter values.

5 Comparison With Subpixel Resolution

A natural question to ask is how using our dissimilarity measure at pixel resolution compares with using the absolute difference measure at subpixel resolution. It is important to notice that even with our slightly defocused images, the intensities of adajacent pixels often differ by thirty gray levels or more. Therefore, the amount of interpolation that is necessary to bring the sampling problem down to the level of quantization noise is quite substantial: The images must be interpolated by a factor of about 10 to 15.

We linearly interpolated our six images by a factor of 10 and ran the stereo algorithm. No appreciable difference in the results was noticed except where objects had nonintegral disparities. For example, in the case of the Clorox bottle, shown in Fig. 10, our measure produced a noticeable disparity wobble between seven and eight pixels, whereas the disparity on the linearly interpolated image was smoother. It should be noted, however, that although the latter disparity map looks smooth, there is actually a substantial variation between the values, ranging from 7.0 to 8.2 pixels. While the disparity map obtained using linear interpolation is certainly as good if not better than the one obtained using our measure, depending on the application, this higher but very expensive accuracy may not be needed. For example, if the goal is segmentation rather than reconstruction, then pixel resolution may be adequate, especially if a postprocessing step can detect and correct the disparity wobble.



(a) (b)

Fig. 9. A straighter door edge was produced using (b) our measure than (a) absolute differences.

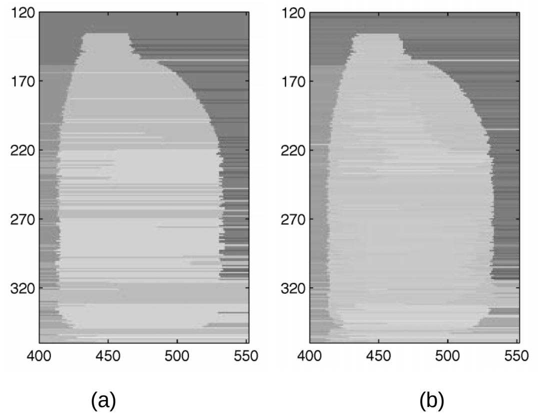


Fig. 10. The disparity map obtained using (a) our measure at pixel resolution (a copy of Fig. 5c) and (b) the absolute difference measure at the resolution of a tenth of a pixel. Both maps exhibit a range of disparities from seven to eight pixels, but the one in Fig. 10b is smoother.

The most compelling reason for working at pixel resolution, at least when using a dynamic programming stereo algorithm, is the increase in speed. The algorithm described in [4] has a running time of approximately $O(mn\Delta \log\Delta)$ (where the image is of size $m \times n$, and Δ is the maximum disparity), which is slightly faster than that of the standard implementation. Even so, interpolating by a factor of 10 increased the computing time by 1,100 percent, from eight seconds to 15 minutes, because both the number of pixels in a scanline and the maximum disparity are proportional to the amount of interpolation.

6 Conclusion

Sampling is an important phenomenon that can contribute significantly to the difference in intensity between corresponding pixels of a stereo pair. When working at pixel resolution, traditional measures of dissimilarity, such as the squared or absolute difference in intensity, do not give a good indication of whether two pixels match, making these measures inadequate for applications in which the dissimilarity values are added to other values in a cost function or are thresholded to determine matching success. We have proposed a dissimilarity measure that is provably insensitive to sampling in the sense that two corresponding pixels always have a dissimilarity near zero (in the noise- and distortion-free case), whenever the camera lenses are slightly defocused to remove aliasing. The measure was shown to improve the results of a stereo algorithm tested on real images.

This dissimilarity measure could be effortlessly integrated into existing stereo algorithms that minimize an objective function that adds dissimilarities to occlusion penalties [1], [2], [5], [7], [8]. In such an algorithm, it could be used either with or without correlation windows. Future work should be aimed at extending the dissimilarity measure from one dimension to two dimensions, to allow its use in motion tracking, not just stereo.

Acknowledgment

This research was supported by the U.S. National Science Foundation under a Graduate Research Fellowship and under contract IRI-9506064, and by the Department of Defense under MURI contract DAAH04-96-1-0007 monitored by ARO and under a subcontract of STTR contract F49620-95-C-0078 monitored by AFOSR.

References

[1] P.N. Belhumeur, "A Binocular Stereo Algorithm for Reconstructing Sloping, Creased, and Broken Surfaces in the Presence of Half-Occlusion," *Proc. Fourth Int'l Conf. Computer Vision*, pp. 431–438, May 1993.
[2] P.N. Belhumeur and D. Mumford, "A Bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 506–512, June 1992.
[3] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo," Technical Report STAN-CS-TR-96-1573, Stanford Univ., July 1996.
[4] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo," *Proc. Sixth Int'l Conf. Computer Vision*, pp. 1,073–1,080, Jan. 1998.
[5] I.J. Cox, S.L. Hingorani, S.B. Rao, and B.M. Maggs, "A Maximum Likelihood Stereo Algorithm," *Computer Vision and Image Understanding*, vol. 63, pp. 542–567, May 1996.
[6] A. Crouzil, L. Massip-Pailhes, and S. Castan, "A New Correlation Criterion Based on Gradient Fields Similarity," *Proc. 13th IAPR Int'l Conf. Pattern Recognition*, pp. 632–636, Aug. 1996.
[7] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and Binocular Stereo," *Int'l J. Computer Vision*, vol. 14, pp. 211–226, Apr. 1995.
[8] S.S. Intille and A.F. Bobick, "Disparity-Space Images and Large Occlusion Stereo," *Proc. Third European Conf. Computer Vision*, pp. 179–186, May 1994.
[9] H.K. Nishihara, "Practical Real-Time Imaging Stereo Matcher," *Optical Eng.*, vol. 23, pp. 536–545, 1984.
[10] D. Scharstein, "Matching Images by Comparing Their Gradient Fields," *Proc. 12th IAPR Int'l Conf. Pattern Recognition*, pp. 572–575, Oct. 1994.
[11] P. Seitz, "Using Local Orientational Information as Image Primitive for Robust Object Recognition," *Proc. SPIE*, vol. 1,199, pp. 1,630–1,639, Nov. 1989.
[12] R. Webster, *Convexity*. Oxford, England: Oxford Univ. Press, 1994.
[13] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," *Proc. Third European Conf. Computer Vision*, pp. 151–158, May 1994.